NBER WORKING PAPER SERIES

DO GOOD REPORTS MEAN HIGHER PRICES? THE IMPACT OF HOSPITAL COMPARE RATINGS ON CARDIAC PRICING

Avi Dor
William Encinosa
Kathleen Carey

Working Paper 22858
http://www.nber.org/papers/w22858

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2016

We thank conference participants at the Biennial American Society of Health Economists Conference, the International Industrial Organization Conference, and the Southern Economics Annual Meetings for comments received on earlier drafts. Jason Hockenberry, Eric Luo, and Sandra Decker also provided invaluable comments. This work was funded by the Agency of Healthcare Research and Quality (AHRQ). The views expressed herein are those of the authors and do not necessarily represent the views or policies of AHRQ, DHHS, or the National Bureau of Economic Research.

Do Good Reports Mean Higher Prices? The Impact of Hospital Compare Ratings on Cardiac Pricing
Avi Dor, William Encinosa, and Kathleen Carey
NBER Working Paper No. 22858
November 2016
JEL No. I11,L11

ABSTRACT

Previous research found that the initiation of Hospital Compare (HC) quality reporting had little impact on patient outcomes. However little is known about its impact on hospital prices, which may be significant since insurers are positioned to respond to quality information when engaging hospitals in price negotiations. To explore this issue we estimate variants of difference-in-difference models allowing HC impacts to vary by levels of quality scores. We separately examine the effects of the three main scores (heart attack, heart failure, and combined mortalities) on transaction prices of two related cardiac procedures: bypass surgery and angioplasty. States which had mandated reporting systems preceding HC form the control group. Analyzing claims data of privately insured patients, we find that HC exerted downward pressure on prices, which we attribute to competitive pressures. However, hospitals ranked "above average" captured higher prices, thereby offsetting the overall policy effect. We conclude that HC was effective at constraining prices without penalizing high performers.

Avi Dor
Department of Health Policy and Management
George Washington University
950 New Hampshire Ave NW, Suite 600
Washington, DC 20052
and NBER
avidor@gwu.edu

Kathleen Carey
Department of Health Policy and Management
Boston University School of Public Health
715 Albany Street, Talbot Building
Boston, MA 02118
kcarey@bu.edu

William Encinosa
Center for Delivery, Organization and Markets
Agency for Healthcare Research and Quality
5600 Fishers Lane
Rockville, MD 20857
and Georgetown University
william.encinosa@ahrq.hhs.gov

I. Introduction

Consumer information in the form of hospital rating systems known as 'reports cards' have been in existence in various forms as early as the 1990s. However, such report cards were made available in a small number of states that generated them independently from each other, while employing a mix of measures and ranking methodologies. In 2005, the Centers for Medicare and Medicaid Services (CMS) launched a uniform on-line national rating system, known as Hospital Compare (HC), aimed at informing consumers and promoting competition among hospitals. Initially, HC consisted of a set of process measures of hospital performance based on general practices.[1] In 2008, these were augmented with outcome-based measures, namely mortality-based hospital rankings which were deemed to be more easily understood by patients, and more effective at motivating hospitals to engage in quality improvement practices (Harris, 2007).

In practice, opinion surveys have shown that consumers were generally unaware or uninterested in these rankings, even as they had become more accessible, and there is little evidence in the empirical literature to suggest that hospital report cards had an impact on consumer choices of hospitals. Although anecdotal evidence suggests that hospital administrators and executives were ostensibly responsive to these rankings, there is similarly little evidence that HC and other report cards have had a significant impact on patient outcomes. We briefly review the literature in the next section.

[1] The set of 26 process measures range from providing aspirin to arriving patients with heart attacks and ACE inhibitors to hospitalized heart failure patients, to discontinuing antibiotic treatment after surgery to avoid resistance; the full listing of instruments is available in https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/HospitalHQA2004_2007200512.pdf

A less explored channel through which public reporting might impact health care markets, and ultimately patients, is hospital pricing. In the private segment of the health care market, private insurers rather than individual patients face the main part of the price. As group purchasers, large insurance firms and managed care organizations who engage in price negotiations with hospitals are more likely to incorporate information from public reporting in their decision-making (Dor, Encinosa, Carey, 2015; Reinhardt, 2006, 2009). However, there is a paucity of evidence on the impact of HC on hospital pricing. An earlier study suggested the initiation of HC reporting contributed to a slowing in the rate of price increases of related hospital procedures (Dor et al., 2015), but it did not address the issue of the responsiveness of prices to rating differences as flagged by Hospital Compare. Yet in a well-functioning market, a hospital's ability to deliver better health outcomes should lead to greater demand and bargaining power for the hospital, and hence be rewarded by higher prices (Brooks et al., 1997). In this paper, we fill the gap by examining the relationship between publicly reported hospital ratings and inpatient prices, allowing for hospital differentiation by relative rankings. Specifically, we conduct an empirical investigation of the impact of the HC measures, as implemented in 2008, on actual prices negotiated and transacted between private insurers and U.S. hospitals. These measures are the HC categorical rankings based on hospital risk-adjusted mortality rates for the three medical conditions made available on-line as of 2008, which included heart attack, heart failure, and pneumonia. We focus on prices of major cardiac procedures related to heart attacks and heart failure. Results suggest that HC exerts downward pressure on prices, but that this effect is offset for hospitals ranked in the highest quality category.

We note that in public discourse there is considerable confusion between hospital billing data, such as those CMS began to release in 2013, and actual payments made to hospitals, namely transaction prices (Meier et al., 2013). Using a claims database for large private insurers we focus our analysis on transaction prices. The rest of the paper proceeds as follows. Section II provides background on hospital report cards and hospital pricing. Section III lays out the analytical framework, including data and sample, estimation strategy, and variable specification. Section IV presents descriptive results for the comparisons of the pre and post Hospital Compare initiation year in control versus treatment states, and the program effects of the intervention. Section V discusses implications. The full regression models are provided in Appendix A.

II. Background

Public Reporting of Hospital Performance

Information about hospital quality performance began appearing in the public domain in the 1990s. An early review of the gains from public disclosure of performance data found that consumers showed little interest in or use of the available information due to various reasons including difficulty in understanding, failure to trust in, and lack of timely access to the data (Marshall et al., 2000). Anecdotal surveys of hospital administrators and executives are more positive, with most reporting increased investment in quality improvement in response to the newly created report cards (Laschober et al., 2006). More recent study of public reporting suggests that information included in hospital report cards may be disconnected from consumer decision-making because of weakness in content, design, and accessibility even as information became more widely disseminated (Sinaiko et al., 2012). Such policy discourse even prompted the California Hospital Association to review and withdraw support for state report cards due to perceived lack of value to users (Teleki and Shannon, 2012).

In some respects, the empirical literature tends to lend support to the view that hospital report cards have limited effect on patient choices and medical outcomes. For instance, Dranove and Sfekas (2008) found that the hospital report cards in New York State had little impact on choices of hospitals by consumers and on hospital market share and Wang et al. (2011) found no significant effects of mortality-based scores in Pennsylvania report cards on hospital volume. Turning to outcomes, Ryan et al. (2012) found that the requirement for hospitals to report process of care measures under the earlier wave of Hospital Compare did not lead to reductions in 30-day mortality rates for heart attack and pneumonia, and had only minimal impact on heart failure mortality. Moreover, choosing reportedly high performing hospitals in New York State did not decrease a patient's chance of dying following coronary intervention (Chen et al., 2012). More recently, Dor, Encinosa, and Carey (2015) began to explore a different channel through which report cards might affect providers and patients, namely by mitigating hospital prices. They found that the introduction of excess mortality measures in Hospital Compare slowed the rate of increase in prices of related cardiac procedures overall. They speculated that the mere injection of quality information, albeit imperfect, into the healthcare market weakens the competitive position of hospitals in hospital-insurer negotiations.[2] However, they did not differentiate between hospitals by their relative rankings, and thus were unable to establish whether better ranked hospitals were able to capture a "rating premium" or if they were penalized by pricing pressures to the same extent as lower-ranked hospitals.

Studies of Cardiac Procedure Pricing

[2] Silber et al [2010] demonstrate that hospital mortality scores are sensitive to model specification. For instance they show that adding volume and hospital staffing to bed ratios that are omitted in the Hospital Compare methodology improved predictability. Note however that from the perspective of evaluating Hospital Compare we are interest in the impact of publically available report cards rather than constructs available to researchers only.

Previous studies examined pricing of cardiac services including coronary artery bypass surgery and angioplasty. Dor, Grossman, and Koroukian (2004) used a bargaining model to derive price models that reflect the dynamics of the hospital-insurer interaction. Results indicated that health maintenance organizations (HMOs) are able to capture larger price discounts from hospitals than more open forms of managed care such as point-of-service (POS) plans and preferred provider organizations (PPOs), a result the authors attributed to the HMO's ability to exercise greater bargaining power due to large-volume purchasing. In a related paper, Dor, Koroukian and Grossman (2004) explored the impact of standardized mortality ratios (SMRs) for coronary artery bypass surgery and for angioplasty on their corresponding prices. As expected, adverse quality had a negative impact on prices; however, this result was not statistically significant. While these scores reflected the type of report card information available in certain states during the period studied (1994-1996), and potentially to some private insurers, they were not widely disseminated; this contrasts with the risk-standardized SMRs (RSMR) that were used in subsequent years to construct the categorical quality ratings reported in Hospital Compare. Broader dissemination of the information in recent years may yield more significant effects of adverse quality on prices than found in the earlier period.

III. Analytical Framework

Data and Study Population

The main patient and price data consist of the Truven Analytics MarketScan Commercial Claims and Encounters database (CCE). This database assembles complete insurance claims for approximately 100 medium-size and large employers previously used in nationally representative population-based studies (Zhou et al., 2005; Hansen and Chang, 2011). The advantage of using the CCE database is that it reports actual transaction prices paid by patients and insurers. Prices are adjusted for local differences by the wage index from the CMS Cost Reports, then inflation adjusted to 2010 dollars. Hospital Compare (HC) ratings come from the CMS Hospital Compare database. In the CMS methodology, they are derived from post discharge risk-standardized 30-day mortality, but are displayed as a categorical score that take one of three values ("Better than," "No Different than," and "Worse than" the U.S. national rate). Other hospital characteristics were obtained from the American Hospital Association Annual Surveys. The full census of inpatients from the Healthcare Cost and Utilization Project was used to create market concentration ratios for hospitalized patients undergoing cardiac procedures. Finally, the Managed Market Surveyor File from InterStudy provided the market area HMO penetration rates.

We extracted claims for hospitalizations for non-elderly employees and dependents undergoing coronary revascularization, comprising coronary artery bypass surgery (CABG) and percutaneous coronary intervention (PCI), formerly referred to as PTCA (angioplasty). In addition to being associated with cardiac conditions featured in HC, CABG and PCI are among the most common major medical procedures in the U.S. healthcare system, with over 1 million procedures performed annually (DeFrances et al., 2008). They also are among the most costly, accounting together for over $3.5 billion in 2007 (CMS, 2010), an amount larger than for any other medical or surgical procedure except for hip and knee replacement (Epstein et al., 2011). CABG and PCI occur relatively frequently compared with other cardiac procedures, and tend to be well defined in claims data for purposes of price estimation.

Our database observes a large proportion of all CABGs and PCI procedures covered by private insurers in 1,288 and 706 hospitals respectively. We merged the above data files for the years 2005-2010. This allowed us to conduct analysis on impacts of the quality measures after their introduction in 2008, as well as a comparison before and after the 2005-2007 phase-in period. After excluding small hospitals with less than 10 procedures and patients who underwent non-cardiac procedures in the same hospitalizations, our sample consisted of 53,765 observations on CABG patients and 24,441 observations on PCI patients. The distribution of patients undergoing CABG and PCI according to quality rankings are displayed in Table 1.

Seven states had reporting systems for hospital quality metrics based on mortality rates following revascularization procedures: California, Florida, Massachusetts, New Jersey, New York, Pennsylvania, and Washington. We included only six states in this group because hospitals in the state of Washington participated in the reporting system on a voluntary basis. We characterize all other states as the intent-to-treat group, since none had report cards of their own prior to the initiation of the federal Hospital Compare. Appendix Table A.1 displays the years in which reporting systems were in effect in each of the control states.

Estimation Strategy

We initially define a simple difference-in-differences (DD) estimator for the treatment effect of Hospital Compare (HC) from the perspective of states without state reporting systems (intent-to-treat states). The change in hospital price due to the initiation of Hospital Compare in all states can be described as

$$\Delta P_{HC} = [E(P_{NR}|\ \textit{post HC treatment}) - E(P_{NR}|\ \textit{pre HC treatment})] - [E(P_{R}|\ \textit{post HC treatment}) - E(P_{R}|\ \textit{pre HC treatment})]$$

Where *NR* is an indicator for intent-to-treat states having no state report card systems prior to HC initiation. At the individual admission level we have

$$P_{ihjt} = a_0 + a_1 HC_{t-1} + a_2 NR_{ijt} + a_3 HC_{t-1} * NR_{ijt} \quad (1)$$

P is price for the individual hospital admission, where i indexes the patient admission, h indexes the hospital, j indexes state, and t indexes year, HC is an indicator for the post implementation period. The effect of HC on the intent-to-treat group of states is given by a_3.[3]

Adjusting for characteristics of the admission in question we obtain the estimating equation

$$P_{ihjt} = a_0 + a_1HC_{t-1} + a_2NR_{ijt} + a_3HC_{t-1}*NR_{ijt} + a_4Z_{ijt} + f_h + f_t + e_{ihjt} \qquad (2)$$

Where in addition to the variables previously defined, Z is a vector of medical characteristics of the admission and insurance type (as described in Appendix Table A.3), f_h and f_t are binary indicators for hospital and year fixed effects, and e_{ihjt} is a random error term. HC enters the model with a one year lag because prices in insurance contracts are fixed in a given year t and can adjust only at the next annual update. Note that while a_3 in equation (1) and (2) can be taken as the treatment effect on price levels, it does not account for differences in reported quality (mortality) scores as flagged in Hospital Compare.

Next, we expand equation (2) to allow the impact of the policy to vary by the intensity of the treatment. More specifically, we will allow the effect of HC to depend on the reported hospital mortality score, Q_h. Noting that Q_h applies to the intent-to-treat states after, but not prior to the initiation of HC, we obtain the following estimating equation:

$$P_{ihjt} = a_0 + a_1HC_{t-1} + a_2NR_{ijt} + a_3HC_{t-1}*NR_{ijt} + a_4HC_{t-1}*Q_{h,t-1} + a_5HC_{t-1}*NR_{ijt}*Q_{h,t-1} + a_6Z_{ijt} + f_h + f_t + e_{ihjt} \qquad (3)$$

[3] This is the familiar case whereby $\Delta = [(a_1 + a_2 + a_3) - a_2] - [a_1 - 0] = a_3$.

Here the second level interactions control for post-HC trend in the intent-to-treat group independently from the mortality scores (a_3), and for the national trend in the HC mortality scores (a_4). The third level interaction gives the incremental program effect (a_5), namely the effect of the HC mortality scores (lagged a year) on prices after the initiation of HC, in intent-to-treat states. The lag in Q accounts for last year's published quality scores in the current year's hospital-insure contract. This estimation strategy is similar to using a continuous variable to define varying policy impacts within the distribution of observations in the treatment group, as found in Finkelstein (2007) [4], and Chou et al., 2014. Although model 3 appears notationally analogous to the familiar difference-in-difference-in-differences (DDD), here the main program effect is given by the same bi-level interaction term as in the basic DD model, rather than the triple interaction term.

To see how the full policy effect is derived, we note the expected value analogue of equation (3):

$$\Delta P_{HC(Q)} = [E(P_{NR}(Q_{NR})|\ \textit{post HC treatment}) - E(P_{NR}|\ \textit{pre HC treatment})] - [E(P_R(Q_R)|\ \textit{post HC treatment}) - E(P_R(Q_R)|\ \textit{pre HC treatment})]$$

Substituting in the parameters of (3) we have

$$\Delta P_{HC(Q)} = [(a_1 + a_2 + a_3 + a_4{*}E(Q|NR=1) + a_5\, E(Q|NR=1)) - a_2] - [(a_1 + a_4{*}E(Q|NR=0)) - 0] =$$

$$a_3 + a_4{*}(E(Q|NR=1) - E(Q|NR=0)) + a_5\, E(Q|NR=1) \qquad (4)$$

Henceforth we refer to eq. (4) as the full DD model and (2) as the naïve DD model. Note that a_3 from equations (3) and (4) are equal if either $a_4=0$ or $[E(Q|NR=1)-E(Q|NR=0)] =0$, and if $a_5=0$.

[4] Finkelstein employed a continuous variable such as the local area senior population share, interacted with a post-treatment dummy to analyze interrupted trends due to the introduction of Medicare. Chou et al use density of obstetrics hospitals as the impact variable in a DD analysis of expansions in maternity benefits in Taiwan.

Thus the implicit premium paid to hospitals for being highly ranked is decomposed into two effects, a direct effect due to scoring levels in the treated states (a_5) and a relative effect when the group is compared with controls.[5]

Note that differencing requires P_{ihjt} to be specified linearly. We estimate the linear models with hospital fixed effects using the generalized method of moments (GMM). This method provides heteroskedasticity-robust standard errors, and is more efficient than robust OLS (Hansen, 2016). Additionally GMM has been shown to be appropriate in panel data estimation when the number of states and individual observations is large, while the number of periods is small, (Hausman and Kuersteiner, 2008). These conditions are met in our data. [6] We ran four models on the CABG and on the PCI samples, including the naive DD model and three full DD models corresponding to the flagged quality scores for heart attacks, heart failure, and for the combined scores. The baseline DD estimates are reported in Appendix Table A.3 and the full DD models are reported in Appendix Table A.4.

There are several reasons to assume that the HC scores are orthogonal to the error term in equations 1 through 4. First, the scores are hospital level variables, whereas prices are measured at the patient level. Second, the HC scores are calculated from a complex transformation of underlying mortality.[7] Third, although the published HC scores are the reference measure in

[5] Empirically the term in the middle disappears, both because a_4 is non-significant and because the Q's for intervention and control states are similarly distributed. In practice we estimate the policy effect as $a_3 + a_5E(Q|NR=1)$. See Table 4.

[6] Hausman and Kuersteiner (2008) define a ''large'' number of states at 50. Our data contain 49 states plus the District of Columbia, and a large sample of individuals, i.e N= 53,765 and N= 24,441 in the CABG and PCI samples.

[7] The Hospital Compare categories are drawn from a highly transformed excess mortality variable. To summarize, prior to being grouped into categories, the underlying mortality scores were generated from the risk-standardized mortality ratio (RSMR) defined as the ratio of predicted to expected mortality rates in the hospital. The denominator adjusts for patient characteristics x only; in the numerator the predictive mortality model incorporates a

public reporting, they are based on the census of Medicare beneficiaries, while pricing is observed for the privately insured.[8] To be sure we also conducted tests for endogeneity using a predictive model for the underlying scores (observable to researchers). As expected we found strong evidence for rejecting endogeneity.[9]

The results from the familiar "naïve" models are used simply as a basis for comparison. These models simply examine whether changes in the states that had report cards of their own similar to Hospital Compare prior to 2008 exhibited less price sensitivity to HC rankings compared with the intent-to-treat states, independently of any tradeoff between price and rating. Finally, we also create a counterfactual to the full model, to test for the validity of the various quasi-experiment designs above. The counterfactual is based on rerunning the models using non-cardiac procedures that should be weakly susceptible to the information from Hospital Compare with respect to pricing. Following Ryan, Nallamothu, and Dimick (2012) and Carey (2015), who employ gastrointestinal diagnoses as the counterfactual for AMI when evaluating the mortality consequences of HC initiation, we first employ surgeries for gastrointestinal cancers as the

---

random hospital-specific effect (α) that accounts for within-hospital correlations of the observed patient outcomes. Accordingly, the excess mortality score is redefined as the ratio of predicted to expected mortality rate, with a random effects term set equal to zero:

$$RS\widehat{M}R_h = \frac{E(y_h|x_h;\beta,\alpha_h)}{E(y_h|x_h;\beta,\alpha_h\!=\!0)}$$

Additionally, the underlying scores were constructed as 3-year moving averages with data drawn from Medicare claims and administrative data. Note that Silber et al showed that hospital mortality scores are sensitive to model specification. However, our interest is in the scoring as publically announced.

[8] For Medicare patients price variation is minimal, since Medicare prices for all inpatient services are administered and essentially set constant for general medical service and diagnosis categories, under the Inpatient Prospective Payment System.

[9] To test for such endogeneity, we performed the Durbin–Wu–Hausman test on the continuous mortality scores for pneumonia, heart attack, and heart failure. The tests consist of saving the residuals from the first stage predictive models as covariates in the pricing models. Hospital staffing variables such as nurses per bed, other full time employees per bed, and log of beds as instrumental variables in the first stage estimates of mortality, we find that for all three mortality outcomes in both CABG and PCI samples, the residuals from the first stage mortality regressions were never statistically significant ($p>0.10$) in the second stage price regressions. Thus, we find no evidence of the mortality scores being endogenous in our price regressions.

comparison to CABG and PCI. A primary example is surgery for colorectal cancer (colectomy). ICD-9 and CPT codes needed to define colectomy related admissions are found in a previous related study (Dor et al., 2012). Secondly, we examine the prices of gastrointestinal hemorrhage cases, since this was one of the few principal diagnoses used in the AHRQ mortality indicators but not adopted in Hospital Compare (but, later adopted by CMS in the HAC Reduction Program; See AHRQ, 2015).[10] We present the GMM estimates for colectomy and gastrointestinal hemorrhage in Table 3. As expected, the HC public reporting does not impact prices for these procedures.

An additional sensitivity analysis that allows for balancing samples in the control and intervention state using propensity scores matching is provided in Appendix B. We find that the results from the matched samples are similar to the results from the full sample analysis. Given the matching estimates are less efficient, we focus on the latter.

Variable Specification

In its on-line rankings Hospital Compare reports separate mortality-based quality scores for heart attacks, heart failure, and pneumonia, using the general categories labeled as above average, at the national average, and below average in each. These rankings are based on 95 percent confidence intervals for the deviation of actual relative to expected mortality rates, using risk adjustment models described in Krumholz et al, 2006a, and 2006b. As a consequence few hospitals were actually ranked below average, while the vast majority of hospitals were classified as average (Table 1; also see Silber et al., 2010). To capture the incremental effect of higher

[10] Beginning in 2015, the Hospital-Acquired Condition (HAC) Reduction Program, mandated by the Affordable Care Act, requires the Centers for Medicare & Medicaid (CMS) to reduce hospital payments by 1 percent for hospitals that rank among the lowest-performing 25 percent with regard to HACs.

quality in the pricing models we created binary indicators for the grouping of hospitals ranked above average versus the grouping of all other hospitals, ranked average or less. We estimated the impact of this variable separately for each of the two Hospital Compare conditions that pertain to cardiac (heart attacks, chronic heart failure), and for a "combined" indicator which was flagged if the hospital received above average rankings in *all* of the HC condition-specific measures (heart attacks, chronic heart failure, pneumonia) during the year. We further estimated separate models for CABG and PCI patients (summarized in Table 3).[11] All models controlled for patient demographics, clinical case complexity, HMO penetration rate, and hospital market concentration (see Appendix Tables A.2-A.3 for the full specification).

IV. Results

Tables 1 and 2 show the distribution of the raw prices and the Hospital Compare quality levels in our data. Table 1 shows that prices for hospitals in the average and below average categories were nearly equal, especially in the case of PCI, while price corresponding with the above average category were substantially higher. Thus, prices generally increased as the HC quality score improved. Table 2 adds the dimension of the comparison between the intent-to-treat states (states with no cardiac report cards prior to HC) and the control group (states with reports prior HC reporting). We observe that, overall, the treatment states experienced smaller increases in CABG and PCI prices around the initiation of HC than the control states. This suggests that simply introducing public reporting had a pro-competitive effect, exerting downward pressure on

[11] In trial regressions we also ran specifications where the below average categories entered the equation separately and results were essentially the same with respect to quality. For example, in a model based on the combined score of all HC conditions which included binary indicators for both the above average and below average categories, the premium for above average in CABG ($5,408 in Table 3) becomes $5,278 (p=0.099); there appears to be a below average "penalty", but it is not statistically significant (-$5,468, p=.39). Similarly, in the equivalent model for PCI, the premium for above average is $965 (p=0.145) and the below average "penalty" is -$1,338 (p=.39).

hospital prices overall. However within the treated group price increases after the implementation of Hospital Compare were substantially higher for hospitals in the above-average category, suggesting that higher rankings would have been rewarded in the marketplace. While these results are unadjusted and descriptive, they provide additional motivation for our hypotheses about the impact of Hospital Compare. We proceed with full analysis below.

Table 3 summarizes the main GMM estimates for the policy variables for CABG and PCI. The bi-level interaction terms for HC and no-report state represents the main program effect in all models. The related coefficients were highly significant in all models, with price reductions ranging from 8,054 to 9,854 for CABG, and 1,364 to 1,756 for PCI. The incremental effect of the high quality ranking is captured by the triple interaction terms in the full DD models; all of the related coefficients were positive. The variation in these coefficients was substantial, with statistical significance found in the models which included the heart failure mortality or the combined scores, while the corresponding coefficients in the heart attack models were not significant. Taken together we interpret these results as indicating that high quality hospitals were not penalized by the initiation of Hospital Compare and may have benefited from a quality premium. We also note the coefficient of the binary Hospital Compare indicator was a highly significant and consistent across all models, reflecting the general increase in CABG and PCI prices between the pre HC and post HC period. Holding the program effects constant, there were no significant differences between no-report states and control states, as would be expected for valid controls.

The bottom panel of Table 3 summarizes coefficients of the counterfactual cases, namely colectomy procedures and Gastrointestinal (GI) treatments. We would not expect to find a strong effect of the cardiac-related HC measure on unrelated GI care and colon cancer surgery. This is confirmed by the estimates. Indeed, none of the program coefficients were significant.[12] As before, the HC coefficient for the pre-post comparison was significant, indicating an increase in prices over the study period independently of the intervention.

In Table 4 the regression coefficients are extracted to demonstrate the overall price-reducing impact of Hospital Compare and the potential offsetting effect of being ranked in the high-quality category (above average) category.[13] The first column reports prices for the baseline case, in the states with no reports (intent-to-treat) in the non-reporting period. The baseline case prices here are risk adjusted using the regression covariates and predicted for the baseline period. As expected, CABG prices are substantially higher than PCI prices. The second and third columns decompose the full policy effect. The second column reflects the change in price due to the introduction of HC at the "typical" hospital with average or below average quality reported. The third column reflects the combined effect of the policy and the premium for reporting high quality.

The second column indicates that HC exerts downward pressure on prices ranging from -16.2% to -19.1% for CABG and -6.3% to -7.3% for PCI across all measures used. From the third column we observe that in general, hospitals in the high quality bracket are able to command higher prices relative to all other hospitals. For instance, in the case of CABG, the second row implies that when the heart failure mortality measure is used, the full effect is 5.7% = -19.1% + 24.8%, where 24.8% represents the quality premium. Note that in this case, the quality premium more than offsets the downward pricing pressure exerted by the introduction of HC. A similar finding is found for PCI. In both the CABG and PCI cases, for the heart attack measure,

[12] For colectomy, the number of patients undergoing the procedure in our data was 4,955, and the mean price was $21,690. For GI hemorrhage, there were 16,924 patients, with a mean price of $11,302.

[13] The term a_4*(E(Q|NR=1) -E(Q|NR=0)) in equation 4 ranged from $1 to $22 and was not statistically significant in any of the models. Therefore it is omitted from the program effect calculations as reported in Table 4.

the quality effect dampens the downward pressure on prices due to HC, but does not fully offset it.

V. Discussion

Previous studies have shown limited impacts of hospital report cards on medical outcomes and consumer choices, calling their value into question. Shifting the focus on hospital prices, the results of this study increases confidence in the value of disseminating report cards while alleviating concerns over markets potential failure to reward higher-performing hospitals. Our results generally suggest that Hospital Compare, the premier source of publicly reported information on hospital quality in the U.S., exerts competitive pressures on hospital pricing contributing to lower prices. However, high quality hospitals were able to capture higher prices, offsetting the effect of Hospital Compare initiation.

Our results have important implications for the future of health care reform. With the implementation of Health Exchanges under the Affordable Care Act (ACA), hospital price negotiations will likely intensify as more participants enter private markets. Despite the growing need, information on pricing remains limited; and while hospital report cards are now accessible to consumers, particularly in the case of cardiac procedures and diagnoses, little is known about the impact of this information on prices ultimately paid by patients and plans.

In addition to recognizing the importance of providing quality information, policy makers have identified a need to provide hospitals with financial incentives to induce delivery of higher quality services to patients. Under revised payment rules now incorporated into the ACA, CMS will adjust overall payments made to hospitals for services provided to Medicare beneficiaries based on adherence to certain quality indicators (Value Based Purchasing). In private markets, however, formal pay-for-performance rules for hospitals are less applicable, and compensation for higher performance is left largely to market forces, through price differentials that are renegotiated annually with insurers. This study suggests that quality report cards can inform and influence hospital-insurer negotiations in the intended direction, thereby increasing consumer welfare.

References

Acemoglu, D., Finkelstein, A. (2008). Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector. *Journal of Political Economy*, 116(5):837-80.

Agency for Healthcare Research and Quality. (2015). AHRQ QI™ Research Version 5.0, Patient Safety Indicators 04, Technical Specifications, Death Rate among Surgical Inpatients with Serious Treatable Complications. Available at: http://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V50/TechSpecs/PSI_04_Death_among_Surgical_Inpatients.pdf Last Accessed: January 5, 2016.

Brooks J.M., Dor A., Wong H.S. (1997). Hospital-insurer bargaining: An empirical investigation of appendectomy pricing. *Journal of Health Economics*, 16(4):417-34.

Busso M., DiNardo J., McCrary J, (2014). New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators. *Review of Economics and Statistics* 96(5): 885-89.

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):978-85.

Caliendo, M., and Kopeinig, S. (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, 22(1), 31-72.

Centers for Medicare & Medicaid Services. (2010). Inpatient hospital payment information for value-driven health care. Available at: http://www.cms.gov/HealthCareConInit/02_Hospital.asp. Last Accessed: August 24, 2011.

Chen L.M., Orav E.J., Epstein A.M. (2012). Public reporting on risk-adjusted mortality after percutaneous coronary interventions in New York State: Forecasting ability and impact on market share and physicians' decisions to discontinue practice. *Circulation: Cardiovascular Quality and Outcomes*, 5(1):70-75.

Chou, S-Y, M. Grossman, J-T Liu (2014). The impact of National Health Insurance on birth outcomes: A natural experiment in Taiwan. *Journal of Development Economics*, Volume 111: 75–91.

DeFrances, C.J., Lucas, C.A., Vuie, V.C., Golosinskiy, A. (2008). 2006 National Hospital Discharge Survey. Hyattsville MD. National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr005.pdf

Dor, A., Encinosa, W.E., Carey, K. (2015). Medicare's hospital compare quality reports appear to have slowed price increases for two major procedures. *Health Affairs*, *34*(1), 71-77.

Dor, A., Koroukian, F., Xu, F., Stulberg, J., Delaney, C., Cooper, G. (2012). Pricing of Surgeries for Colon Cancer: Patient Severity and Market Factors. *Cancer,* 118(23):5741-748.

Dor, A., Grossman, M., Koroukian, S.M. (2004). Hospital transaction prices and managed-care discounting for selected medical technologies. *The American Economic Review, 94*(2), pp. 352-356.

Dor, A., Koroukian, S.M., Grossman, M. (2004). Managed care discounting: Evidence from the MarketScan database. *Inquiry,* 41(2):159-169.

Dranove, D., Sfekas, A. (2008). Start spreading the news: A structural estimate of the effects of New York hospital report. *Journal of Health Economics,* 27 (2008) 1201–1207.

Epstein, A., Polsky, D., Yang, F., Yang, L., Groeneveld, P. W. (2011). Coronary revascularization trends in the United States, 2001-2008. *Journal of the American Medical Association,* 305(17):1769.

Finkelstein, A. (2007). The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare. *Quarterly Journal of Economics,* 122(1):1-37.

Harris, G. (2007). Government Report Rates Hospitals on Their Heart Treatment. *The New York Times,* p. A13.

Krumholz, H.M., Wang, Y., Mattera, J.A., Wang, Y., Han, L.F., Ingber, M.J., Roman, S., Normand. (2006a). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*, 113(13), 1683-1692.

Krumholz, H.M., Wang, Y., Mattera, J.A., Wang, Y., Han, L.F., Ingber, M.J., Roman, S., Normand, S.T. (2006b). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*, 113(13), 1693-1701.

Hansen, B.E. (2016). Econometrics. University of Wisconsin. Accessed on June 8, 2016: http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf

Hansen D.L., Chang S. (2011). Health Research Data for the Real World: The Thomson Reuters Marketscan Databases. White Paper, Pharmaceutical Division, Thomson Reuters. Available at: http://thomsonreuters.com/products_services/healthcare/healthcare_products/pharmaceuticals/white_papers/. Last Accessed: August 24, 2011.

Hausman J, and Kuersteiner G, (2008). Difference in difference meets generalized least squares: Higher order properties of hypotheses tests. *Journal of Econometrics*, 144(2), 371-391.

Laschober, M., Maxfield, M., Felt-Lisk, S., Miranda, D.J. (2006). Hospital response to public reporting of quality indicators. *Health Care Financing Review*, 28(3), 61-76.

Leuven, E., Sianesi, B., 2012. PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing.

Marshall M.N., Shekelle P.G., Leatherman S., Brook R.H. (2000). The public release of performance data – what do we expect to gain? A review of the evidence. *Journal of the American Medical Association,* 283(14):1866-1874.

Meier, B., Mcginty, J.C., & Creswell, J. (2013). Hospital Billing Varies Wildly, U.S. Data Shows. *The New York Times*, May 8. p. A1.

Reinhardt U.E. (2006). The pricing of U.S. hospital services: Chaos behind a veil of secrecy. *Health Affairs,* 25(1):57-69.

Reinhardt U.E. (2009). How do hospitals get paid? A primer. *The New York Times*, January 23. http://economix.blogs.nytimes.com/2009/01/23/how-do-hospitals-get-paid-a-primer/; Last Accessed: August 12, 2012.

Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score", The American Statistician 39(1), 33-38.

Ryan A.M., Nallamothu B.K., Dimick J.B. (2012). Medicare's public reporting initiative on hospital quality had modest or no impact on mortality from three key conditions. *Health Affairs,* 31(3):585-590.

Silber J.H., Rosenbaum P.R., Brachet T.J., Ross R.N., Bressler L.J., Even-Shoshan O., et al., (2010). The Hospital Compare mortality model and the volume-outcome relationship. *Health Services Research*, 45(5 Pt 1):1148–67.

Sinaiko A., Eastman D., Rosenthal M. (2012). How report cards on physicians, physician groups, and hospitals can have greater impact on hospital choices. *Health Affairs,* 31(3):602-611.

Teleki S., Shannon M. (2012). In California, quality reporting at the state level is at a crossroads after hospital group pulls out. *Health Affairs,* 31(3):642-646.

Wang J., Hockenberry J, S-Y Chou , Yang M (2011). Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania, 30(2): 392–407.

Zhou F., Harpaz R, Jumaan A.O., Winston C.A., Shefer A. (2005). Impact of varicella vaccination on health care utilization. *Journal of the American Medical Association,* 294(7):797-802.

Table 1: Hospital Prices By Quality Level as reported in Hospital Compare

CABG Sample (N=20,774)		
Hospital Quality Reported: *Heart Attack, Heart Failure, and Pneumonia Mortalities*	Sample	Mean Price[1] During HC Reporting Period
Above average in at least one score[2]	4.3%	$70,097 (87,251)
Average	92.2%	58,276*** (70,857)
Below average in at least one score	3.5%	56,116*** (36,120)
PCI Sample (N=39,002)		
Hospital Quality Reported: *Heart Attack, Heart Failure, and Pneumonia Mortalities*	Sample	Mean Price[1] During HC Reporting Period
Above average in at least one score[2]	5.3%	$29,179 (18,864)
Average	91.0%	25,955*** (19,049)
Below average in at least one score	3.7%	25,945*** (16,608)

Note: 30-day mortality reported in Hospital Compare.

1. Mean prices are in 2010 dollars over the reporting period 2008-2010. Standard deviations are in parentheses.
2. Above average for all three years in at least one of the three quality measures. Similarly, below average pertains to all three years in at least one of the three measures.

Source: MarketScan 2005-2010.

***Statistically different from above average at the 99% level.

Table 2: Hospital Prices by State Groupings and Hospital Compare Status

CABG Sample				
Did the State have quality reports before Hospital Compare (HC) reporting?	Hospital Reported Quality Post-HC	Prices		
		Pre-HC	Post-HC	Period Differences
Yes (19.8%)	At most average (90.7%)	$57,072	$71,658	25.6%***
	Above average (9.3%)	$60,571	$70,118	15.8%
	All	$57,374	$71,505	24.6%
No (80.2%)	At most average (96.7%)	$48,055	$54,506	13.4%***
	Above average (3.3%)	$45,377	$70,074	54.4%
	All	$47,954	$54910	14.5%
PCI Sample				
Did the State have quality reports before Hospital Compare (HC) reporting?	Hospital Reported Quality Post-HC	Prices		
		Pre-HC	Post-HC	Period Differences
Yes (19.1%)	At most average (89.1%)	$20,867	$24,600	17.9%***
	Above average (11.9%)	$24,224	$26,839	10.8%
	All	$21,187	$24,912	17.6%
No (80.9%)	At most average (96.2%)	$23,820	$26,356	10.6%***
	Above average (3.8%)	$20,827	$33,824	62.4%
	All	$23,682	$26,532	12.0%

Source: Authors calculations based on the MarketScan 2005-2010 inpatient claims files.

Notes: Mean prices are in 2010 dollars. Pre-reporting period is 2005-2007. Post-reporting period is 2008-2010. As in Table 1, above average means above average for all three years in at least one of the three HC mortality measures. Below average is below average for all three years in at least one of the three measures. Details on states with pre-HC reporting are in Appendix Table A.1.

***The "at most average" difference is statistically different from the "above average" difference at the 99% level.

Table 3: Impact of Hospital Compare Reporting on Private Hospital Prices

	CABG				PCI			
	Naive DD	Full DD Model			Simple DD	Full DD Model		
		heart failure mortality	heart attack mortality	combined scores		heart failure mortality	heart attack mortality	combined scores
Hospital Compare Reporting	14,892*** (2,700)	15,486*** (2,967)	14,972*** (2,964)	14,787*** (3,069)	2,972*** (590)	3,104*** (633)	2,913*** (630)	2,831*** (636)
No-Report State	454 (2,554)	524 (2,553)	442 (2,565)	430 (2,552)	209 (606)	212 (606)	192 (606)	205 (606)
Hospital Compare *No-Report State	-8,054*** (2,788)	-9,854*** (3,042)	-8,350*** (3,101)	-9,295*** (3,123)	-1,364** (560)	-1,756*** (610)	-1,516** (610)	-1,649*** (602)
Hospital Compare *Above Average Quality	--	-4,355 (4,633)	-639 (4,792)	279 (3,123)	--	-859 (1,070)	293 (1,095)	421 (683)
Hospital Compare *No-Report State *Above Average Quality	--	12,786** (6,333)	2,842 (5,992)	5,408* (3,201)	--	2,557** (1,272)	1,774 (1,308)	1,231* (659)
	Colectomy				Gastrointestinal hemorrhage			
	DD Model	heart failure mortality	heart attack mortality	combined scores	DD Model	heart failure mortality	heart attack mortality	combined scores
Hospital Compare Reporting	9,935** (3,929)	8,966*** (2,994)	8,275*** (2,999)	8,805*** (3,051)	5,206*** (1,419)	5,011*** (1,479)	5,135*** (1,453)	4,940*** (1,578)
No-Report State	-1,997 (4,300)	-5,005* (2,820)	-5,003* (2,821)	-4,950* (2,828)	1,018 (1,188)	1,018 (1,193)	988 (1,201)	1,029 (1,195)
Hospital Compare *No-Report State	-3,925 (3,706)	-1,390 (2,589)	-331 (2,575)	-675 (2,675)	-1,676 (1,341)	-1,375 (1,325)	-1,737 (1,347)	-1,512 (1,419)
Hospital Compare *Above Average Quality	--	-6,252* (3,354)	6,765 (7,340)	-940 (3,413)	--	1,553 (4,401)	744 (3,106)	1,118 (2,808)
Hospital Compare *No-Report State *Above Average Quality	--	6,154 (4,139)	-14,045* (7,824)	-2,025 (3,576)	--	-2,439 (4,440)	841 (3,307)	-700 (2,867)

Notes: Hospital fixed effects GMM using the covariates of Appendix Table A.4, with heteroskedasticity-robust standard errors in parentheses. Hospital Compare and mortality scores are lagged a year. DD= Difference-in-Difference estimates without the quality report.
***Statistically different from zero at the 99% level.
**Statistically different from zero at the 95% level.
*Statistically different from zero at the 90% level.

Table 4: Impact of Medicare Hospital Compare Reporting on Private Hospital Prices: Program and Quality Effects

Reported mortality measure	Estimated CABG Prices		
	Baseline Price *(No public reporting)*	***Program Effect*** *(HC reporting of average or below average quality)*	***Quality "Premium"*** *(HC reporting of better than average quality)*
Full DD Results			
Heart attack	$51,539	$51,539 – 8,350*** (-16.2%)	+ $2,842 (+5.5%)
Heart failure	$51,539	$51,539 – 9,854*** (-19.1%)	+ $12,786** (+24.8%)
Combined	$51,539	$51,539 – 9,295*** (-18.0%)	+ $5,408* (+10.5%)
		HC reporting of any quality	
Simple DD Results	$51,484	$51,484 – 8,054*** (-15.6%)	
	Estimated PCI Prices		
	No public reporting	*HC reporting of average or below average quality*	*HC reporting of better than average quality*
Full DD Results			
Heart attack	$24,200	$24,200 – 1,516** (-6.3%)	+ $1,774 (+7.3%)
Heart failure	$24,200	$24,200 – 1,756*** (-7.3%)	+ $2,557** (+10.6%)
Combined	$24,200	$24,200 – 1,649*** (-6.8%)	+ $1,231* (+5.1%)
		HC reporting of any quality	
DD Results	$24,472	$24,472 – 1,364** (-5.6%)	

Note: HC=Hospital Compare reporting of 30-day mortality. Prices are in 2010 dollars, estimated from the GMM hospital fixed effect regressions , controlling for the differential impact of lagged HC on prices between States with and without other public reporting of CABG and PCI outcomes. ``No Public Reporting" prices are predicted from the regressions assuming no HC and no other State public reports. Source: MarketScan 2005-2010.

***Statistically different from zero at the 99% level.

** Statistically different from zero at the 95% level

* Statistically different from zero at the 90% level.

Appendix A

Appendix Table A.1: Hospital State Report Card History from 2005

State	Report Card	Year(s) to Which Report Cards Pertain
New York	CABG	2003-2005, 2004-2006, 2005-2007, and 2006-2008
	PTCA (Angioplasty)	2003-2005, 2004-2006, 2005-2007, 2006-2008
Pennsylvania	CABG Surgery; Valve Surgery	2005, 2005-2006, 2006-2007, 2007-2008 and 2008-2009
New Jersey	CABG	2004, 2006, 2007, and 2008
California	CABG	2004- 2005, 2005-2006, 2007, and 2007-2008
Massachusetts	CABG	2005, 2006, 2007, 2008, and 2009
	PTCA (Angioplasty)	2005, 2006, 2007, 2008, and 2009
Florida	CABG	2005, 2006, 2007, 2008, 2009
	PTCA (Angioplasty)	2005, 2006, 2007, 2008, 2009
Washington	CABG	2010, voluntary

Notes: Washington is not included in the report card states for purposes of this study.

Appendix Table A.2: Patient Sample Descriptive Statistics

	CABG	PCI
Hospital Price	53,765 (62,504)	24,441 (18,816)
Hospital Compare Reporting (yes, no)	0.306 (0.461)	0.252 (0.434)
No State Report (yes, no)	0.802 (0.398)	0.809 (0.393)
Hospital Compare*No State Report	0.237 (0.426)	0.187 (0.390)
Hospital Compare*Above Average Heart Attack Quality	0.041 (0.198)	0.037 (0.189)
Hospital Compare*No State Report*Above Average Heart Attack Quality	0.030 (0.171)	0.025 (0.155)
HMO Market Penetration	0.210 (0.108)	0.217 (0.108)
HRR cardiac Herfindahl Index	0.295 (0.257)	0.321 (0.275)
Stent	--	0.818 (0.386)
Two Vessels Bypassed	0.302 (0.459)	--
Three Vessels Bypassed	0.386 (0.487)	--
Four or More Vessels Bypassed	0.132 (0.339)	--
Age	55.7 (8.2)	55.4 (6.7)
Female	0.232 (0.422)	0.247 (0.431)
Union	0.257 (0.437)	0.278 (0.448)
HMO Insured	0.210 (0.407)	0.232 (0.422)
Arrhythmias	0.184 (0.388)	0.107 (0.309)
Diabetes	0.203 (0.402)	0.111 (0.314)
Catheterization	0.104 (0.306)	0.154 (0.360)
AMI	0.817 (0.387)	0.946 (0.227)
Stroke	0.118 (0.323)	0.018 (0.133)
Three or more chronic conditions	0.045 (0.207)	0.009 (0.096)
Pacemaker	0.018 (0.132)	0.014 (0.117)
Valve Replacement	0.094 (0.292)	0.001 (0.032)
N	20,774	39,002

Source: MarketScan 2005-2010. Standard deviations in parentheses. Hospital Compare is lagged a year.

Appendix Table A.3: Simple Difference-in-Difference Estimates of the Impact of Hospital Compare Reporting on Private Hospital Prices (GMM Estimation, Hospital Fixed Effects)

	CABG	PCI
Hospital Compare Reporting	14,892*** (2,700)	2,972*** (590)
No State Report	454 (2,554)	209 (606)
Hospital Compare* No State Report	-8,054*** (2,788)	-1,364** (560)
HRR Cardiac HHI	1,509 (3,369)	970 (906)
HMO Market Penetration	3,459 (6,170)	-1,549 (1,675)
Two vessels bypassed	2,368** (1,039)	--
Three or more vessels bypassed	2,313** (1,150)	--
Stent	--	1,007*** (250)
Age 54-59	-445 (1,105)	-475** (210)
Age 60-64	-1,777 (1,093)	-678*** (211)
Female	2,303** (1,099)	-308 (208)
Union	-2,189** (944)	-1,742*** (219)
HMO-insured	-5,361*** (1,101)	-2,286*** (241)
Arrhythmias	9,629*** (1,336)	3,629*** (382)
Diabetes	755 (909)	1,566*** (299)
Catheterization	7,377*** (1,282)	436 (297)
Valve Replacement	14,092*** (3,081)	28,174*** (8,243)
Pacemaker	18,089*** (4,635)	12,832*** (1,388)
AMI	-10,181*** (2,581)	-5,797*** (629)
Stroke	7,899*** (1,179)	8,401*** (1,143)
Three or more chronic conditions	21,163*** (3,551)	11,732*** (1,703)

Notes: GMM, with heteroskedasticity-robust standard errors in parentheses. Time fixed effects not shown. Hospital Compare indicator included with a one year lag. * $p < .1$, ** $p < .05$, *** $p < .01$.

Appendix Table A.4: Full DD Estimates of the Impact of Hospital Compare Reporting on Private Hospital Prices, (GMM Estimation, Hospital Fixed Effects)

	CABG			PCI		
	heart failure mortality	heart attack mortality	combined scores	heart failure mortality	heart attack mortality	combined scores
Hospital Compare Reporting	15,486*** (2,967)	14,972*** (2,964)	14,787*** (3,069)	3,104*** (633)	2,913*** (630)	2,831*** (636)
No State Report	524 (2,553)	442 (2,565)	430 (2,552)	212 (606)	192 (606)	205 (606)
Hospital Compare* No State Report	-9,854*** (3,042)	-8,350*** (3,101)	-9,295*** (3,123)	-1,756*** (610)	-1,516** (610)	-1,649*** (602)
Hospital Compare*Above Average Quality	-4,355 (4,633)	-639 (4,792)	279 (3,123)	-859 (1,070)	293 (1,095)	421 (683)
Hospital Compare*No State Report* Above Average Quality	12,786** (6,333)	2,842 (5,992)	5,408* (3,201)	2,557** (1,272)	1,774 (1,308)	1,231* (659)
HRR Cardiac HHI	1,532 (3,371)	1,473 (3,367)	1,430 (3,370)	921 (906)	984 (906)	947 (906)
HMO Market Penetration	3,386 (6,162)	3,481 (6,169)	3,466 (6,163)	-1,472 (1,672)	1,530 (1,675)	1,504 (1,673)
Two Vessels Bypassed	2,403** (1,041)	2,367** (1,039)	2,386** (1,040)	--	--	--
Three or More Vessels Bypassed	2,330** (1,154)	2,311** (1,150)	2,329** (1,152)	--	--	--
Stent	--	--	--	1,003*** (250)	1,003*** (250)	994*** (250)
Age 54-59	-452 (1,105)	-438 (1,105)	-429 (1,103)	-478** (211)	-476** (210)	-473** (210)
Age 60-64	-1,787 (1,093)	-1,778 (1,093)	-1,784 (1,093)	-677*** (211)	-676*** (211)	-674*** (211)
Female	2,308** (1,099)	2,303** (1,099)	2,302** (1,098)	-303 (208)	-303 (208)	-303 (208)
Union	-2,203** (947)	-2,194** (945)	-2,195** (945)	-1,746*** (219)	-1,741*** (219)	-1,738*** (219)
HMO-insured	-5,413*** (1,100)	-5,373*** (1,098)	-5,409*** (1,102)	-2,294*** (241)	-2,296*** (242)	-2,299*** (242)
Arrhythmias	9,564*** (1,340)	9,625*** (1,336)	9,600*** (1,337)	3,617*** (382)	3,624*** (382)	3,623*** (382)
Diabetes	732 (911)	748 (909)	745 (909)	1,572*** (299)	1,569*** (299)	1,567*** (299)
Catheterization	7,349*** (1,283)	7,372*** (1,281)	7,373*** (1,282)	443 (297)	434 (297)	442 (297)
Valve Replacement	14,004*** (3,072)	14,096*** (3,081)	14,096*** (3,080)	28,194*** (8,240)	28,132*** (8,239)	28,158*** (8,242)
Pacemaker	18,079*** (4,634)	18,089*** (4,634)	18,147*** (4,635)	12,835*** (1,389)	12,835*** (1,389)	12,830*** (1,389)
AMI	-10,243*** (2,572)	-10,179*** (2,581)	-10,219*** (2,576)	-5,799*** (629)	-5,800*** (629)	-5,793*** (629)
Stroke	7,955*** (1,178)	7,904*** (1,179)	7,958*** (1,177)	8,383*** (1,143)	8,384*** (1,143)	8,388*** (1,143)
Three or More Chronic Conditions	21,147*** (3,545)	21,167*** (3,551)	21,158*** (3,548)	11,708*** (1,704)	11,695*** (1,704)	11,694*** (1,703)

Notes: GMM with heteroskedasticity-robust standard errors clustered at the hospital in parentheses. Time fixed effects not shown. Hospital Compare and mortality scores are with a one year lag. ***Statistically different from zero at the 99% level. **Statistically different from zero at the 95% level. *Statistically different from zero at the 90% level.

Appendix B: A test for misspecification error.

Two confounding effects due to treatment assignment may arise when estimating average treatment effect on the treated (ATT), due to sample selection bias and group matching. While selection bias is highly unlikely in our case (patients do not select their location based on which states provided report cards) matching may be in issue both between groups and within groups overtime. As a sensitivity test, here we rerun our difference-in-difference models subsetting to samples that have been matched with propensity scores. In particular, in the CABG sample, we match the 4,096 observations in the reporting states to 4,096 in the no reporting states using nearest neighbor matching, without replacement and with common support. In the PCI sample, we match 7,195 observations in the reporting states to 7,195 in the no reporting states. We use the Stata program "psmatch2" to construct the propensity scores as the propensity to be a reporting state, based on all the patient and market characteristics used in all our models (Leuven and Sianesi, 2012). Comparing matching to no matching before the regressions, the bias is reduced from 15.9 to 7.8 in the PCI model, and from 16.6 to 7.3 in the CABG model. The bias is the difference of the sample means in the reporting and no reporting (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the reporting and no reporting groups (Rosenbaum and Rubin, 1985). Applying the GMM fixed effects difference-in-difference estimators to the matched samples, we obtained the results shown in Table B.1. These results do not differ much from the patterns found in the full sample results of Table 3.

Note that Busso, DiNardo, and McCrary (2014) show that matching with propensity scores is preferable to sample reweighting, however, propensity scores are sensitive to model specification, and in our case is not clear which variables should account for common support, e.g., patient level variables or all variables including hospital and market area characteristics

(Caliendo and Kopeinig, 2008; Leuven and Sianesi, 2012). To assess the potential bias more generally we expand the expected value of the full DD estimate to include the vector of all such variables.

Restating equation [4] we have:

$$\mathrm{P_{ihjt}} = \mathrm{a_0} + \mathrm{a_1 HC_t} + \mathrm{a_2 NR_{ijt}} + \mathrm{a_3} Q_{h,t-1} + a_4\, \mathrm{HC_t} * \mathrm{NR_{ijt}} + \mathrm{a_5 Q}_{h,t-1} * \mathrm{NR_{ijt}} + \mathrm{a_6} HC_t * Q_{h,t-1} + a_7 * HC_t * NR_{ijt} * Q_{h,t-1} + \mathrm{a_8} * \mathrm{Z_{ijt}} + \mathrm{e_{ihjt}}$$

Under this model, the average impact of hospital compare is given by:

$$\Delta\mathrm{HC} = [\mathrm{E(P|nr, Post)} - \mathrm{E(P|nr, Pre)}] - [\mathrm{E(P|r, post)} - \mathrm{E(P|r, pre)}]$$

$$= \Big[\big(a_0 + a_1 + a_2 + a_3 E(Q_{nr}|post) + a_4 + a_5 E(Q_{nr}|post) + a_6 E(Q_{nr}|post) + a_7 E(Q_{nr}|post) + a_8 E(Z_{nr}|post)\big) - \big(a_0 + a_3 E(Q_{nr}|pre) + a_5 E(Q_{nr}|pre) + a_8 E(Z_{nr}|pre)\big)\Big] - \Big[\big(a_0 + a_1 + a_3 E(Q_r|post) + a_6 E(Q_r|post) + a_8 E(Z_r|post)\big) - \big(a_0 + a_3 E(Q_r|pre) + a_8 E(Z_r|pre)\big)\Big]$$

$$= a_3\{[E(Q_{nr}|post) - E(Q_{nr}|pre)] - [E(Q_r|post) - E(Q_r|pre)]\} + a_4 + a_5[E(Q_{nr}|post) - E(Q_{nr}|pre)] + a_6[E(Q_{nr}|post) - E(Q_r|post)] + a_7 E(Q_{nr}|post) + a_8\{[E(Z_{nr}|post) - E(Z_{nr}|pre)] - [E(Z_r|post) - E(Z_r|pre)]\}$$

Similarly, using equation 3 (Hospital Compare quality scores are not reported prior to HC implementation), we have

$$\Delta\mathrm{HC} = \Big[\big(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 * E(Q_{nr}|post) + \alpha_5 E(Q_{nr}|post) + \alpha_6 E(Z_{nr}|post)\big) - \big(\alpha_0 + \alpha_2 + \alpha_6 E(Z_{nr}|pre)\big)\Big] - \Big[\big(\alpha_0 + \alpha_1 + \alpha_4 E(Q_r|post) + \alpha_6 E(Z_r|post)\big) - \big(a_0 + \alpha_6 E(Z_{nr}|pre)\big)\Big]$$

$$= \Big[\alpha_1 + \alpha_3 + \alpha_4 * E(Q_{nr}|post) + \alpha_5 E(Q_{nr}|post) + \alpha_6\big(E(Z_{nr}|post) - E(Z_{nr}|pre)\big)\Big] - \Big[\alpha_1 + \alpha_4 E(Q_r|post) + \alpha_6\big(E(Z_r|post) - E(Z_r|pre)\big)\Big]$$

$$= \alpha_3 + \alpha_4[E(Q_{nr}|post) - E(Q_r|post)] + \alpha_5 E(Q_{nr}|post) + \alpha_6\Big[\big(E(Z_{nr}|post) - E(Z_{nr}|pre)\big) - \big(E(Z_r|post) - E(Z_r|pre)\big)\Big]$$

Thus, misspecification due to balancing would not be an issue if

$$[E(Z_{nr}|post) - E(Z_{nr}|pre)] - [E(Z_r|post) - E(Z_r|pre)] \approx 0$$

Calculating this term in our data we get -263 for CABG and 754 for PCI in the full difference-in-difference model for the combined heart failure/heart attack score. Taking account of this component will not alter the basic empirical finding of a negative program effect (Table 4). Moreover, from a two sided Z-test neither term was statistically different from zero (p=.95, p=0.92 respectively).

Table B.1: Impact of Hospital Compare Reporting on Private Hospital Prices (matched propensity score samples)

	CABG				PCI			
	Naive DD	Full DD Model			Simple DD	Full DD Model		
		heart failure mortality	heart attack mortality	combined scores		heart failure mortality	heart attack mortality	combined scores
Hospital Compare Reporting	15,245*** (3,246)	15,858*** (3,484)	15,246*** (3,417)	14,803*** (3,650)	3,488*** (729)	3,702*** (776)	3,458*** (757)	3,289*** (782)
No-Report State	2,767 (3,501)	2,913 (3,507)	2,757 (3,504)	2,699 (3,495)	-219 (750)	-228 (750)	-238 (750)	-217 (750)
Hospital Compare *No-Report State	-11,591*** (3,944)	-15,048*** (3,964)	-11,349*** (4,346)	-13,095*** (4,136)	-1,779** (752)	-2,396*** (832)	-2,018** (827)	-2,275*** (790)
Hospital Compare *Above Average Quality	--	-43,863 (4,586)	-56 (4,786)	1,713 (3,807)	--	-1,238 (1,079)	112 (1,093)	566 (796)
Hospital Compare *No-Report State *Above Average Quality	--	20,916* (11,947)	2,430 (7,039)	5,562 (4,353)	--	3,197* (1,764)	1,899 (1,617)	1,445* (758)

Notes: Matched samples: CABG N=8,198 and PCI N=14,390. Observations in "no reporting "states were matched 1 to 1 to observations in reporting states by nearest neighbor propensity scores. Hospital fixed effects GMM using the covariates of Appendix Table A.4, with heteroskedasticity-robust standard errors in parentheses. Hospital Compare and mortality scores are lagged a year. DD= Difference-in-Difference estimates without the quality report.
***Statistically different from zero at the 99% level.
**Statistically different from zero at the 95% level.
*Statistically different from zero at the 90% level.