NBER WORKING PAPER SERIES

MONEYBALL IN MEDICARE:
HETEROGENEOUS TREATMENT EFFECTS

Edward C. Norton
Emily J. Lawton
Jun Li

Working Paper 27948
http://www.nber.org/papers/w27948

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2020

This work was supported by the National Institute on Aging (P01-AG019783). No conflicts of interest or financial disclosures are reported by any authors. Elena Prager and Lina Maria Elleg rd provided insightful conference discussions. We are grateful for comments by seminar participants at ASHEcon and the Nordic Health Economics Study Group (2019). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27948.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Edward C. Norton, Emily J. Lawton, and Jun Li. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Moneyball in Medicare: Heterogeneous Treatment Effects
Edward C. Norton, Emily J. Lawton, and Jun Li
NBER Working Paper No. 27948
October 2020
JEL No. I11,I18

ABSTRACT

One of the most important changes to the United States health care system over the last two decades is the emergence of pay-for-performance as a way to encourage hospitals and other providers to improve quality of care. Unlike fee-for-service, these value-based purchasing programs measure aspects of quality and financially reward hospitals that are outstanding or at least improving in their care. Prior research has shown that hospitals often improve more when the marginal financial incentives are larger. However, the exact relationship between marginal financial incentives and year-over-year improvement in measures remains unclear. In this study, we use national 2015 2018 data on approximately 2,700 hospitals to estimate how hospitals respond to pay-for-performance incentives in the Hospital Value-Based Purchasing (HVBP) Program. We show that this relationship is non-linear, has strong serial correlation, is somewhat similar for safety-net hospitals as non-safety-net hospitals, and is proportional to the size of the Medicare patient population.

Edward C. Norton
Department of Health Management and Policy
Department of Economics
University of Michigan
School of Public Health
1415 Washington Heights, M3108 SPHII
Ann Arbor, MI 48109-2029
and NBER
ecnorton@umich.edu

Emily J. Lawton
Department of Health Management and Policy
University of Michigan
School of Public Health
1415 Washington Heights
Ann Arbor, MI 48109
United States
ejlawton@umich.edu

Jun Li
Syracuse University
Maxwell School of Citizenship
and Public Affairs
Department of Public Administration
and International Affairs
314 Lyman Hall
Syracuse, NY 13244
jli208@maxwell.syr.edu

I. Introduction

The Centers for Medicare and Medicaid Services (CMS) in the United States has devoted substantial effort to move away from traditional fee-for-service and toward value-based reimbursement. As a result, CMS has implemented a number of value-based purchasing programs, which provide financial incentives to hospitals to improve quality of care and reduce episode spending. It is not known, however, whether these value-based purchasing programs achieve their goals, and if so, how hospitals respond to the financial incentives over time (Ryan et al., 2015; Figueroa et al., 2016; Glied and Sacarny, 2018; Banerjee et al., 2019).

This paper studies the Hospital Value-Based Purchasing (HVBP) Program. Implemented in fiscal year 2013, the HVBP program rewards or penalizes hospitals based on performance on a variety of measures, including patient experience, clinical process, clinical outcomes including patient survival, and Medicare spending per beneficiary. A recent study showed that hospitals facing larger financial incentives at the margin for a particular quality measure were often more likely to improve performance on that measure in the following year (Norton et al., 2018). However, this study used a simple linear relationship to predict one-year improvement, which may not fully capture the nuances of the incentives in the program or dynamic responses from hospitals.

The relationship between improvement and incentives in the program may be complex and non-linear for several reasons. There may be diminishing returns to incentive size or a kink dividing hospitals with zero versus positive incentives. There may be heterogeneity in hospitals' response to incentives because some hospitals, such as safety-net hospitals, face more limited resources to improve quality of care. There may also be non-linearities over time because interventions may take longer than one year to come to fruition. Furthermore, the relationship could depend on the form of the marginal incentive, i.e., either the marginal total performance score or marginal future reimbursement. The *marginal total performance score* reflects how much higher the hospital can expect to score in the program with improvement, while the *marginal future reimbursement* captures how many more dollars can be expected. The former is a percentage score, while the latter is in absolute dollars. The latter is weighted by Medicare revenue and therefore tends to be larger for large hospitals.

Understanding the relationship between incentives in the program and quality improvement among hospitals is critical for policy refinement. For instance, if hospitals respond non-linearly, then the incentive amount per hospital should be based on where the hospitals are in relation to the optimum point on the production frontier. Similarly, if hospitals respond to absolute dollars as opposed to percentage score, then the program may need additional targeting for smaller hospitals. Moreover, if there is a jump in the effect of the marginal incentive on changes in hospital performance on measures once the marginal incentive becomes non-zero, it may be important to modify the program so that more hospitals have a non-zero marginal incentive for improvement on targeted measures.

To assess for non-linear effects, lags, and different responses for safety-net hospitals compared to non-safety-net hospitals, we use national data from 2015−2018 on approximately 2,700 hospitals. Specifically, we test whether hospitals facing larger financial incentives for four specific measures (acute myocardial infarction survival, congestive heart failure survival, pneumonia survival, and Medicare spending per beneficiary) improve on those measures in the next year. We focus on these measures because they are highly salient to patients and the health care community, commonly used to evaluate health systems performance, and have been a stable component of the program since their introduction.

The results are striking. The relationship is decidedly non-linear for all four measures, with both an extensive margin for having any positive incentive and an intensive margin for the magnitude. The pattern of lagged effects indicates that hospitals respond most strongly to incentives in the immediate prior year and less so with longer passage of time. The differences between safety-net and non-safety-net hospitals are somewhat modest, although the magnitude of the marginal incentive coefficient in the scaled marginal TPS models for survival measures is generally larger for safety-net hospitals than for non-safety-net hospitals. Marginal future reimbursement (weighted by Medicare revenue) is a slightly more powerful predictor of change in hospital performance on measures than the marginal total performance score (unweighted).

In summary, the results confirm that financial incentives matter in HVBP and that hospitals respond to higher financial incentives by improving on those measures. For policy, this has implications for the HVBP program design. Even small financial incentives create a positive response, so it would be important to make sure that all (or nearly all) hospitals have a positive financial incentive to improve. This is not generally true of the program. Our findings also show that measures are highly serially correlated and that year-to-year fluctuations in the measures are less important. Finally, safety-net and non-safety-net hospitals respond in a similar way, meaning that separating the program for these two types of hospitals will have negligible benefits.

II. Theoretical Framework

Under pay-for-performance (P4P), hospitals that perform better on some measures will be paid more. Because P4P programs inherently rely on economic incentives to drive quality improvement, for these P4P programs to achieve their desired behavioral responses, it is imperative that hospitals respond to these incentives (Norton et al., 2018; Hoffman and Yakusheva, 2020; Li and Norton, 2019). In contrast, if hospitals are not responsive to the financial incentives, then the premise of the P4P programs fails, and public policy should change.

In mathematical terms, the change in an outcome measure *y* under P4P should be some function of the marginal incentive for that measure (defined in two ways as marginal future reimbursement (MFR) or scaled marginal total performance score (MTPS), both discussed in more detail below), as well as various hospital characteristics $\boldsymbol{x}$. For measure *m* measured for hospital *h* in year *t* with a lag of *k* years, the basic relationship is

$$\Delta y_{mht} = f(MarginalIncentive_{mht-k}, \boldsymbol{x_{ht}}) \tag{1}$$

To visualize this complicated relationship, imagine graphing the change in quality of care on the *y*-axis as a function of the marginal future reimbursement on the *x*-axis (see Figure 1). Because CMS wants a high return on its investment, CMS hopes that hospitals would be in the northwest corner of the graph, that is, have a high change in quality of care for only a small financial incentive. In contrast, CMS wants to avoid paying large incentives and only getting little or no change in quality in return. That is, it wants to avoid having hospitals in the southeast part of the graph.

We can glean other insights from Figure 1. If hospitals are aligned in a straight line from the origin (to the northeast) then each hospital has the same ratio of marginal financial incentive to change in quality of care. That means that they have the same marginal cost effectiveness ratio. Small incentives lead to small improvements in quality of care, and large incentives lead to proportionately larger improvements in quality of care.

If instead of a straight-line relationship the function is concave, then it is progressively more expensive to induce larger changes in quality of care. The economically efficient way to run the program with a concave function would be to give small financial incentives to hospitals. If instead of a concave function, the function was convex, then the economically efficient way to run the program would be to only give large financial incentives which would induce large increases in quality of care.

Another possibility for the relationship could be a discontinuous jump at the origin, with small financial incentives discontinuously inducing modest increases in quality, and then perhaps a concave function for positive values. Behavioral economics has found that in some cases, small incentives are materially different than zero incentive (Thornton, 2008). Alternatively, another possibility is that there is no relationship at all. Perhaps the program is too confusing or hospitals have too many other things to worry about, but it could be that hospitals ignore the incentives and if by random luck they happen to improve measured quality of care anyway, then they are happy to collect their bonus payment. In that case, the hospitals would be scattered along the x-axis with no apparent relationship.

It is critical to know the empirical relationship between the financial incentives that hospitals face and how they respond by improving quality of care. Therefore, this paper seeks to measure the relationship between these two important variables. If we find a tight plot of hospitals around a line, then there is a strong relationship with low heterogeneous effects. In contrast, if we instead find lots of dispersion, this would indicate a weak relationship, poor measurement, problem of small numbers, or high heterogeneous treatment effects.

Theory does not dictate the exact functional form of the relationship between change in measures and the marginal incentive $f(MarginalIncentive_{mht-k}, \boldsymbol{x_{ht}})$. In our prior paper (Norton et al., 2018) we assumed that the function $f(\cdot)$ was linear, that there was only a one-year lag between the measurement of the marginal incentive and change in hospital performance on measures, and that the incentive was weighted by Medicare revenue (emphasizing marginal future reimbursement rather than marginal total performance score). In this paper, we empirically test each of those assumptions, while allowing for different effects by safety-net and non-safety-net hospitals.

First, the function could be non-linear. We consider and compare several different functional forms. The function should be monotonic, meaning that an increase in the marginal incentive would (weakly) lead to greater improvement in the corresponding measure, but never a decrease in performance. However, there could be diminishing returns to financial incentives such that the relationship is logarithmic. There could be a discrete jump in the effect of the marginal incentive on changes in hospital performance on measures. For example, if hospitals only paid attention to the incentive if it was non-zero, but not if it was zero, then this could have a discrete jump. Often, financial relationships are log-linear, meaning that a percentage increase in the financial variable leads to a linear change in the dependent variable. Therefore, we compare five different functional forms for $f(MarginalIncentive_{mht-k})$.

If positive: $f(MarginalIncentive_{mht-k}) = \alpha_1 1(MarginalIncentive_{mht-k} > 0)$

Linear: $f(MarginalIncentive_{mht-k}) = \beta_1 MarginalIncentive_{mht-k}$

If positive & linear: $f(MarginalIncentive_{mht-k}) =$

$$\zeta_1 1(MarginalIncentive_{mht-k} > 0) + \zeta_2 MarginalIncentive_{mht-k}$$

If positive & quadratic: $f(MarginalIncentive_{mht-k}) =$

$$\gamma_1 1(MarginalIncentive_{mht-k} > 0) + \gamma_2 MarginalIncentive_{mht-k} + \gamma_3 MarginalIncentive^2_{mht-k}$$

Natural log (estimated on the sub-sample with positive marginal incentive):

$$f(MarginalIncentive_{mht-k}) = \delta_1 \ln(MarginalIncentive_{mht-k})$$

We compare the fit across different specifications in several ways. For each measure, we graph the relationship between the change in hospital performance on that measure and the marginal incentive on both the linear and log scale. We look for a break at zero and whether the relationship over the positive range of the marginal incentive was flat, linear, or log-linear. We also use information criteria (AIC and BIC) to compare the model fit because these different functional forms are not nested.

Second, the timing is likely to matter. In our prior paper, we assumed that the marginal incentive in year *t* affected the change in measure performance over the next year, from year *t* to year *t*+1. There are several reasons why this relationship may have a different timing. There is a roughly two-year lag between when data on patients are collected (then the data are cleaned and analyzed, resulting in achievement and improvement points, and ultimately a financial bonus) and when the percentage bonus is applied to future Medicare payments. Therefore, it takes time for a hospital to become aware of its own performance and how its performance could affect future payments.

Further complicating the timing is the possibility that hospital performance is highly serially correlated. If hospitals perform similarly year over year, and they know this, then they would have a good sense of their marginal incentive. Consequently, the year-to-year variation in performance is less important, and the time lag is less important, and we would expect hospitals to respond similarly each year because the marginal incentive is similar each year.

To explore the role of timing, we conduct two empirical analyses. We compute the correlation of the marginal incentive across years, with up to a three-year lag. If hospital performance on measures is similar year after year, then the correlation should be fairly high. In contrast, if the measures are poorly designed and are fairly random (perhaps due to small sample sizes) then the correlation across years will be small (Friedson et al, 2019). We also run regressions that include the marginal incentive with lags, up to three lags. Therefore, we run models with the following general specification.

$$\Delta y_{mht} = \beta_0 + \beta_1 MarginalIncentive_{mht-1} + \beta_2 MarginalIncentive_{mht-2} + \beta_3 MarginalIncentive_{mht-3} + \beta \boldsymbol{x_{ht}} + \boldsymbol{\varepsilon}_{mht} \qquad (2)$$

Third, we explore heterogeneous treatment effects. Specifically, we test whether safety-net hospitals respond differently to the HVBP program financial incentives compared to non-safety-net hospitals. Safety-net hospitals treat a disproportionate share of uninsured and low-income patients. They also tend to be larger, more likely to be teaching hospitals, and less likely to receive bonuses under the HVBP program (Norton et al., 2018). The precise definition of safety-net hospital used for this empirical study is in the data section.

We are interested in knowing if safety-net hospitals are just as responsive to financial incentives as non-safety-net hospitals. On the supply side, safety-net hospitals may be more resource constrained, making it harder to respond to financial incentives to improve quality of care. However, receiving a financial bonus could be more important for safety-net hospitals if revenue has declining marginal utility. Therefore, it is an empirical question of whether safety-net hospitals are more or less responsive to the financial incentives. To test this, we conduct all analyses separately by safety-net status.

Fourth, conceptually there are two different ways to measure the size of the financial incentive. The first way is scaled marginal TPS, which reflects each hospital's original marginal TPS (that is, the effect of a one-decile improvement in a measure on the total performance score) then multiplied by the hospital's Medicare days as a percentage of inpatient days. The TPS treatment effect is scaled since, compared with other hospitals, hospitals that have a larger Medicare patient share may be more responsive to a given marginal TPS value. The second way is MFR, which reflects each hospital's original marginal TPS weighted by the size of the Medicare reimbursement that the hospital receives. Under the program, the TPS affects reimbursement through getting translated into a percent bonus or penalty (determined by performance) on each Medicare reimbursed case. Thus, given their size (or more precisely, their amount of Medicare revenue), large hospitals have relatively larger MFR than marginal TPS when compared to small hospitals. In essence, the difference in these incentives is the difference between relative and absolute—with scaled marginal TPS being the relative incentive and MFR being the amount of incentive in absolute dollars.

It is not clear *a priori* which of these incentives (MFR or scaled MTPS) is what hospitals respond to for improvement in measure performance. On one hand, hospitals may respond to marginal TPS if they respond to percentage effects. On the other hand, hospitals may respond to marginal MFR if they respond to absolute dollars. The difference being that if a large and small hospital each had a 1% financial bonus, the large hospital would earn more money. Therefore, we estimate some models with MFR and some with scaled marginal TPS.

$$\Delta y_{mht} = \theta_0 + \theta_1 MFR_{mht-1} + \theta \boldsymbol{x}_{\boldsymbol{ht}} + \boldsymbol{\varepsilon}_{mht}$$

$$\Delta y_{mht} = \vartheta_0 + \vartheta_1 MTPS_{mht-1} + \vartheta \boldsymbol{x}_{\boldsymbol{ht}} + \boldsymbol{\varepsilon}_{mht}$$

We compare θ_1 to ϑ_1 in magnitude and statistical significance. We also compare the goodness of fit of the models using information criteria (AIC and BIC) because the models are not nested.

III. Empirical Measures

A. Performance Improvement in HVBP

We focus on four out of more than twenty measures in HVBP to assess hospital improvement under the program. These four are survival for acute myocardial infarction (AMI), survival for congestive heart failure (CHF), survival for pneumonia (PN), and Medicare spending per beneficiary (MSPB). We focus on these measures for several reasons. First, they were consistently included in the HVBP program for fiscal years 2015–2018 and represent a considerable contribution to overall performance. Second, mortality-based measures are salient to the general population and reflect widely-used and well-understood measures of quality. In contrast, other measures in the HVBP program were phased out, and often removed due to a ceiling effect, where many hospitals achieved the maximum score and could not further improve. Furthermore, measures introduced in later years may also be less natural targets for hospital improvement efforts because they may be less familiar and require significant reallocation efforts.

B. Effect of a one-decile improvement in measure performance

To estimate the incentives to improve for each hospital, we estimate the effect of achieving a one-decile improvement in a specific measure on each hospital's future scaled total performance score and Medicare reimbursement (Norton et al., 2018). A one-decile improvement in a measure is standardized across measures. It is large enough to be meaningful and it is roughly equal to the mean year-to-year improvement for hospitals that do improve. A one-decile improvement can be estimated without individual data. Because we know each hospital's score for each of the measures in all domains in every year, we can simulate what would happen if that measure improved by one decile in the respective year. For hospitals in the top decile, we simulate them moving to the top of the distribution.

In this way, we can estimate the effect of a modest improvement in each measure. How large is a one-decile improvement? It depends on the measure and which part of the distribution the hospital is in. A one-decile improvement in the tails of the distribution generally means a larger change in absolute performance rate. For example, for AMI survival, hospitals in the lowest quintile would need to improve mortality by 0.7 percentage points (on a baseline of 16 percent mortality in fiscal year 2015), while those in the second and fourth quintiles would only need to improve by 0.3 percentage points, those in the middle quintile by 0.2 percentage points, and those in the top quintile by 1.4 percentage points. Given the change in measure performance, we can estimate the change in achievement and improvement points, which leads to increases in TPS, percent bonus, and dollars. Although all hospitals in this simulation will improve their measure performance, the change in dollars may be small or zero if there is no corresponding change in points.

In summary, the marginal future reimbursement is the product of five derivatives that collectively calculate the effect of a one-decile change in a measure on future reimbursement.

For each of the four spending and survival measures (*m*), we estimate the following for each hospital *h* in year *t*:

$$MFR_{mht} = \frac{dMeasure_{mht}}{d1decile_{mht}} \times \mathrm{E}\left[\frac{dPoints_{mht}}{dMeasure_{mht}}\right] \times \frac{dTPS_{ht}}{dPoints_{mht}} \times \frac{d\%Bonus_{ht}}{dTPS_{ht}} \times \frac{d\$_{ht}}{d\%Bonus_{ht}}$$

C. Scaled Marginal Total Performance Score and Marginal Future Reimbursement

Our conceptual framework predicts that having a larger financial incentive to improve should be associated with a greater year-to-year improvement in that measure. Therefore, using fiscal year 2015–2018 data, we estimate the change in measure performance (from the base year to the subsequent year) as a function of the size of the marginal incentive. We operationalize this in two ways. We use the calculated scaled marginal total performance score as well as the calculated marginal future reimbursement, where both marginal incentives are from a one-decile improvement in measure performance. Certainly, hospitals may improve by more (or less) than one decile, but a one-decile improvement standardizes improvements across different measures.

We then estimate several models for each marginal incentive. Within each model specification, we analyze the impact of improving the performance of a specific survival or spending measure by one-decile, accounting for all years in which this measure was included in the HVBP program. Accordingly, we combine and analyze fiscal year 2015–2018 data for all hospitals that met the minimum number of eligible cases to receive a measure score on that specific measure for each of these four years.

Our initial model estimates the change in a measure *m* from the base year *t* to the subsequent year in hospital *j* as a function of the base year marginal incentive (MFR or scaled MTPS) (due to a one-decile change in the measure). The model has one observation per hospital per base year. The models are run separately by safety-net status and include hospital fixed effects. Hospital-level control variables ($\boldsymbol{x}$) include the value-based purchasing adjustment factor in the base year, performance on the given measure in the base year, Medicare days as a percentage of total inpatient days (for the MFR models, since the MTPS variable was scaled by this value), teaching affiliation, number of beds, for-profit status, and integration status.

$$\Delta Measure_{mht} = \beta_0 + \beta_1 MarginalIncentive_{mht} + \boldsymbol{x}'_{ht}\beta + \varepsilon_{mht}$$

We also estimate non-linear versions of the initial measure-specific models. First, we estimate the change in the specific measure as a function of whether the hospital has a positive marginal incentive in a given year (due to a one-decile change in performance on that measure). Second, among those hospitals with a positive marginal incentive, we estimate the change in the measure as a function of the natural log of the marginal incentive. Third, we include a quadratic term for the marginal incentive when estimating the change in measure performance as a function of the marginal incentive. Fourth, we use a fully flexible specification where we categorize hospitals into five groups according to baseline 2015 performance on the measure and then run models of hospitals' measure performance improvement in subsequent years as a function of their baseline performance category. This approach was a more flexible way of looking at that relationship, allowing for more non-linearities as well as the capacity to observe whether the pattern was monotonic, linear, U-shaped, or exhibited distinct nonlinearities. As a sensitivity analysis, we also categorized hospitals into ten groups for this analysis (not shown), with substantively similar results.

For each model, our hypothesis is that the coefficient on the marginal future reimbursement or the scaled marginal total performance score is greater than zero for the survival measures because a higher score on the survival measures indicates better performance, while our hypothesis is that the coefficient is negative for the spending measure since a lower score indicates better performance (i.e., less spending) on that measure.

Finally, for each of the survival and spending measures, we estimate models with lags. This model specification estimates the change in the given measure in the most recent year as a function of the marginal incentives corresponding to each of the three preceding years. This provides important information on the lag structure pertaining to hospital measure performance improvement in response to marginal incentives.

IV. Data

To estimate the effect of a one-decile improvement in measure performance, we need national, hospital-level data. Specifically, we need HVBP performance data for each program measure in every year for all hospitals subject to the HVBP program. We also need hospital-specific values in each year for calculating case-mix adjusted annual Medicare base payment rates to determine hospitals' marginal future reimbursement. Additionally, for every hospital, we need data on hospital characteristics and previous payment adjustments resulting from the HVBP program to control for these factors in the analyses.

We used the CMS Hospital Compare data to identify all hospitals that were continuously subject to the HVBP program in fiscal year 2015–2018 and that met the minimum case requirements to receive a measure score in all four years for at least one specific survival or spending measure. This produced a national sample of around 2,674 hospitals operating in all states except Maryland, which is not subject to the HVBP program (see Table 1). We then used these data to evaluate the effect of marginal incentives on hospital-level performance improvement on a specific survival or spending measure. In each measure-specific analysis, we restricted the sample to the subset of hospitals that met the minimum case requirements for that specific measure across all study years. Therefore, the number of observations differs by measure. Table 2 displays the total number of observations (over the 2015–2017 base years) for each measure-specific analysis.

We calculated the measure-specific outcome and independent variables by hospital and year. The outcome variable is each hospital's year-to-year change in performance on the given measure. Our independent variables of interest are the scaled marginal total performance score and the marginal future reimbursement from a one-decile improvement in performance on the survival and spending measures.

To obtain Medicare reimbursement information, we used annual CMS Final Rule and Impact Files. We obtained the annual, hospital-specific, actual HVBP payment adjustments from the year-specific Table 16B FR file. Additionally, we used the case-mix adjusted annual Medicare base payment rates to calculate hospitals' marginal future reimbursement. Because HVBP payment adjustments only affect the Medicare Part A base payment rates, we excluded Medicare reimbursements for disproportionate share hospital, indirect medical education, and new technology payments.

A potential source of heterogeneity examined in this study relates to hospitals' safety-net status. Because safety-net status is not officially designated by CMS, we use a definition from recent literature and construct our measure of safety-net status using the Impact File. Accordingly, we define safety-net hospitals based on two methods (Chatterjee and Joynt, 2014, Gilman et al., 2014, Gilman et al., 2015, Mohan et al., 2013), where one component of our measure is the percentage of disproportionate share hospital (DSH) adjustment and the other is the amount of Medicare uncompensated care per claim (UCC). CMS uses DSH percentage to reimburse hospitals serving larger proportions of patients covered by Medicare supplementary Social Security Income or Medicaid. Following fiscal year 2014, Medicare's DSH payments were reduced to 25 percent, and the remaining 75 percent transitioned to uncompensated care. Each Medicare DSH hospital received an uncompensated care payment based on 75 percent of the estimated DSH payment, share of uninsured patients, and a relative amount of UCC as compared to a national sample of DSH hospitals. Given the transition to UCC under the Affordable Care Act, we consider both DSH and UCC as two dimensions of safety-net status. Consistent with recent literature, we calculated the average value of the percentage of DSH and the average amount of Medicare UCC across all data years by hospital. Then, we characterize hospitals as safety-net if their DSH or UCC average is in the top quartile of DSH or UCC, respectively, across our national sample. Based on the most recent data for our study time frame, this results in 846 safety-net hospitals and 1,828 non-safety-net hospitals.

Finally, our study also incorporates a number of control variables in the analysis. To control for hospital factors that may affect quality, changes to quality, and resources to invest in change, we also used the Impact File to obtain data on annual hospital characteristics. To obtain hospital integration status, we used the fiscal year 2015 and 2016 Hospital Cost Report Information System. Using these data, we defined hospital integration with post-acute care providers by whether a hospital owned a skilled nursing facility (SNF), inpatient rehabilitation facility (IRF), or home health agency (HHA). To examine hospital characteristics such as teaching affiliation and for-profit status, we supplemented the Impact File and Hospital Cost Report Information System with the American Hospital Association Annual Survey from 2013. Hospitals were defined as having teaching affiliation if they met at least one of the following criteria in the AHA Survey. These criteria were: (1) approval to participate in residency or internship training by the Accreditation Council for Graduate Medical Education, (2) medical school affiliation as reported to the American Medical Association, (3) member of the Council of Teaching Hospitals of the Association of American Medical Colleges, or (4) internship or residency approved by the American Osteopathic Association.

V. Results

The relationship between changes in survival or spending and the financial incentives are decidedly non-linear. Our preferred model has an indicator for any positive incentive and a continuous linear measure of the incentive on change in performance on the measure (see Tables 3 through 6). Having any positive incentive improves measure performance in all three survival measures and the spending measure, at least for non-safety-net hospitals. In addition, a linear increase in the incentive increases the survival rate. Allowing for the non-linear jump for any positive incentive greatly improves the model fit. This was also confirmed with graphs that plotted the year-over-year change in measure against the marginal incentive (not shown). Quadratic terms do not improve the fit. That is, linear models fit about as well as quadratic models with a squared term. Log-linear models on a restricted sample of those with any positive incentive also fit well overall for the survival measures. In summary, non-linear models fit better than the simple linear models from the prior study.

Additionally, there is a significant relationship between hospitals' baseline 2015 performance category for a given measure and subsequent performance improvement on that measure. This relationship is generally monotonic, with hospitals in lower baseline 2015 performance categories improving more on average (see Table 7).

There is a strong serial correlation between the marginal future reimbursement across years, and also the scaled marginal total performance score across years (see Table 8). The correlation for a one-year lag generally ranges from about 0.5 to 0.9. The correlations for two or three years are smaller, but mostly still above 0.2. This indicates a strong serial correlation from year to year in the incentives. This result suggests that measures are not random and driven by outliers.

We also found no evidence that a two-year (or three-year) lag was a consistently stronger predictor than a one-year lag. This suggests that hospitals' responses are relatively quick and not over long timeframes. When we ran regressions with lags of up to three years, the coefficients were not stable across the lags (Table 9). In some results, the single-lag coefficient was most statistically significant, but in other results the second or third lag was strongest. The coefficients were sometime opposite signs, which is consistent with high correlation across covariates. In summary, we conclude that one-year lags is empirically best and captures the duration of hospital responses.

Overall, safety-net hospitals respond in somewhat similar ways as non-safety-net hospitals (see Tables 3, 4, 5, 6, and 9). All types of hospitals seem to respond about the same to having a positive financial incentive in the HVBP program as well as to marginal future reimbursement incentives. However, the magnitude of the marginal incentive coefficient in the scaled marginal TPS models for survival measures is generally larger for safety-net hospitals than for non-safety-net hospitals, suggesting a larger response to scaled marginal TPS incentives among safety-net hospitals on key measures.

Among the two possible marginal incentives (MFR or scaled MTPS), both had similar patterns of sign, magnitude, and significance, overall, although the magnitude of the marginal incentive term was larger in the scaled MTPS models. For three of the four measures (all but CHF), the marginal future reimbursement provided as good or better overall fit, suggesting that MFR more accurately reflects the incentives faced by hospitals. This implies that the amount of money at stake is important, and that larger hospitals, or those with more money at stake, will respond more strongly.

VI. Conclusions

We studied the empirical relationship between the financial incentives inherent in the pay-for-performance HVBP program for hospitals and changes in hospital performance on HVBP measures. In general, hospitals that face larger financial incentives respond more by increasing survival for the three conditions that are measured and by lowering Medicare spending per beneficiary. We also found that this relationship is not linear. There is a large jump at zero when the incentives become positive.

The incentives are also highly serially correlated from year to year, and models with lags do not show a consistent pattern. Therefore, models with a one-year lag seem to capture the time horizon of the responses from hospitals. Safety-net hospitals respond somewhat similarly to non-safety-net hospitals, although the magnitude of the marginal incentive coefficient in the scaled marginal TPS models for survival measures is generally larger for safety-net hospitals than for non-safety-net hospitals. Finally, hospitals respond slightly more strongly to the marginal future reimbursement, which is weighted by amount of Medicare payments, than to scaled marginal total performance score. This suggests that when responding to the program, the total amount of dollars matters for hospitals.

Together, our results indicate that hospitals respond strongly to financial incentives in the HVBP Program, by improving performance on survival and Medicare spending per beneficiary measures, which are vital to policymakers and patients. This response is otherwise linear, except for a sizable jump around zero when the marginal incentives become positive. Yet our findings also indicate that for at least some measures (i.e., acute myocardial infarction), over 30 percent of hospitals face no incentives to improve. Thus, policymakers should consider refining the program to expose more hospitals to non-zero marginal incentives, even if the resulting incentive size is small.

REFERENCES

Banerjee, S., D. McCormick, M.K. Paasche-Orlow, M.Y. Lin, A.D. Hanchate. 2019. "Association between degree of exposure to the Hospital Value Based Purchasing Program and 30-day mortality: experience from the first four years of Medicare's pay-for-performance program." *BMC Health Services Research* 19(1): 921.

Chatterjee, P., K.E. Joynt. 2014. "Do cardiology quality measures actually improve patient outcomes?" *J. Am. Heart Assoc*. 3(1).

Figueroa, J.F., Tsugawa, Y., Zheng, J., Orav, E.J., Jha, A.K. 2016. "Association between the value-based purchasing pay for performance program and patient mortality in US hospitals: observational study." *BMJ* 353, i2214.

Friedson, A.I., W.C. Horrace, A.F. Marier. 2019. "So many hospitals, so little information: How Hospital-Value Based Purchasing is a game of chance." *Southern Economic Journal* 86(2):773–799.

Gilman, M., E.K. Adams, J.M. Hockenberry, I.B. Wilson, A.S. Milstein, E.R. Becker. 2014. "California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs." *Health Affairs* 33(8): 1314–1322.

Gilman, M., J.M. Hockenberry, E.K. Adams, A.S. Milstein, I.B. Wilson, E.R. Becker. 2015. "The financial effect of value-based purchasing and the hospital readmissions reduction program on safety-net hospitals in 2014: a cohort study." *Annals of Internal Medicine* 163(6): 427–436.

Glied, S., A. Sacarny. 2018. "Is the US health care system wasteful and inefficient? A review of the evidence." *Journal of Health Politics Policy and Law* 43(5): 739–765.

Hoffman, G.J., O. Yakusheva. 2020. "Association between financial incentives in Medicare's Hospital Readmission Reduction Program and hospital readmission performance." *JAMA Open Network* 3(4).

Mohan, A., J. Grant, M. Batalden, D. McCormick. 2013. "The health of safety net hospitals following Massachusetts health care reform; changes in volume, revenue, costs, and operating margins from 2006 to 2009." *Int. J. Health Services* 43: 321–335.

Li, J., E.C. Norton. 2019. "Pay-for-Performance and Long-Term Care." In Oxford Research Encyclopedia of Economics and Finance. Oxford University Press.

Norton, E.C., J. Li, A. Das, L.M. Chen. 2018. "Moneyball in Medicare." *Journal of Health Economics* 61:259–273.

Ryan, A.M., Burgess Jr, J.F., Pesko, M.F., Borden, W.B., Dimick, J.B., 2015. "The early effects of Medicare's mandatory hospital pay-for-performance program." *Health Services Research* 50(1):81–97.

Thorton, R. L., 2008."The Demand for, and Impact of, Learning HIV Status." *American Economic Review*. 98(5):1829-1863.

Figure 1.
Theoretical relationship between the change in hospital quality over time and the marginal future reimbursement.

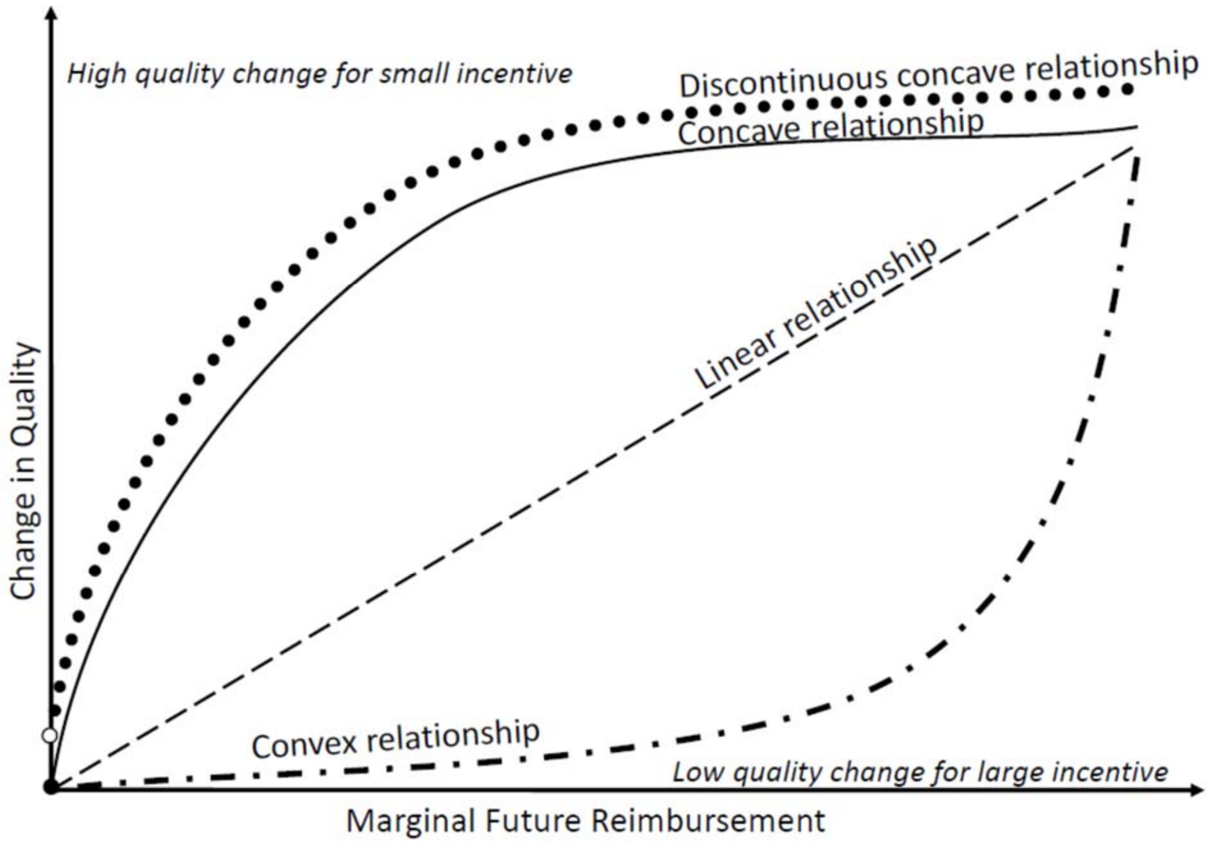

Table 1: Hospital-Year Level Descriptive Statistics (Representing 846 Safety-Net Hospitals and 1,828 Non-Safety-Net Hospitals)

Hospital Characteristics	Safety-Net Hospitals			Non-Safety-Net Hospitals		
	Yes (%)	No (%)		Yes (%)	No (%)	
Hospital Size:						
<100 Beds (%)	19.0	81.0		33.4	66.6	
100-299 Beds (%)	43.2	56.8		50.1	49.9	
300+ Beds (%)	37.8	62.2		16.5	83.5	
Teaching Affiliation	51.5	48.5		35.0	65.0	
For-Profit Hospital	22.3	77.7		19.2	80.9	
Integrated with:						
Skilled Nursing Facility	18.8	81.2		15.6	84.4	
Skilled Nursing Facility, Inpatient Rehabilitation Facility, or Home Health Agency	52.4	47.6		51.6	48.4	
	Mean	**Minimum**	**Maximum**	**Mean**	**Minimum**	**Maximum**
Value-Based Purchasing Adjustment Factor[1]	-0.01	-1.39	3.33	0.17	-1.50	3.86
Medicare Days as a Percent of Total Inpatient Days	29.5	0.4	63.9	41.8	0.0	82.4

1. The adjustment factor reflects the incentive bonus or penalty used to adjust payments to a given hospital under the Hospital Value-Based Purchasing (VBP) Program. This value is calculated as a function of the given hospital's HVBP Total Performance Score. In particular, it reflects (1) the product of the Base Operating DRG Payment Amount Reduction, the quotient of the Total Performance Score divided by 100, and the Exchange Function Slope (2) minus the Base Operating DRG Payment Amount Reduction.

The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures).

Table 2: Hospital-Year Level Descriptive Statistics, by Variable of Interest and Measure (Representing 2,674 Hospitals in Total)

	Safety-Net Hospitals					Non-Safety-Net Hospitals				
	Hospital-Year Sample Size	Mean	Minimum	Maximum	Positive Marginal Incentive (%)	Hospital-Year Sample Size	Mean	Minimum	Maximum	Positive Marginal Incentive (%)
Measure Performance in Base Year										
Acute Myocardial Infarction Survival (0-100%)	1,311	85.95	80.31	89.96	---	3,063	86.08	79.93	89.73	---
Congestive Heart Failure Survival (0-100%)	1,971	88.42	83.32	93.02	---	4,608	88.07	82.37	92.89	---
Pneumonia Survival (0-100%)	2,091	88.70	78.41	92.84	---	4,974	88.59	79.04	92.70	---
Medicare Spending Per Beneficiary (Ratio)[1]	2,535	0.99	0.66	1.33	---	5,484	0.98	0.71	1.33	---
Change in Measure Performance from Base Year to Subsequent Year										
Acute Myocardial Infarction Survival	1,311	0.38	-2.91	3.38	---	3,063	0.35	-3.43	3.61	---
Congestive Heart Failure Survival	1,971	0.09	-4.93	2.98	---	4,608	0.08	-6.08	3.90	---
Pneumonia Survival	2,091	0.23	-4.10	6.11	---	4,974	0.21	-5.48	5.13	---
Medicare Spending Per Beneficiary	2,535	0.11	-16.56	25.75	---	5,484	-0.05	-19.65	33.65	---
Scaled Marginal TPS (for 1 Decile Measure Performance Change)[2]										
Acute Myocardial Infarction Survival	1,311	0.24	0.00	1.69	68.3%	3,063	0.31	0.00	2.09	63.5%
Congestive Heart Failure Survival	1,971	0.36	0.00	2.50	92.1%	4,608	0.50	0.00	4.32	94.1%
Pneumonia Survival	2,091	0.36	0.00	3.00	89.7%	4,974	0.53	0.00	4.27	91.3%
Medicare Spending Per Beneficiary	2,535	1.09	0.00	9.82	97.9%	5,484	1.54	0.00	11.95	98.2%
Marginal Future Reimbursement (for 1 Decile Measure Performance Change) ($)										
Acute Myocardial Infarction Survival	1,311	21,680	0	364,091	68.3%	3,063	15,258	0	313,285	63.5%
Congestive Heart Failure Survival	1,971	22,975	0	332,313	92.1%	4,608	18,353	0	511,045	94.1%
Pneumonia Survival	2,091	18,947	0	579,986	89.7%	4,974	15,323	0	186,194	91.3%
Medicare Spending Per Beneficiary	2,535	53,507	0	760,988	97.9%	5,484	46,828	0	882,163	98.2%

The sample reflects all hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures).

1. This measure reflects the ratio of a hospital's average Medicare Per Beneficiary Spending amount to the median Medicare Per Beneficiary Spending amount across all hospitals.

2. Each hospital's *Scaled* Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Table 3: Change in Acute Myocardial Infarction (AMI) Survival Rate performance (from the base year to the subsequent year) as a function of marginal incentive (from a one-decile improvement in the measure), by whether hospital meets safety-net classification criteria

Change in AMI Survival Rate	Safety-Net Hospitals (N=1,311 hospital-years unless otherwise indicated)					Non-Safety-Net Hospitals (N=3,063 hospital-years unless otherwise indicated)				
	Scaled Marginal Total Performance Score[1]					Scaled Marginal Total Performance Score[1]				
1(Marginal Incentive>0)	0.371*** (0.052)		0.291*** (0.060)	0.319*** (0.092)		0.337*** (0.033)		0.282*** (0.037)	0.323*** (0.058)	
Marginal Incentive		0.776*** (0.146)	0.472*** (0.161)	0.317 (0.403)			0.450*** (0.069)	0.204*** (0.076)	0.041 (0.189)	
Quadratic: Marginal-Squared				0.137 (0.353)					0.107 (0.113)	
Ln(Marginal Incentive) (N=895 Safety-Net; N=1,944 Non-Safety-Net)					0.234*** (0.069)					0.175*** (0.042)
	Marginal Future Reimbursement (Ten-Thousands of Dollars)					Marginal Future Reimbursement (Ten-Thousands of Dollars)				
1(Marginal Incentive>0)	0.371*** (0.052)		0.277*** (0.058)	0.239*** (0.062)		0.337*** (0.033)		0.268*** (0.036)	0.280*** (0.040)	
Marginal Incentive		0.057*** (0.011)	0.042*** (0.011)	0.063*** (0.016)			0.063*** (0.009)	0.039*** (0.009)	0.031** (0.015)	
Quadratic: Marginal-Squared				–0.001 (0.001)					0.001 (0.001)	
Ln(Marginal Incentive) (N=895 Safety-Net; N=1,944 Non-Safety-Net)					0.194*** (0.053)					0.131*** (0.038)

These results report the coefficient and its standard error from regressing measure rate differences from the base year to the subsequent year on marginal incentive (effect of one-decile improvement on scaled TPS or future reimbursement). The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Cluster-robust standard errors in parentheses that adjust for clustering at the hospital level
*** p<0.01, ** p<0.05, * p<0.10

Table 4: Change in Congestive Heart Failure (CHF) Survival Rate performance (from the base year to the subsequent year) as a function of marginal incentive (from a one-decile improvement in the measure), by whether hospital meets safety-net classification criteria

Change in CHF Survival Rate	Safety-Net Hospitals (N=1,971 hospital-years unless otherwise indicated)					Non-Safety-Net Hospitals (N=4,608 hospital-years unless otherwise indicated)				
	Scaled Marginal Total Performance Score[1]					**Scaled Marginal Total Performance Score[1]**				
1(Marginal Incentive>0)	0.057		–0.147	–0.070		0.324***		0.250***	0.226***	
	(0.103)		(0.116)	(0.131)		(0.070)		(0.076)	(0.082)	
Marginal Incentive		0.321***	0.378***	0.084			0.186***	0.119**	0.190*	
		(0.076)	(0.085)	(0.202)			(0.050)	(0.054)	(0.107)	
Quadratic: Marginal-Squared				0.182*					–0.034	
				(0.108)					(0.042)	
Ln(Marginal Incentive) (N=1,816 Safety-Net; N=4,338 Non-Safety-Net)					0.139***					0.077**
					(0.044)					(0.034)
	Marginal Future Reimbursement (Ten-Thousands of Dollars)					**Marginal Future Reimbursement (Ten-Thousands of Dollars)**				
1(Marginal Incentive>0)	0.057		0.051	0.029		0.324***		0.330***	0.347***	
	(0.103)		(0.111)	(0.118)		(0.070)		(0.074)	(0.077)	
Marginal Incentive		0.003	0.002	0.012			0.008	–0.002	–0.012	
		(0.006)	(0.007)	(0.016)			(0.006)	(0.006)	(0.010)	
Quadratic: Marginal-Squared				–0.001					0.000*	
				(0.001)					(0.000)	
Ln(Marginal Incentive) (N=1,816 Safety-Net; N=4,338 Non-Safety-Net)					–0.027					–0.077***
					(0.033)					(0.025)

These results report the coefficient and its standard error from regressing measure rate differences from the base year to the subsequent year on marginal incentive (effect of one-decile improvement on scaled TPS or future reimbursement). The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Cluster-robust standard errors in parentheses that adjust for clustering at the hospital level
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 5: Change in Pneumonia (PN) Survival Rate performance (from the base year to the subsequent year) as a function of marginal incentive (from a one-decile improvement in the measure), by whether hospital meets safety-net classification criteria

Change in PN Survival Rate	Safety-Net Hospitals (N=2,091 hospital-years unless otherwise indicated)					Non-Safety-Net Hospitals (N=4,974 hospital-years unless otherwise indicated)				
	Scaled Marginal Total Performance Score[1]					**Scaled Marginal Total Performance Score[1]**				
1(Marginal Incentive>0)	0.287***		0.273***	0.027		0.175***		0.080	–0.005	
	(0.074)		(0.091)	(0.099)		(0.060)		(0.068)	(0.075)	
Marginal Incentive		0.175	0.037	0.967***			0.221***	0.196***	0.431***	
		(0.112)	(0.126)	(0.248)			(0.051)	(0.057)	(0.118)	
Quadratic: Marginal-Squared				–0.514***					–0.097**	
				(0.151)					(0.045)	
Ln(Marginal Incentive) (N=1,876 Safety-Net; N=4,539 Non-Safety-Net)					0.171***					0.159***
					(0.055)					(0.039)
	Marginal Future Reimbursement (Ten-Thousands of Dollars)					**Marginal Future Reimbursement (Ten-Thousands of Dollars)**				
1(Marginal Incentive>0)	0.287***		0.221***	0.182**		0.175***		0.003	–0.048	
	(0.074)		(0.080)	(0.084)		(0.060)		(0.069)	(0.075)	
Marginal Incentive		0.027***	0.018**	0.039***			0.078***	0.078***	0.124***	
		(0.008)	(0.008)	(0.014)			(0.009)	(0.011)	(0.020)	
Quadratic: Marginal-Squared				–0.001**					–0.004***	
				(0.000)					(0.001)	
Ln(Marginal Incentive) (N=1,876 Safety-Net; N=4,539 Non-Safety-Net)					0.185***					0.231***
					(0.043)					(0.029)

These results report the coefficient and its standard error from regressing measure rate differences from the base year to the subsequent year on marginal incentive (effect of one-decile improvement on scaled TPS or future reimbursement). The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Cluster-robust standard errors in parentheses that adjust for clustering at the hospital level
*** p<0.01, ** p<0.05, * p<0.10

Table 6: Change in Medicare Spending Per Beneficiary (MSPB) performance (from the base year to the subsequent year) as a function of marginal incentive (from a one-decile improvement in the measure), by whether hospital meets safety-net classification criteria

Change in MSPB Performance	Safety-Net Hospitals (N=2,535 hospital-years unless otherwise indicated)					Non-Safety-Net Hospitals (N=5,484 hospital-years unless otherwise indicated)				
	Scaled Marginal Total Performance Score[1]					**Scaled Marginal Total Performance Score[1]**				
1(Marginal Incentive>0)	–1.628* (0.842)		–1.728* (0.883)	–1.673* (0.917)		–1.689*** (0.616)		–1.933*** (0.627)	–1.919*** (0.645)	
Marginal Incentive		–0.019 (0.137)	0.053 (0.147)	–0.007 (0.323)			0.084 (0.056)	0.135** (0.058)	0.120 (0.136)	
Quadratic: Marginal-Squared				0.010 (0.053)					0.002 (0.018)	
Ln(Marginal Incentive) (N=2,483 Safety-Net; N=5,383 Non-Safety-Net)					0.211 (0.194)					0.108 (0.108)
	Marginal Future Reimbursement (Ten-Thousands of Dollars)					**Marginal Future Reimbursement (Ten-Thousands of Dollars)**				
1(Marginal Incentive>0)	–1.628* (0.842)		–1.666** (0.841)	–1.639* (0.844)		–1.689*** (0.616)		–1.680*** (0.618)	–1.649*** (0.619)	
Marginal Incentive		0.015 (0.010)	0.017* (0.010)	0.002 (0.021)			–0.007 (0.008)	–0.004 (0.008)	–0.022 (0.014)	
Quadratic: Marginal-Squared				0.000 (0.000)					0.000* (0.000)	
Ln(Marginal Incentive) (N=2,483 Safety-Net; N=5,383 Non-Safety-Net)					0.016 (0.131)					–0.040 (0.081)

These results report the coefficient and its standard error from regressing measure rate differences from the base year to the subsequent year on marginal incentive (effect of one-decile improvement on scaled TPS or future reimbursement). The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (n=8,022 observations across the three base years across all measures). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Cluster-robust standard errors in parentheses that adjust for clustering at the hospital level
*** p<0.01, ** p<0.05, * p<0.10

Table 7: Change in measure performance (from the base year to the subsequent year) as a function of baseline performance in 2015 group, by measure and whether hospital meets safety-net classification criteria

Change in Measure Performance	Safety-Net Hospitals				Non-Safety-Net Hospitals			
(Baseline Group: Group 1)	Change in AMI Performance	Change in CHF Performance	Change in PN Performance	Change in MSPB Performance	Change in AMI Performance	Change in CHF Performance	Change in PN Performance	Change in MSPB Performance
Group 2	–0.270***	0.022	–0.331***	–0.218	–0.304***	–0.309***	–0.384***	–0.440***
	(0.068)	(0.062)	(0.060)	(0.191)	(0.043)	(0.044)	(0.043)	(0.109)
Group 3	–0.521***	–0.146**	–0.552***	–0.350*	–0.516***	–0.429***	–0.589***	–0.443***
	(0.066)	(0.059)	(0.055)	(0.203)	(0.043)	(0.042)	(0.040)	(0.120)
Group 4	–0.601***	–0.289***	–0.682***	–0.499**	–0.666***	–0.569***	–0.767***	–0.739***
	(0.066)	(0.063)	(0.057)	(0.216)	(0.043)	(0.042)	(0.041)	(0.116)
Group 5	–0.772***	–0.330***	–0.852***	–0.946***	–0.838***	–0.739***	–0.986***	–1.147***
	(0.066)	(0.063)	(0.056)	(0.219)	(0.044)	(0.042)	(0.041)	(0.129)

These results report the coefficient and its standard error from regressing measure rate differences from the base year to the subsequent year on the 2015 baseline performance category. The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for the given survival or spending measure (n=8,022 observations across the three base years across all measures). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

Cluster-robust standard errors in parentheses that adjust for clustering at the hospital level

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 8: Scaled Marginal Total Performance Score and Marginal Future Reimbursement Correlations, FYs 2015-2018

		Safety-Net Hospitals						Non-Safety-Net Hospitals					
		Scaled Marginal Total Performance Score[1]			Marginal Future Reimbursement			Scaled Marginal Total Performance Score[1]			Marginal Future Reimbursement		
Measure	**Year**	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**
AMI Survival Rate	2015												
	2016	0.66			0.83			0.60			0.67		
	2017	0.17	0.43		0.40	0.52		0.16	0.43		0.31	0.47	
	2018	0.19	0.45	0.70	0.51	0.60	0.78	0.16	0.38	0.67	0.27	0.40	0.70
CHF Survival Rate	2015												
	2016	0.18			0.57			0.30			0.58		
	2017	0.05	0.23		0.53	0.64		0.11	0.23		0.58	0.51	
	2018	0.01	0.21	0.44	0.55	0.62	0.78	0.11	0.20	0.48	0.65	0.53	0.71
PN Survival Rate	2015												
	2016	0.64			0.56			0.60			0.66		
	2017	0.27	0.41		0.37	0.50		0.31	0.45		0.53	0.65	
	2018	0.25	0.42	0.49	0.43	0.65	0.61	0.28	0.46	0.58	0.53	0.65	0.72
Medicare Spending per Beneficiary Ratio	2015												
	2016	0.54			0.91			0.58			0.89		
	2017	0.43	0.62		0.86	0.90		0.55	0.62		0.86	0.89	
	2018	0.40	0.53	0.58	0.83	0.88	0.90	0.45	0.52	0.62	0.83	0.84	0.91

The sample reflects all hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (2,674 hospitals).

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Table 9: 2017-2018 measure change as a function of marginal incentives over time

	Safety-Net Hospitals						Non-Safety-Net Hospitals					
	Scaled Marginal Total Performance Score[1]			Marginal Future Reimbursement (Ten-Thousands of Dollars)			Scaled Marginal Total Performance Score[1]			Marginal Future Reimbursement (Ten-Thousands of Dollars)		
Measure	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**	**2015**	**2016**	**2017**
AMI Survival Rate	–0.140	–0.200	0.506**	–0.028	–0.005	0.031***	–0.204**	–0.017	0.161*	–0.040**	0.009	0.031***
	(0.168)	(0.167)	(0.218)	(0.024)	(0.016)	(0.012)	(0.086)	(0.087)	(0.097)	(0.019)	(0.018)	(0.010)
CHF Survival Rate	0.224**	–0.327*	0.033	0.039*	–0.020	–0.002	0.144**	–0.221**	0.071	0.013	–0.006	0.007
	(0.099)	(0.193)	(0.116)	(0.021)	(0.022)	(0.010)	(0.065)	(0.089)	(0.054)	(0.018)	(0.016)	(0.007)
PN Survival Rate	–0.144	–0.119	0.063	–0.009	0.003	0.003	–0.027	–0.138**	0.051	–0.014	0.004	0.024**
	(0.108)	(0.125)	(0.124)	(0.018)	(0.021)	(0.008)	(0.069)	(0.065)	(0.055)	(0.015)	(0.018)	(0.010)
Medicare Spending per Beneficiary Ratio	–0.179	0.005	–0.104	–0.038	–0.029	0.030	0.004	–0.135	–0.009	0.020	0.016	–0.035**
	(0.279)	(0.239)	(0.223)	(0.051)	(0.033)	(0.022)	(0.139)	(0.105)	(0.093)	(0.039)	(0.021)	(0.015)

These results report the coefficient and its standard error from regressing 2017-2018 measure rate differences on the given marginal incentive (effect of one-decile improvement on TPS or future reimbursement) over time. The sample reflects hospitals that met the criteria to receive a measure score in all 4 years for given survival or spending measure (2,674 hospitals). Patient experience, clinical process, and mortality measures are reported as percentages, and higher scores indicate better performance. Lower values in Patient Safety Composite Score, CLABSI Score, Elective Delivery Prior to 39 Completed Weeks Gestation, Healthcare-Associated Infections, and Medicare-Spending-per-Beneficiary Ratio indicate better quality.

1. Each hospital's Scaled Marginal TPS value reflects its Marginal TPS value multiplied by its Medicare days as a percentage of inpatient days.

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10