Hospitals and Nursing Homes

Has Pay-for-Performance Decreased Access for Minority Patients?

Andrew M. Ryan

Objective. To examine whether the CMS and Premier Inc. Hospital Quality Incentive Demonstration (PHQID), a hospital-based pay-for-performance (P4P) and public quality reporting program, caused participating hospitals (1) to avoid treating minority patients diagnosed with acute myocardial infarction (AMI), heart failure, and pneumonia and (2) to avoid providing coronary artery bypass graft (CABG) to minority patients diagnosed with AMI.

Data Sources. One hundred percent Medicare inpatient claims, denominator files, and provider of service files from 2000 to 2006.

Study Design. We test for differences in the conditional probability of receiving care at PHQID hospitals for AMI, heart failure, and pneumonia before and after implementation of the PHQID between white and minority patients. We also test for differences in the conditional probability that white and minority patients diagnosed with AMI receive CABG in hospitals participating, and not participating, in the PHQID before and after the implementation of the PHQID.

Data Extraction Methods. Data were obtained from CMS.

Principal Findings. We find little evidence that the PHQID reduced access for minority patients: only “Other Race” beneficiaries had a significant reduction in adjusted admissions to PHQID hospitals in the postperiod, and only for AMI. Only marginally significant ($p < .10$) evidence of a reduction in CABG was found, also occurring for Other Race beneficiaries.

Conclusions. Despite minimal evidence of minority patient avoidance in the PHQID, monitoring of avoidance should continue for P4P programs.

Key Words. Incentives, disparities, hospitals

The misalignment of incentives between payers and providers is a common critique of the U.S. health care system (Institute of Medicine [IOM] 2001). Both public quality reporting and pay-for-performance (P4P) represent efforts by payers to align providers behind the objectives of higher quality and, potentially, lower cost care, resulting in higher value (IOM 2006).

However, public quality reporting and P4P are subject to their own critiques. In addition to multitasking, the allocation of resources toward a subset of measured activities and away from unmeasured activities (Holmstrom and Milgrom 1991; Eggleston 2005), patient avoidance is frequently raised as perhaps the most deleterious unintended consequence of public quality reporting and P4P (Werner, Asch, and Polsky 2005; Epstein 2006; Casalino et al. 2007; Chien et al. 2007; Hood 2007). Patient avoidance, or “cream-skimming,” occurs when providers determine that it is in their interest to avoid treating a patient who is likely to reduce their performance on a publicly reported or financially incentivized quality measure. This behavior has been documented extensively in the public quality reporting literature (Burack et al. 1999; Dranove et al. 2003; Narins et al. 2005; Epstein 2006). In a survey among surgeons in the New York cardiac report card program, 62 percent of respondents indicated that they did not operate on at least one high-risk coronary artery bypass graft (CABG) patient over the prior year “primarily because of public reporting” (Burack et al. 1999). In a later survey, 83 percent of New York surgeons “agreed or strongly agreed that patients who might benefit from angioplasty may not receive the procedure as a result of public reporting of physician-specific patients’ mortality rates” (Narins et al. 2005).

Theoretically, the incentives for providers to engage in strategic patient avoidance can be counteracted by risk adjustment, in which case mix severity is held equal across providers through statistical adjustment. However, risk adjustment suffers from a number of potential problems. First, risk adjustment must be based on observable patient characteristics, but providers are likely to have additional information about patients which is not captured by risk adjustment (Dranove et al. 2003). Also, as noted by Dranove et al., risk adjustment, even if “correct in expectation terms but incomplete . . . may not compensate risk-averse providers sufficiently for the downside of treating sick patients” (Dranove et al. 2003). Supporting this premise, one study of New York surgeons found that “85 percent believed that the risk-adjustment model used in the Percutaneous Coronary Interventions in New York State(s) (sic) 1998–2000 report is not sufficient to avoid punishing physicians who perform higher-risk interventions” (Narins et al. 2005).

In the absence of complete risk adjustment, providers may engage in statistical discrimination: the application of perceived group characteristics to

Address correspondence to Andrew M. Ryan, M.A., Ph.D., Department of Public Health, Weill Medical College–Cornell University, 402 East 67th Street, LA-215, New York, NY 10065; e-mail: amr2015@med.cornell.edu.

individuals. As noted by McGuire et al. (2008), statistical discrimination “appears to be a potent (if more difficult to observe) source of discrimination in health care use” (p. 532). Statistical discrimination may result in provider avoidance of racial and ethnic minority (henceforth “minority”) patients, who may be perceived to have higher unmeasured risk than non-Hispanic white (henceforth “white”) patients, in P4P and public quality reporting programs. Even if unobserved severity is not correlated with race, physicians and hospitals may inaccurately perceive it to be, and thus attempt to avoid patients on that basis. Thus, the combination of two factors, (1) providers’ desire to avoid patients on the basis of unmeasured severity and (2) providers’ perception that minority patients have greater unmeasured risk, could lead to avoidance of minority patients as a result of the incentives of P4P.

The limited empirical evidence on this topic has been conducted in the context of public quality reporting. A study by Werner, Asch, and Polsky (2005) found that disparities in rates of CABG procedures increased for black and Hispanic patients, relative to white patients, after the implementation of the New York CABG public quality reporting program. Because similar increases in disparities were not observed for percutaneous transluminal coronary angioplasty (PTCA) and cardiac catheterization, which were not publicly reported, the study suggests that public quality reporting in New York caused the observed increases in disparities for CABG. Despite concerns raised by several authors (Casalino et al. 2007; Chien et al. 2007; Hood 2007), there is no empirical evidence on the effect of P4P programs on patient avoidance. Evidence from the United Kingdom’s National Health Services P4P program suggests that provider practices that served lower-income patients had lower quality scores (Doran et al. 2006) and that the program has not successfully combated racial and ethnic disparities related to diabetes care (Millett et al. 2007). However, these studies do not examine patient avoidance or other potential unintended consequences of P4P. As noted by (Chin et al. 2007), “Disturbingly, there is essentially no literature evaluating the effectiveness or potential harms of these policies (P4P and other performance incentives) on disparities” (p. 25S).

The current investigation addresses this gap in the literature by evaluating whether the CMS and Premier Inc. Hospital Quality Incentive Demonstration (PHQID), a hospital-based P4P and public quality reporting program, caused participating hospitals (1) to avoid treating minority (non-white) patients diagnosed with acute myocardial infarction (AMI), heart failure, and pneumonia and (2) to avoid providing CABG to minority patients diagnosed with AMI. The study also examines the mechanisms through which minority patients may have been avoided in the PHQID.

CONTEXT OF STUDY: THE PREMIER HOSPITAL QUALITY INCENTIVE DEMONSTRATION

The first phase of the PHQID, beginning in the fourth quarter of 2003 and ending in the third quarter of 2006, paid a 2 percent bonus on Medicare reimbursement rates to hospitals performing in the top decile of a composite quality measure for each incentivized condition (heart failure, AMI, community-acquired pneumonia, hip and knee replacement [not evaluated in this study], and CABG) and a 1 percent bonus for hospitals performing in the second decile. Penalties were administered to hospitals with exceptionally poor performance. Bonus payments were disbursed based on composite quality measures, consisting predominately of process measures but including some outcome measures, for each incentivized condition. Participation in the PHQID was voluntary: of the 421 hospitals asked to participate, 266 (63 percent) chose to do so (Lindenauer et al. 2007). Hospitals' eligibility to participate was based on their subscription to Premier's *Perspective* database, a database used for benchmarking and quality improvement activities, as of March 31, 2003.¹

MECHANISMS OF HOSPITAL PATIENT AVOIDANCE IN THE PHQID

There are three distinct mechanisms through which patients are admitted to hospitals: through the emergency department (ED), by referral (typically by physicians), and through transfers (typically from other hospitals or long-term care facilities). Thus, if hospitals in the PHQID avoided minority patients, they must have done so through one or more of these sources of admission. First, while evidence suggests that physicians themselves were not given financial incentives for quality performance in the PHQID (Damberg et al. 2007), hospitals may have discouraged community physicians, through a variety of means, from referring minority patients (Einbinder and Schulman 2000; Hargraves, Stoddard, and Trude 2001). Second, hospitals could have decreased minority admissions by creating greater barriers to admission through the ED. Among Medicare beneficiaries, minority patients tend to be more likely than white patients to be admitted through the ED (see Table 1), and by decreasing ED capacity (Burton 2009), hospitals could effectively decrease minority admissions. PHQID hospitals could also have discouraged transfers of high risk, and minority, patients: African Americans have been found to have a decreased probability of transfer in other settings (Gurwitz et al. 2002).

Table 1: Beneficiary Characteristics by Condition Cohort and Race

	<i>White</i>	<i>Black</i>	<i>Hispanic</i>	<i>Other Race</i>	<i>Non-White</i>
Region Analysis					
AMI					
<i>n</i>	1,190,474	139,785	24,149	31,048	194,982
Age (mean)	75.7	71.7	74.3	73.4	72.3
Female (%)	47.9	56.1	49.3	44.6	53.4
Comorbidities (mean)	1.69	1.94	1.89	1.76	1.90
Admission through ED (%)	63.7	70.3	71.6	62.4	69.2
Admission through referral (%)	18.5	16.6	13.6	20.3	16.8
Admission through transfer (%)	17.8	13.1	14.8	17.3	14.0
Heart failure					
<i>n</i>	2,361,893	561,146	65,163	62,897	689,206
Age (mean)	77.6	69.7	74.0	73.6	70.5
Female (%)	56.0	58.6	56.0*	52.7	57.8
Comorbidities (mean)	1.92	2.07	2.11	1.97	2.07
Admission through ED (%)	67.1	72.6	75.2	65.2	72.1
Admission through referral (%)	28.4	23.8	21.0	29.6	24.1
Admission through transfer (%)	4.5	3.6	3.8	5.2	3.8
Pneumonia					
<i>n</i>	2,226,619	288,399	50,732	66,648	405,779
Age (mean)	76.6	70.7	74.2	74.7	71.8
Female (%)	54.3	55.1	53.5	49.0	53.9
Comorbidities (mean)	2.63	2.66	2.60	2.49	2.62
Admission through ED (%)	69.1	73.8	74.1	64.7	72.3
Admission through referral (%)	25.9	19.8	19.2	27.7	21.0
Admission through transfer (%)	4.9	6.4	6.8	7.6	6.6
CABG Analysis					
AMI					
<i>n</i>	1,239,570	120,407	25,265	28,813	174,485
Age (mean)	74.9	70.9	73.5	72.4	71.5
Female (%)	45.5	54.8	46.3	42.6	51.6
Comorbidities (mean)	1.68	1.95	1.86	1.76	1.90

Notes. Test performed is *t*-test of difference versus white.

Comorbidities is the count of Elixhauser comorbidities.

**p* > 0.05.

AMI, acute myocardial infarction; CABG, coronary artery bypass graft; ED, emergency department.

METHODS

Data

We use several sources of Medicare data from 2000 to 2006 in this analysis: 100 percent inpatient claims, denominator files, and provider of service files. Inpatient claims are used to identify the principal diagnoses for which beneficiaries are admitted and secondary diagnoses and type of admission for risk adjustment. The Medicare denominator file is used to include additional risk

adjusters and to determine beneficiary zip code of residence. Data from the Medicare provider of service file are used to identify hospital structural characteristics. Only short-term, acute care hospitals are included in the analysis. To align the panel with the start of the PHQID, the study period spans the 6-year interval from the fourth quarter of 2000 through the third quarter of 2006.

Analysis

To evaluate the avoidance of minority patients with AMI, heart failure, and pneumonia, using patient and hospital zip codes, we identify hospital referral regions² (HRRs) in which PHQID hospitals operate and identify fee-for-service (FFS) Medicare beneficiaries living in these HRRs. We include only FFS Medicare beneficiaries living in these HRRs in order to identify patients in the market areas of PHQID hospitals. This includes 7,068,953 admissions from 3,981,516 Medicare beneficiaries diagnosed with AMI, heart failure, or pneumonia who lived in one of the 118 HRRs in which PQHID hospitals operated between the fourth quarter of 2000 and the third quarter of 2006. Using individual-level models, we estimate the probability that patients living in these HRRs receive care at PHQID hospitals as a function of beneficiary characteristics, minority status, an indicator for the PHQID implementation period (after Q3 in 2003), and an interaction between minority status and the PHQID period indicator. For patient i with diagnosis j in HRR l at time t :

$$\begin{aligned} \Pr(\text{Premier Admission})_{ijlt} = & b_{0j} + b_{1j}X_{ijlt} + b_{2j} \text{minority}_{ijlt} + b_{3j} \text{post}_t \\ & + b_{4j} \text{HRR}_l + \delta_{1j} \text{minority}_{ijlt} \times \text{post}_t + e_{ijlt}, \end{aligned} \quad (1)$$

where X is a vector of individual risk adjusters (age, gender, 30 dummy variables for the Elixhauser comorbidities [Elixhauser et al. 1998], type of admission [emergency, urgent, elective], and season of admission), *minority* is a vector of indicators for racial and ethnic groups (black, Hispanic/Latino, and other [consisting of Asian/Pacific Islanders and Native Americans]), *post* is an indicator for the post-PHQID implementation period, and *HRR* is a vector of HRR fixed effects. The equation is estimated separately for each diagnosis (AMI, heart failure, and pneumonia). The vector δ_1 contains the difference-in-differences estimates: negative coefficients in δ_1 would indicate that minority patients living in HRRs served by PQHID hospitals became less likely to receive care at these hospitals after the PHQID began. The equation is also respecified where (1) the effect of race is estimated jointly for all minority patients and (2) a time trend and its square are substituted for the *post* indicator

(allowing for the examination of annual effects). This series of models is henceforth referred to as the “regional analysis.”

To examine the mechanisms through which minority patients are potentially avoided by PHQID hospitals, for each condition, we estimate the conditional probability of hospital admission via each distinct source of admission (ED, referral, or transfer) to either PHQID or non-PHQID hospitals before and after the PHQID was implemented. We estimate a multinomial logit model with six outcomes (three sources of admission $\times$ PHQID or non-PHQID admissions) that includes beneficiaries living in HRRs that are served by PHQID hospitals, controlling for patient severity as described above. This specification does not include hospital fixed effects. For each condition for each source of admission, we then test minority avoidance by estimating the condition difference-in-difference-in-differences (DIDID) (between whites and minorities, PHQID and non-PHQID hospitals, before and after the PHQID). This analysis is called the “source of admission analysis.” The results from this analysis are only reported for the conditions and racial and ethnic groups for which patient avoidance is observed in the regional analysis.

Finally, to examine whether the PHQID led to a reduction in CABG for minority patients diagnosed with AMI, we model the probability of receiving CABG among all Medicare beneficiaries diagnosed with AMI from the fourth quarter of 2000 through the third quarter of 2006. This “CABG analysis” is performed separately from the regional analysis because a hospital’s prospective CABG patients can be defined as patients who are diagnosed with AMI. Consequently, it is unnecessary to define a hospital’s potential patients based on where they live in relation to hospitals, as is done in the regional analysis.

The CABG analysis includes 1,414,055 AMI admissions from 1,270,001 Medicare beneficiaries in 1,063 acute care hospitals. Hospitals that did not perform CABG in both the pre- and post-PHQID implementation periods are excluded. We estimate the conditional probabilities that white and minority patients receive CABG in hospitals participating, and not participating, in the PHQID before and after the commencement of the PHQID. For patient i in hospital k at time t :

$$\begin{aligned} \Pr(CABG)_{ikt} = & b_0 + b_1 X_{ikt} + b_2 minority_{ikt} + b_4 post_t \\ & + b_5 hospital_k + b_6 PHQID_{kt} + \delta_1 minority_{ikt} \times post_t \\ & + \delta_2 minority_{ikt} \times PHQID_{kt} + e_{ikt} \end{aligned} \quad (2)$$

where X and $minority$ are defined as before, $hospital$ is a vector of hospital fixed effects, $post$ is a postimplementation indicator, equal to 1 in or after the

4th quarter of 2003 for all hospitals, and *PHQID* is equal to 1 in or after the 4th quarter of 2003 for participating hospitals. To test the effect of the *PHQID* on the provision of CABG for minority patients, we derive DIDID estimates: the conditional difference in CABG rates from the pre- to postperiod for *PHQID* hospitals relative to non-*PHQID* hospitals and for minority patients relative to white patients. The equation is also respecified where a time trend and its square are substituted for the *post* indicator.

Even if evidence emerged that the *PHQID* caused participating hospitals to decrease the provision of CABG to minority patients relative to white patients, it is possible that minority access to care did not decrease. Because PTCA is a substitute for CABG (Cutler and Huckman 2003), minority patients may have received PTCA at increased rates that offset decreases in CABG. To examine this question, we perform the exact analysis as described above with the exception of modeling the probability of patients receiving PTCA instead of CABG.

With the exception of the multinomial logit models, we use linear probability specifications to estimate each equation. This estimation strategy is employed primarily for ease in calculating model marginal effects in the presence of the complex interaction terms. Further, marginal effects from linear and nonlinear models are typically quite similar, particularly when calculated at the mean of model variables (Wooldridge 2002, p. 454; Angrist and Pischke 2009, p. 107).³ Finally, the use of the linear specification allows us to compare our results directly with Werner, Asch, and Polsky (2005), who used a virtually identical estimation approach in their investigation of minority patient avoidance in the New York CABG public quality reporting program. All models are also estimated using logit models, and coefficient inference is nearly identical.

To treat heteroskedasticity arising from patient clustering within HRRs and hospitals, cluster-robust standard errors are estimated (Williams 2000) at the HRR level for the regional analysis and at the hospital-level for the source of admission and the CABG analysis. All analysis is performed using *Stata* 10.0 (StataCorp. 2007).

RESULTS

Table 1 shows descriptive statistics of the analytic sample by diagnosis and race for both the regional analysis and CABG analysis. It shows that, for AMI and heart failure, among those living in one of the 118 HRRs in which *PHQID* hospitals operate, black, Hispanic, Other Race, and nonwhite beneficiaries are younger than white beneficiaries and have more comorbidities (as measured

by the count of Elixhauser comorbidities). Also, black beneficiaries diagnosed with AMI and heart failure are more likely to be female than their white counterparts. Among beneficiaries diagnosed with pneumonia, black, Hispanic, Other Race, and nonwhite beneficiaries are again younger than white beneficiaries and, with the exception of Other Race beneficiaries, have more comorbidities than white beneficiaries. Table 1 also shows a similar pattern of findings for beneficiaries diagnosed with AMI in the CABG analysis.

Table 2 shows the adjusted admission rates to PHQID hospitals, controlling for patient characteristics, before and after the implementation of the PHQID as well as the pre–post differences and difference-in-differences (DID) estimates derived from individual-level linear probability models (equation [1]) for AMI, heart failure, and pneumonia. For each racial and ethnic cohort of beneficiaries, there are no significant pre–post differences in adjusted admission rates to PHQID hospitals for any diagnosis. Also, for black, Hispanic, and all nonwhite beneficiaries, the DID of the adjusted admission rate is not significant. For Other Race beneficiaries, adjusted admission rates to PHQID hospitals dropped in the postperiod, and the DID is significant ($p < .05$). However, the examination preintervention trends (see Appendix SA1) indicates a sharp reduction in adjusted AMI admissions at PHQID hospitals for Other Race beneficiaries that preceded the start of the intervention, casting doubt about whether the PHQID was the causal event that lead to these reductions in admissions. The DID estimates for heart failure and pneumonia are not significant for Other Race beneficiaries. Supplemental analysis examining whether PHQID hospitals that were close to the thresholds for quality bonuses were more likely to avoid minority patients did not find evidence of this effect (see Appendix SA2).

The source of admission analysis indicates that a reduction in emergency room admissions at PHQID hospitals in the postperiod was the factor contributing to the largest reduction in admissions to PHQID hospitals for Other Race beneficiaries (see Appendix SA3). However, the differences-in-differences-in-differences (DIDID) estimate of this reduction was not significant at $p < .10$.

Table 3 shows the adjusted CABG rates before and after the implementation of the PHQID, the pre–post differences in CABG rates, DID estimates ($[\text{PHQID Post} - \text{PHQID Pre}] - [\text{Non-PHQID Post} - \text{Non-PHQID Pre}]$), and DIDID estimates ($\text{DID Nonwhite} - \text{DID White}$) derived from the individual-level linear probability model (equation [2]) for patients diagnosed with AMI. Table 3 indicates that adjusted CABG rates decreased from the pre- to the postperiod for each racial and ethnic cohort in both PHQID and non-PHQID hospitals, reflecting an overall substitution towards PTCA

Table 2: Adjusted Admission Rates to PHQID Hospitals for Patients Living in Hospital Referral Regions Served by PHQID Hospitals, by Race, for AMI, Heart Failure, and Pneumonia Patients before and after the PHQID

	<i>Pre-PHQID</i> (Q4 2000–Q3 2003) (%)	<i>Post-PHQID</i> (Q4 2004–Q3 2006) (%)	<i>Post-Pre</i> <i>Difference</i> (%)	<i>Difference-in-Differences</i> (vs. White) (%)
AMI				
White	20.8	21.1	0.2 (0.2)	—
Black	20.9	21.4	0.4 (0.3)	0.2 (0.3)
Hispanic	17.6	17.9	0.3 (0.9)	0.1 (0.9)
Other race	21.3	20.0	−1.3 (0.7)	−1.5** (0.6)
Nonwhite	20.6	20.7	0.2 (0.3)	−0.1 (0.3)
Heart failure				
White	17.5	17.7	0.2 (0.1)	—
Black	18.3	18.8	0.5 (0.3)	0.3 (0.3)
Hispanic	15.8	16.1	0.3 (0.4)	0.1 (0.4)
Other race	17.0	16.7	−0.3 (0.5)	−0.5 (0.5)
Non-white	17.9	18.3	0.4 (0.3)	0.2 (0.3)
Pneumonia				
White	16.3	16.3	0.0 (0.1)	—
Black	16.6	16.8	0.2 (0.2)	0.2 (0.2)
Hispanic	13.8	14.5	0.8 (0.8)	0.8 (0.8)
Other race	15.1	15.3	0.3 (0.4)	0.2 (0.4)
Non-white	16.0	16.3	0.3 (0.2)	0.3 (0.2)

Notes. Heteroskedasticity robust standard error shown in parentheses.

p values not shown for predicted levels of pre- and postadmission rates.

Admission rates are adjusted for patient characteristics (age, gender, 30 dummy variables for the Elixhauser comorbidities type of admission [emergency, urgent, elective], and season of admission). Adjusted rates are shown at the mean of model variables.

DID (difference-in-differences) = (Postwhite − Prewhite) − (Postminority − Preminority).

***p* < 0.05,

AMI, acute myocardial infarction; PHQID, Premier Inc. Hospital Quality Incentive Demonstration.

Table 3: Adjusted CABG Rates for Medicare Beneficiaries Diagnosed with AMI in PHQID and Non-PHQID Hospitals by Race

	Non-PHQID (%)		PHQID (%)		DID (Non-PHQID vs. PHQID) (%)		DIDID (vs. White) (%)	
	Pre (Q4 2000-Q3 2003)	Post (Q4 2004-Q3 2006)	Pre (Q4 2000-Q3 2003)	Post (Q4 2004-Q3 2006)	Post-Pre	Post-Pre		
White	13.8	12.3	14.2	12.8	-1.5 (0.1)	-1.3 (0.4)	0.2 (0.4)	—
Black	10.2	9.2	10.5	9.4	-1.0 (0.2)	-1.1 (0.5)	-0.1 (0.6)	-0.3 (0.5)
Hispanic	13.6	12.0	13.9	12.4	-1.6 (0.5)	-1.5 (0.8)	0.1 (0.8)	-0.1 (0.7)
Other	15.4	13.1	15.7	12.1	-2.3 (0.5)	-3.6 (0.9)	-1.4 (0.9)	-1.5* (0.9)
Non-white	11.5	10.3	11.8	10.2	-1.2 (0.2)	-1.6 (0.5)	-0.4 (0.5)	-0.6 (0.4)

Notes: Heteroskedasticity robust standard error shown in parentheses.

p values not shown for predicted levels of pre and post CABG rates.

CABG rates are adjusted for patient characteristics (age, gender, 30 dummy variables for the Elixhauser comorbidities type of admission [emergency, urgent, elective], and season of admission). Adjusted rates are shown at the mean of model variables and at the mean fixed effect for PHQID and non-PHQID hospitals.

DID (difference-in-differences) = (PHQID Post - PHQID Pre) - (Non-PHQID Post - Non-PHQID Pre).

DIDID (difference-in-difference-in-differences) = DID Non-white - DID white, where DID Non-white could be any minority group.

* $p < 0.10$.

AMI, acute myocardial infarction; CABG, coronary artery bypass graft; PHQID, Premier Inc. Hospital Quality Incentive Demonstration.

(see Appendix SA4). However, DID estimates for each racial and ethnic cohort are small and insignificant, and DIDID estimates are marginally significant ($p < .10$) only for Other Race beneficiaries, indicating minimal evidence of PHQID hospital reduction in the provision of CABG to minority patients. Further, analysis of preintervention trends (see Appendix SA1) indicates that the reduction in CABG at PHQID hospitals for Other Race beneficiaries preceded the start of the PHQID.

Sensitivity analysis that replaces the receipt of PTCA for the receipt of CABG as the dependent variable indicates that, although adjusted PTCA rates increased markedly from the pre- to the postperiod for each racial and ethnic cohort in both PHQID and non-PHQID hospitals, no DID and DIDID estimates were significant (see Appendix SA4). This suggests that increases in PTCA rates did not vary across racial and ethnic cohorts and that PHQID hospitals were not more likely to substitute PTCA for CABG for minority patients.

DISCUSSION

Two related analyses included in this paper indicate that the Premier Hospital Quality Incentive Demonstration has made only a minimal, in any, impact on access to care for racial and ethnic minority patients. The regional analysis, evaluating whether minority patients living in market areas that are served by hospitals participating in the PHQID are treated at PHQID hospitals, shows a significant reduction in the proportion of patients that were treated in PHQID hospitals after the program was implemented only for Other Race Medicare beneficiaries with AMI admissions. Worth noting is that only one of the twelve separate statistical tests that were performed was rejected at $p < .05$, raising the possibility that the significant finding was due to chance. Further, the estimate of the reduction of AMI admissions for Other Race beneficiaries, relative to white beneficiaries, is only 1.5 percentage points, representing a small impact, and the reduction occurred largely before the commencement of the PHQID, casting doubt on whether the PHQID caused this reduction.

The CABG analysis shows little evidence that minority patients diagnosed with AMI became less likely, relative to whites, to receive CABG at PHQID hospitals after the PHQID was implemented. The only marginally significant evidence of a reduction in CABG occurred for Other Race beneficiaries: reductions in CABG rates for black, Latino, and all nonwhite patients as a result of the PHQID were not evident. As in the regional analysis, much of the reduction in CABG for Other Race beneficiaries occurred before the

PHQID began, casting doubt on whether the PHQID caused this reduction. Sensitivity analysis shows that racial variations in the increase in the provision of PTCA for minority patients treated at PHQID hospitals were not observed.

Overall, findings from this study indicate that concerns raised in recent literature that P4P may further contribute to racial and ethnic disparities by reducing access to care (Casalino et al. 2007; Chien et al. 2007; Hood 2007) have not been realized in a recent large-scale hospital-based P4P demonstration. Given the evidence that P4P programs in general (Armour et al. 2001; Town et al. 2005; Petersen et al. 2006; Rosenthal and Frank 2006; Pearson et al. 2008), and of the PHQID specifically (Glickman et al. 2007; Ryan 2009), have not substantially impacted provider behavior to improve quality, it is perhaps not surprising that the negative unintended consequences of the PHQID are also limited.

To our knowledge, this is the first study to empirically test the effect of P4P on access to care for minority patients. The results from the CABG analysis conflict with those found by Werner et al. in their analysis of the effect of public quality reporting on CABG rates. The Werner et al. study employed a similar DIDID estimation strategy (examining whether minority patients experienced a decrease in CABG rates, relative to whites, in New York (which had the public reporting program) relative to comparison states, before and after the public reporting program began). They found that the New York public reporting program decreased CABG rates by 2.0 percentage points for blacks and by 3.4 percentage points for Hispanics (both significant at $p < .01$). The magnitude of these effects is stronger than those observed in the current study, perhaps due to the fact that mortality rates were not publicly reported at either the hospital or surgeon level in the PHQID, potentially decreasing the incentives to avoid minority patients.

Another potential reason why patient avoidance on the basis of race was not observed in this study is that physicians and hospitals in the PHQID did not have sufficient incentives to avoid patients on the basis of unobserved severity. Providers likely incur psychic costs as a result of avoiding patients because of financial incentives (McGuire and Pauly 1991), and the expected financial benefits of doing so may have been too small to justify patient avoidance. This is particularly likely given the fact that quality was determined primarily by process performance (which is less sensitive to patient risk, and consequently patient avoidance) and given the relatively small magnitude of the performance incentives. Also, providers may not have perceived race to be related to unobserved severity: while Table 1 suggests that minorities tend to have more observed severity than whites (which is largely accounted for in

risk adjustment), minority patients may not have been perceived to have greater unmeasured severity.

Furthermore, evidence of minority patient avoidance may not have been observed due to the practice of exception reporting. Absent avoiding the treatment of minority patients, hospitals that feared that minority patients would decrease their quality scores could simply not count minority patients toward their quality performance. Known as exception reporting, hospitals in the PHQID had complete discretion to exclude patients from counting toward their quality performance. Some evidence suggests that exception reporting in public quality reporting is associated with increasing measured process quality (Doran et al. 2008; Ryan et al. 2009). Consequently, PHQID hospitals may have disproportionately excluded minority patients from counting toward their quality scores if they thought that this would increase their scores. Research in this area in the United Kingdom's P4P program is mixed: Doran et al. (2008) found some evidence that exception reporting was more likely for members of low-income households while also finding evidence that exception reporting was *less* likely for racial and ethnic minority patients in the United Kingdom. Further research should examine the effects of exception reporting in P4P and public quality reporting programs and to determine, on net, whether it is a desirable element of these programs.

This study has a number of limitations. First, in the regional analysis, the defining of PHQID market areas by HRRs may be an inadequate means to determine potential patients for PHQID hospitals. However, given that a minimum of 65 percent of Medicare hospitalizations occur in HRRs in which beneficiaries live (Wennberg and Cooper 1996), the misidentification of potential patients is likely not a major issue in this study. Second, the DIDID estimates in the CABG analysis are somewhat imprecise, making inference uncertain and potentially subject to Type II error. However, the effects observed in this study are consistently small while the direction of the effects is inconsistent, suggesting that large standard errors are not chiefly responsible for the null inferences. Third, the PHQID is a relatively small, voluntary pilot program consisting of less than 300 hospitals that are not representative of the population of U.S. hospitals (Ryan 2009). As a result, the findings from this study may not be generalizable to other P4P programs. Fourth, while the PHQID applied to all patients, only its effect on Medicare patients was examined in this analysis. While the use of Medicare data standardizes insurance status across patients, it limits the generality of the findings to nonelderly patients, who may have a different unobserved risk profile than elderly patients. Furthermore, the classification of Medicare beneficiaries' race and

ethnicity is not entirely reliable; specifically, recent research shows that many Asian and Hispanic beneficiaries are misidentified as being white (Eicheldinger and Bonito 2008). This misidentification could have attenuated actual differences between whites and minorities that were examined in this study.

CONCLUSION

While this study found little evidence overall that the PHQID increased disparities in access for minority patients, careful monitoring of the effects of P4P on disparities in other programs is warranted. Other P4P programs, perhaps those that more strongly incentivize outcome performance, have stronger financial rewards or penalties, have more extensive public reporting, or programs that incentivize both hospital and physician performance, may create stronger incentives for the avoidance of minority patients. Future research should examine these issues and explore the effects of severity-based patient avoidance in P4P programs more generally. At the very least, this study explicates an analytic approach that can be employed to measure the avoidance of minority patients in other P4P programs.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This work has been supported by the Jewish Healthcare Foundation under the grant “Achieving System-wide Quality Improvements—A collaboration of the Jewish Healthcare Foundation and Schneider Institutes for Health Policy.” I would like to thank Stanley Wallack, Christopher Tompkins, Jennifer Meagher, and Lawrence Casalino for helpful comments on this paper.

Disclosures: None.

Disclaimers: None.

NOTES

1. See the following for details: <http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/faqs-year1-3.jsp#eligible>
2. Hospital Referral Regions divide the United States into 306 areas that reflect health care markets; HRRs were developed by Dartmouth Atlas researchers. See <http://www.dartmouthatlas.com>
3. According to Angrist and Pischke: “While a nonlinear model may fit the conditional expectation function for limited dependent variables more closely than a linear model, when it comes to marginal effects, this probably matters little” (p. 107).

REFERENCES

- Angrist, J. D. A., and J.-S. A. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Armour, B. S., M. M. Pitts, R. Maclean, C. Cangialose, M. Kishel, H. Imai, and J. Etchason. 2001. "The Effect of Explicit Financial Incentives on Physician Behavior." *Archives of Internal Medicine* 161: 1261–6.
- Burack, J. H., P. Impellizzeri, P. Homel, and J. N. Cunningham. 1999. "Public Reporting of Surgical Mortality: A Survey of New York State Cardiothoracic Surgeons." *Annals of Thoracic Surgery* 68: 1195–200.
- Burton, T. M. 2009. Doctors Fight Plan to Limit Beds in ER. *Wall Street Journal*, February 28.
- Casalino, L. P., G. C. Alexander, L. Jin, and R. T. Konetzka. 2007. "General Internists' Views on Pay-for-Performance and Public Reporting of Quality Scores: A National Survey." *Health Affairs* 26 (2): 492–9.
- Chien, A. T., M. H. Chin, A. M. Davis, and L. P. Casalino. 2007. "Pay for Performance, Public Reporting, and Racial Disparities in Health Care: How Are Programs Being Designed?" *Medical Care Research and Review* 64: 283S–304S.
- Chin, M. H., A. E. Walters, S. C. Cook, and E. S. Huang. 2007. "Interventions to Reduce Racial and Ethnic Disparities in Health Care." *Medical Care Research and Review* 64 (5): 7S–28S.
- Cutler, D. M., and R. S. Huckman. 2003. "Technological Development and Medical Productivity: The Diffusion of Angioplasty in New York State." *Journal of Health Economics* 22 (2): 187–217.
- Damberg, C. L., M. E. Sorbero, A. Mehrotra, S. S. Teleki, S. Lovejoy, and L. Bradley. 2007. "An Environmental Scan of Pay for Performance in the Hospital Setting: Final Report. Prepared for United States Department of Health and Human Services" [accessed on April 28, 2009]. Available at <http://aspe.hhs.gov/health/reports/08/payperform/PayPerform07.html>
- Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh, and M. Roland. 2006. "Pay-for-Performance Programs in Family Practices in the United Kingdom." *New England Journal of Medicine* 355 (4): 375–84.
- Doran, T., C. Fullwood, D. Reeves, H. Gravelle, and M. Roland. 2008. "Exclusion of Patients from Pay-for-Performance Targets by English Physicians." *New England Journal of Medicine* 359 (3): 274–84.
- Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy* 111: 555–88.
- Eggleston, K. 2005. "Multitasking and Mixed Systems for Provider Payment." *Journal of Health Economics* 24: 211–23.
- Eicheldinger, C., and A. Bonito. 2008. "More Accurate Racial and Ethnic Codes for Medicare Administrative Data." *Health Care Financing Review* 29 (3): 27–42.
- Einbinder, L. C., and K. A. Schulman. 2000. "The Effect of Race on the Referral Process for Invasive Cardiac Procedures." *Medical Care Research and Review* 57: 162–80.

- Elixhauser, A., C. Steiner, D. R. Harris, and R. N. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.
- Epstein, A. J. 2006. "Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence." *Medical Care Research and Review* 63: 403–26.
- Glickman, S. W., F. S. Ou, E. R. DeLong, M. T. Roe, B. L. Lytle, J. Mulgund, J. S. Rumsfeld, W. B. Gibler, E. M. Ohman, K. A. Schulman, and E. D. Peterson. 2007. "Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction." *Journal of the American Medical Association* 297 (21): 2373–80.
- Gurwitz, J. H., R. J. Goldberg, J. A. Malmgren, H. V. Barron, A. J. Tiefenbrunn, P. D. F. Frederick, and J. M. Gore. 2002. "Hospital Transfer of Patients with Acute Myocardial Infarction: The Effects of Age, Race, and Insurance Type." *American Journal of Medicine* 112 (7): 528–34.
- Hargraves, J. L., J. J. Stoddard, and S. Trude. 2001. "Minority Physicians' Experiences Obtaining Referrals to Specialists and Hospital Admissions." *Medscape General Medicine* 3 (4): 10.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal Agent Analyses—Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law Economics and Organization* 7: 24–52.
- Hood, R. G. 2007. "Pay-for-Performance—Financial Health Disparities and the Impact on Healthcare Disparities." *Journal of the National Medical Association* 99 (8): 953–8.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Institute of Medicine. 2006. *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: National Academy Press.
- Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler. 2007. "Public Reporting and Pay for Performance in Hospital Quality Improvement." *New England Journal of Medicine* 356: 486–96.
- McGuire, T. G., J. Z. Ayanian, D. E. Ford, R. E. M. Henke, K. M. Rost, and A. M. Zaslavsky. 2008. "Testing for Statistical Discrimination by Race/Ethnicity in Panel Data for Depression Treatment in Primary Care." *Health Services Research* 43 (2): 531–51.
- McGuire, T. G., and M. V. Pauly. 1991. "Physician Response to Fee Changes with Multiple Payers." *Journal of Health Economics* 10 (4): 385–410.
- Millett, C., J. Gray, S. Saxena, G. Netuveli, K. Khunti, and A. Majeed. (2007). "Ethnic Disparities in Diabetes Management and Pay-for-Performance in the UK: The Wandsworth Prospective Diabetes Study." *PLoS Med* 4 (6): e191. DOI:10.1371/journal.pmed.0040191
- Narins, C. R., A. M. Dozier, F. S. Ling, and W. Zareba. 2005. "The Influence of Public Reporting of Outcome Data on Medical Decision Making by Physicians." *Archives of Internal Medicine* 165: 83–7.
- Pearson, S. D., E. C. Schneider, K. P. Kleinman, K. L. Coltin, and J. A. Singer. 2008. "The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001–2003." *Health Affairs* 27 (4): 1167–76.

- Petersen, L. W., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145: 265–72.
- Rosenthal, M. B., and R. G. Frank. 2006. "What is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63: 135–57.
- Ryan, A. M. 2009. "The Effects of the Premier Hospital Quality Incentive Demonstration on Mortality and Hospital Costs in Medicare." *Health Services Research* 64 (3): 821–42.
- Ryan, A. M., C. P. Tompkins, J. F. Burgess, and S. S. Wallack. 2009. "The Relationship between Performance on Medicare's Process Quality Measures and Mortality: Evidence of Correlation, Not Causation." *Inquiry* 46: 274–90.
- StataCorp. 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.
- Town, R., R. Kane, P. Johnson, and M. Butler. 2005. "Economic Incentives and Physicians' Delivery of Preventive Care—A Systematic Review." *American Journal of Preventive Medicine* 28: 234–40.
- Wennberg, J. E., and M. M. Cooper. 1996. *The Dartmouth Atlas of Health Care 1996*. Chicago: American Hospital Publishing.
- Werner, R. M., D. A. Asch, and D. Polsky. 2005. "Racial Profiling—The Unintended Consequences of Coronary Artery Bypass Graft Report Cards." *Circulation* 111 (10): 1257–63.
- Williams, R. L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56 (2): 645–6.
- Wooldridge, J. A. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1. Time Tends in PHQID Admissions and CABG Procedures by Race, 2001–2006.

Appendix SA2. Effect of Proximity to PHQID Bonus Thresholds on the Avoidance of Minority Patients.

Appendix SA3. Source of Admission Analysis.

Appendix SA4. Effect of PHQID on Disparities in Receipt of PTCA.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.