Financial and Quality Impacts of the Medicare Physician Group Practice Demonstration

Gregory Pope, John Kautter, Musetta Leung, Michael Trisolini,
Walter Adamache, and Kevin Smith

RTI International

Objective: To examine the impact of the Medicare Physician Group Practice (PGP) demonstration on expenditure, utilization, and quality outcomes.

Data Source: Secondary data analysis of 2001–2010 Medicare claims for 1,776,387 person years assigned to the ten participating provider organizations and 1,579,080 person years in the corresponding local comparison groups.

Study Design: We used a pre-post comparison group observational design consisting of four pre-demonstration years (1/01–12/04) and five demonstration years (4/05–3/10). We employed a propensity-weighted difference-in-differences regression model to estimate demonstration effects, adjusting for demographics, health status, geographic area, and secular trends.

Principal Findings: The ten demonstration sites combined saved $171 (2.0%) per assigned beneficiary person year (p<0.001) during the five-year demonstration period. Medicare paid performance bonuses to the participating PGPs that averaged $102 per person year. The net savings to the Medicare program were $69 (0.8%) per person year. Demonstration savings were achieved primarily from the inpatient setting. The demonstration improved quality of care as measured by six of seven claims-based process quality indicators.

Conclusions: The PGP demonstration, which used a payment model similar to the Medicare Accountable Care Organization (ACO) program, resulted in small reductions in Medicare expenditures and inpatient utilization, and improvements in process quality indicators. Judging from this demonstration experience, it is unlikely that Medicare ACOs will initially achieve large savings. Nevertheless, ACOs paid through shared savings may be an important first step toward greater efficiency and quality in the Medicare fee-for-service program.

Keywords: Medicare Physician Group Practice demonstration, accountable care organization, shared savings, cost efficiency, quality of care

Introduction

With continuing interest in improving Medicare quality of care and controlling its costs, Congress and the Centers for Medicare & Medicaid Services (CMS) are exploring alternative approaches to Medicare reform. Accountable Care Organizations (ACOs), implemented in Medicare as the Shared Savings Program or MSSP, are one of the key reform initiatives in the traditional Medicare fee-for-service (FFS) program that were authorized under The Patient Protection and Affordable Care Act of 2010 (ACA) (Berwick, 2011; Fineberg, 2012; Fisher, McClellan, & Safran, 2011). The MSSP was built directly on its predecessor, the Medicare Physician Group Practice (PGP) demonstration, which was Medicare's first physician pay-for-performance initiative (Centers for Medicare & Medicaid Services, 2011). Like the MSSP, the PGP demonstration explicitly established financial incentives for both cost savings and quality improvement through a shared savings model (Kautter, Pope, Trisolini, & Grund, 2007). This article presents the results of the comprehensive CMS-funded evaluation of the PGP demonstration, which had unique access to demonstration participants and data (Kautter *et al.*, 2012).

The PGP demonstration relied on the physician group as the organizational means to improve the quality and efficiency of care. PGPs shared savings they created in the care of beneficiaries assigned to them with the Medicare program and retained more of the savings the higher their measured quality of care. PGPs faced the business risk of investments to improve quality and efficiency without any upfront payments from Medicare and the risk of foregone FFS revenues. However, that financial risk was mitigated by CMS' continuance of FFS payments, its use of provider-specific base

costs as a starting point for measuring savings, and its lack of penalties for underperformance.

CMS implemented the 5-year PGP demonstration on April 1, 2005. The demonstration's "base year" for measuring quality and efficiency improvements was calendar year 2004, and the five "performance years" ran consecutively from April to March starting in 2005. There were 10 participants in the demonstration: Forsyth Medical Group (North Carolina), Middlesex Health System (Connecticut), Dartmouth-Hitchcock Clinic (New Hampshire), Geisinger Clinic (Pennsylvania), University of Michigan (Michigan), St. John's (now Mercy) Health System (Missouri), Park Nicollet Health Services (Minnesota), Marshfield Clinic (Wisconsin), Billings Clinic (Montana), and Everett Clinic (Washington State). The participating organizations were all large, each having at least 200 affiliated physicians. Half were located in predominantly rural or small city areas, and no participant was located in the core of a large city. Two participants were faculty group practices within academic medical centers; five belonged to an integrated delivery system consisting of at least one hospital in addition to the physician group; two were freestanding physician group practices; and one was a hospital sponsored physician network of small groups and individual physician practices. In the remainder of this article, we refer to the participating PGPs as PGP 1 through PGP 10. To mask their identities, this numbering is not in the order the PGPs are listed above.

The PGP demonstration was not designed to test specific interventions; therefore, participating sites had complete autonomy in determining strategies to achieve higher quality care and expenditure savings. Site visits and interviews with PGP staff conducted by the evaluation team, together with annual demonstration implementation reports from each of the PGPs, indicated that these strategies were not uniformly designed, defined, or implemented across the 10 PGPs, although there were a number of commonalities. In general, the strategies could be classified as either process interventions, which were implemented throughout a larger system, or program interventions, which often targeted a specific population and required patient or beneficiary enrollment.

We found that process interventions were widespread and included patient registries and electronic medical records, information system interventions (e.g., automated alert systems in medical records), medication reconciliation programs, educational interventions for physicians and staff regarding evidence-based care guidelines, and reporting or feedback to encourage adherence to care protocols.

Some PGPs had disease specific registries that identified patients with specific conditions and generated lists of patients that should be followed for some form of care management. The most comprehensive patient registry was found at PGP 9. Staff at this PGP indicated that the PGP demonstration had acted as a catalyst for the development of their registry, which was populated through automatic feeds from several different databases maintained throughout their health system. The registry was used to track patients and identify gaps in care, enabling appropriate care to be provided in a timely fashion. The PGPs were at various stages of development on their electronic medical records (EMRs) during the demonstration. A few PGPs had fully developed EMRs, whereas one PGP did not have any EMR. The remaining PGPs had some form of EMR, but still in development.

Information system interventions related to the use of a registry or EMR were used to improve care provided to the beneficiary. Examples of these interventions were the visit planner at PGP 9, an intervention list at PGP 6, and alert systems built in to the EMRs at groups such as PGP 5 and PGP 6.

The core role of these types of interventions was to recognize gaps in care and ensure that they were provided during the next patient encounter. They also prevented the occurrence of prolonged gaps in the future. The visit planner at PGP 9 was printed for the provider to review and use at the point of care. It listed the services required for the visit and reminded the physician and other providers to supply specific services during the encounter. Several groups, including PGP 1, PGP 5, and PGP 6, used EMRs with automated alert systems to ensure that appropriate care was provided.

PGP 1 found that medication reconciliation activities needed to be better integrated into the patient visit process. The definition of medication reconciliation and activities believed to be required for successful medication reconciliation varied across the PGPs participating in the demonstration. However, there was general consensus that medication reconciliation is important for avoiding adverse outcomes, particularly in the Medicare population, which includes many beneficiaries taking several different medications.

Several groups found that providing periodic feedback to physicians and staff on quality and performance measures improved metrics throughout their systems. Examples of feedback mechanisms existed at several groups, including PGP 2 where intranet feedback systems were in place for individual physicians to view their quality metrics online. Issues with quality metrics or physician performance were reviewed by PGP management teams.

In addition to process re-design interventions, groups implemented several clinically based care management programs that targeted specific patient populations. These can be classified as programs that targeted specific diseases or conditions or programs that targeted a subset of beneficiaries based on cost or patient complexity. Most of the PGPs implemented disease related programs that were expected to generate cost savings, such as congestive heart failure care management programs to reduce hospitalizations and readmissions. Other programs addressed anticoagulation therapy, diabetes, chronic obstructive pulmonary disease, cancer, psychiatric conditions, coronary artery disease, and hypertension. Care management programs for these conditions most often involved education for patient self-management techniques and periodic patient follow-up and assistance with scheduling of appointments and coordination of care. The programs also encouraged adherence to prescribed treatment and attempted to detect and arrest any deterioration of health status before it necessitated expensive interventions such as hospitalization.

Additional programs existed at several PGPs that were not disease-based. Several of these programs focused on patients with multiple chronic diseases or patients who were high cost or high risk. The "Gold Star Population" at PGP 2 for example, was defined as a population that had three or more select comorbidities, seven or more evaluation and management visits, or had been hospitalized with charges of $10,000 or more. Once identified, this population received either complex care coordination or a more formal health coaching intervention. PGP 7 and PGP 10 provided general care coordination services once a patient was discharged and was receiving home care services or other post-acute care services.

Methods

Study Population

The study population consisted of Medicare FFS beneficiaries between January 2001 and March 2010. For each PGP and year, the intervention group was comprised of beneficiaries who received a plurality of "office or other outpatient" evaluation and management (E&M) services from the PGP. For each PGP and year, all counties containing at

least 1% of the assigned intervention beneficiaries formed the comparison area. Non-intervention beneficiaries with one or more "office or other outpatient" E&M visits (and no E&M visits at PGP providers) were randomly sampled from these areas to form comparison groups for each PGP, balancing the number of intervention beneficiaries in each county.[1]

Study Design

A nonrandomized pre-post comparison group observational design was used in the study consisting of four years (January 2001–December 2004) prior to PGP intervention activities and five years (April 2005–March 2010) afterward. The design had elements of both repeated measures and cross-sectional designs because individual beneficiaries could qualify for their group as many as 9 times. Medicare 100% administrative claims and enrollment data for each year in the study period were used to create the study's analytic dataset.

To adjust for potential differences between the intervention and comparison groups, we estimated propensity scores for each PGP site and study year. A propensity score is the predicted probability that a beneficiary was a member of the PGP's assigned beneficiaries, conditional on observed covariates. We plotted the distribution of propensity scores, checked for adequate overlap between intervention and comparison groups, and removed beneficiaries with probabilities of intervention group membership that were less than 0.10 or greater than 0.90.

Outcomes

The study had three primary outcomes. First, Medicare expenditures overall and for six cost components (inpatient hospital, skilled nursing facility, institutional (hospital) outpatient, Part B physician/supplier,

home health, and durable medical equipment). The demonstration did not include hospice payments. Second, health care utilization measured by annual hospital stays and emergency department (ED) visits. Third, claims-based quality measures for diabetes mellitus (HbA1c management, lipid measurement, nephropathy care, eye exams), congestive heart failure (left ventricular ejection fraction testing), coronary artery disease (lipid profile), and preventive care (breast cancer screening). Demonstration participants also reported medical-records-based quality indicators for the intervention group, but we do not analyze them here because those indicators were not available for the comparison group.

Statistical Analysis

A repeated cross section difference-in-differences regression model (Imbens & Wooldridge, 2009) was used to estimate demonstration effects. The model contained an indicator for the demonstration period (April 2005–March 2010), an indicator distinguishing PGP beneficiaries from comparison group beneficiaries, and the interaction between these two indicators. The interaction term estimates the average annual effect of the PGP demonstration during the demonstration period. The model also included indicators for each study year as well as for county, the Hierarchical Condition Categories' (HCC) concurrent risk score to measure disease severity, and demographic covariates (age/sex group, race/ethnicity, Medicare-Medicaid dual eligible status, Medicare eligibility by end stage renal disease (ESRD), and original eligibility for Medicare by disability).

The equation for the general statistical model specification is:[2]

---

[1] In some counties, the maximum number of comparison beneficiaries available was less than the number of intervention beneficiaries.

[2] We also specified a time trend model that allowed for separate pre-demo and post-demo slopes. That model estimated an annual impact of the Demonstration that was similar to our preferred model. We preferred the fixed effects model because it provides the most parsimonious overall summary of the annual effect of the Demonstration.

$$E_{iy} = a + b_1{}^*D + b_2{}^*P + b_3{}^*D^*P + b_j{}^*X_j + b_k{}^*X_k + b_m{}^*X_m + e \qquad (1)$$

where:

E_{iy} = the annualized Medicare expenditure amount for beneficiary i in year y,

a = an intercept term,

D = an indicator coded 1 for PGP assigned beneficiaries and 0 for comparison beneficiaries,

P = a period indicator coded 1 for the Demonstration performance years (PY1–PY5) and 0 for the pre-Demonstration period (2001–2004),

X_j = a vector of j beneficiary-level covariates,

X_k = a vector of k year indicators coded for each of the years from 2002 to PY5, with 2001 serving as the reference year,

X_m = a vector of indicators for individual counties within each geographic area,[3]

$b_1, b_2, b_3, b_j, b_k,$ and b_m are regression coefficient vectors, and

e = a residual term.

In Equation (1), the coefficient of primary interest is b_3. This interaction coefficient estimates the average annual effect of the Demonstration on annual expenditures during the Demonstration performance years compared to comparison group beneficiaries during that period. Coefficient b_1 adjusts for constant annual differences between the groups that persist throughout the study period, and b_2 estimates increased expenditures during the performance years that were common to both groups.

Model specification varied by outcome measure. Expenditure analyses were estimated by weighted least squares, logistic regression

was used for the binary quality of care measures, and utilization results were estimated by a two-part model consisting of separate equations for the probability of any utilization as well as utilization among service users. Each analysis was conducted for all 10 PGPs combined, as well as separately by PGP. Individual observations in the data are not independent because many beneficiaries appear in multiple years. We used the cluster option in Stata 12.0 to correct standard errors for this clustering. Observations were weighted by inverse propensity scores and by the fraction of each year that beneficiaries were eligible for Medicare (Schafer & Kang, 2008).

For the expenditure analyses by subgroup, four additional terms were added to the difference-in-differences model described above: 1) main effect for the subgroup, 2) two-way interaction of the subgroup by assigned beneficiary status, 3) two-way interaction of the subgroup by demonstration period, and 4) three-way interaction of subgroup by assigned beneficiary status by demonstration period. The 3-way interaction was used to identify subgroup effects (Pocock *et al.*, 2002). These models were estimated on the full sample of assigned and comparison group beneficiaries, not on the subsamples consisting of subgroup members only.

Results

Study Population

Exhibit 1 presents descriptive statistics for the intervention and comparison group across all 10 PGPs. The person years (N) across all PGP sites and nine years is 1,776,387 for the intervention group and 1,579,080 for the comparison group, for an overall study sample size of 3,355,467. The distributions in Exhibit 1 are not weighted by the propensity scores. The intervention and comparison groups during the pre-demonstration

---

[3] We did not include site indicators in our pooled sites model because our model already contained a large set of county indicators, which provide for more fine-grained geographic control. As a sensitivity analysis, we added nine site dummies to our model and re-ran it. The savings estimate did not change at all.

Exhibit 1. Descriptive Statistics for PGP Demonstration Evaluation Intervention and Comparison Groups

	Intervention Group (N = 1,776,387)		Comparison Group (N = 1,579,080)	
	Pre-Demo Period	**Demo Period**	**Pre-Demo Period**	**Demo Period**
Age (%)				
0–64	12.9	17.0	13.3	16.5
65–74	41.0	39.3	40.6	39.4
75–84	34.1	31.4	33.7	31.2
85+	12.0	12.3	12.5	12.9
Sex (%)				
Male	41.9	42.1	41.1	42.1
Female	58.1	57.9	58.9	57.9
Race (%)				
White	96.9	95.8	95.1	94.5
Black	1.8	2.4	3.4	3.5
Other	1.2	1.8	1.5	2.0
Dual Eligible (%)				
Medicare & Medicaid	12.6	15.6	13.7	15.6
Medicare–only	87.4	84.4	86.3	84.4
Risk Score (mean)	0.91	1.03	0.86	0.93

NOTE: [1]The sample size (N) is person years.

SOURCE: Authors' analysis of 2001–2010 Medicare administrative data.

period (January 2001–December 2004) and demonstration period (April 2005–March 2010) are similar on the demographic characteristics shown in Exhibit 1. Mean HCC risk scores differed by group and time period. The intervention group had a higher risk score in the pre-demonstration period than the comparison group (0.91 vs. 0.86), suggesting a somewhat sicker intervention population for the PGPs prior to the start of the demonstration. Mean risk scores increased between the pre-demonstration and demonstration periods in both the intervention and comparison groups, but at a higher rate in the intervention group.

Medicare Expenditures

Estimated Demonstration savings for pooled PGPs, by individual PGP, by cost component, and by subpopulation are shown in Exhibit 2.

The overall impact of the demonstration across all PGP sites was a savings of $171 per assigned beneficiary person year during the demonstration performance period (standard error = $22, 95% confidence interval = $127 to $215, p<0.001). This represents a savings of 2.0 percent of assigned beneficiary expenditures. CMS paid performance bonuses to the participating PGPs that averaged $102 per assigned beneficiary person year across the five demonstration years and all 10 PGPs (Centers for Medicare & Medicaid Services, 2011). Hence, we estimate that the PGP demonstration generated net savings to the Medicare program of $69, or 0.8 percent, per demonstration assigned beneficiary person year. (Detailed results for overall expenditures are in an Appendix Exhibit.)

The PGP-specific estimates and their 95% confidence intervals are displayed in Exhibit 3 as well as being reported in Exhibit 2. The plot shows

Exhibit 2. Covariate Adjusted Difference in Differences in Mean Annualized Medicare Expenditures between the PGP Intervention Group and the Comparison Group—Demonstration Period Yearly Average

	Estimated Demonstration Savings	Standard Error	P-Value
Medicare Expenditures, Pooled PGPs	−$171	$22	<0.001
Medicare Expenditures by PGP			
PGP 1	323	79	<0.001
PGP 2	−188	64	0.003
PGP 3	−229	94	0.015
PGP 4	87	74	0.244
PGP 5	−310	59	<0.001
PGP 6	−818	53	<0.001
PGP 7	−26	102	0.798
PGP 8	−142	69	0.041
PGP 9	21	49	0.675
PGP 10	120	91	0.191
Medicare Expenditures by Cost Components			
Inpatient, Facility	−228	18	<0.001
Hospital Inpatient	−176	16	<0.001
Skilled Nursing Facility	−68	8	<0.001
Outpatient/Professional	25	12	0.043
Hospital Outpatient	85	7	<0.001
Physician/Supplier	−39	7	<0.001
Home Health	−22	3	<0.001
Durable Medical Equipment	0	3	0.934
Medicare Expenditures by Subpopulations			
Chronic Conditions			
Cancer	−181	90	0.044
Congestive heart failure	−687	105	<0.001
Diabetes	−456	58	<0.001
Chronic obstructive pulmonary disease	−522	87	<0.001
Acute ischemic heart disease	−602	221	0.006
Stroke	−775	190	<0.001
Vascular disease	−535	95	<0.001
Any of 7 above conditions	−337	39	<0.001
Medicare/Medicaid dual eligibility	−90	67	0.177

(Continued)

Exhibit 2 Continued. Covariate Adjusted Difference in Differences in Mean Annualized Medicare Expenditures between the PGP Intervention Group and the Comparison Group—Demonstration Period Yearly Average

	Estimated Demonstration		
	Savings	Standard Error	P-Value
Aged but originally entitled to Medicare by disability	−361	108	0.001
End stage renal disease	497	663	0.454
Currently entitled to Medicare by disability	65	67	0.331
Upper 10% risk score	−1,922	190	<0.001
Upper 25% risk score	−1,254	100	<0.001
Hospitalized	−402	82	<0.001

NOTES: [1]Estimates are derived from multivariate regression models, including demographic and geographic covariates, pre-existing trends, and the risk score.

[2]Regression estimates are weighted by each person/year's inverse propensity score and the fraction of each year eligible for Medicare.

[3]A negative value for demonstration savings indicates a savings, a positive value a dis-savings.

The Physician/Supplier cost component includes physician/professional expenditures in the inpatient setting.

SOURCE: Authors' analysis of 2001–2010 Medicare administrative data.

Exhibit 3. Forest Plot of Annual Per Person Savings Estimate by PGP
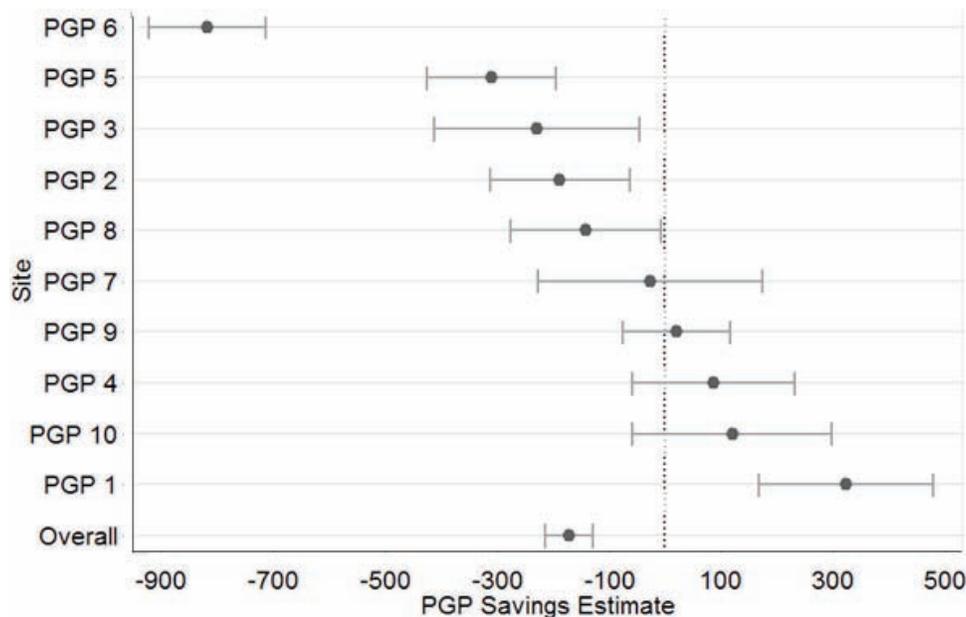
NOTE: The horizontal bars are the 95 percent confidence intervals of the savings estimates.
SOURCE: Authors' analysis of 2001–2010 Medicare Administrative Data.

that the combined expenditure effect was largely attributable to PGP 6, while the effects for the other nine practices were distributed around zero.

Across all 10 PGPs, demonstration savings were achieved primarily from the inpatient facility setting (savings = $228, p<0.001). The estimated demonstration impact on total outpatient/professional expenditures indicates slight dis-savings (dis-savings = $25, p=0.043), possibly indicating some degree of substitution

of outpatient for inpatient services among the demonstration PGPs.

The demonstration generated statistically significant per person year savings for beneficiaries diagnosed with major chronic conditions, with risk scores in the upper 10 percent or upper 25 percent of the intervention group, who were hospitalized and who were currently entitled for Medicare by age, but originally entitled by disability. There were no statistically significant demonstration effects for beneficiaries diagnosed with ESRD, beneficiaries currently entitled to Medicare by disability, or beneficiaries dually eligible for Medicare and Medicaid.

Utilization

Demonstration impacts on utilization derived from the hospital stay and ED visit two-part model regressions are shown in Exhibit 4. For all PGP assigned beneficiaries combined, the probability of a hospital stay and ED visit fell, respectively, 0.0048 more (p<0.01) and 0.0060 more (p<0.01) than for the comparison beneficiaries. For beneficiaries with at least one hospitalization, the reduction in number of hospitalizations was 0.0056 (p<0.10), and there was no statistically significant effect on ED visits among beneficiaries with at least one visit. The total demonstration effect for all PGP sites combined was a 0.0089 reduction (p<0.01) in hospitalizations and a 0.0137 reduction (p<0.01) in ED visits. In other words, the Demonstration reduced the annual rate of hospitalizations per 1,000 person years from 364 to 355, and the rate of ED visits from 633 to 619.

Quality of Care

For each of the seven Medicare claims-based quality measures for which we had data from both the intervention and comparison groups, we measured performance rates and estimated logistic regressions to adjust performance for covariates and pre-existing trends. Pooled performance rates

across all PGPs are presented in Exhibit 5. By the last year of the demonstration period, the PGP assigned beneficiaries had a higher quality of care (i.e., they received the recommended care) and larger improvements over time compared to their comparison group for all seven quality indicators. This was true even after adjusting for covariates and pre-existing trends: our demonstration effect indicator ("adjusted" difference-in-differences coefficient in Exhibit 5) shows that the demonstration had a significant positive impact on the quality of care patients received in six of the seven indicators and in all four of the diabetes indicators in particular. Quality process improvements attributable to the demonstration ranged from a 0.69 percentage point higher performance rate for HbA1c testing for diabetics to a 5.04 percentage point higher performance rate for medical attention for nephropathy for diabetics.

Analysis of individual PGP quality measure results (not shown in Exhibit 5) showed no inverse relationship between individual PGP quality performance and individual PGP cost savings performance. PGP 6, with the best cost savings performance, had higher quality of care, even after adjusting for covariates and pre-existing trends, for five of the seven quality indicators. PGP 1, with the worst cost savings performance, had higher quality of care, even after adjusting for covariates and pre-existing trends, for two of the seven quality indicators.

Qualitative Analysis

A qualitative analysis of the PGP demonstration implementation experience, conducted jointly by the authors and PGP staff, identified four promising opportunities for improving service delivery that can complement payment policy interventions for achieving larger Medicare savings and greater quality improvements (Trisolini, Aggarwal, Leung, Pope, & Kautter, 2008). These could be tested

Exhibit 4. PGP Demonstration Impacts on Utilization: Absolute Difference-in-Differences Demonstration Effects

PGP	Hospitalizations				Emergency Department Visits			
	Mean Hospitalizations	Probability > 0	Number, Given > 0	Total Effect	Mean visits	Probability > 0	Number, Given > 0	Total Effect
Pooled	0.364	-0.0048***	-0.0056*	-0.0089***	0.633	-0.0060***	-0.0086	-0.0137***
1	0.315	0.0051*	0.0044	0.0086	0.690	-0.0034	0.0195	0.0047
2	0.357	-0.0097***	-0.0193	-0.0208***	0.544	-0.0048*	0.0125	-0.0011
3	0.306	-0.0051	-0.0112	-0.0110*	0.589	0.0093**	0.0269	0.0275
4	0.382	0.0023	-0.0120	-0.0022	0.679	0.0068*	-0.0054	0.0085
5	0.406	-0.0063***	0.0015	-0.0085*	0.657	-0.0093***	-0.0001	-0.0138
6	0.330	-0.0148***	-0.0330***	-0.0335***	0.665	-0.0193***	-0.0472***	-0.0523***
7	0.350	-0.0031	0.0158	0.0026	0.543	-0.0109***	-0.0377*	-0.0355***
8	0.372	-0.0081***	0.0190*	-0.0023	0.560	-0.0024	0.0204	0.0055
9	0.359	0.0016	-0.0035	0.0001	0.676	-0.0049*	-0.0080	-0.0122
10	0.454	0.0005	0.0123	0.0072	0.720	-0.0049*	-0.0107	-0.0105

NOTES: [1]Probabilities of at least one hospitalization and of at least one Emergency Department visit were estimated using logit regression models.

[2]Number of hospitalizations conditional on at least one hospitalization and number of ED visits conditional on at least one visit were estimated with zero-truncated negative binomial models.

[3]Logistic and negative binomial regression models included demographic and geographic covariates, pre-existing trends, and the risk score.

[4]Regression estimates are weighted by each person/year's inverse propensity score and the fraction of each year eligible for Medicare.

*** = significant at 1% level; ** = significant at 5% level; * = significant at 10% level.

SOURCE: Authors' analysis of Medicare claims and enrollment data for 2001 to 2010.

Exhibit 5. Claims-based quality measures of performance over time, with difference-in-differences results

| | | Intervention Groups (PGPs) | | | Comparison Groups (CGs) | | | Percentage-Point Change in Quality Performance | | | | Post-Demonstration Difference-in-Differences | | |
| | | | | | | | | PGPs | | CGs | | | | p-value for adjusted difference*** |
		2002	Base Year (BY)	Performance Year 5 (PY5)	2002	Base Year (BY)	Performance Year 5 (PY5)	Pre-Demo PGP Difference (BY-2002)	Post-Demo PGP Difference (PY5-BY)	Pre-Demo CG Difference (BY-2002)	Post-Demo CG Difference (PY5-BY)	Unadjusted*	Adjusted**	
DM-1	HbA1c Testing	90.9%	91.3%	93.0%	86.2%	88.5%	86.6%	0.41	1.66	2.26	-1.89	3.55	0.69	0.000
DM-4	DM LDL-C Testing	76.9%	82.1%	86.3%	72.9%	78.6%	82.0%	5.18	4.20	5.64	3.45	0.75	2.77	0.000
DM-6	Medical Attention for Nephropathy	72.8%	74.8%	84.3%	65.7%	67.7%	76.4%	2.04	9.43	1.99	8.70	0.74	5.04	0.000
DM-7	DM Eye Exam	69.6%	70.6%	73.4%	64.3%	65.1%	66.4%	1.03	2.85	0.80	1.33	1.52	0.92	0.039
HF-2	Left Ventricular Ejection Fraction Testing	86.5%	85.5%	90.3%	84.3%	87.7%	88.7%	-0.99	4.73	3.43	0.96	3.77	0.42	0.53
CAD-5	Lipid Profile	68.9%	73.6%	78.5%	67.7%	71.4%	75.9%	4.65	4.95	3.71	4.53	0.42	3.15	0.000
PC-5	Breast Cancer Screening	76.9%	76.0%	78.2%	72.5%	70.5%	71.5%	-0.81	2.17	-1.91	0.91	1.26	2.71	0.000

NOTES: * Difference-in-differences estimates based on [(PGPPY5-BY)–(CGPY5-BY)].

** Adjusted using multivariate logistic regression analysis on propensity-weighted PGP and CG samples, controlling for pre-existing trends and covariates.

*** p-values from beta-coefficients of the multivariate logistic regression models.

SOURCE: Author's analysis of 2001–2010 Medicare claims and enrollment data.

further in future demonstration projects, and variations of them are being explored by a number of the new ACOs.

The first opportunity is increasing patient engagement. The PGPs believed that involving patients more deeply in pre-visit processes and self-management support has the potential to improve quality while containing costs. The goals are to make physician visits more effective and accurate in the treatment that can be provided and to enable complementary services to be provided in a more timely fashion if reimbursement can be made available. Much of day-to-day chronic disease care can be provided by patients themselves or by family members. This includes adherence to prescribed medications; consistent attendance at regular physician visits; active communication with physicians and nurses regarding symptoms and problems; prompt attendance for ordered testing services; and maintaining diet and exercise programs as consistently as possible.

The second opportunity is expanding support for care management programs. Many of the PGPs intensified their care management efforts through daily telemonitoring programs, nurse telephone management, patient education, and other interventions. The PGP demonstration incentives provided one way of funding these programs through performance payments for demonstrated cost savings. PGPs were also interested in exploring direct incentives, such as per-member per-month capitated reimbursement for heart failure case management, which could fund a range of non-visit services, such as telephonic nurse case management.

The third opportunity is improving care transitions. Health care providers have historically given too little emphasis to care transitions, partially because clinical responsibilities and associated reimbursements are often divided between providers; however, the demonstration incentives reward PGPs for reducing overall Medicare spending so they have a financial incentive to better manage the many care transitions that may be required for treatment of chronic diseases. A number of PGPs tested new transition management programs that applied to patients with particular diagnoses or for particular types of transitions, such as from hospital to home. Some PGPs also explored management of other types of transitions, such as from hospital to nursing home. Since those organizations are often separate corporations, they typically have not shared data on patients effectively in the past, and communication regarding care transitions has often been incomplete.

The fourth opportunity is expanding the roles of non-physician providers. The PGPs studied redesigning primary care practice to increase the use of non-physicians, such as through greater use of planned visits; integrating care management into clinical practice, such as delegating some types of patient testing or exams (e.g., diabetic foot exams) to non-physicians; expanding patient education; and providing greater data support to physicians to enhance the quality and cost-effectiveness of their clinical work. Physician buy-in to these efforts was sometimes a challenge, but many of the PGPs had success in implementing the new non-physician roles. If the new roles are well-structured, and the staff well-trained, then physicians may view them as complementing the care they provide and enabling them to concentrate on the elements of care that most need their expertise.

Discussion

Our results show a small, but statistically significant, reduction in the level of medical expenditures resulting from the PGP demonstration, which used a shared savings pay-for-performance model similar to the Medicare MSSP model for

ACOs implemented in 2012. We also found that significant utilization reductions and quality improvements across multiple measures resulted from the PGP demonstration. These impacts provide some counterbalance to concerns raised about the potential impact of the PGP demonstration based on its interim evaluation (Sebelius, 2009) that was published during the demonstration period (Berenson, 2010; Iglehart, 2011); however, these results also indicate that achieving large savings in Medicare expenditures may require additional reforms.

As previously mentioned, there was a higher risk score growth among intervention than comparison beneficiaries during the PGP demonstration period. The risk scores are calculated from the diagnoses that providers record on claims they submit to Medicare for reimbursement. Some have questioned whether risk score (diagnosis) coding changes have affected PGP demonstration savings estimates (Colla *et al.*, 2012; MedPAC, 2009; Wilensky, 2011). To investigate this hypothesis, we compared changes in risk scores and mortality rates among the provider organization intervention and comparison beneficiaries during the demonstration period. The mortality rate is a measure of population health status that is independent of the claims diagnoses used to calculate the risk scores, and is not subject to manipulation by providers. The correlation of risk score and mortality growth rates across the 20 intervention and comparison groups was 0.82, indicating that risk score changes were strongly associated with mortality changes. We conclude that our demonstration savings estimates are unlikely to be overestimated because of diagnosis coding changes that affected the risk scores we used to control for health status changes.

Colla *et al.* (2012) report a PGP demonstration annual savings impact of $114 per person year, which was based on their "low variation conditions" (LVC) risk adjustment methodology.[4] These authors however found a substantial increase in savings when using an HCC risk adjustment methodology,[5] finding an annual savings impact of $496 per person year. It is not clear to us how these authors estimated such a high savings of $496 when using an HCC methodology, which is almost three times as high as the savings reported in our study ($171), and which is also substantially higher than the savings calculated by CMS during the PGP demonstration (Centers for Medicare & Medicaid Services, 2011).[6] Interestingly, the preferred savings result of Colla *et al.* (2012) based on their LVC risk adjustment methodology ($114) almost lies in the confidence interval of our preferred savings result based on an HCC risk adjustment methodology (CI = $127 to $215). Further, Colla *et al.* (2012) conclude that "Most of the savings were concentrated among dually eligible beneficiaries." This is also inconsistent with our results, which show an annual savings impact for non-dual eligibles of $186 (p<0.001) and $90 (p=0.177) for dual eligibles.[7]

Based on our savings estimates and the demonstration performance payments, more than half of gross savings were returned to the participating PGPs as incentive payments. Net Medicare program savings were correspondingly reduced to an estimated 0.8 percent. Thus, judging from the experience of the PGP

---

[4] This LVC risk adjustment methodology restricts diagnoses to stroke, acute myocardial infarction, hip fracture, and colorectal cancer.

[5] An HCC risk adjustment methodology is also used in our study.

[6] CMS calculations of savings were based on the PGP Demonstration payment design (Kautter *et al.*, 2007), which also used an HCC risk adjustment methodology.

[7] Certainly, coordination of care across the Medicare and Medicaid programs is an important goal to improve efficiency and quality of care for both programs. Although the PGP interventions did include dual eligibles, the PGP interventions were not specifically focused on dual eligibles like, for example, Medicare Advantage dual eligible special needs plans (SNPs). In addition, the PGP interventions did not include care coordination across the Medicare and Medicaid programs, which is an important aspect to improving outcomes for dual eligibles.

demonstration, it is unlikely that the similar MSSP will initially achieve large reductions in Medicare program expenditures.

Moreover, the participants in the PGP demonstration were large, sophisticated organizations that volunteered for the demonstration based in part on their previous experience and existing infrastructure in managing care and the expectation of doing well under the demonstration. The MSSP is also a voluntary program, but the expertise of the typical MSSP participant likely falls short of the PGP participants. On the other hand, many of the PGP participants were located in relatively low expenditure areas, and thus had limited opportunities to demonstrate cost savings. If the MSSP is able to attract participants from high-expenditure areas, the opportunities for Medicare savings may be greater than in the PGP demonstration. Additionally, in the long run, quality improvements such as those we found for the PGPs could also result in cost savings.

Moreover, the quality improvements we found in this evaluation probably understate the total quality improvement effect of the PGP demonstration. Quality of care was measured for this evaluation by seven claims-based process measures of quality that were available for both the PGPs and the comparison groups (CGs), but these do not cover all aspects of quality of care. A number of medical record-based measures were also included in the financial reconciliation protocol used by CMS to determine performance bonus payments for the PGPs in the demonstration, but they were not included in this evaluation since the lack of available medical-records data for the CGs precluded difference in differences analysis for impact evaluation. All of the PGPs achieved improvements for most of those medical-records based quality measures during the

demonstration. Future evaluations should explore ways to include CG data for a broader array of quality measures, including medical record-based measures of intermediate outcomes (e.g., HbA1c levels, LDL cholesterol levels, and blood pressure levels) and final outcomes (e.g., rates of progression for diabetics to complications such as retinopathy, nephropathy, and neuropathy, rates of progression to cardiovascular diseases, and mortality rates).

What are the barriers to achieving larger Medicare savings and quality improvements in a shared savings model? The organizations participating in the PGP demonstration could not be expected to change their entire business model in response to a time-limited and Medicare-only initiative. More fundamentally, the uncertain prospect of a share of future savings limited PGPs' willingness to make costly new investments. Foregone FFS revenues limited the return on investment for PGPs that included hospitals, since reduced inpatient utilization was the major source of savings. Also, the PGP demonstration established incentives at the level of the physician group that did not necessarily filter down to individual physicians. Although several organizations shared Medicare bonuses with individual physicians, many continued to compensate doctors primarily on the basis of providing more visits and generating more billings. For example, PGP 6, which generated the highest amount of savings during the PGP demonstration, used the performance payments for general organizational development, rather than for individual physician payments, which they believed reinforced their group-oriented, nonprofit organization culture.

The "upside only" nature of shared savings payment provides an inducement for organizations to participate voluntarily, as Medicare, Medicaid, and private insurers organize ACOs and move

away from traditional FFS reimbursement, but it also mitigates the motivation for ACOs to undertake more dramatic service delivery system reforms (Berenson, 2010). Over time, payment models for ACOs that include some downside financial risk, such as the MSSP's "two-sided" model or even global capitation, may become more widespread and they could have larger impacts on cost savings and quality improvement.

In decades past, some managed care organizations pursued a strategy of subjecting physician groups to high-powered financial incentives such as capitation. That experience was generally seen to be unsuccessful as many physician organizations were unable to manage the financial risk or were perceived by the public to be reducing quality of care in response to the financial incentives (Robinson, 2001). With these lessons learned, the current approach in the ACO program is focusing on milder initial financial incentives similar to those used in the PGP demonstration (shared savings with upside incentives only) and larger organizations better able to bear risk (Medicare ACOs must treat a minimum of 5,000 Medicare beneficiaries). Moreover, based on the PGP Demonstration experience, MSSP risk adjustment has been modified to be based on a prospective rather than concurrent HCC model, and to minimize increases in average risk scores over time due to diagnosis coding intensity for beneficiaries continuously assigned to a particular ACO. It is too early to tell which of these or other related approaches now being tested will be most successful in reforming the Medicare program—ACOs, medical homes, bundled episode payment, pay-for-performance, various approaches to capitation, value-based payment, or Medicare Advantage. The next few years will be ones of considerable experimentation with all of these approaches, which are not mutually exclusive. Many of these newly-developed payment approaches may find their niche, either alone or in combination.

Disclaimer

Correspondence

Gregory Pope, M.S., RTI—Health Care Financing and Payment Program, 1440 Main Street, Suite 310, Waltham, Massachusetts, 02451, gpope@rti.org, Tel. (781) 434-1742, Fax (781) 434-1701.

Acknowledgment

References

Berenson, R. A. (2010). Shared Savings Programs for Accountable Care Organizations: A Bridge to Nowhere? *The American Journal of Managed Care*, *16*(10), 721–726. PubMed

Berwick, D. M. (2011). Making Good on ACOs' Promise—The Final Rule for the Medicare Shared Savings Program. *The New England Journal of Medicine*, *365*, 1753–1756. PubMed http://dx.doi.org/10.1056/NEJMp1111671

Centers for Medicare & Medicaid Services (2011, July). Medicare Physician Group Practice

Demonstration: Physician Groups Continue to Improve Quality and Generate Savings Under Medicare Physician Pay-For-Performance Demonstration. Retrieved from https://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/PGP_Fact_Sheet.pdf

Colla, C. H., Wennberg, D., Meara, E., Skinner, J., Gottlieb, D., Lewis, V., Fisher, E. (2012). Spending Differences Associated with the Medicare Physician Group Practice Demonstration. *Journal of the American Medical Association*, *308*(10), 1015–1023. PubMed http://dx.doi.org/10.1001/2012.jama.10812

Fineberg, H. V. (2012). A Successful and Sustainable Health System—How to Get There from Here. *The New England Journal of Medicine*, *366*(11), 1020–1027. PubMed http://dx.doi.org/10.1056/NEJMsa1114777

Fisher, E. S., McClellan, M., & Safran, D. (2011). Building the Path to Accountable Care. *The New England Journal of Medicine*, *365*(26), 2445–2447. PubMed http://dx.doi.org/10.1056/NEJMp1112442

Iglehart, J. K. (2011). Assessing an ACO Prototype—Medicare's Physician Group Practice Demonstration. *The New England Journal of Medicine*, *364*(3), 198–200. PubMed http://dx.doi.org/10.1056/NEJMp1013896

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86. http://dx.doi.org/10.1257/jel.47.1.5

Kautter, J., Pope, G. C., Trisolini, M., & Grund, S. (2007). Medicare Physician Group Practice Demonstration Design: Quality and Efficiency in Pay-for-Performance. *Health Care Financing Review*, *29*(1), 15–29. PubMed

Kautter, J., Pope, G. C., Leung, M., Trisolini, M., Adamache, W., & Smith, K. … Schwartz, M. (2012, September). Evaluation of the Medicare Physician Group Practice Demonstration. Final Report Prepared for Centers for Medicare and Medicaid Services Under Contract Number HHSM-500-2005-00029I. Retrieved from http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/Downloads/PhysicianGroupPracticeFinalReport.pdf

MedPAC (Medicare Payment Advisory Committee) (2009, June). Report to Congress: Improving Incentives in the Medicare Program. Retrieved from http://www.medpac.gov/documents/jun09_entirereport.pdf

Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Statistics in Medicine*, *21*, 2917–2930. PubMed http://dx.doi.org/10.1002/sim.1296

Robinson, J. C. (2001). The End of Managed Care. *Journal of the American Medical Association*, *285*(20), 2622–2628. PubMed http://dx.doi.org/10.1001/jama.285.20.2622

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*, 279–313. PubMed http://dx.doi.org/10.1037/a0014268

Sebelius, K. (2009). Report to Congress: Physician Group Practice Demonstration Evaluation Report. Retrieved from http://www.cms.

gov/Medicare/Demonstration-Projects/
DemoProjectsEvalRpts/downloads/PGP_
RTC_Sept.pdf

Trisolini, M., Aggarwal, J., Leung, M., Pope, G. C., & Kautter, J. (2008). The Medicare Physician Group Practice Demonstration: Lessons Learned on Improving Quality and Efficiency in Health Care. New York: The Commonwealth Fund.

Wilensky, G. R. (2011). Lessons from the Physician Group Practice Demonstration—A Sobering Reflection. *The New England Journal of Medicine*, *365*(18), 1659–1661. PubMed http://dx.doi.org/10.1056/NEJMp1110185

Appendix Exhibit. Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures (standard errors and p-values for statistical significance are shown below coefficient estimates in dollars; regression is estimated for assigned and comparison beneficiaries on 2001–PY5 data)

Variable	All PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
N	3,355,467	207,218	448,090	154,686	238,103	431,700	571,974	217,858	302,177	504,703	278,958
R²	0.584	0.533	0.567	0.602	0.596	0.593	0.566	0.591	0.602	0.566	0.632
Assigned beneficiary	191	171	519	−232	−191	29	163	62	−130	7	1,382
s.e.	15	53	43	66	50	38	33	69	46	33	63
p-v	0.000	0.001	0.000	0.000	0.000	0.453	0.000	0.371	0.005	0.838	0.000
Post*AB (Demo effect)	−171	323	−188	−229	87	−310	−818	−26	−142	21	120
s.e.	22	79	64	94	74	59	53	102	69	49	91
p-v	0.000	0.000	0.003	0.015	0.244	0.000	0.000	0.798	0.041	0.675	0.191
Risk score	7,677	7,250	8,295	7,492	7,343	7,467	7,304	8,698	7,746	7,138	8,076
s.e.	14	53	44	56	47	37	42	58	54	32	45
p-v	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2002	379	170	553	349	327	595	282	612	249	311	451
s.e.	17	60	51	84	64	45	37	85	56	38	81
p-v	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2003	633	376	718	390	539	805	639	852	569	527	934
s.e.	18	63	54	80	67	47	39	85	56	40	79
p-v	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2004	1,065	866	1,143	782	1,081	1,173	909	1,349	953	1,057	1,565
s.e.	19	66	57	84	66	49	41	91	58	42	81
p-v	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Post-demonstration period	1,479	1,101	1,654	1,136	1,251	1,699	1,615	1,500	1,268	1,358	2,053
s.e.	23	81	69	102	79	57	54	112	71	50	96
p-v	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PY2	143	96	233	170	282	50	148	179	226	20	150
s.e.	22	74	67	97	78	58	52	100	70	48	91
p-v	0.000	0.194	0.000	0.078	0.000	0.387	0.005	0.073	0.001	0.685	0.101

(Continued)

Appendix Exhibit Continued. Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures (standard errors and p-values for statistical significance are shown below coefficient estimates in dollars; regression is estimated for assigned and comparison beneficiaries on 2001–PY5 data)

Variable	All PGPs		PGP 1		PGP 2		PGP 3		PGP 4		PGP 5		PGP 6		PGP 7		PGP 8		PGP 9		PGP 10	
		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.
		p-v		p-v		p-v		p-v		p-v		p-v		p-v		p-v		p-v		p-v		p-v
PY3	243		10		356		139		286		185		191		472		325		220		251	
	23		79		70		100		80		64		57		104		73		51		97	
		0.000		0.901		0.000		0.165		0.000		0.004		0.001		0.000		0.000		0.000		0.010
PY4	485		289		580		468		686		367		385		857		526		331		642	
	25		83		72		105		84		68		63		109		77		53		101	
		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000
PY5	378		286		486		486		656		439		93		647		189		268		563	
	25		85		71		106		85		70		61		110		80		54		98	
		0.000		0.001		0.000		0.000		0.000		0.000		0.126		0.000		0.018		0.000		0.000
Male (0=no; 1=yes)	−212		−157		−277		−143		−253		−153		−231		−309		−178		−164		−271	
	13		48		37		55		43		34		31		59		39		29		52	
		0.000		0.001		0.000		0.009		0.000		0.000		0.000		0.000		0.000		0.000		0.000
Age group (0-54)	−360		−579		−229		−476		−356		−309		−730		455		−529		−357		−404	
	30		133		87		144		89		73		83		164		99		63		116	
		0.000		0.000		0.009		0.001		0.000		0.000		0.000		0.005		0.000		0.000		0.000
Age group (55-64)	−466		−473		−315		−532		−695		−432		−498		−445		−646		−386		−561	
	32		127		95		128		94		77		90		156		123		60		127	
		0.000		0.000		0.001		0.000		0.000		0.000		0.000		0.004		0.000		0.000		0.000
Age group (75-84)	−102		−133		91		−128		−5		−335		−134		68		−44		−140		−65	
	13		48		38		57		43		36		31		59		40		30		53	
		0.000		0.006		0.017		0.025		0.914		0.000		0.000		0.250		0.267		0.000		0.215
Age group (> 85)	−206		−576		270		−452		−61		−587		−374		246		38		−269		−169	
	20		69		61		88		69		55		47		93		64		47		83	
		0.000		0.000		0.000		0.000		0.372		0.000		0.000		0.008		0.552		0.000		0.043

(Continued)

Appendix Exhibit Continued. Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures (standard errors and p-values for statistical significance are shown below coefficient estimates in dollars; regression is estimated for assigned and comparison beneficiaries on 2001–PY5 data)

Variable	All PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
Medicaid status (0=no; 1=yes)	−294	−660	−179	−333	63	−428	−380	−694	−59	−204	125
s.e.	24	100	74	108	66	57	61	120	93	48	102
p-v	0.000	0.000	0.016	0.002	0.338	0.000	0.000	0.000	0.526	0.000	0.221
Originally disabled (0=no; 1=yes)	−251	−14	−114	−331	−196	−392	−322	−406	−372	−97	−251
s.e.	31	115	98	133	95	76	77	167	134	61	126
p-v	0.000	0.906	0.243	0.013	0.040	0.000	0.000	0.015	0.006	0.111	0.047
ESRD status (0=no; 1=yes)	−5,196	−3,511	−10,537	−1,192	−2,839	−5,804	−5,148	−7,883	−1,942	−2,038	−7,924
s.e.	186	740	654	705	594	525	495	802	575	473	495
p-v	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Race = black	242	658	−759	452	−134	104	439	−177	535	48	253
s.e.	48	501	288	303	72	245	451	196	130	210	92
p-v	0.000	0.189	0.008	0.137	0.063	0.672	0.331	0.367	0.000	0.819	0.006
Race = Asian	−771	−750	−210	−559	−804	103	−1,080	−580	−1,209	−554	−981
s.e.	89	465	352	163	584	482	263	409	188	335	231
p-v	0.000	0.107	0.551	0.001	0.168	0.831	0.000	0.157	0.000	0.098	0.000
Race = other race	−288	−24	−73	−184	−27	−364	−235	−732	−185	−145	−540
s.e.	61	158	201	238	263	206	162	223	158	187	178
p-v	0.000	0.878	0.716	0.438	0.918	0.076	0.146	0.001	0.241	0.438	0.002
CMS Hierarchical Condition Category (HCC) = cancer	102	654	−315	557	−488	72	387	−200	447	−36	329
s.e.	26	96	73	113	86	66	69	111	90	58	93
p-v	0.000	0.000	0.000	0.000	0.000	0.275	0.000	0.072	0.000	0.532	0.000
HCC = diabetes	−157	−84	−263	−490	30	−152	−205	−181	−207	14	−88
s.e.	17	64	50	73	52	41	42	80	59	38	67
p-v	0.000	0.194	0.000	0.000	0.571	0.000	0.000	0.024	0.000	0.721	0.192

(Continued)

Appendix Exhibit Continued. Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures (standard errors and p-values for statistical significance are shown below coefficient estimates in dollars; regression is estimated for assigned and comparison beneficiaries on 2001–PY5 data)

Variable	All PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
HCC = AMI	743	1,579	621	1,160	615	−1,055	2,358	144	912	1,163	614
s.e.	60	255	191	273	206	139	174	279	227	121	190
p-v	0.000	0.000	0.001	0.000	0.003	0.000	0.000	0.606	0.000	0.000	0.001
HCC = CHF	436	524	374	361	541	347	404	627	418	26	1,488
s.e.	31	108	95	155	100	76	76	142	113	64	116
p-v	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.683	0.000
HCC = stroke	440	−85	584	259	−144	−41	1,883	479	1,768	131	187
s.e.	50	177	168	205	148	111	160	228	202	103	167
p-v	0.000	0.633	0.001	0.207	0.329	0.715	0.000	0.035	0.000	0.204	0.261
HCC = vascular disease	226	150	306	−155	695	−26	408	20	35	109	861
s.e.	27	99	85	116	97	56	72	111	102	58	100
p-v	0.000	0.129	0.000	0.182	0.000	0.641	0.000	0.861	0.732	0.060	0.000
HCC = COPD	58	192	−89	260	30	40	100	5	146	51	334
s.e.	25	86	74	115	77	61	69	112	101	51	99
p-v	0.020	0.025	0.231	0.024	0.701	0.512	0.147	0.967	0.148	0.312	0.001

NOTES: 1. Dependent variable is Medicare annualized expenditures.

2. The regression is estimated on 2001 to PY5 data (2001 to 2010 data) for PGP assigned and comparison group beneficiaries (simulated assigned and comparison group beneficiaries before PY1), selected as described in the text.

3. Regression is weighted by Medicare eligibility fraction and by beneficiary propensity scores as described in the text.

4. Demonstration impact is estimated by the coefficient of (assigned beneficiary)*(performance year). Negative coefficients indicate savings and positive coefficients indicate dis-savings on a per person per year basis.

5. P-values for statistical significance of regression coefficient estimates presented below coefficient estimates. A p-value of '0' indicates that the coefficient is significantly different from zero at better than the 0.1% level of significance. A p-value of, for example, 0.006 indicates a 0.6% level of significance, 0.015 indicates a 1.5% level of significance, a p-value of 0.077 indicates a 7.7% level of significance, a p-value of 0.325 indicates a 32.5% level of significance, etc.

6. Regression models also include dummy variables for county of residence of beneficiaries and a constant term (not shown in exhibit). The 2001 and PY1 year dummy variables are omitted to avoid collinearity.

7. Statistical significance levels (p-values) and coefficient standard errors are adjusted for beneficiary-level clustering.

8. Results do not reflect the Demonstration PY5 risk score cap.

SOURCE: Author's analysis of 2001–2010 Medicare administrative data.