

Spending Differences Associated With the Medicare Physician Group Practice Demonstration

Carrie H. Colla, PhD

David E. Wennberg, MD, MPH

Ellen Meara, PhD

Jonathan S. Skinner, PhD

Daniel Gottlieb, MS

Valerie A. Lewis, PhD

Christopher M. Snyder, PhD

Elliott S. Fisher, MD, MPH

TO IMPROVE CARE AND SLOW cost growth, payers are increasingly turning to new payment models, including accountable care organizations (ACOs). The Centers for Medicare & Medicaid Services (CMS) has launched 3 ACO programs—Pioneer, the Shared Savings Program, and the Advance Payment Model—which differ slightly in their details but share a common approach: participating organizations can share in savings if they meet quality and cost targets for their assigned beneficiaries.^{1,2}

Accountable care organizations were included in the Affordable Care Act in part because simulations suggested that CMS could achieve savings from these models,^{3,4} and an earlier program, the Physician Group Practice Demonstration (PGPD), appeared to be effective. In this demonstration, 10 participating physician groups were eligible for up to 80% of any savings they generated (after crossing a 2% savings threshold) if they were also able to demonstrate improvement on 32 quality measures, including the adequacy of preventive care (eg, colorectal cancer screening) and the effectiveness of chronic disease management (eg, percentage of diabetic patients with most recent low-density li-

Context The Centers for Medicare & Medicaid Services (CMS) recently launched accountable care organization (ACO) programs designed to improve quality and slow cost growth. The ACOs resemble an earlier pilot, the Medicare Physician Group Practice Demonstration (PGPD), in which participating physician groups received bonus payments if they achieved lower cost growth than local controls and met quality targets. Although evidence indicates the PGPD improved quality, uncertainty remains about its effect on costs.

Objective To estimate cost savings associated with the PGPD overall and for beneficiaries dually eligible for Medicare and Medicaid.

Design Quasi-experimental analyses comparing preintervention (2001-2004) and post-intervention (2005-2009) trends in spending of PGPD participants to local control groups. We compared estimates using several alternative approaches to adjust for case mix.

Setting Ten physician groups from across the United States.

Patients and Participants The intervention group was composed of fee-for-service Medicare beneficiaries (n=990 177) receiving care primarily from the physicians in the participating medical groups. Controls were Medicare beneficiaries (n=7 514 453) from the same regions who received care largely from non-PGPD physicians. Overall, 15% of beneficiaries were dually eligible for Medicare and Medicaid.

Main Outcome Measure Annual spending per Medicare fee-for-service beneficiary.

Results Annual savings per beneficiary were modest overall (adjusted mean \$114, 95% CI, \$12-\$216). Annual savings were significant in dually eligible beneficiaries (adjusted mean \$532, 95% CI, \$277-\$786), but were not significant among nondually eligible beneficiaries (adjusted mean \$59, 95% CI, \$166 in savings to \$47 in additional spending). The adjusted mean spending reductions were concentrated in acute care (overall, \$118, 95% CI, \$65-\$170; dually eligible: \$381, 95% CI, \$247-\$515; nondually eligible: \$85, 95% CI, \$32-\$138). There was significant variation in savings across practice groups, ranging from an overall mean per-capita annual saving of \$866 (95% CI, \$815-\$918) to an increase in expenditures of \$749 (95% CI, \$698-\$799). Thirty-day medical readmissions decreased overall (-0.67%, 95% CI, -1.11% to -0.23%) and in the dually eligible (-1.07%, 95% CI, -1.73% to -0.41%), while surgical readmissions decreased only for the dually eligible (-2.21%, 95% CI, -3.07% to -1.34%). Estimates were sensitive to the risk-adjustment method.

Conclusions Substantial PGPD savings achieved by some participating institutions were offset by a lack of saving at other participating institutions. Most of the savings were concentrated among dually eligible beneficiaries.

JAMA. 0;0(10):1015-1023

www.jama.com

poprotein cholesterol level < 130 mg/dL; to convert from mg/dL to mmol/L, multiply by 0.0259).⁵⁻⁸ According to public

reports, all 10 organizations met the quality benchmarks required to be eligible for savings⁹ and some achieved sufficient sav-

Author Affiliations: Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire (Drs Colla, Wennberg, Meara, Skinner, Lewis, and Fisher and Mr Gottlieb); Norris Cotton Cancer Center, Manchester, New Hampshire (Dr Colla); Department of Economics, Dartmouth College, Hanover, New Hampshire (Drs Skinner and Snyder); and

National Bureau of Economic Research, Hanover, New Hampshire (Drs Meara, Skinner, and Snyder).

Corresponding Author: Carrie H. Colla, PhD, Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, 35 Centera Pkwy, Lebanon, NH 03766 (Carrie.H.Colla@Dartmouth.edu).

For editorial comment see p 1038.

ings to receive bonuses. Overall, CMS estimated that PGPD participants reduced spending by \$137 million over the program's 5 years.⁹

Some question whether the magnitude of savings could have been overestimated due to the approach adopted for risk adjustment.^{10,11} The CMS used hierarchical condition categories (HCCs), which use claims-based diagnoses to determine a score for each beneficiary that is used for risk adjustment.¹² The observation that HCC scores increased more rapidly at some PGPD sites than in controls raised concerns that the program's apparent savings may have been due to changes in coding practices rather than improved care.¹¹

Second, nothing is yet known about the overall effect of the PGPD on vulnerable populations, specifically those eligible for both Medicare and Medicaid.^{13,14} Dually eligible beneficiaries are overwhelmingly poor, have little social support,^{15,16} and consume a disproportionate share of Medicare and Medicaid spending because of their multiple, severe health conditions and often co-occurring psychiatric disorders.¹⁷⁻²¹ Vulnerable populations such as the dually eligible are of particular concern because the potential impact of the ACO payment model on their care is uncertain. On the one hand, high-need populations could benefit the most from improved care coordination and chronic disease management. Alternatively, their limited social resources and complex health conditions could lead physician groups to focus instead on other, less challenging populations.

In this article, we estimate the magnitude of savings achieved by the PGPD program for all beneficiaries and for both dually and nondually eligible beneficiaries, while testing the sensitivity of the findings to different risk adjustment approaches.

METHODS

We used Medicare administrative data to analyze changes in spending and diagnostic coding for beneficiaries assigned to each of the 10 PGPD participants and their local control groups.⁸ A benefi-

ciary was assigned to a PGPD medical group if its physicians delivered the predominance of that beneficiary's care; control groups comprised beneficiaries who resided in the same counties as PGPD beneficiaries but received their care from non-PGPD physicians. We used a quasi-experimental design comparing trends in spending among PGPD participants and controls. This difference-in-difference design nets out fixed differences between PGPD participants and controls, as well as removes concurrent trends in local health markets. Site-specific savings estimates were combined, weighted by the number of assigned beneficiaries, to estimate the overall differences in payments associated with the demonstration.

Data

We used Parts A (hospital) and B (physician services) Medicare fee-for-service administrative claims data for all physician groups from 2001 through 2009. For years 2001-2005, we used 20% of the Medicare population and from 2006-2009 we used 100% of Medicare claims (2010 data are not yet available.) This study was approved by the Dartmouth College Institutional Review Board, which also determined that informed consent was not required.

Study Population

We assigned beneficiaries to the 10 PGPD participants and control populations using CMS reported methods.^{7,8} Beneficiaries were weighted according to the person-months in Medicare to appropriately address program exit (death) and entry. We denoted 2001-2004 as before and 2005-2009 as after the intervention. Our 2006 cohort size and estimates of CMS bonus payments resemble those reported by CMS (eTable 1 available at <http://www.jama.com>).¹¹ We repeated analyses on subsets of dually eligible and nondually eligible beneficiaries.

Outcome Variables

Our primary outcome measure is Medicare payments per person-year summed across all services (using the gross do-

mestic product [GDP] deflator to adjust payments to 2009 dollars).^{22,23} Following CMS methods, we capped annual spending at \$100 000 per beneficiary.⁸ We also stratified annual Medicare spending for each beneficiary into major categories (eg, acute care hospital, skilled nursing, professional services). We further stratified physician services using the Berenson-Eggers Type of Service (BETOS) categories (eg, evaluation and management, procedures, imaging, diagnostic tests).

To provide insight into potential mechanisms of savings or whether efforts to control costs could have led to higher readmissions or emergency department visits, we report readmissions to the hospital within 30 days for any cause (stratified by medical and surgical hospitalizations) and visits to the emergency department (ED). Each was measured as an individual indicator of whether a patient experienced a given event in the year (ED visit or readmission for those hospitalized).

Control Variables

All models were adjusted for age, sex, and race (black/other), and interactions between these variables. Additionally we adjusted for federal disability and Medicaid eligibility status and race-specific income at the zip code level (proportion under the federal poverty line and proportion in a high-income group, defined within race at the 85th percentile).²⁴ Means of these variables are listed in eTable 2.

The official evaluation analyzing spending growth in the PGPD used the methodology of HCCs to risk adjust, which determines a score predicting spending based on the individual's demographic characteristics and the presence or absence of claims-based diagnoses.^{12,25} Hierarchical condition categories risk adjustment may be sensitive to diagnostic and coding practices for 2 reasons. First, it is sensitive to the practice intensity of physicians—the more visits, procedures, and tests delivered, the more opportunities there are to add diagnoses to the claims used to create HCCs. Second, coding diagno-

ses on a claim involves subjective judgment. For example, a patient receiving a follow-up visit for hypertension who also has osteoarthritis could have either or both diagnoses coded on the claim. Including a second hypertension diagnosis has no effect on HCC, but adding osteoarthritis does. Variations in diagnostic testing, decisions about whether to attribute a new symptom (eg, joint pain) to a disease (eg, osteoarthritis), or intentional decisions to ensure the recording of all conditions can cause HCC scores for patients with identical illness levels to vary.²⁶⁻²⁸

We therefore considered an alternative clinical risk adjuster less subject to diagnostic intensity or coding practices, the combined annual rates of hip fracture, stroke, colorectal cancer, and acute myocardial infarction (AMI) in participants and controls, averaged across enrollees in each site. These low-variation conditions (LVCs) require an acute care hospitalization and therefore more closely reflect the true disease burden for these conditions.^{29,30} Prior research has found these conditions to be indicators of incident events.^{29,30} Furthermore, these measures predict mortality and health care expenditures at the regional level.³¹ We identified LVCs using Medicare hospital claims and diagnoses (eTable 3).^{32,33} Annual rates for each LVC were calculated in the participant and control group for each of the 10 local areas by year.

Statistical Analysis

We compared the changes over time in payments for PGPD participants to those for local control beneficiaries to estimate the payment differences associated with participation in the PGPD. We used Stata 12 MP (StataCorp) to complete statistical analyses. This difference-in-difference research design mitigates confounding factors that could affect measured differences in payments or health status between participant and control groups. By comparing changes over time between the participant and control group, we also implicitly adjusted for broader trends

in health care spending or Medicare beneficiary population health common to both groups.

For each outcome described above (overall spending, spending by category, and quality measures), we fit the following linear regression model:

$$E(Y_{ijt}) = \beta_{0j} + (\beta_{1j}\text{Participant}_{ijt}) + (\beta_{2jt}\text{Area}_j \times \text{Year}_t) + (\beta_{3j}\text{Participant}_{ijt} \times \text{After}_t),$$

for which Y_{ijt} is a given outcome (ie, spending) for patient i , residing in site j , at time t , $\text{Participant}=1$ if a patient was assigned to a PGPD participant, and β_{2jt} reflects year-specific effects for each PGPD area (10 areas \times 9 years=90) to control for local and time-specific factors unrelated to the PGPD that could affect payments. The coefficients of interest were the 10 site-specific interaction terms (β_{3j}) between Participants and the period after the PGPD was implemented, 2005-2009. To distinguish changes in the way PGPD sites treated patients from changes in the underlying health status of assigned patients, we further adjusted for demographic and clinical risks, using the LVC risk-adjustment approach described above. We compared the sensitivity of our results across risk adjustment methods (TABLE 1).

We adjusted for intraclass correlation within each of the 10 PGPD areas, controls vs PGPD participants, and within beneficiary over time using techniques developed to address correlation within nonnested groups, multiway clustering of standard errors.³⁴ In our data, this approach yielded standard error estimates similar to those obtained using Huber³⁵-White³⁶ sandwich estimates clustering by site and group (participants or controls). We estimated the cumulative association of the PGPD with payment differences as the weighted average of the 10 independent site-specific effects, β_{3j} , weighting estimates by the relative population share of each region. The significance threshold for all 2-sided t tests is .05. Further methodologic details are provided in the eAppendix available at <http://www.jama.com>.

RESULTS

The participant and control groups' mean age, proportion disabled, proportion dying annually, average number of comorbidities, and prevalence of each comorbidity were similar at baseline (eTable 2). Control group beneficiaries were slightly more likely to be women, Medicaid eligible, and black. Demographics of the participant and control groups did not change appreciably between the preintervention and postintervention periods, suggesting PGPD participants did not systematically target specific demographic groups for either enrollment or disenrollment.

The FIGURE depicts unadjusted annual means of spending in each year for beneficiaries assigned to PGPD physician groups and local controls. This figure illustrates that trends in the participants and controls were similar in the pre-PGPD period. The Figure also illustrates that for all enrollees, the reduction in growth of spending for non-dually eligible beneficiaries was modest. Overall, average annual Medicare payments per beneficiary in PGPD participating sites increased by \$1206 (15.2%) between the preintervention and postintervention periods and \$1230 (16.5%) for controls. After adjustment, per capita annual savings estimates were modest (\$114, 95% CI, \$12-\$216, $P=.03$, Table 1) (full regression results are available upon request). This result reflects the average of significant annual savings in the dually eligible beneficiaries (\$532, 95% CI, \$277-\$786, $P<.001$) and nonsignificant savings in the nondually eligible beneficiaries (\$59, 95% CI, \$166 in savings to \$47 in additional spending, $P=.28$).

Savings estimates were sensitive to the approach to risk adjustment used. These analyses are presented in Table 1. The true per beneficiary savings attributable to the PGPD therefore are likely to lie between the conservative mean LVC-adjusted result (\$114) and the mean HCC-adjusted result (\$496, 95% CI, \$468-\$524, Table 1), which may be more susceptible to coding biases.¹¹

Changes in participant and control group health status during the study period differed depending on the measurement method used. The baseline mean HCC score was 1.05 for PGPD participants and 1.03 for the local controls (eTable 2). Mean HCC scores increased to 1.18 for PGPD participants, a 12.4% increase, and to 1.12 for con-

Table 1. Spending Changes Associated With the Physician Group Practice Demonstration Overall and by Site^a

Beneficiary Type	Participant 2001-2004 Spending Annually per Beneficiary, Mean (95% CI) US \$ ^b	Estimated Change in Spending Associated With PGPD Annually per Beneficiary, Estimate (95% CI), US \$	
		Adjusted by Low Variation Condition (LVC) Rate ^c	Adjusted by Hierarchical Clinical Category (HCC) Score ^d
All PGPD participants			
All	7915 (7830 to 7999)	-114 (-216 to -12)	-496 (-524 to -468)
Dually eligible	10 495 (10 211 to 10 780)	-532 (-786 to -277)	-751 (-790 to -712)
Nondually eligible	7549 (7461 to 7636)	-59 (-166 to 47)	-404 (-428 to -380)
Billings Clinic			
All	7196 (6890 to 7501)	-309 (-373 to -245)	-103 (-116 to -90)
Dually eligible	9350 (8199 to 10 501)	-331 (-623 to -39)	-271 (-335 to -207)
Nondually eligible	6950 (6637 to 7264)	-278 (-319 to -236)	-24 (-37 to -11)
Dartmouth-Hitchcock Clinic			
All	8418 (8173 to 8662)	132 (39 to 226)	-665 (-705 to -625)
Dually Eligible	12 040 (11 067 to 13 013)	-397 (-826 to 32)	-1310 (-1349 to -1271)
Nondually eligible	8018 (7769 to 8266)	111 (17 to 206)	-492 (-528 to -456)
Everett Clinic			
All	7667 (7239 to 8094)	116 (-26 to 259)	466 (445 to 486)
Dually eligible	10 639 (9412 to 11 866)	287 (111 to 462)	407 (376 to 438)
Nondually eligible	7066 (6618 to 7514)	125 (2 to 248)	177 (164 to 189)
Forsyth Medical Group			
All	7300 (7017 to 7582)	-276 (-457 to -95)	-571 (-586 to -557)
Dually eligible	10 803 (10 002 to 11 604)	-742 (-955 to -528)	-522 (-552 to -492)
Nondually eligible	6532 (6238 to 6826)	-194 (-403 to 16)	-185 (-196 to -173)
Geisinger Clinic			
All	7294 (7067 to 7522)	252 (166 to 337)	-745 (-787 to -704)
Dually eligible	8843 (8150 to 9536)	79 (-165 to 323)	-376 (-422 to -330)
Nondually eligible	7020 (6782 to 7258)	297 (216 to 378)	-471 (-498 to -443)
Marshfield Clinic			
All	7284 (7113 to 7455)	-642 (-725 to -559)	-1119 (-1151 to -1087)
Dually eligible	8739 (8161 to 9317)	-987 (-1209 to -765)	-1797 (-1839 to -1756)
Nondually eligible	7095 (6917 to 7272)	-520 (-636 to -405)	-1266 (-1300 to -1231)
Middlesex Health System			
All	8785 (8477 to 9093)	749 (698 to 799)	93 (66 to 121)
Dually eligible	12 447 (11 315 to 13 579)	598 (194 to 1002)	462 (416 to 508)
Nondually eligible	8343 (8027 to 8659)	701 (635 to 768)	169 (143 to 195)
Park Nicollet Clinic			
All	7070 (6796 to 7344)	-16 (-98 to 65)	-65 (-76 to -55)
Dually eligible	10 051 (8932 to 11 170)	-1610 (-1708 to -1512)	-1058 (-1105 to -1010)
Nondually eligible	6737 (6460 to 7014)	188 (114 to 262)	-49 (-63 to -35)
St John's Clinic			
All	7152 (6954 to 7350)	-70 (-205 to 64)	-29 (-38 to -20)
Dually eligible	9426 (8787 to 10 064)	78 (-40 to 197)	254 (245 to 264)
Nondually eligible	6810 (6604 to 7016)	-102 (-226 to 21)	-133 (-143 to -124)
University of Michigan Faculty Group Practice			
All	12 714 (12 234 to 13 193)	-866 (-918 to -815)	-1155 (-1174 to -1137)
Dually eligible	17 511 (15 923 to 19 100)	-2499 (-2627 to -2371)	-2072 (-2098 to -2045)
Nondually eligible	12 043 (11 545 to 12 542)	-717 (-776 to -657)	-620 (-635 to -606)

Abbreviation: PGPD, Physician Group Practice Demonstration.

^aThis table is based on author analyses of Medicare claims files, 2001-2005 (20% sample), 2006-2009 (100% sample).

^bSpending capped at \$100 000 annually per beneficiary and inflated to 2009 US dollars using the gross domestic product deflator.

^cA negative number in this column represents savings. Estimates derived from a linear model adjusting for area-year indicators, age, black race, woman, Medicaid eligibility, and disability. The model adjusts for zip code-level rates of poverty and high income. The model adjusts for the rate of low-variation conditions for each of the 10 local areas for each year separately for treatment and control groups. Low-variation condition rate is the number of individuals experiencing the conditions hip fracture, stroke, colon cancer, and acute myocardial infarction per thousand Medicare beneficiaries.

^dA negative number in this column represents savings. Estimates derived from a linear model adjusting for area-year indicators, age, black race, woman, Medicaid eligibility, and disability. The model adjusts for zip code-level rates of poverty and high income. The model adjusts for the individual beneficiary's hierarchical condition categories.

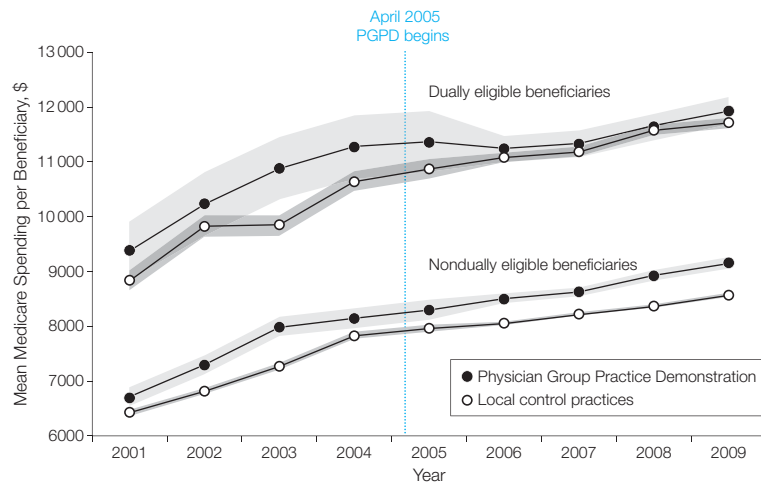
trols, an 8.7% increase. After regression adjustment, we found a significant positive association between PGPD participation and mean HCC score changes over time (0.03 increase in HCC score during the intervention period, 95% CI, 0.029-0.32, $P < .001$). The differential changes in HCC score were not mirrored in measures of risk less susceptible to potential manipulation such as age or mortality rates, which both went down during the intervention period.

Across participating systems, estimated savings and the effect of risk adjustment approaches varied markedly, with mean LVC-adjusted estimates ranging from savings of \$866 annually per beneficiary at the University of Michigan (95% CI, \$815-\$918) to greater expenditures by \$749 (95% CI, \$698-\$799) relative to controls at Middlesex (Table 1). Only 4 sites saved a significant amount across all beneficiaries (University of Michigan, Marshfield, Billings, and Forsyth), whereas 3 sites had no significant change and 3 sites increased expenditures relative to controls during the PGPD. Dartmouth-Hitchcock and Geisinger only exhibited savings under the HCC risk-adjustment approach; both had relatively large increases in HCC scores relative to their control group (Table 1).

Models stratified by the type of service demonstrate that significant savings occurred across all patients in acute care (\$118, 95% CI, \$65-\$170) and home health care (\$17, 95% CI, \$7-\$28, TABLE 2). Further, analysis revealed that in sites where savings occurred on acute care, hospitalization rates declined during the PGPD.

The Figure illustrates unadjusted growth in Medicare spending separated for dually eligible and nondually eligible beneficiaries. Within the dual beneficiary population, the rate of growth was similar in the intervention and control groups before the intervention. Between the preintervention and postintervention periods, the spending growth rate for dual beneficiaries treated by PGPD participants was 9.7% compared with a 15.3% increase

Figure. Medicare Spending per Beneficiary: Physician Group Practice Demonstration Participants and Local Controls



Data points represent annual per beneficiary spending, capped at \$100 000 and inflated to 2009 dollars using the gross domestic product deflator. Author analyses of Medicare claims files, 2001-2005 (20% sample), 2006-2009 (100% sample). Shaded areas indicate 95% confidence intervals.

among those treated by local control practices. As noted, this translates into mean \$532 in annual per beneficiary savings in the dually eligible beneficiaries (95% CI, \$277-\$786, $P < .001$, Table 1), or a 5% decrease in Medicare spending for the dually eligible patient. Savings in the dually eligible were less sensitive to the risk-adjustment approach (Table 1).

Much of these mean savings were achieved through a reduction in acute care hospitalizations (\$381, 95% CI, \$247-\$515, Table 2), procedures (\$55, 95% CI, \$15-\$94), and home health care (\$28, 95% CI, \$64 in savings to \$8 in additional spending). The reductions in spending were roughly similar across diagnosis groups, suggesting that savings may have been achieved through better care management overall rather than through disease-specific interventions.

The proportion of the assigned patient population that was dually eligible ranged from 11% in Billings and Middlesex to 20% in Forsyth, with a mean of 15% across all sites. Annual baseline spending on dually eligible beneficiaries ranged from \$8739 in Marshfield Clinic to \$17 511 in the University of Michigan Faculty Group Practice

(Table 1). These 2 sites achieved substantial mean savings in the dually eligible beneficiaries (Marshfield, \$987, 95% CI, \$765-\$1209 or 11%, University of Michigan, \$2499, 95% CI, \$2371-\$2627 or 14%). Park Nicollet achieved substantial savings in the dually eligible beneficiaries (\$1610, 95% CI, \$1512-\$1708) but also experienced increased spending in the nondually eligible (\$188, 95% CI, \$114-\$262) and so on average did not produce savings.

There was no overall association between the PGPD and the probability of ED visits in either the full PGPD population or among dually eligible beneficiaries (TABLE 3). These averages, however, mask significant reductions in ED visits in the sites that produced the largest savings in dually eligible beneficiaries, Marshfield, Park Nicollet, and the University of Michigan. The PGPD was associated with lower medical 30-day readmissions on average across the 10 sites and lower readmissions for both medical and surgical admissions in the dually eligible beneficiaries (Table 3 and eTable 4).

COMMENT

We found modest estimates of overall savings associated with the PGPD, but

larger savings among the dually eligible, a vulnerable patient population. Our estimates indicate that on average, the PGPD saved a mean of \$114 annually per beneficiary assigned to a physician group in an ACO-like model. This overall result masks substantial heterogeneity in results across partici-

pating institutions and by population subgroup. Among dually eligible beneficiaries, PGPD physician groups achieved a mean annual per capita savings of \$532, or 5%, while savings among nondually eligible beneficiaries were not statistically significant. Savings were achieved in large part

through reductions in hospitalizations.

The association between the PGPD incentive structure and payment differences varied widely by site, with some sites producing large reductions in spending growth in response to the shift away from fee for service while others experienced increased spending compared with local physician groups. Spending reductions did not appear to be associated with lower quality of care, whether reflected in their previously reported quality scores,³⁷ or with measures of readmission rates and ED visits.

The variation both in levels and changes in risk-adjusted spending across the participating organizations was remarkable. We know little about why some succeeded and others failed to achieve savings. One hypothesis is that organizations beginning with higher spending levels have greater opportunities to achieve savings. The University of Michigan had the highest mean baseline spending (\$12 714 overall, \$17 511 on dually eligible beneficiaries) and achieved the greatest per beneficiary savings. However, 2 relatively low spending systems, Marshfield and Park Nicollet, also experienced substantial savings among dually eligible beneficiaries.

Other factors may have contributed to achieving higher levels of performance in some sites, such as governance models; internal leadership; physician engagement strategies; the degree of coherence of electronic health records and other health information technological tools; and the specific approaches adopted for chronic disease management, care transitions, and quality improvement.³⁸

It is not possible to analyze the specific contributions of disease management and care coordination programs in the PGPD, and thus conclusions are largely speculative.³⁸ Still, we may conjecture that the size of the institution could affect the incentives to implement fundamental changes in the delivery system—the larger the system, the more likely preexisting informa-

Table 2. Spending Changes Associated With Physician Group Practice Demonstration by Spending Category^a

Beneficiary Type	Participant 2001-2004 Spending Annually per Beneficiary, Mean (95% CI), US \$ ^b	Estimated Change in Spending Associated With PGPD Annually per Beneficiary, Estimate (95% CI), US \$ ^c
Acute care hospitalization		
All	3251 (3199 to 3304)	-118 (-170 to -65)
Dually eligible	4292 (4118 to 4466)	-381 (-515 to -247)
Nondually eligible	3104 (3050 to 3158)	-85 (-138 to -32)
Procedures		
All	1113 (1102 to 1125)	-3 (-13 to 7)
Dually eligible	1206 (1165 to 1247)	-55 (-94 to -15)
Nondually eligible	1100 (1088 to 1112)	0 (-13 to 14)
Home health care		
All	322 (314 to 330)	-17 (-28 to -7)
Dually eligible	473 (445 to 501)	-28 (-64 to 8)
Nondually eligible	301 (293 to 309)	-14 (-24 to -4)
Tests		
All	296 (294 to 298)	-2 (-8 to 5)
Dually eligible	359 (351 to 366)	-16 (-23 to -8)
Nondually eligible	287 (285 to 290)	-1 (-9 to 7)
Durable medical equipment		
All	459 (447 to 470)	31 (10 to 53)
Dually eligible	748 (705 to 791)	-15 (-41 to 12)
Nondually eligible	418 (406 to 429)	34 (6 to 33)
Evaluation and management		
All	844 (838 to 849)	14 (2 to 27)
Dually eligible	1147 (1127 to 1168)	-14 (-41 to 12)
Nondually eligible	801 (795 to 806)	19 (6 to 33)
Imaging		
All	381 (377 to 384)	-2 (-9 to 6)
Dually eligible	397 (388 to 407)	-5 (-17 to 7)
Nondually eligible	378 (375 to 382)	-2 (-12 to 8)
Long term		
All	323 (309 to 337)	1 (-10 to 13)
Dually eligible	650 (592 to 709)	-1 (-42 to 40)
Nondually eligible	276 (263 to 290)	5 (-7 to 17)
Skilled nursing		
All	497 (481 to 512)	-4 (-20 to 12)
Dually eligible	772 (717 to 828)	5 (-40 to 50)
Nondually eligible	458 (442 to 473)	-2 (-22 to 18)

Abbreviation: PGPD, Physician Group Practice Demonstration

^aThis table is based on author analyses of Medicare claims files, 2001-2005 (20% sample), 2006-2009 (100% sample).

^bSpending capped at \$100 000 annually per beneficiary and inflated to 2009 US dollars using the gross domestic product deflator.

^cA negative number in this column represents savings. Estimates derived from a linear model adjusting for area-year indicators, age, black race, woman, Medicaid eligibility, and disability. The model adjusts for zip code-level rates of poverty and high income. The model adjusts for the rate of low-variation conditions for each of the 10 local areas for each year separately for treatment and control groups. Low-variation conditions rate is the number of individuals experiencing the conditions hip fracture, stroke, colon cancer, and acute myocardial infarction per thousand Medicare beneficiaries.

tion systems are in place and the greater the absolute dollar Medicare performance bonus for a given proportional reduction in Medicare costs. We do find evidence for this hypothesis, as cost savings and the number of physicians in each network were correlated ($\rho=0.52$, $P=.12$) for overall savings and in savings in the dually eligible beneficiaries ($\rho=0.63$, $P=.049$).

Dually eligible beneficiaries have historically proven to be a difficult group to manage because of high illness burden, low socioeconomic status, and lack of social supports. Our results suggest that while some care management or coordination programs have failed to demonstrate savings,³⁹⁻⁴² ACOs and similar shared-savings contracts have the potential to improve care for this high-cost group. In response to the contingent shared-savings incentives in the PGPD, participating physician groups reported creating chronic condition management programs, patient registries, case coordination teams, and instituting electronic medical records.³⁸ We might expect these programs, aimed at coordinating care across clinicians and supporting care for chronic conditions, to have the largest influence on the dually eligible population. However, it does appear that much of the cost savings in the dually eligible beneficiaries came in the first few years of the program (Figure); later years showed more rapid growth in spending relative to controls, possibly owing to the limited time horizon of the PGPD program. Although current Medicare ACO programs are initially planned to last 3 to 5 years (Shared Savings Program and Pioneer), they are renewable after the initial time period.

Our results stand in contrast to the modest savings reported in the Massachusetts Alternative Quality Contract (an early ACO model), which appear to have been achieved largely by focusing referrals on lower cost providers, rather than through reductions in utilization.⁴³ While the Alternative Quality Contract applied to a younger and comparatively much healthier commercially insured population, the high-

Table 3. Changes in Utilization-Based Quality Measures Associated With the Physician Group Practice Demonstration^{a,b}

	Participant 2001-2004 Mean, % (95% CI)	Estimated Change Associated With PGPD, % (95% CI)
Emergency department visit rate		
All	30.9 (30.7 to 31.2)	0.06 (−0.11 to 0.24)
Dually eligible	46.0 (45.3 to 46.7)	−0.10 (−0.52 to 0.32)
Nondually eligible	28.8 (28.6 to 29.0)	0.14 (−0.04 to 0.32)
30-Day medical readmission rate		
All	15.8 (15.4 to 16.3)	−0.67 (−1.11 to −0.23)
Dually eligible	17.3 (16.2 to 18.3)	−1.07 (−1.73 to −0.41)
Nondually eligible	15.5 (15.1 to 16.0)	−0.58 (−1.08 to −0.07)
30-Day surgical readmission rate		
All	9.3 (8.9 to 9.8)	−0.17 (−0.59 to 0.25)
Dually eligible	13.0 (11.6 to 14.4)	−2.21 (−3.07 to −1.34)
Nondually eligible	8.8 (8.3 to 9.3)	0.14 (−0.29 to 0.57)

^aThis table is based on author analyses of Medicare claims files, 2001-2005 (20% sample), 2006-2009 (100% sample).

^bEstimates derived from a linear model adjusting for area-year indicators, age, black race, woman, Medicaid eligibility, and disability. The model adjusts for zip code-level rates of poverty and high income. The model adjusts for the rate of low-variation conditions for each of the ten local areas for each year separately for treatment and control groups. Low-variation conditions rate is the number of individuals experiencing the conditions hip fracture, stroke, colon cancer, and acute myocardial infarction per thousand Medicare beneficiaries.

risk group in their study did achieve the largest savings. Our results from the PGPD suggest that participants found ways to achieve savings through improving care and reducing expensive services such as hospitalizations. This article highlights the potential benefits of the ACO model for patients with serious or complex illness, a group for whom improved quality and coordination is especially important.

This study has important limitations. First, we did not have access to the exact methods CMS used to calculate savings. Published reports, however, provided reasonable guidance and our application of these methods resulted in estimates that are similar in magnitude and direction to those published by CMS (eTable 1). Second, we acknowledge that the LVC approach to risk adjustment could have underestimated or overestimated savings had there been other real changes in health status between periods not associated with our measures of LVCs. There are a number of factors that can affect risk-adjusted savings estimates (eg, patient selection, differential Medicare Advantage enrollment, pay for performance increasing coding intensity). By providing a range of estimates we hope to present the reader with plausible bounds on effect sizes. While preven-

tive efforts, incentivized through pay for performance in the PGPD, may affect health status, research suggests that many years of continuous preventive treatment are likely to be required to reduce the incidence of AMI and stroke. An exception is colorectal cancer screening, which we expect would increase the number of colorectal cancer surgeries in the short term. Thus, while far from perfect, the use of LVCs is a reasonable and potentially less biased alternative to HCC adjustment.

Third, our significance test for the overall cost savings from the PGPD assumes that the treatment effects in each region are independent. If spillover effects occurred whereby one PGPD learned from the experiences of others how to reduce expenditures, our confidence intervals would be biased downward. An alternative approach that adjusts only for the correlation over time for a given beneficiary (as was used to estimate savings from the Blue Cross Blue Shield Alternative Quality Contract),⁴³ would imply that our confidence intervals are far too conservative.

Finally, our data sources only inform us about Medicare spending in the fee-for-service population. We did not measure any possible spillover effects among those enrolled in Medicare Ad-

vantage, nor did we measure Medicaid costs for the dually eligible beneficiaries, which may substitute in part for a reduction in Medicare costs.^{44,45} However, most of the evidence on cost-shifting between Medicaid and Medicare has emphasized policy changes in Medicaid that incur costs for Medicare, rather than the reverse.⁴⁶ Medicare covers acute care services for the dually eligible, while Medicaid covers Medicare premiums, cost sharing, and long-term (custodial) nursing home services. If Medicare spending is reduced, the cost sharing portion paid by Medicaid would also likely decrease. A shared savings model could, however, shift costs from Medicare to Medicaid for those who are institutionalized if reduced hospitalizations resulted in more Medicaid-paid nursing home days, rather than Medicare-paid hospital or skilled nursing facility days (paid for by Medicare after a preceding 3-day hospital stay). However, we did not observe reductions in Medicare skilled nursing spending in the dually eligible, savings for the noninstitutionalized dually eligible beneficiaries were similar to those we report, and we found no evidence of an increase in institutionalization among dually eligible beneficiaries in the PGPD compared with controls.

Our data sources also limited the information we had on quality, and we only measured utilization-based indicators of quality. Despite modest cost savings to the Medicare program overall, quality metrics in the Demonstration improved for every institution. We did not measure any clinical or patient reported outcomes but all PGPD sites were required to collect quality information data for payment purposes, and all sites significantly improved quality of care during the demonstration period.⁹ Because limiting care is an important concern particularly for vulnerable groups, further work could more carefully consider how spending reductions affect other quality measures.

Our results suggest that the ACO reforms included in the Affordable Care

Act, such as the Pioneer and the Medicare Shared Savings Programs, have at least the potential to slow spending growth, particularly for costly patients.^{21,47} The remarkable degree of heterogeneity across participating sites underscores the importance of timely evaluation of current payment reforms and a better understanding of the institutional factors that lead to either success or failure in effecting changes in health care practices.

Author Contributions: Drs Colla and Gottlieb had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Wennberg, Meara, Gottlieb, Snyder, Fisher, Colla.

Acquisition of data: Skinner, Fisher.

Analysis and interpretation of data: Wennberg, Meara, Skinner, Gottlieb, Snyder, Colla.

Drafting of the manuscript: Gottlieb, Fisher, Colla.

Critical revision of the manuscript for important intellectual content: Wennberg, Meara, Skinner, Gottlieb, Snyder, Fisher, Colla.

Statistical analysis: Skinner, Gottlieb, Snyder, Colla.

Obtained funding: Skinner, Fisher.

Administrative, technical, or material support: Fisher, Colla.

Study supervision: Meara, Colla.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This research was funded by grant NIA P01AG19783 from the National Institute on Aging, the Dartmouth Atlas Project (supported by a consortium led by the Robert Wood Johnson Foundation), and the Commonwealth Fund.

Role of the Sponsor: None of the funders had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Online-Only Material: The eAppendices are available at <http://www.jama.com>.

Additional Contributions: We thank Harold Sox, MD, of Dartmouth Medical School for helpful comments on the manuscript for which he received no compensation.

REFERENCES

- Centers for Medicare & Medicaid Services. Pioneer accountable care organization model: general fact sheet. <http://innovations.cms.gov/Files/fact-sheet/Pioneer-ACO-General-Fact-Sheet.pdf>. Accessed August 14, 2012.
- Berwick DM. Making good on ACOs' promise—the final rule for the Medicare shared savings program. *N Engl J Med*. 2011;365(19):1753-1756.
- Fisher ES, McClellan MB, Bertko J, et al. Fostering accountable health care: moving forward in Medicare. *Health Aff (Millwood)*. 2009;28(2):w219-w231.
- Congressional Budget Office. Budget options, volume 1, health care. <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/99xx/doc9925/12-18-healthoptions.pdf>. December 2008. Accessed May 11, 2012.
- Trisolini M, Kautter J, Pope GC, Bapat B, Olmsted E, Urato M. Physician Group Practice Demonstration quality measurement and reporting specifications. http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/Quality_Specs

_Report.pdf. July 29, 2005. Accessed August 14, 2012.

6. Sebelius K; Centers for Medicare & Medicaid Services. Report to Congress: Physician Group Practice Demonstration evaluation report. http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/PGP_RTC_Sept.pdf. 2009. Accessed May 11, 2012.

7. Kautter J, Pope GC, Trisolini M, Grund S. Medicare physician group practice demonstration design: quality and efficiency pay-for-performance. *Health Care Financ Rev*. 2007;29(1):15-29.

8. Kautter J, Pope GC, Trisolini M, Bapat B, Olmsted E, Klosterman M. Physician Group Practice Demonstration bonus methodology specifications. http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/Downloads/PGP_Payment.pdf. December 20, 2004. Accessed May 11, 2012.

9. Centers for Medicare & Medicaid Services. Medicare Physician Group Practice Demonstration fact sheet. https://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/PGP_Fact_Sheet.pdf. July 2011. Accessed May 11, 2012.

10. Wilensky GR. Lessons from the Physician Group Practice Demonstration—a sobering reflection. *N Engl J Med*. 2011;365(18):1659-1661.

11. Medicare Payment Advisory Commission (MedPAC). Report to the Congress: improving incentives in the Medicare program. http://www.medpac.gov/documents/jun09_entirereport.pdf. June 2009. Accessed May 11, 2012.

12. Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev*. 2004;25(4):119-141.

13. Starfield B, Gervas J, Mangin D. Clinical care and health disparities. *Annu Rev Public Health*. 2012;33:89-106.

14. Pollack C, Armstrong K. Accountable care organizations and health care disparities. *JAMA*. 2011;305:1706-1707.

15. Rousseau D, Lawton E, Langston J, Connolly J, Howard J. Dual eligibles: Medicaid enrollment and spending in Medicare beneficiaries in 2007. <http://www.kff.org/medicaid/upload/7846-02.pdf>. 2010. Accessed May 11, 2012.

16. Moon S, Shin J. Health care utilization among Medicare-Medicaid dual eligibles: a count data analysis. *BMC Public Health*. 2006;6:88.

17. Jacobson G, Neuman T, Damico A. Medicare's role for dual eligible beneficiaries. <http://www.kff.org/medicaid/upload/8138-02.pdf>. April 2012. Accessed May 11, 2012.

18. Young K, Garfield R, Musumeci M, Clemans-Cope L, Lawton E. *Medicaid's Role for Dual Eligible Beneficiaries*. <http://www.kff.org/medicaid/upload/7846-03.pdf>. April 2012. Accessed May 11, 2012.

19. Kasper J, O'Malley Watts M, Lyons B. Chronic disease and co-morbidity among dual eligibles: implications for patterns of Medicaid and Medicare service use and spending. <http://www.kff.org/medicaid/upload/8081.pdf>. July 2010. Accessed May 11, 2012.

20. Kronick R, Bella M, Gilmer T. The faces of Medicaid III: refining the portrait of people with multiple chronic conditions. http://chcs.org/user_docs/Faces_of_Medicaid_III.pdf. October 2009. Accessed August 14, 2012.

21. Coughlin TA, Waidmann TA, Phadera L. Among dual eligibles, identifying the highest-cost individuals could help in crafting more targeted and effective responses. *Health Aff (Millwood)*. 2012;31(5):1083-1091.

22. Huskamp HA, Newhouse JP. Is health spending slowing down? *Health Aff (Millwood)*. 1994;13(5):32-38.

23. The World Bank. World development indicators, GDP deflator. <http://data.worldbank.org/indicator/NY>

- .GDP.DEFL.KD.ZG. 2011. Accessed May 11, 2012.
24. United States Census. US Census Bureau webpage. http://www2.census.gov/census_2000/datasets/. 2000. Accessed May 11, 2012.
 25. Centers for Medicare & Medicaid Services. CMS-HCC Risk adjustment model. http://www.cms.gov/Medicare/Health-Plans/MedicareAdvgtgSpecRateStats/Risk_adjustment.html. Accessed May 11, 2012.
 26. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. *N Engl J Med*. 2010;363(1):45-53.
 27. Welch HG, Sharp SM, Gottlieb DJ, Skinner JS, Wennberg JE. Geographic variation in diagnosis frequency and risk of death among Medicare beneficiaries. *JAMA*. 2011;305:1113-1118.
 28. United States Government Accountability Office. Medicare Advantage: CMS should improve the accuracy of risk score adjustments for diagnostic coding practices. <http://gao.gov/assets/590/587637.pdf>. January 2012. Accessed May 11, 2012.
 29. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. *N Engl J Med*. 1989;321(17):1168-1173.
 30. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. *N Engl J Med*. 1994;331(15):989-995.
 31. Skinner J, Wennberg JE. Regional inequality in Medicare spending: the key to Medicare reform. In: Garber A, ed. *Frontiers of Health Economics*. Boston, MA: MIT Press; 2000. <http://www.nber.org/chapters/c9830.pdf>. Accessed May 11, 2012.
 32. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part I: the content, quality, and accessibility of care. *Ann Intern Med*. 2003;138(4):273-287.
 33. Buntin MB, Colla CH, Deb P, Sood N, Escarce JJ. Medicare spending and outcomes after postacute care for stroke and hip fracture. *Med Care*. 2010;48(9):776-784.
 34. Cameron AC, Gelbach JB, Miller DL. Robust inference with multiway clustering. *J Bus Econ Stat*. 2011;29:238-249.
 35. Huber P. *The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions*. Berkeley: University of California Press; 1967.
 36. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*. 1980;48:817-838.
 37. Center for Medicare & Medicaid Services. Medicare Physician Group Practice Demonstration fact sheet. http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/PGP_Fact_Sheet.pdf. July 2011. Accessed May 11, 2012.
 38. Trisolini M, Aggarwal J, Leung M, Pope G, Kautter J. The Medicare Physician Group Practice Demonstration: lessons learned on improving quality and efficiency in health care. http://www.commonwealthfund.org/usr_doc/1094_Trisolini_Medicare_phys_group_practice_demo_lessons_learned.pdf?section=4039. February 2008. Accessed August 14, 2012.
 39. Goetzel RZ, Ozminkowski RJ, Villagra VG, Duffy J. Return on investment in disease management: a review. *Health Care Financ Rev*. 2005;26(4):1-19.
 40. Peikes D, Chen A, Schore J, Brown R. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA*. 2009;301(6):603-618.
 41. Schore JL, Brown RS, Cheh VA. Case management for high-cost Medicare beneficiaries. *Health Care Financ Rev*. 1999;20(4):87-101.
 42. Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is \$1 billion a year a good investment? *Am J Manag Care*. 2007;13(12):670-676.
 43. Song Z, Safran DG, Landon BE, et al. Health care spending and quality in year 1 of the alternative quality contract. *N Engl J Med*. 2011;365(10):909-918.
 44. Bubolz T, Emerson C, Skinner J. State spending on dual eligibles under age 65 shows variations, evidence of cost shifting from Medicaid to Medicare. *Health Aff (Millwood)*. 2012;31(5):939-947.
 45. Konezka RT, Spector W, Shaffer T. Effects of nursing home ownership type and resident payer source on hospitalization for suspected pneumonia. *Med Care*. 2004;42(10):1001-1008.
 46. Grabowski DC. Medicare and Medicaid: conflicting incentives for long-term care. *Milbank Q*. 2007;85(4):579-610.
 47. Medicare Payment Advisory Commission (MedPAC). Coordinating care for dual-eligible beneficiaries. http://www.medpac.gov/chapters/Jun11_Ch05.pdf. June 2011. Accessed May 11, 2012.