REPORT TO CONGRESS

PHYSICIAN GROUP PRACTICE DEMONSTRATION EVALUATION REPORT

Kathleen Sebelius
Secretary of Health and Human Services
2009

Table of Contents

List of Tables and Figures

EXECUTIVE SUMMARY

Section 412 of the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act of 2000 (BIPA) mandated the Physician Group Practice (PGP) Demonstration. BIPA also required four reports to the Congress that assess the impacts of the Demonstration on expenditures, access, and quality. This report is the last of these mandated reports and uses data from the 10 participating practices from the first two years of the Demonstration. BIPA specified the basic features of the Demonstration model by requiring incentive payments that are a portion of the Medicare savings, additional performance payments for improvements on process and outcome measures, payments on a fee-for-service (FFS) basis, and the use of geographic areas to define the scope of the Demonstration. Using this design framework, the Demonstration began on April 1, 2005 and will run for five years through March 31, 2010.

The 10 participating physician group practices (referred to as the "PGPs" in this report) are large group practices, ranging from 232 to 1,291 affiliated physicians. Two PGPs are freestanding physician group practices; two are faculty group practices within academic medical centers; five belong to an integrated delivery system, which consists of at least one hospital in addition to the physician group (and may include other health care providers such as home health agencies or nursing homes); and one is a physician network that is sponsored by a hospital affiliate and is comprised of 60 small and individual physician practices. PGPs can receive performance payments of up to 80 percent of the Medicare expenditures that are saved in a

performance year. To receive the full performance payment, PGP sites must meet or exceed performance targets established for ambulatory care quality measures.

This report assesses the BIPA required impacts using two years of data. Because of this short time frame, additional study will be needed to more adequately observe the implementation and refinement of interventions, and for their full impacts to be realized. Because some changes external to the Demonstration cannot be controlled, it may be difficult to determine pure Demonstration effects. Keeping these qualifications in mind, the preliminary impacts of the Demonstration are summarized below. This report is based on the study and analyses provided by RTI International under contract to the Centers for Medicare & Medicaid Services (CMS). In the discussion and analyses that follow, the year prior to the start of the Demonstration, calendar year 2004, is termed the "base year" (BY). This year is used as the benchmark to measure savings and improvements in quality in performance year 1 (PY1; April 1 2005 to March 31, 2006) and in performance year 2 (PY2; April 1 2006 to March 31, 2007).

Impacts to be Addressed Under BIPA

Access

Because the PGP Demonstration retains the structure of the Medicare FFS system, there is no enrollment process for beneficiaries. Therefore, beneficiaries are retroactively "assigned" (termed Assigned Beneficiaries in this report) to the PGP sites if they received the plurality of their office or other outpatient evaluation and management (E&M) services during a year at the

site. Traditional Medicare service coverage, co-pay and deductible structures, freedom of provider choice, and FFS payments to providers were all maintained during the Demonstration. While the number of Assigned Beneficiaries remained almost constant during the Demonstration, the frequency of their E&M visits per year increased slightly from 5.4 in the BY to 5.5 in PY2. The consistency in the number of beneficiaries assigned and the frequency of the E&M visits suggest that access to care did not change during the first two years of the Demonstration. Since care management programs and redesigned care processes were available to Medicare beneficiaries, it may be reasoned that the Demonstration improved access.

Quality

The PGPs have reported implementing or enhancing a variety of care management programs that focus on improving the efficiency and quality of health care. These programs include chronic disease management programs, high risk/high cost care management, transitional care management, end-of-life/palliative care programs, and initiatives designed to standardize and improve the quality of care. In addition, information technology such as electronic medical records, patient disease registries, and patient monitoring systems, are being used by the PGPs with the goals of improving practice efficiency and quality of care delivered to patients, and to better understand the utilization of services by the Medicare FFS population.

In performance years 3 (PY3: April 1, 2007 to March 31, 2008) and thereafter, quality is measured using 32 ambulatory care quality measures covering five condition modules. The performance on 7 measures is calculated from Medicare claims, while sampling techniques are

used to assess performance on 25 chart-based measures. In PY2, a total of 27 measures were used: 10 for diabetes mellitus, 10 for congestive heart failure, and 7 for coronary artery disease. Overall, including both claims and chart-based measures, all participating PGPs achieved benchmark performance on at least 25 of the 27 Demonstration process quality indicators in PY2. Five PGPs achieved benchmark performance on all 27 measures in PY2, compared to two that achieved benchmark performance on all 10 measures used in PY1. Between the BY and PY2, the PGP groups demonstrated improvement by increasing their quality scores an average of 9 percentage points on the diabetes mellitus measures, 11 percentage points on the heart failure measures, and 5 percentage points on the coronary artery disease measures. Despite differences in organizational structures, the PGPs were able to attain similar levels of quality performance measures.

Expressed as a percentage, the 7 claims-based quality indicators can be compared between the PGPs and their local comparison groups (CGs). Since only the PGPs are completing the chart based measures, they cannot be compared to the local CGs. The PGP geographic service areas consist of all counties from which the PGP derives at least 1 percent of its Assigned Beneficiaries. With a few exceptions, the local CGs are comprised of the complement of beneficiaries who are not assigned to the PGPs. Between the BY and PY2, 4 of 7 claims based indicators (lipid measurement, urine protein testing, left ventricular ejection fraction testing, and lipid profile) showed greater improvement among PGP Assigned Beneficiaries than among the CGs. This improvement is statistically significant at the 5 percent level, which is expressed in statistical notation as $p < .05$. The differences in the 3 other indicators (HbA1c management, eye exam, and breast cancer screening) between the PGP and

CG beneficiaries were not statistically significant. The finding that participating PGPs improved their claims-based quality process indicators more than their comparison group remained true even after adjusting for pre-Demonstration trends in the claims-based quality indicators.

Expenditures

In the first two performance years, the net savings of the Demonstration to the Medicare Trust Funds were $2,260,000. Under the Demonstration, financial performance is evaluated by comparing each PGP's Actual Medicare expenditures to the PGP's Target Expenditures. A two percent band or corridor around the Target Expenditures is needed to account for normal fluctuations in measuring Medicare expenditures that can result from changes in the number of beneficiaries, imprecision in the measurement methods, and other random events. Differences between the Target and Actual Expenditures within the two percent corridor are not counted as savings or losses. Target Expenditures are determined for each PGP using the PGP's BY expenditures (2004), adjusted for expenditure growth of the local comparison group (CG) and changes in patient characteristics, as measured by risk scores.

If a PGP's Actual Expenditures are less than 98 percent of its Target Expenditures, the PGP is deemed to have saved Medicare expenditures and earns a performance payment. Four PGPs were in this category in the first two years of the Demonstration. They achieved total savings of $26,907,000 and received performance payments of $21,163,000 (up to 80 percent of total savings). These sites are referred to as the "4 PGPs earning performance payments in PY2" in subsequent analyses.

If Actual Expenditures fall within 98 percent and 102 percent of the target, the PGP is deemed to not have saved Medicare expenditures and is not eligible for performance payments. Four PGPs were in this category. Finally, if Actual Expenditures exceed 102 percent of Target Expenditures, the PGP will receive no performance payments and is considered to have achieved negative Medicare savings. Two PGPs were in this category and they had a total of $3,484,000 in negative savings. The PGPs not earning performance payments are referred to as the "6 PGPs not earning performance payments in PY2" in subsequent analyses.

After accounting for performance payments and negative savings, the net savings to Medicare were $2,260,000 in the first two years of the Demonstration. Since most of the savings are returned to the sites as performance payments, the net savings to the Medicare Trust funds in the first two years of the Demonstration were minimal when expressed as a percentage of all Target Expenditures.

Ignoring performance payment offsets, Actual Expenditures were $120 per person or 1.2 percent less than Target Expenditures per beneficiary for the combined 10 PGPs in PY2. This reduction is statistically significant ($p < .01$). Note that the difference between Target Expenditures and Actual Expenditures in this analysis is not subject to the two percent corridor. However, a similar concept, statistical significance, is used to identify meaningful results. The majority of the financial savings occurred in outpatient, not inpatient, services. On average, outpatient expenditures were $83 per person year less than expected, while inpatient expenditures were $25 per person year less than expected and not statistically significant.

Analyses

A major evaluation issue is: what has driven the performance results? Why did four PGPs earn performance payments in PY2, and six did not? The Demonstration is not designed to test specific interventions; therefore, participating sites have complete autonomy in determining strategies that will provide higher quality care and expenditure savings. During site interviews conducted in 2005-2006, the PGPs attributed their savings model to many factors, including: 1) the organizational structure and prior investments made by the PGPs, 2) new investments in care management programs and redesigned care processes made as a result of the Demonstration, 3) more intensive diagnostic coding of beneficiaries resulting from care management processes and quality reporting incentives, and 4) changes in local market conditions. The first three factors and one additional factor, pre-existing expenditure trends, were studied in order to understand why some PGPs earned performance payments, and some did not. The fourth factor, changes in local market conditions, is accounted for in the demonstration design. The change in the growth rate of CG expenditures reflects changes from the base year in the local markets and is used to compute Target Expenditures for each PGP site. These analyses are described below:

Organizational Structure

In PY 2, the 4 PGPs earning performance payments are characterized as being either affiliated with an academic medical center or a freestanding physician group practice. No performance payments were earned by the five PGPs belonging to an integrated delivery system

(defined as a system that includes hospital ownership, but not affiliated with an academic medical center) and the one physician network PGP that is sponsored by a hospital affiliate. The presence of a hospital was hypothesized as a potential deterrent to achieving savings under the Demonstration, since these systems may be unable to reduce avoidable admissions or use lower cost care substitutes without affecting their inpatient revenue. Based on the PY2 result, there is some evidence to support this hypothesis, since the 6 PGPs not earning performance payments had owned hospitals in their delivery systems. Contrary to this view, though, are the two academic medical centers, which also had integrated hospitals, and received performance payments. Hospitals with high occupancy, such as the academic medical centers, may be able to replace reductions in Medicare inpatient revenue with private pay admissions.

Care Management Programs / Redesigned Care Processes

All PGP sites have stated that they have implemented disease and/or care management programs to reduce expense, while improving the quality of care. These clinically based care management programs can be characterized as being Disease Specific, which target beneficiaries with certain diagnosis, or General Care Coordination Programs, which use enrollment criteria that is not disease specific. The 4 PGPs that earned performance payments in PY2 had lower inpatient and outpatient expenditures than the 6 PGPs not earning performance payments. While lower costs are consistent with the expectation about care management, sufficient data was not available to test this hypothesis using a rigorous analysis. Consequently, measuring the specific contribution of care management programs and redesigned care processes to cost savings, and evidence of their impact is largely anecdotal.

More Intensive Diagnostic Coding

Target Expenditures are computed for each PGP by inflating the PGP BY expenditures by the expenditure growth rates of the CGs, and then adjusting for changes in beneficiary characteristics. Beneficiary characteristics are reflected by risk scores, which are determined with a Hierarchical Condition Categories (HCC) model that was adapted for the Demonstration. Since the HCC risk scores are derived from the diagnostic coding on claims, the PGPs may have a financial incentive to more fully code, or code more intensely, as that can impact performance payments.

Nationally, there has been a pattern of increased intensity and specificity of coding among FFS providers, particularly physicians, over the last ten years. This trend was expected to appear in both the PGPs and the CGs. If the upward trend in risk scores is the same in the PGP assigned and CG beneficiaries, Demonstration savings and performance payments are unaffected. The HCC model uses age and gender, disease or condition, and other patient level characteristics to assign a beneficiary to HCC categories. Between the BY and PY2, the disease component of the HCC scores for the PGP sites grew at an average of 8.2 percent, compared to an average of 5.2 percent for the CGs.

Risk score growth can be affected by a number of factors. It is possible that PGP sites are attracting beneficiaries with particular diseases, who would benefit from care and disease management programs. Alternatively, if the PGPs are more carefully and intensively coding

than the CGs even though patients may be similar in their underlying morbidity, performance payments may be based less on an underlying difference in population risk and more on improved patient tracking and increased delivery of services for those with chronic conditions. Distinguishing between case mix change and change related to coding initiatives is difficult and challenging to quantify when they occur simultaneously. Several analyses were conducted to assess the influence of coding related to Demonstration incentives; however, the results were inconclusive for the period under study. Future evaluation work is planned to address whether trends in risk scores become more similar over time, as coding practices reach the point that they capture nearly all disease each year.

Pre-existing Expenditure Trends

An important evaluation issue is whether the financial results would have been obtained without the Demonstration. To help answer this question, the potential savings that would have occurred using data prior to the start of the Demonstration were compared to the savings achieved during the Demonstration period.

For the 4 PGPs earning performance payments in PY2, Actual Expenditures were $334 per person year lower (3.5 percent lower) than Target Expenditure, which is statistically significant ($p < .01$). But this performance was almost matched in the pre-Demonstration period. An analysis that takes into account the pre-Demonstration experience in 2004 found that Actual Expenditures were $291 per person year lower (3.9 percent lower) than the simulated Target Expenditures, which is statistically significant ($p < .01$). The difference between Actual and

Target Expenditures in the pre-Demonstration period ($291) and the PY2 Demonstration period ($334) was only $43 per person year, and was not statistically significant. The similar level of expenditure savings, both in the pre-Demonstration and Demonstration periods, suggests that these PGPs exhibited favorable cost trends prior to the Demonstration – trends that might have continued had the Demonstration not occurred. Apparently, these PGPs were more successful in controlling their expenditure growth than other providers in their local market area, and this appeared to help them in achieving the shared savings objectives under the Demonstration. One interpretation of these trends is that these sites had a cost-saving infrastructure in place prior to the Demonstration, which may be one of the reasons why they elected to participate in the Demonstration. The analyses could not determine the extent to which savings were influenced by pre-existing expenditure trends or resulted from a response to the financial incentives of the Demonstration.

In the Demonstration period, Actual Expenditures of the 6 PGPs not earning performance payments in PY2 were not statistically significant different from Target Expenditures. In the simulated pre-Demonstration period, Actual Expenditures were $91 higher (1.6 percent higher) than Target Expenditures and statistically significant ($p < .10$). On average, the 6 PGPs not earning performance payments in PY2 were trending above their local market expenditures prior to the Demonstration. In general, their performance improved in the Demonstration period, but not sufficiently to share in savings under the Demonstrations performance payment methodology.

Future Refinements

The PGP design may be used in other CMS demonstrations, or continued in the PGP Demonstration. To this end, several refinements in the design are summarized below.

Comparison Groups

Simplifying the process for calculating the benchmark used for measuring savings would be an important element in refining the PGP Demonstration. The expenditure growth rate and the risk scores of the comparison groups are particularly important, as they are used in establishing Target Expenditures for computing savings. Determining a geographically defined comparison group for measuring cost and quality performance remains technically challenging, data intensive, and administratively burdensome. Possible alternatives to constructing a local comparison group include using growth factors based on the: 1) local area, such as an MSA, 2) statewide average, 3) national average, or 4) a combination of these three. The growth factor could also take into account the current spending level of the group with less generous factors used in areas with high spending. PGPs participating in a demonstration could also be comparatively benchmarked against each other. Further analyses are needed to determine the impacts of using these alternative benchmarks.

Adjusting for Medicare Payment Policy Changes

Target Expenditures are computed for each PGP by increasing the PGP BY expenditures by the expenditure growth rates of the CGs adjusting for changes in beneficiary characteristics. There are no adjustments for changes in Medicare payment formulas or policies. It had been assumed that these could be ignored because of the geographic matching of the PGPs and their comparison groups. This assumption is reasonable as long as the PGP and comparison groups are similarly affected by such changes. Perhaps the most likely source of a differential payment effect is the Indirect Medical Education (IME) and Disproportionate Share (DSH) payments to hospitals made under the Inpatient Prospective Payment System (IPPS). To test this assumption, Demonstration savings and the resulting performance payments were re-estimated by removing IME and DSH payments to hospitals from expenditures. This analysis found that this exclusion can influence the savings calculation and the resulting performance payments. The potential change in performance payments is not so pronounced overall, but is more re-distributive across the PGPs, with some earning more and some less. Consequently, the assumption that geographic matching obviates the need to adjust for Medicare payment system changes may need to be reassessed in future Demonstration designs, depending upon the desired incentives.

Conclusion

The innovation of the PGP Demonstration model is that provider groups are given a financial incentive to provide more efficient, higher quality care. Performance payments are computed with the standard Medicare FFS claims processing system, and requires no additional

data submission on the part of participating practices other than the sample of chart-based quality measures. While financial risk is mitigated by the continuance of FFS payments, providers are at risk for infrastructure improvements.

The improvement in the quality measure processes and reporting in the first two years of the Demonstration suggest that access has been improved while providing high quality care. The effect of the Demonstration on promoting expenditure savings is less certain. A major evaluation issue is: what has driven the performance results? Why did four PGPs earn a performance payment, and six did not? Because the Demonstration was not structured to test specific interventions, and the beneficiaries are assigned retroactively, it is difficult to identify a specific protocol or action that explains performance.

The most distinguishing factors between the PGPs that earned and did not earn performance payments were the pre-existing expenditure trends and the organizational structures. Various analyses could not determine the precise reasons for the difference in financial performance. Disease management, care coordination programs, and information systems have been implemented by all PGPs. The specific contributions of the disease management and care coordination programs in reducing Medicare expenditures could not be rigorously determined and are largely anecdotal. With all PGPs having exceptionally high marks for their quality reporting and processes, the 4 PGPs earning performance payments in PY2 had virtually the same quality performance as those that did not earn performance payments. The increase in the PGP risk scores over those of comparison populations appeared to be a characteristic of all PGPs; therefore, this factor does not explain earnings.

Since the purpose of the Demonstration is to test incentives, the reaction to the incentives is perhaps the most meaningful result for policy implications The two incentives, to improve quality reporting and processes, and to reduce expenditures are overlapping. The fact that all PGPs improved their quality reporting and processes without knowing in advance that they would be financially rewarded through performance payments suggests that quality improvements can be gained by the reasonable expectation of receiving a reward. Thus, the PGPs have demonstrated that physician practices can change their behavior in response to expected returns and that the PGP model may be a useful tool for bringing about quality improvement in FFS Medicare.

Since expenditures were trending lower years before the Demonstration started, reductions in expenditures can be attributed to the continuation of pre-existing trends, as well as to any response to the financial incentives of the Demonstration. The analyses could not determine the extent to which each influenced savings. Over the next three years of the Demonstration, the stability of the PGPs earning performance payments, changes in the number of PGPs that earn performance payments, and the magnitude of those payments should give insight into how these 10 physician practices react to the financial motivation of a shared savings model. Additional performance data and analyses from the remaining years of the Demonstration should help clarify these issues.

SECTION 1: INTRODUCTION

The Physician Group Practice (PGP) Demonstration was mandated by Section 412 of the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act (BIPA) of 2000. Section 412 requires a report to Congress "not later than two years after the enactment of this section, and biennially thereafter for six years." The reports are to "address the impact of the use of those authorities on expenditures, access, and quality…" The first and second reports to Congress (in 2002 and 2004) described the design and pre-implementation status. The third report (2006) described in detail the demonstration design, the performance payment model, PGP site characteristics, care redesign initiatives, and the quality metrics. Quality and expenditure performance data were not available for the third report. This fourth report assesses the impacts required by BIPA provisions: expenditures, access, and quality using data from the first two years of operations, which are performance year 1 (PY1: April 2005 to March 2006) and performance year 2 (PY2: April 2006 to March 2007). The base year (BY) for measuring performance improvement is calendar year 2004.

This report is organized as follows: Section 2 provides background on the participants and the demonstration design; Section 3 presents and analyzes the results from PY1 and PY2; and Section 4 discusses potential refinements that may need to be explored further if this Demonstration model were to be expanded to other initiatives. An Appendix of tables follows Section 4. This report is based on the study and analyses provided by RTI International under

contract to the Centers for Medicare & Medicaid Services (CMS). RTI International was also the implementation contractor for the PGP Demonstration.

SECTION 2: BACKGROUND

Key Design Features

Section 412 of BIPA requires the Demonstration to focus on "the efficiencies of furnishing heath care in a group-practice setting as compared to the efficiencies of furnishing health care in other health care delivery systems." BIPA also specified the basic features of the Demonstration model by requiring incentive payments to participants that are "equal to a portion of the Medicare savings realized for such year relative to a performance target." Additional features referenced in BIPA include: 1) providing performance payments for process and outcome improvements; 2) continuing payment on a fee-for-service (FFS) basis, 3) excluding Medicare+Choice (predecessor program to Medicare Advantage) enrollees, and 4) using geographic areas to define the scope of the Demonstration. Because FFS benefits and payment structures were not altered, a patient attribution rule was the only option for identifying beneficiaries to be included in the Demonstration.

The design of the Demonstration is described in detail in several papers and reports, including: "Medicare Physician Group Practice Demonstration Design: Quality and Efficiency Pay-for-Performance," by Kautter, Pope, and Trisolini (2007); the 2006 report to Congress, entitled "Physician Group Practice Demonstration First Evaluation Report;" a GAO report, entitled "Care Coordination Programs Used in Demonstration Show Promise, but Wider Use of

Payment Approach May Be Limited (2008);" and "The Medicare Physician Group Practice Demonstration: Lessons Learned on Improving Quality and Efficiency," by Trisolini, Aggarwal, Leung, Pope, and Kautter (2008). The major design elements are summarized in Table 2-1.

One design feature is the two percent threshold or corridor that is used to measure savings. Within this corridor, two percent above Target Expenditures, and two percent below, no savings or losses are recognized. The 2002 PGP Demonstration Design Report considered the need for a threshold and modeled the impact of sample size and cost variability with threshold percentages ranging from two to four percent. The authors concluded: "A bonus threshold avoids paying a bonus for small differences in actual versus target expenditures that could be due to chance. Choosing an appropriate bonus threshold involves trading off the probabilities of paying deserved bonuses versus not paying undeserved bonuses. Based on our simulations, we recommend a bonus threshold of 2.0% (Pope, 2002)."

Table 2-1 Summary of PGP Design Implementation Features

Attribute	Summary
Participants	The 10 Demonstration participants are large medical practice organizations with diverse organizational structures including free-standing multi-specialty group practices, faculty group practices, integrated delivery systems, and a physician network made up of small and individual physician practices. These large practices have the infrastructure or the ability to implement or redesign systems that can change the delivery of care.
Patient Attribution	Because the PGP Demonstration retains the structure of the Medicare FFS system, there is no enrollment process. Therefore, beneficiaries are retroactively "assigned" to the PGP sites, i.e., they are the "Assigned Beneficiaries (ABs)" if they received the plurality of their office or other outpatient evaluation and management (E&M) services at a PGP site during a year. In PY2, the average number of office or other outpatient E&M visits was 5.5 per Assigned Beneficiary in the PGPs, which represented 85 percent of their total E&M charges for the year. The remaining 15 percent were furnished by other non-PGP local area providers.
Comparison Group	Each PGP has a comparison group (CG) that is used to compute Target Expenditures using the CG risk adjusted expenditure growth rates within their local market. The CG beneficiaries are drawn from the PGP's geographic service area, which consists of all counties in which the PGP derives at least 1 percent of its Assigned Beneficiaries. The beneficiaries in the CG are the complement of the PGP Assigned Beneficiaries within an area. Beneficiaries having any office or other E&M services at the PGP in a performance year are excluded from the CG. Beneficiaries assigned at the PGP in a prior year are also excluded from the CG.
Savings	Target Expenditures are the per capita Base Year (BY) expenditures of the PGP multiplied by the growth rate of CG expenditures. Savings are defined as the: [Target Expenditures minus Actual Expenditures] that exceed two-percent of Target expenditures multiplied by the number of AB person years. The two percent corridor is used to allow for normal variation in Medicare expenditures. Both PGP and CG expenditures are adjusted for changes in the risk characteristics of their beneficiaries. Expenditures are the actual Part A and Part B Medicare payments shown on claims, and exclude deductibles and co-insurance. Medicare Advantage enrollees are excluded from both groups. No adjustment is made for changes in the Medicare payment system over time, as it is assumed that such changes will affect both groups similarly within a geographic area.
Performance Payments	Medicare retains 20 percent of any savings and the remaining 80 percent is shared with the physician groups. Performance payments are derived from the pool of shareable savings based on the groups' levels of financial and quality performance. If all quality targets are met, shared savings can be entirely paid to a PGP as a performance payment. Lower rates of compliance on quality measures will result in lower performance payments. An increasing share of the performance payments are based on the proportion of quality targets met. In PY1, 30 percent was based on quality performance, 40 percent in PY2, and 50 percent in PY3 and PY4. For example, in PY1, 70 percent of

	the maximum performance payment was earned regardless of quality performance, and 30 percent was earned proportional to the quality performance percentages reached. Beginning in PY2, the PGPs agreed to put their Physician Quality Reporting Initiative (PQRI) payments (1.5 percent of allowed charges) at risk using the percentages from their Demonstration quality target performance.
Quality Measures	Quality is measured using 32 ambulatory care quality measures covering five condition modules. The performance on 7 measures can be calculated from Medicare claims, while sampling techniques were used to assess performance on the 25 chart-based measures. To develop a composite quality score, each quality measure is weighted, with claims-based measures weighted higher than clinical record-based measures. PGPs earn points for each quality measure benchmark. Benchmarks can be one of: 75 percent absolute performance, the applicable HEDIS Medicare Advantage plan proxy, or a quality improvement target. The points are summed up and compared to the total points possible in the performance year, which was 45 points in PY2.

Participant Characteristics

Organizational Characteristics

The 10 PGPs are large group practices, ranging from 232 to 1,291 affiliated physicians and caring for over 220,000 Medicare beneficiaries. Table 2-2 summarizes the organizational characteristics of the PGPs. Two participants are faculty group practices within academic medical centers; five belong to an integrated delivery system, which consists of at least one hospital in addition to the physician group (and may include other health care providers such as home health agencies or nursing homes); two are freestanding physician group practices; and one is a physician network that is sponsored by a hospital affiliate and comprised of 60 small and individual physician practices. The physician network PGP is unique in that it is not a group practice, as are the other sites. Rather it is a management services organization that provides services to improve quality, medical management, public reporting, contracting, and information management services to multiple independent physician practices, each of which was offered the choice to join the Demonstration.

Most of the PGPs have experience with capitated managed care. Seven PGPs either own or had owned an HMO. Also, some of the participating provider groups have been delegated care management responsibilities by managed care insurers. As a result, most PGP participants built care management infrastructures prior to their involvement in the Demonstration.

Table 2-2
PGP Demonstration Participants: Organizational Characteristics and Selected PY2 Results

Participant	Organizational Characteristics				PY 2 Results		
	Organizational structure	PGP includes hospital?	Not for profit?	Number of providers	Percentage of Quality Points Earned	Assigned Beneficiaries	Earned Performance Payment in PY2?
Academic Medical Centers							
Dartmouth-Hitchcock Clinic	Faculty/community group practice	Yes	Yes	907	98 %	30,600	Yes
University of Michigan Faculty Group Practice	Faculty practice	Yes	Yes	1,291	100	19,200	Yes
Freestanding group practices							
Marshfield Clinic	Group practice	No	Yes	1,039	100	38,700	Yes
The Everett Clinic	Group practice	No	No	250	96	9,700	Yes
Integrated Delivery Systems							
Billings Clinic	Group practice	Yes	Yes	232	98	13,400	No
Geisinger Clinic	Group practice	Yes	Yes	833	100	25,400	No
Forsyth Medical Group	Group practice	Yes	Yes	250	100	14,000	No
Park Nicollet Clinic	Group practice	Yes	Yes	648	98	19,000	No
St. John's Clinic	Group practice	Yes	Yes	522	100	31,700	No
Network model							
Middlesex Health System[2]	Network model	Yes	Yes[2]	293	96	17,700	No

NOTES:
[1] HMO may be owned by associated health system, not demonstration participant per se.
[2] The participant is Integrated Resources for the Middlesex Area (IRMA), which is a for profit venture of the not-for-profit Middlesex Health System.

SOURCE: RTI International

Geographic Location

Table 2-3 shows the PGP participants and their geographic characteristics. Three PGPs are located in the Northeast, one is in the South, four are located in the Midwest, and two are in the West. Half of the participants are located in predominantly rural areas, which include scattered small cities or towns. Three PGP sites are located in suburban small city areas, one is located in a smaller urban area, and one is located in a suburban area adjacent to a large city. No participant is located in a large urban core city.

Table 2-3 PGP Participants: Geographic Location

Region / Participant	Service area	Urban / Rural Characteristic
Northeast		
Dartmouth-Hitchcock Clinic	New Hampshire / Eastern Vermont	Rural, small city
Middlesex Health System	South-Central Connecticut	Suburban, small city
Geisinger Clinic	Central-Northeast Pennsylvania	Rural, small city
South		
Forsyth Medical Group	Northwest North Carolina	Small urban city
Midwest		
Marshfield Clinic	North-Central Wisconsin	Rural, small city
St. John's Clinic	South-central Missouri / Northwest Arkansas	Rural, small city
Park Nicollet Clinic	South-Central Minnesota	Suburban, large metropolitan
University of Michigan Faculty Group Practice	Southeastern Michigan	Suburban, small city
West		
Billings Clinic	South-Central Montana/Northwestern Wyoming	Rural, small city
The Everett Clinic	West-Central Washington	Suburban, small city

SOURCE: RTI International

Demonstration Incentives

As added by BIPA, Section 1866A(a)(1) of the Social Security Act requires the PGP Demonstration to test and expand the use of incentives that (A) "encourage coordination of care furnished to individuals under the programs of Medicare Parts A and B…; (B) encourage investment in administrative structures and processes to ensure efficient service delivery; and (C) reward physicians for improving health care outcomes." Since the primary goal was to test incentives, Demonstration participants had complete flexibility in determining the methods of achieving savings, improving quality, coordinating care, investing in administrative structures, and rewarding physicians.

Performance payments are structured so that Demonstration participants have an incentive to reduce expenditures and improve quality. To receive a performance bonus, expenditures must be less than 98 percent of Target Expenditures. These expenditures include all Medicare Part A and Part B expenditures, except hospice. It may be reasoned that to reduce expenditures, a practice might reduce not only their own professional services, but also reduce or substitute lower cost services over which they have direction. Since a site can receive up to 80 percent of these savings, the incentive to reduce expenditures should be strong. Lowering the sharing proportion would undoubtedly weaken the incentive. The PGPs are not at risk if their Actual Expenditures exceed 102 percent of Target Expenditures, except that such losses must be offset against any future period performance payment.

The largest expenditure component for Medicare beneficiaries is typically inpatient services; therefore, this is the most likely area to reduce expenditures, at least in the short run. However, the presence of a hospital within the PGP organization may interfere with this strategy. For the two freestanding group practice PGPs, there is undoubtedly a strong incentive to reduce inpatient admissions, since the organization will not lose revenue as a result of deferred admissions.

For the eight participants that own or are affiliated with a hospital, the strength of the incentive to reduce inpatient admissions is not as clear. These participants may or may not be concerned about the loss of inpatient admissions and the associated revenues, depending on the available inpatient capacity. If excess capacity exists and admissions are reduced, an organization may lose more revenue than could be gained by the Demonstration performance payment. Thus, the Demonstration incentive to reduce expenditures for these participants may not be as strong as for the PGP participants without hospitals. If there is no excess inpatient capacity, these participants may not suffer revenue reductions by admitting fewer Medicare beneficiaries. In this case, it may be reasoned that systems having high occupancy hospitals may have a stronger incentive under the Demonstration to reduce inpatient care, as they may be able to reduce avoidable admissions or use lower cost care substitutes without affecting their inpatient revenue.

The appropriateness of the performance payment is important if the incentives are to work correctly. The PGPs are responsible for controlling all expenditures, which are defined to be all Part A and Part B Medicare payments shown on the claims. Services rendered by a practice may represent only a small fraction of the Total Expenditures. For example, a PGP may

refer a patient to a hospital that is not owned or controlled by the PGP. Total Expenditures would include the cost of inpatient services, while the actual services provided by the PGP may be only the professional fees. While a performance payment may not be a large percentage of the Total Expenditures assigned to the PGP, it can be a significant percentage of the expenditures (Medicare revenue) rendered by a practice. Thus, the PGP has the potential to earn a substantial performance payment as a proportion of its own Medicare revenues, which should strengthen the financial incentive of the Demonstration.

The Demonstration also provides a strong incentive to improve quality, as the maximum performance payment can only be earned if all quality measures benchmarks are met or exceeded. Otherwise the PGPs may lose a portion of their performance payment. In PY1, 30 percent of the performance payments were based on quality measure performance; in PY2, 40 percent; and in PY3 and PY4, 50 percent will be based on quality measure performance. The 2007 PQRI incentive payments were made to the PGPs using the same quality measure proportions calculated under the Demonstration. These payments combined with the incentives of the Demonstration should act to improve quality performance.

PGP Strategies

Two goals of the Demonstration stated in the BIPA legislation were to: (A) "encourage coordination of care furnished to individuals under the programs of Medicare Parts A and B…;" and (B) "encourage investment in administrative structures and processes to ensure efficient service delivery." To meet these goals, as well as to maintain and improve their performance on the quality indicators, the PGP participants indicated several main strategies, including the: (1)

implementation of standardized, evidence-based care models and protocols, (2) better adherence to quality of care protocols on the part of both patients and physicians through disease management interventions, (3) implementation of patient registries, used to track patients and identify any gaps in care, (4) provider education and feedback including data profile reports comparing individual providers to their peers or other benchmarks, and (5) implementation of health information technology, including patient registries, enhanced electronic medical record systems, and patient monitoring systems to improve care.

A unique characteristic of these programs is their integration into physician practices. Such integration may provide additional resources to the clinical team, such as a case manager or support from a health information technology system, which can be combined with information from the medical record. This information can help provide a more complete picture of the patient's health, from which to direct care. In addition, patients may have more immediate access to clinical staff if an intervention with the patient is necessary.

All PGP sites have stated that they have implemented disease and/or care management programs to reduce expense, while improving the quality of care. These clinically based care management programs can be characterized as being Disease Specific or related to General Care Coordination. Disease Specific Programs target a subset of beneficiaries based on diagnosis. It may be reasoned that the PGP sites would focus on implementing disease-related programs that can generate the most short-term quality improvement and cost savings, such as programs related to CHF, diabetes, or vaccinations. General Care Coordination Programs cast a wider net for enrollment criteria, since enrollment is not disease based. Several programs focus on patients with multiple conditions or patients that are high cost or high risk**.**

Table A-1 in the Appendix summarizes the care coordination programs across the PGP sites that were identified at site visits and discussions with the PGPs. Table A-3 in the Appendix shows selected excerpts from the proceedings of the Commonwealth Funds' 2006 PGP conference contained in "Medicare Physician Group Practices: Innovations in Quality and Efficiency." These excerpts give additional information and illustrate some of the programs and changes that the PGP sites have undertaken. The entire conference proceedings document should be accessed to see the programs and strategies stated by the PGP sites (see reference section for url).

The Commonwealth Report on Lessons Learned (Trisolini, 2008) summarized conference proceedings in which the PGPs identified strategies they used to improve care. A summary from these proceedings is shown below, as they help identify the actions taken by the PGPs in response to the Demonstration. Motivating physicians was integral to implementing these strategies. The Report references seven steps that are necessary for change. These steps included establishing a sense of urgency, building a coalition, sharing a vision, empowering staff, communicating goals, establishing processes and sustaining "change by emphasizing new patient management techniques and treatments." The Report gives several examples of these strategies, which are shown below:

> Information Systems
>
> Registries enable providers to gain access to a broader set of point-of-care information for review prior to a health visit. This type of information is useful for providing accurate care and facilitating discussions between patients and providers during health care visits. St. John's Health System uses its disease registry to generate "visit planner" documents prior to each health care visit that help physicians plan and structure the visit. The visit

planner serves as a checklist for physicians to ensure that needed tests and services are provided.

Another strategy is to develop registries that capture patient co-morbidities and integrate registries initially developed for individual diseases. This integrated information allows for more complex care management vs. individual disease management. The University of Michigan and St. John's Health System have both developed registries for chronic diseases instead of focusing on a single condition, and they have found them to be valuable tools for improving coordinated care and delivery.

Use of Teamwork

Staff often prefers providing health care as a team in which members work together to determine optimal care for a patient. PGPs have found that working in teams results in an increase in their own and patient satisfaction as well as improvement in quality measures. Each team member makes a different professional contribution. In a team-based system, physicians can shift some responsibilities to physician assistants or other non-physician providers, which can substantially decrease their workloads. A common concern, however, is that physicians may hesitate to shift some of their work because they are uncertain of the abilities of other staff. The implementation of a team-based system may therefore involve changing the general physician mindset. Another challenge is the often limited opportunity for teams to sit together and discuss care.

Strong Leadership

The implementation of new interventions requires strong leadership from both the physician and administrative sides. Administrative leaders are important for generating interest in management systems and securing resources. Physician champions for these types of interventions are also very important since they are involved in all patient care interventions. The best advocates are often the peers of individuals who will be affected by the program. Thus, physician champions are usually the best candidates for communicating the benefits of new clinical interventions.

Care Transitions

A care transition occurs when a patient is transferred from one provider to another. This may involve transfers within a facility, transfers between facilities within a larger integrated delivery system, or transfers in or out of a particular facility. For elderly patients with chronic conditions, poor transition management can result in hospital readmissions or increased visits to emergency facilities.

Health care providers historically have given too little emphasis on care transitions, partially because clinical responsibilities and associated reimbursements are often divided between providers. The demonstration incentives reward PGPs for reducing overall Medicare spending, however, 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 are testing new transition management programs that may apply to patients with particular diagnoses or those undergoing particular types of transitions, such as the transition from hospital to home. Preventing hospital readmissions through timely outpatient follow-up care by physicians has been a particular focus of these programs since it has the potential to reduce costs and also patient morbidity.

In addition, demonstration staff is also exploring management of other types of transitions, such as those from hospitals to nursing homes. 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. Coordinating care among the multiple specialist physicians who may treat high-risk patients is also a potential area for improvement, since they may not communicate well about treatments and prescriptions a patient has received.

SECTION 3: RESULTS FROM PERFORMANCE YEARS 1 AND 2

The Demonstration began on April 1, 2005 and will operate for five years through March 31, 2010. This report summarizes findings from the first two performance years on expenditures, access, and quality as required by the BIPA legislation. The "base year" (BY) or reference point for measuring quality and efficiency improvements is calendar year 2004, and the five "performance years" (PYs) are the annual periods that begin in April of each year. This section is organized by the impact topic required in BIPA. The data and analyses for this report were provided by RTI International under contract to CMS.

Impact on Access

Because the FFS benefits and payment structures were not altered, access to care was not expected to change as a result of the Demonstration. To assess changes in access, relevant statistics for the beneficiaries assigned to the PGPs were generated in Table 3-1, and these statistics are discussed below.

Assigned Beneficiaries: An attribution algorithm assigns a beneficiary to a PGP, i.e., "Assigned Beneficiaries (ABs)," if the plurality of office or other outpatient Evaluation and Management (E&M) visits are furnished by that PGP. The total number of beneficiaries assigned to the 10 PGP sites has been fairly stable across all sites, starting in the BY with 223,000, and declining slightly to 220,000 in PY2. The number of

beneficiaries assigned to the PGPs in PY2 had a range of between 9,715 and 38,743 and the average number of beneficiaries per site was 21,958. While the total number of ABs has changed only slightly over the two year period, about one-third of them leave the cohort and are replaced each year.

Evaluation & Management Visit Characteristics: The average number of E&M visits per year were 5.4 per beneficiary in the BY. This average increased slightly to 5.5 visits in PY2.

Proportion of allowed charges: On average, the mean proportion of the allowed charges for office or other outpatient E&M visits provided to the Assigned Beneficiaries at the PGPs was 85 percent of the total outpatient E&M charges. This statistic suggests that the attribution model is appropriate, as this level of E&M visits should provide sufficient opportunities for PGPs to manage and coordinate the care provided to their Assigned Beneficiaries.

Medicare Eligibility: In the BY, PY1 and PY2, about 82 percent of ABs were Medicare eligible due to age. In PY2, only a small percentage of beneficiaries, 0.8 percent, were Medicare eligible by reason of having end-stage renal disease (ESRD). The proportion of beneficiaries entitled to Medicare due to a disability increased from about 16.5 percent in the BY to 18.0 percent in PY2.

In conclusion, both the average number of outpatient E&M visits (5.5 in PY2) and the overall proportion of outpatient E&M charges (85 percent) were essentially constant between the

BY and PY2. This consistency and the stability of the number of AB (220,000 in PY2) suggest that access to care did not change under the Demonstration. Since beneficiaries had access to care management programs and redesigned care processes that were established or extended by the participating sites, it may be reasoned that one effect of the Demonstration was to improve access.

Table 3-1 PGP Assigned Beneficiary Characteristics

Characteristic	BY	PY1	PY2
Number of Assigned Beneficiaries	223,000	224,000	220,000
Average across 10 sites	22,320	22,389	21,958
Range across sites:			
Minimum	8,383	9,313	9,715
Maximum	44,609	42,017	38,743
Office or outpatient E&M services			
Mean number of office or other outpatient E&M visits at the PGP	5.4	5.4	5.5
	Percentages		
Mean percentage of allowed charges for office or other outpatient E&M visits provided at the PGP	85.0	85.0	85.0
Mean percent of beneficiaries by Medicare eligibility			
Aged	82.8	82.0	81.3
ESRD	0.7	0.8	0.8
Disability	16.5	17.2	18.0
	100.0	100.0	100.0
Medicaid eligibility			
Percentage Medicaid eligible for at least one month during the year	13.5	14.2	15.1
Hospice status			
Percentage of beneficiaries enrolled in hospice during the year	1.4	1.7	1.8

SOURCE: RTI analysis of 2004 through 2007 using 100 percent Medicare Claims Files and Enrollment Dataset

Impact on Quality

When fully transitioned in PY3, health care quality will be measured using 32 ambulatory care quality indicators covering five condition modules. In PY1, 10 diabetes measures were in effect. In PY2, 10 Congestive Heart Failure (CHF) measures along with 7 Coronary Artery Disease (CAD) measures were added. In PY3, 5 hypertension / preventive care measures will be added, bringing the total quality measures to 32. The performance on 7 measures can be calculated from Medicare claims, while sampling techniques are used to assess performance on the 25 chart review measures. Table A-2 in the Appendix lists all quality measures used in the Demonstration.

A composite quality score is used to summarize each PGP's quality reporting performance. The numerator of the composite score is the number of points earned for the quality measures which meet or exceed the benchmark. Benchmarks are defined for each quality measure within each PGP and can be one of three statistics: 1) the higher of 75 percent compliance or the Medicare mean for the Health Effectiveness Data and Information Set (HEDIS) measure; 2) a 10 percent reduction in gap between administrative baseline and 100 percent compliance; or 3) the 70th percentile Medicare HEDIS level. The denominator is constructed from the total number of possible points available for that performance year. In computing the aggregate score, claims-based measures are given a weight of 4 and chart-based measures are given a weight of 1.[1] In PY1, the total number of possible points was 22, which is comprised of the 4 claims-based measures weighted by 4 (16 points in total) and 6 chart-based

[1] This weighting was developed to take advantage of the higher statistical precision associated with the larger claims sample sizes.

measures weighted by 1 (6 points in total), for a total of 22 possible points. In PY2, the total number of possible points was 45, which is comprised of the 6 claims-based measures weighted by 4 (24 points in total) and 21 chart-based measures weighted by 1 (21 points in total), for a total of 45 possible points. Table A-2 in the Appendix also shows the number of points that may be earned for meeting the benchmarks.

The following quality reporting performance was observed during the Demonstration through PY2:

Performance on All Measures (chart and claims-based)

PY1: In PY1, 10 Diabetes Mellitus (DM) quality measures were in effect. All PGPs achieved benchmark performance for at least 7 of these 10 quality measures, and two physician groups achieved benchmark performance on all 10 measures. In PY1, the average performance score across the PGPs was 19.5 points out of a possible 22 points, or 88.6 percent.

PY2: In PY2, the 10 diabetes measures remained in effect, and 10 Congestive Heart Failure (CHF) measures along with 7 Coronary Artery Disease (CAD) measures were added. All PGPs achieved benchmark performance on at least 8 of the 10 DM quality measures. All PGPs achieved 100 percent benchmark performance on all of CHF and CAD quality measures. Overall, the PGPs achieved benchmark performance on 25 or more quality measures, and five PGPs achieved benchmark performance on all 27

measures. In PY2, the average performance score across the PGPs was 44.3 points out of a possible 45 points, or 98.4 percent.

Improvement over BY: The PGPs demonstrated improved quality of care on the chronic conditions measured in both PY1 and PY2. Between the BY and PY2, the PGPs increased their quality scores an average of 9 percentage points on the DM measures, 11 percentage points on the HF measures, and 5 percentage points on the CAD measures.

Physician Quality Reporting Incentive (PQRI) Payments in PY2: The Tax Relief and Health Care Act of 2006 required CMS to establish a quality data reporting system and to pay an incentive for satisfactorily reporting data on quality measures beginning in 2007. The PGPs agreed to use their quality measure proportions calculated under the Demonstration to determine their PQRI incentive payments in 2007. Each PGP group received at least 96 percent of their PQRI incentive payment based on their performance on PY2 quality measures, and PQRI incentive payments totaled $2.9 million. The PQRI payments earned for PY2 varied between $98,000 and $500,000 across the 10 PGPs with the PQRI payment exceeding the performance payments (of zero) at 6 PGPs. The addition of the PQRI incentive payments undoubtedly enhanced the groups' financial incentive to reach and exceed quality measure targets. In PY1, there were no PQRI payments and only two PGPs achieved benchmark performance on all 10 measures. In PY2, when PQRI payments were introduced, five groups achieved benchmark performance on all 27 measures. In conclusion, PQRI payments combined with the Demonstration's incentives worked to improve the quality performance.

Comparison of Claims-based Measures

Each CG is comprised of beneficiaries who reside in the geographic areas served by the PGPs. Within the Demonstration, only the expenditure and risk score growth rates of the CGs are used to compute the Target Expenditures of the PGPs. However, their claims, and in particular, their claims-based quality measure performance, can also be analyzed and compared to the PGPs. Thus, the impact of the Demonstration on quality can also be examined by comparing the values of the seven claims-based quality measures for each PGP and its CG. This comparison should be used with some caution, since only the PGPs know that their quality reporting is being monitored and their performance payments may be based on the results. Thus, in theory, only the PGPs have a financial incentive to improve their quality processes and reporting over the BY. The proportions reached and the improvement exhibited by the CGs may be interpreted as showing the change that would have occurred in the absence of the Demonstration.

Section 1 of Table 3-2 shows the percentages reached of the seven claims-based quality measures in PY2. PGP performance exceeded that of the CGs on all measures. While PC-5 Breast Cancer Screening will be phased into performance results in PY3, its measure is still available for comparison in PY2. The largest differences between the PGPs and CGs quality performance were found for DM-6 (urine protein testing) and for PC-5 (breast cancer screening), at 11.7 percentage points and 6.4 percentage points, respectively. The smallest differences were for HF-2 (left ventricular ejection fraction testing) and DM-1 (HbA1c management), 2.5 percentage points and 2.8 percentage points, respectively.

Section 2 of Table 3-2 shows that the PGPs also exhibited more improvement than CGs in all but one category between the BY and PY2. However, improvement in three of the seven measures was not statistically significantly different from the improvement observed for the CGs. The largest improvements by the PGPs were in DM-6 (urine protein testing) and HF-2 (LVF testing following hospitalization). In both measures, the PGPs improved their scores by almost 5 percentage points over the CGs. Even after adjusting for pre-Demonstration trends in the claims-based quality indicators, the PGPs improved their claims-based quality process indicators more than their comparison groups.

Table 3-2 Performance of Claims-Based Measures in PY2

		Section 1			Section 2		
		PY2 Percentages Reached			Percentage Point Change in PY2 over BY		
Measure		PGP	CG	Difference: PGP - CG	PGP	CG	Difference: PGP - CG
DM-1	HbA1c Management	92.2	89.4	2.8	0.9	1.0	−0.1
DM-4	Lipid Measurement	86.2	80.9	5.3	4.1	2.3	1.8 ***
DM-6	Urine Protein Testing	83.2	71.5	11.7	8.4	3.8	4.6 ***
DM-7	Eye Exam	72.8	66.9	5.9	2.2	1.8	0.4
HF-2	Left Ventricular Ejection Fraction Testing	90.3	87.8	2.5	4.8	0.1	4.7 ***
CAD-5	Lipid Profile	78.1	74.5	3.6	4.5	3.1	1.4 **
PC-5	Breast Cancer Screening	77.1	70.7	6.4	1.1	0.2	0.9

* = statistically significant at the 10% level.
** = statistically significant at the 5% level.
*** = statistically significant at the 1% level.
Note: PC-5 is not part of performance calculations until PY3.

SOURCE: RTI analysis of Demonstration data

Impact on Expenditures

In the Demonstration, savings are defined for each PGP as the difference between Target Expenditures and Actual Expenditures that exceeds a two-percent threshold of Target Expenditures, multiplied by the number of AB person years. Positive savings suggests that the Demonstration has reduced Medicare expenditures (as Actual Expenditures are lower than 98 percent of Target Expenditures), and negative savings (Actual Expenditures are more than 102 percent of Target Expenditures) suggests that the Demonstration resulted in higher expenditures. If Target Expenditures less Actual Expenditures are within +/- 2 percent of Target Expenditures, no positive or negative savings are attributed to a PGP. While negative savings do not place a PGP at financial risk, they offset positive savings in calculating overall net savings associated with the Demonstration.

A portion of the savings, up to 80 percent, can be earned as performance payments. If all quality targets are met, a PGP can earn the entire amount, i.e., 80 percent of savings, as a performance payment. An increasing share of the performance payments are based on the proportion of quality targets met. In PY1, 30 percent was based on quality performance, 40 percent in PY2, and 50 percent in PY3 and PY4. For example, in PY1, 70 percent of the maximum performance payment was earned regardless of quality performance, and 30 percent was earned proportional to the quality performance percentages reached. Table 3-3 shows the savings and performance payments in the first two years of the Demonstration, and these are summarized below.

PY1: The net savings to the Medicare Trust Funds were $677,000. Savings of $9,530,000 were achieved by two PGPs. Offsetting these savings were the associated performance payments of $7,323,000, and negative savings of $1,530,000 at two other PGPs. Six PGPs had actual expenditures lower than their targets but were within the two percent corridor and thus did not qualify for performance payments.

PY2: The net savings to the Medicare Trust Funds were $1,583,000. Savings of $17,377,000 were achieved by four PGPs, two of which earned performance payments in PY1. Offsetting these savings were $13,840,000 in performance payments to the four PGPs and negative savings of $1,954,000 from one PGP. Three PGPs had actual expenditures lower than their targets but were within the positive two percent corridor, and thus did not qualify for performance payments, and two PGPs had expenditures that exceeded the target but were within the negative two percent corridor.

Combined PY1 and PY2: In the first two years, the net savings to the Medicare Trust Funds were $2,260,000. The Demonstration achieved savings of $26,907,000. Offsetting these savings were performance payments of $21,163,000, which were made to four PGPs, and losses of $3,484,000 at two PGPs. Since most of the savings are returned to the sites as performance payments, the net savings to the Medicare Trust funds in the first two years of the Demonstration were minimal when expressed as a percentage of all Target Expenditures.

Table 3-3 Summary of Savings and Performance Payments (Amounts in 000's)

PGP Site (Names Blinded)	PY 1 Savings	PY 1 Earned Performance Payments	PY 2 Savings	PY 2 Earned Performance Payments	Combined Years Savings	Combined Years Earned Performance Payments
Savings and Payments						
PGPs Earning Performance Payments in PY2						
PGP 1	$ 6,035	$ 4,565	$ 7,227	$ 5,782	$ 13,262	$ 10,347
PGP 2	3,495	2,758	1,549	1,239	5,044	3,997
PGP 4	0		8,437	6,690	8,437	6,690
PGP 6	0		164	129	164	129
Total	**$ 9,530**	**7,323**	**$ 17,377**	**13,840**	**$ 26,907**	**21,163**
PGPs Not Earning Performance Payments in PY2						
PGP 3	0		0		0	
PGP 5	0		0		0	
PGP 7	0		0		0	
PGP 8	0		0		0	
PGP 9	-214		0		-214	
PGP 10	-1,316		-1,954		-3,270	
Total	**-1,530**		**-1,954**		**-3,484**	
Total Expenditure Savings	**8,000**		**15,423**		**23,423**	
Less: Performance Payments	**-7,323**		**-13,840**		**-21,163**	
Net Trust Funds Savings	**$ 677**		**$ 1,583**		**$ 2,260**	

SOURCE: Compiled from RTI analysis of Demonstration data.

Analyses

The remainder of this section presents analyses that were performed to help understand how PGP savings were generated. Using claims and site interview data, four analyses were done. The first analysis compares expenditures by condition. The second analysis examines expenditures by service component. In the third analysis, risk scores and coding were examined in order to assess the extent to which diagnostic coding practices may have influenced savings. The fourth analysis assesses the impact of expenditure trends observed prior to the start of the Demonstration. The last section draws a conclusion about what drives financial performance based on the results from the four analyses described above, as well other factors that may be associated with financial performance. In these analyses, the four PGPs that earned performance payments in PY2 are referred to as the "4 PGPs earning performance payments in PY2." They are distinguished from the six that did not, which are referred to as the "6 PGPs not earning performance payments in PY2."

Subpopulation Analysis by Condition

Target Expenditures minus Actual Expenditures (for all Part A and Part B services) were estimated for various subpopulations, including beneficiaries with at least one of eight high cost conditions and other selected beneficiary categories. Some high cost conditions have been stated as targets for care management programs; therefore, the expenditure profiles of these conditions may help explain some Demonstration financial impacts, particularly of the 4 PGPs earning performance payments in PY2. Table 3-4 displays the expenditures by condition for PY2 on an annual, per person basis, i.e., "per person year (PPY)." Note that the difference between Target

Expenditures and Actual Expenditures in this and subsequent analyses are not subject to the 2 percent corridor. However, statistical significance is used to identify meaningful results. There is also considerable overlap among the subgroups because a beneficiary can be in more than one subgroup. "Positive" results indicate that Actual Expenditures were lower than Target Expenditures for the subpopulation and "negative" results indicate that Actual Expenditures exceeded Target Expenditures.

Across the 10 PGPs, Actual Expenditures were lower than Target Expenditures for beneficiaries with diabetes mellitus (DM: $224 per person year lower), coronary artery disease (CAD: $555 per person year lower),[2] and chronic obstructive pulmonary disease (COPD: $423 per person year lower). The results for the diabetes and CAD subgroups were somewhat expected, as these diseases were mentioned by most PGPs as care / disease management targets sites. The COPD results were somewhat surprising as few, if any, PGPs mentioned targeting COPD for care / disease management. However, it is a high-cost group with a significant hospital admission rate; therefore, it could be affected by PGP care management programs or other factors affecting expenditures for high-cost beneficiaries.

For the 4 PGPs earning performance payments in PY2, Actual Expenditures were $334 per person year less than Target Expenditures, which is statistically significant ($p < .01$), compared to Actual Expenditures exceeding Target Expenditures by $23 per person year for the 6 PGPs not earning performance payments, a difference which is not statistically significant. Not surprisingly, the 4 PGPs earning performance payments in PY2 had statistically significant

[2] As defined for this analysis, CAD includes primarily symptomatic coronary artery disease, i.e., acute myocardial infarction, unstable angina, stable angina, and old myocardial infarction. Asymptomatic chronic CAD is not included in the definition.

lower costs for many conditions: diabetes mellitus, CAD, COPD, stroke, and heart arrhythmia. The Demonstration impact on the Congestive Heart Failure (CHF) subgroup was not statistically different from zero for either the 4 PGPs earning performance payments or the 6 PGPs not earning performance payments in PY2. This finding was surprising, since nearly all PGPs stated that CHF patients were a major focus of their efforts to reduce expenditure growth.

The 4 PGPs earning performance payments in PY2 also had lower expenditures for beneficiaries with risk scores in the upper 25 percent ($777 per person year less), beneficiaries entitled to Medicare due to disability ($615 per person year less), and hospitalized beneficiaries ($954 per person year less). These expenditure differences are all statistically significant ($p < .01$). The lower costs of hospitalized beneficiaries were not surprising because this is a relatively large and high-cost group where efficiencies can be gained. Inpatient admissions were targeted by some PGPs, and in particular for transition management programs. Although lower, the Actual Expenditures for Medicaid enrollees in the 4 PGPs earning performance payments in PY2 were not statistically different from the Target Expenditures.

Table 3-4 Target Minus Actual Expenditures per Person Year by Subgroup, PY 2 (positive number = favorable)

Subgroup attributes	All 10 PGPs	4 PGPs Earning Performance Payments in PY2	6 PGPs Not Earning Performance Payments in PY2
All Beneficiaries	$ 120 ***	$ 334 ***	$ −23
Beneficiaries with conditions:			
1. Congestive Heart Failure	103	378	−81
2. Diabetes	224 **	340 *	146
3. Coronary Artery Disease	555 **	797 *	393
4. Cancer	−40	54	−102
5. Chronic Obstructive Pulmonary Disease	423 **	1,093 ***	−23
6. Stroke	137	1,053 *	−473
7. Vascular disease	134	385	−34
8. Heart arrhythmia	182	577 **	−81
9. Any of the above diagnoses	126 *	371 ***	-37
Other Attributes:			
10. Entitled to Medicare by disability	228	615 ***	−30
11. Any of the 70 risk adjustment diagnoses	161 **	459 ***	−38
12. None of the 70 risk adjustment diagnoses	-5	−9	−3
13. Hospitalization (conditioned on IP costs > 0)	554 ***	954 ***	288
14. No Hospitalization (conditioned on IP costs = 0)	60 ***	173 ***	−15
15. Risk score in upper 25 percent	231	777 ***	−133
16. Medicaid enrollee	112	316	−24

Statistical significance is a two-tailed test of difference from zero.
* = statistically significant at the 10% level.
** = statistically significant at the 5% level.
*** = statistically significant at the 1% level.

SOURCE: RTI analysis of Demonstration data.

Expenditures by Service Component

Table 3-5 shows Target Expenditures minus Actual Expenditures by service component for PY2. Over all 10 PGPs, outpatient expenditures were $83 per person year less than target and statistically significant ($p < .01$). In contrast, inpatient expenditures were only $25 per person year less than expected, and the difference was not statistically significant.

Among the 4 PGPs earning performance payments in PY2, inpatient hospital expenditures were $113 per person year lower than target, and outpatient expenditures were $185 per person year lower than target. Hospital outpatient expenditures were $126 per person year lower than target and was the largest contributor to the total outpatient expenditure result. This was not surprising since the PGPs stated that their care management programs were typically oriented towards keeping patients out of the hospital. Apparently the 4 PGPs earning performance payments in PY2 were more successful at this strategy than the others. Skilled nursing facility expenditures exceeded Target Expenditures, but the difference was not statistically significant. For the 6 PGPs not earning performance payments in PY2, Actual Expenditures were not statistically different from Target Expenditures in any service component, except for home health and durable medical equipment, two relatively minor expenditure categories.

Table 3-5 Target minus Actual Expenditures per Person Year by Component, PY2 (positive number = favorable)

Expenditure component	Average of physician group practices: All 10 PGPs	4 PGPs Earning Performance Payments in PY2	6 PGPs Not Earning Performance Payments in PY2
Total	$ 120 ***	$ 334 ***	$ −23
Hospital inpatient	25	113 **	−35
Skilled nursing facility	−13	−17	−11
Total outpatient	83 ***	185 ***	16
Physician/supplier	30 **	42 *	23
Hospital outpatient	39 **	126 ***	−20
Home Health	21 ***	18 *	22 ***
Durable medical equipment	−1	−19	12 *

* = statistically significant at the 10% level.
** = statistically significant at the 5% level.
*** = statistically significant at the 1% level.
SOURCE: RTI analysis of Demonstration data.

Impact of Risk Scores and Coding

The Hierarchical Condition Categories (HCC) model was adapted to adjust expenditure estimates for beneficiary risk.[3] There is an implicit assumption in the design of the Demonstration that medical conditions of beneficiaries are coded similarly in both the PGPs and the CGs. There has been a pattern of increased intensity and specificity of coding among FFS

[3] The Hierarchical Condition Categories (HCC) model, used to risk adjust payments in Medicare Advantage, uses 70 groups of ICD-9-CM codes, each with similar disease characteristics and costs, as markers of morbidity and predictors of costs. Claims are the source of diagnoses for the model. Additional categories or cells are also included in the model for gender, age, disability, and Medicaid entitlement. The relative costs of the HCC groups are used to generate risk scores for each beneficiary. A concurrent model was developed for the Demonstration, which uses BY payment weights to determine performance year target expenditures.

providers nationwide over the last 10 years or more. This trend was expected for both PGP and CG providers. If the upward trend in risk scores is the same for PGP and CG beneficiaries, comparisons of PGP and CG savings are unaffected. Since the risk model is dependent on the diagnosis codes placed on claims, it is possible to increase HCC risk scores by changing coding practices as well as by treating a sicker patient mix.

Within the Demonstration, the PGPs have a financial incentive to more fully code, or code more intensely, because of the potential impact on performance payments. Risk scores can be raised by increased screening for, and recording of disease, coding disease with close attention to level of severity allowable, and being sure that codes for chronic disease are consistently recorded on claims. Qualifying diagnoses from medical services provided by both PGP and non-PGP providers are used to risk-adjust Demonstration Assigned Beneficiaries. Thus, diagnoses from participating PGPs do not fully determine the risk scores of their Assigned Beneficiaries.

Table 3-6 shows that between the BY and PY2, the disease component of the HCC scores for the PGP sites grew at an average of 8.2 percent, compared to an average of 5.2 percent for the CGs. For the 4 PGPs earning performance payments in PY2, risk scores grew on average 3.2 percent more than the CGs, while the scores of the 6 PGPs not earning performance payments in PY2 grew on average 2.8 percent more than the CGs. Without PGP 10, the only site with risk score growth lower than its comparison group, the remaining 5 PGPs not earning performance payments in PY2 had higher risk score growth, 4.1 percent, than the 4 PGPs earning performance payments in PY2, 3.2 percent growth.

Table 3-6 Percentage Point Change in Disease Component of Risk Score from BY to PY2

PGP Site (Names Blinded)	Risk Score Percentage Change		
	PGP	CG	Difference: PGP - CG
4 PGPs Earning Performance Payments in PY2			
PGP 1	9.5	9.1	0.4
PGP 2	7.7	5.0	2.7
PGP 4	11.0	4.3	6.7
PGP 6	2.7	-0.2	2.9
Average Percentage Change	**7.7**	**4.5**	**3.2**
6 PGPs Not Earning Performance Payments in PY2			
PGP 3	13.5	5.6	7.9
PGP 5	8.8	5.3	3.4
PGP 7	6.4	4.7	1.7
PGP 8	8.5	4.1	4.4
PGP 9	10.0	7.1	2.9
PGP 10	4.0	7.4	-3.4
Average Percentage Change	**8.5**	**5.7**	**2.8**
Average without PGP 10	**9.4**	**5.4**	**4.1**
Overall PGP Average	**8.2**	**5.2**	**3.0**

Source: RTI Analysis of Demonstration data

As shown in Table 3-6, the relative PGP risk score growth rate was not directly correlated with shared savings and performance payments across sites. PGP 3 had the largest percentage point change in risk scores, 7.9 percentage points, but did not earn a performance payment. PGP 1 had the lowest difference in risk score growth rate, 0.4 percent, but did earn a performance payment. PGP 10 was the only site to have a negative growth rate compared to its CG, and it incurred the only negative savings in PY2.

Risk score growth can be affected by a number of factors. It is possible that PGP sites are attracting beneficiaries with particular diseases, beneficiaries who would benefit from care

and disease management programs. Alternatively, if the PGPs are more carefully and intensively coding than the CGs even though patients may be similar in their underlying morbidity, performance payments may be based less on an underlying difference in population risk and more on improved patient tracking and increased delivery of services for those with chronic conditions. Distinguishing between case mix change and change related to coding initiatives is very difficult and challenging to quantify when they occur simultaneously.

Several analyses were conducted[4] to assess the influence of coding related to Demonstration incentives; however, the results were inconclusive for the period under study. There remains the possibility that a large component of the coding growth is related to changes in patient mix even though it is known that coding was an important focus for most of the participating PGPs. The coding growth in the comparison groups shows a shift in coding patterns that is occurring generally even without the incentives of the Demonstration.

A question is whether the PGP trend in coding growth will continue over time. This can only be studied and answered over a longer demonstration time frame. Coding incentives can produce growth of risk scores for a few years but opportunity for new growth over the prior period will likely diminish as coding practices reach the point that they capture nearly all disease each year. The effect on adjusted expenditures of any incentive-driven coding may have been present, but not in a consistent way and may run its course.

[4] These include changes in proportion of patients with a diagnosis in the model, change in proportion of patients coded with particular diseases and changes in coding growth before vs. during the demonstration.

Pre-existing Expenditure Trends

An important evaluation question is whether the savings of the 4 PGPs earning performance payments in PY2 were prompted by the incentives of the Demonstration. To answer this question, the results that would have occurred in the absence of the Demonstration were simulated using data from prior to the start of the Demonstration. The simulation method essentially mimicked the financial reconciliation calculation used to determine savings and performance payments within the Demonstration, except that the base year was 2002, and not 2004. Target Expenditures are the per person year expenditures rolled forward from 2002 using the comparison group growth factors, adjusted for the change in risk scores. These beneficiary level results were then aggregated to the PGP level and compared to the actual results in PY2 to identify expenditure differences that occurred after the Demonstration began in Table 3-7. The results of these analyses are described below:

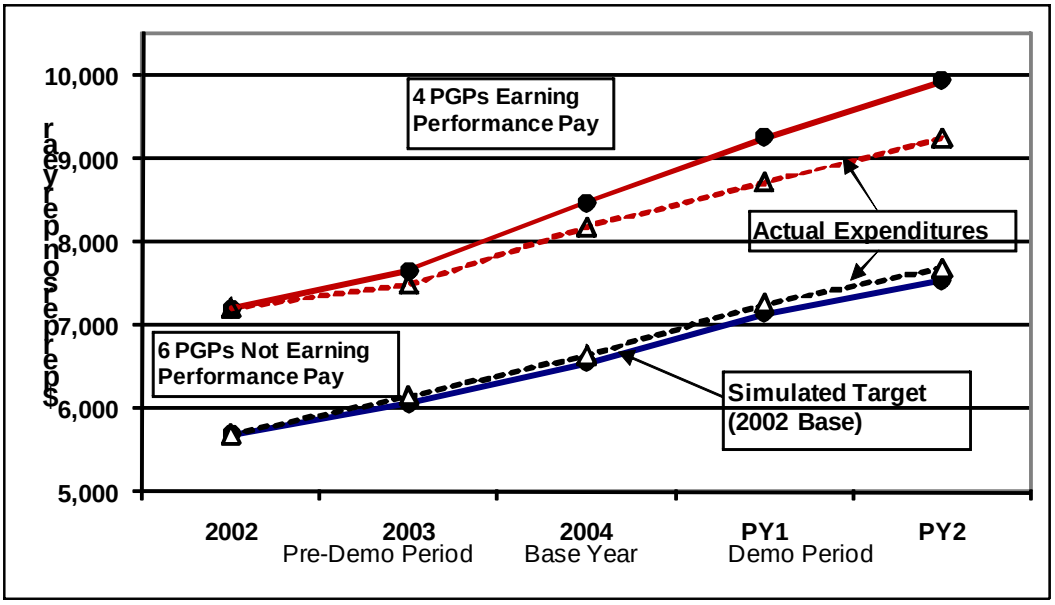
Simulated Pre-Demonstration Trends:

4 PGPs earning performance payments in PY2

The red lines (top set of lines) in Figure 3-1 compare the actual Assigned Beneficiary expenditures to simulated Target Expenditures per person year from 2002 to PY2. Note that expenditures are measured relative to the first year of the pre-Demonstration period, 2002, rather than relative to the Demonstration base year of 2004. Across the 4 sites, the Actual Expenditures of the Assigned Beneficiaries were trending below Target Expenditures prior to the

Demonstration. By 2004, the Demonstration base year, Actual Expenditures of the Assigned Beneficiaries were $8,180 or 3.4 percent lower than the simulated Target Expenditures of $8,471, a difference of $291 per person per year.

Figure 3-1 Expenditure Trends



NOTE: Unweighted average of PGP expenditures for each group.
SOURCE: RTI analysis of Medicare claims and enrollment data 2002-PY2.

6 PGPs not earning performance payments in PY2

The black lines (bottom set of lines) in Figure 3-1 show the cost trends for this group. Compared to the simulated 2002 base year, the Actual Expenditures of the Assigned Beneficiaries after two years (to 2004) were $6,636 or 1.4 percent higher than the simulated Target Expenditures of $6,546, a difference of $90 per person per year.

Comparison of Simulated Cost Trends to Actual PY 2 Performance:

Table 3-7 compares the simulated expenditures for the last pre-Demonstration year, 2004, and Demonstration expenditures for PY2. These periods are analogous because results are measured two years after the base year in both cases. The table is stratified by PGPs earning and not earning performance payments in PY2. The statistical significance of the savings differences were estimated by "bootstrapping" standard errors.[5] In PY2, the PGP Actual Expenditures aggregated across all 10 sites were $120 or 1.2 percent lower than Target Expenditures and were statistically significant ($p < .01$). In the simulated pre-Demonstration period, the PGP expenditures for the 10 sites were $62 or 0.6 percent lower than Target Expenditures, but not statistically significant. The results of this analyses follow:

4 PGPs earning performance payments in PY2

These sites exhibited favorable cost trends prior to the Demonstration. In the Demonstration, Actual Expenditures were lower than Target Expenditure by $334 per person year (3.5 percent lower) than Target Expenditures in PY2, which is statistically significant ($p < .01$). But this performance was almost matched in the pre-Demonstration period, where Actual Expenditures were lower than simulated Target Expenditures by $291 per person year and was statistically significant ($p <$

[5] The difference of Assigned Beneficiary from Target Expenditures were repeatedly estimated on samples drawn (with replacement) from the overall sample of assigned beneficiaries. The variation in these estimates is used to calculate a standard error and statistical significance representing the uncertainty in the estimate.

.01). The difference between expenditures in the pre-Demonstration and Demonstration periods was only $43 per person year, and was not statistically significant.

Table 3-7
Target Minus Actual Expenditures in Demonstration, and Pre-Demonstration, PY 2

Expenditures	All 10 PGPs	4 PGPs Earning Performance Payments in PY2	6 PGPs Not Earning Performance Payments in PY2
Target – Actual Expenditures per person year			
Demonstration PY2	120***	334 ***	−23
Simulated pre-demonstration 2004 year	62	291 ***	−91 *
Demonstration minus pre-demonstration	58	43	68
Target – Actual Expenditures as a Percentage of Target Expenditures			
Demonstration PY2	1.2**	3.5 ***	−0.4
Simulated pre-demonstration 2004 year	0.6	3.9 ***	−1.6 **
Demonstration minus pre-demonstration	0.6	−0.4	1.3

NOTES:

1. Target minus Assigned Beneficiary expenditures are estimated using the demonstration financial reconciliation algorithm except that the 2 percent threshold is not applied to the difference between target and Assigned Beneficiary expenditures in calculating physician group practice performance payments.
2. The pre-Demonstration Period is 2004 with a base year of 2002.
3. Demonstration Performance Year 2 (PY2) is April 2006 to March 2007, with a base year of 2004.

* = statistically significant at the 10% level.
** = statistically significant at the 5% level.
*** = statistically significant at the 1% level.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2002 to Performance Year 2.

6 PGPs not earning performance payments in PY2

Actual Expenditures were not statistically different from Target Expenditures in PY2. Simulated pre-Demonstration Assigned Beneficiary expenditure growth was greater than Target Expenditure growth on average for these PGPs. Relative to this unfavorable pre-Demonstration trend, the 6 PGPs not earning performance payments in PY2 showed some improvement in the Demonstration period, but the improvement was not statistically significant from 0 (their Demonstration minus pre-Demonstration results are positive but not statistically significant), and it was not sufficient for them to earn PY2 performance payments.

Pre-Demonstration Expenditure Patterns:

The analysis above suggests that the expenditure savings found in the Demonstration were no different than those realized before the Demonstration. The expenditure patterns in the pre-Demonstration period may be relevant in understanding why some PGPs earned performance payments. Using the analytical method described above, the relative expenditures and targets for the 3 years prior to the Demonstration (2002 through 2004) were simulated for all PGPs.

Expenditure growth for the 4 PGPs that earned performance payments in PY2 were found to be 2.6 percent lower on average than their local markets in 2002 and 2003. Such cost efficiencies could not have resulted from the Demonstration, since the

Demonstration had not started and was in a development stage at that time. They could have resulted from other attributes, such as their practice style or their infrastructure.

Just prior to the start of the Demonstration in the 2004 base year, expenditures for the 4 PGPs that earned performance payments in PY2 decreased by 5.5 percent on average compared to their local market area growth rate. The lower expenditure growth rates observed in this pre-Demonstration year, -5.5 percent, have yet to be exceeded within the Demonstration, which were -3.5 percent in PY2. In addition to normal year to year variation, several reasons may explain why the pre-Demonstration expenditure growth rate was greater than that found in the Demonstration. One explanation is that they resulted from changes made in anticipation of the Demonstration's payment incentives. Another is that the changes would have been implemented regardless of plans to participate in the Demonstration. This explanation is supported by the fact that in the 2005-2006 site interviews, some sites indicated that they were in the process of reducing costs in response to local market demands, or as part of a strategy of providing value to purchasers.

One of the 6 PGPs that did not earn a performance payment had lower expenditure growth (-4.4 percent lower) in the base year and lower expenditure growth prior to 2003, but higher expenditure growth in the Demonstration. This was also the only PGP that experienced a negative growth rate in the risk scores in PY2 (see Table 3-6). Expenditure growth for 5 of the 6 PGPs that did not earn performance payments were 1.1 percent higher than their local market prior to 2003 and 3.3 percent higher in the year

prior to the Demonstration. Expenditure growth began to decrease only after the Demonstration started, although not enough to result in any savings.

Conclusion:

As designed, the PGP Demonstration rewarded sites for expenditure trends that were favorable compared to trends of their local markets. However, the 4 PGPs earning performance payments in PY2 exhibited favorable cost trends prior to the Demonstration – trends that might have continued had the Demonstration not occurred. Apparently, these PGPs were more successful in controlling their expenditure growth than other providers in their local market area, and this appeared to help them in achieving the shared savings objectives under the Demonstration. One interpretation of these trends is that these sites had a cost-saving infrastructure in place prior to the Demonstration, which may be one of the reasons why they elected to participate in the Demonstration. The analyses could not determine the extent to which savings were influenced by pre-existing expenditure trends or resulted from a response to the financial incentives of the Demonstration. On average, the 6 PGPs not earning performance payments in PY2 were trending above their local market expenditures prior to the Demonstration. In general, their performance improved in the Demonstration period, but not sufficiently to share in savings under the demonstrations performance payment methodology.

What drives performance?

A major evaluation issue is: what has driven the performance results? Why did 4 PGPs earn a performance payment, and 6 did not? Because the Demonstration was not structured to test specific interventions, and the beneficiaries are assigned retroactively, it is difficult to identify a specific protocol or action that explains performance. In the preceding sections, differences between the two PGP groups, those that earned performance payments and those that did not, were explored. Earning performance payments were most likely explained by having lower expenditures prior to the start of the Demonstration. The increase in risk scores over those of comparison populations appears to be a characteristic of the PGPs earning and not earning performance payments. While the higher growth rates in risk scores may not explain the financial performance differences between the two PGP groups, it obviously affects performance payments. Without higher risk scores, only one PGP would have earned a performance payment.

Other factors, including quality performance, conditions, care modalities, organizational structure, and the number of Assigned Beneficiaries potentially may explain the expenditure differences between the two PGP groups. These factors are discussed below in the context of how they may help explain performance. To help identify associations, Table 2-2 shows additional information for the PGP groups that earned (first two groups) and those that did not earn performance payments (last two groups).

Quality Performance: Higher quality reporting and improvement should be correlated with performance payments, as the Demonstration has incentives to promote both results. There were virtually no differences in the quality measure performance across all PGPs. In PY2, the

quality measure performance was 100 percent for half the practices (45 out of 45 quality points reached) and the lowest level was 95.6 percent (43 out of 45 quality points reached). Despite differences in organizational structures, the PGPs were able to attain similar levels of quality performance measures.

While an outstanding result, the success on quality measures appears to be independent of performance payments. The independence of the quality performance may be due to the timing differences associated with each incentive. The actions needed to achieve quality performance are likely more concrete and easier to integrate within a system than strategies that deal with financial performance. For example, treatment protocols may be changed to adhere to the quality metrics in a onetime change, which may continue from year to year, and affect every patient. Adherence to the financial performance incentive may require more elaborate systems, and more time to generate results.

Conditions: The Subpopulation Analysis by Condition showed that the 4 PGPs earning performance payments in PY2 had statistically significant lower costs for many conditions: diabetes mellitus, CAD, COPD, stroke, and heart arrhythmia. Some of the conditions exhibited very large cost differences between the two groups, such as COPD which was about $1,000 less costly per beneficiary with COPD in the 4 PGPs earning performance payments than in the 6 that did not. The costs for other conditions, such as cancer, vascular disease, and CHF were not significantly different between the two groups. For beneficiaries entitled to Medicare due to disability and beneficiaries with risk scores in the upper 25 percent, the 4 PGPs earning performance payments had lower costs. These results suggest that the 4 PGPs earning

performance payments had lower cost treatment practices for some conditions, holding other factors constant.

Care Modality: Disease management and care coordination programs have been documented at each site. Since both PGP groups have incorporated similar programs and protocols, the implementation of these strategies do not differentiate financial performance. All PGP sites have stated that they have implemented disease and/or care management programs to reduce expense, while improving the quality of care. These clinically based care management programs can be characterized as being Disease Specific or related to General Care Coordination. Disease Specific Programs target a subset of beneficiaries based on diagnosis, while General Care Coordination Programs cast a wider net for enrollment criteria, since enrollment is not disease based. The 4 PGPs that earned performance payments in PY2 had lower inpatient and outpatient expenditures than the 6 PGPs not earning performance payments. While lower costs are consistent with the expectation about care management, sufficient data was not available to test this hypothesis using a rigorous analysis. Consequently, measuring the specific contribution of care management programs and redesigned care processes to cost savings, and evidence of their impact is largely anecdotal.

Organizational Structure: In PY 2, the 4 PGPs earning performance payments are characterized as being either affiliated with an academic medical center or a freestanding physician group practice. No performance payments were earned by the five PGPs belonging to an integrated delivery system (a system that includes a hospital) and the one physician network PGP that is sponsored by a hospital affiliate. The presence of a hospital was hypothesized as a potential issue for achieving savings under the Demonstration, since these systems may be

unable to reduce avoidable admissions or use lower cost care substitutes without affecting their inpatient revenue. Based on the PY2 result, there is some evidence to support this hypothesis, since the 6 PGPs not earning performance payments had owned hospitals in their delivery systems. However, both academic medical center PGPs earned performance payments and have integrated hospitals. It may also indicate that more than two years are needed for these organizations to organize and generate shareable savings. The capacity utilization of affiliated hospitals may be important, because it is easier to replace lower Medicare inpatient utilization with private-pay utilization if occupancy rates are high.

Number of beneficiaries: The number of Assigned Beneficiaries does not predict earning performance payments. As shown on Table 2-2, some PGPs with a large number of Assigned Beneficiaries did not earn performance payments, while some a smaller number did. For example, the University of Michigan had less than 20,000 Assigned Beneficiaries and earned a performance payment in PY2, while the St. John's Clinic had more than 30,000 Assigned Beneficiaries and did not earn a performance payment.

Conclusion: Two years into the demonstration, performance trends and organizational structure appear to be associated with financial performance. The number of Assigned Beneficiaries does not appear to be related to the ability of a PGP site to earn a performance payment. Lower costs for some conditions and care modalities (lower inpatient and outpatient expenditures) are associated with earning performance payments. While these differences are consistent with the hypotheses that the disease management and coordinated care programs were involved in achieving these results, their impact could not be rigorously determined and remains largely anecdotal. Since all PGPs had outstanding improvements in quality measure reporting

and processes, this factor does not differentiate financial performance among the 10 PGPs. This achievement undoubtedly helped to improve access, patient care, and quality for all beneficiaries involved in the PGP demonstration.

SECTION 4: POTENTIAL REFINEMENTS OF THE MODEL

The innovation of the PGP Demonstration model is that provider groups are given a financial incentive to provide more efficient, higher quality care. Performance payments are computed with the standard Medicare FFS claims processing system, and requires no additional data submission on the part of participating practices other than the sample of chart-based quality measures. While financial risk is mitigated by the continuance of FFS payments, providers are at risk for infrastructure improvements.

The basic goals stated in BIPA have been tested during the first two years of the PGP Demonstration. Over the period, the PGP sites that earned performance payments reduced expenditures by $26.9 million. Offsetting these savings were performance payments of $21,163,000, which were made to four PGPs, and losses of $3,484,000 at two PGPs. After these offsets, the net savings to the Medicare Trust Funds were about $2.3 million in the first two years of the Demonstration. All PGPs improved their quality, as measured by the performance indicators. The analyses could not determine the extent to which savings of the 4 PGPs that earned performance payments in PY2 were influenced by pre-existing expenditure trends or resulted from a response to the financial incentives of the Demonstration. Two performance years may be too short a time frame in which to observe the full impacts of the Demonstration. Additional study is needed to more adequately observe the implementation and refinement of interventions, and for their full impacts to be realized. Given these initial outcomes, the

Demonstration design elements referenced in Section 2 are revisited below to address future refinements of the PGP savings model.

PGP Participants

Small practices: The model has been tested in large physician group practices, integrated delivery systems, and one physician network. However, most physicians are organized as solo practitioners or members of small group practices. While the financial model is transparent to participating providers and would impose no special burden on small practices, the provision of quality data that require medical chart abstraction may impose high costs and burdens on small practices. Medicare's "Care Management Performance" Demonstration is testing the feasibility of obtaining reporting of PGP-Demonstration-like quality indicators from smaller practices, and the recent Electronic Health Record Demonstration is using a similar quality measure reporting methodology. With lower volumes, savings estimates for small practices are more likely to be subject to large variations. In order to have sufficiently sized populations on which to measure financial performance, smaller practices would need to come together, possibly via a physician network or other organization. The wide applicability of the model to smaller practices may depend on the formation of "network model" organizations that aggregate the experience of many small physician practices.

Large practices: The Demonstration model could be offered to other large physician organizations, such as those having a minimum of 150 or more practitioners participating in Medicare. Also, physicians groups would be required to have a strong primary care and patient care management focus. They could be free-standing physician groups or groups that are part of

integrated delivery systems including hospitals and other institutional providers. If the model were offered on a voluntary basis, self-selected participation by physician groups that expected to do well financially under the Demonstration model should be anticipated.

Patient Attribution

An attribution model is used retrospectively to assign beneficiaries to the Demonstration sites. Without knowing these beneficiaries in advance, an organization may not be able to exert enough control over beneficiaries who retain freedom of provider choice and have no incentives to choose high-quality or efficient providers or to restrain their use of services. However, this is countered by the frequency – an average of 5.5 visits – with which the groups see assigned patients throughout the year. Site managers and coordinators were interviewed regarding their views on the appropriateness of the beneficiary assignment methodology during a series of site visits to all 10 PGPs. In general, most PGPs found the assignment methodology to be a reasonable approach that resulted in a set of beneficiaries for whom they could be held accountable for cost and quality performance.

The two PGPs that are affiliated with academic medical centers offered a different perspective on the assignment algorithm. They have found that E&M services provided by specialists and surgeons accounted for a significant number of their Assigned Beneficiaries, due to the high proportion of referral services that they provide. As a result, they did not believe they had overall control of the care for a number of their Assigned Beneficiaries using the present attribution model. For future consideration, they offered an alternative attribution model that

uses E&M services provided by only primary care physicians.[6] Preliminary empirical analysis using this alternative assignment method suggests that it would not materially have had an impact on the number of beneficiaries assigned to the PGPs.

Comparison Group

The expenditure growth rate and the risk scores of the comparison group are particularly important, as they are used in establishing the benchmark for determining PGP savings. Appropriately defining a geographically defined comparison group for measuring cost and quality performance remains technically challenging, data intensive, and administratively burdensome. Lacking randomized patient assignment, the comparability of the PGP and CG populations is not certain, especially when a participating provider has a large market share or is unusual in the context of its local market area (e.g., an academic medical center in a rural area). If participation in a Demonstration model became widespread, the identification of non-participating comparison groups might be problematic.

Simplifying the process of calculating the benchmark used for measuring savings would be an important refinement. The choice of a benchmark requires a trade-off between two competing goals. First, the less comparable a population, the less valid is the benchmark. Second, the more rigorously-defined the comparison group, the greater are the information needs and the amount of processing time and effort to compute the expenditure benchmark. Comparison groups may include all FFS beneficiaries in a local market area, such as an MSA,

[6] Note that the PGP Demonstration model does not specify physician type in the plurality algorithm.

FFS beneficiaries residing in the PGP's state, all FFS beneficiaries nationwide, or a combination of these three. The growth factor could also take into account the current spending level of the group with less generous factors used in areas with high spending. PGPs participating in a demonstration could also be comparatively benchmarked against each other. Complex definitions of comparison groups may increase administrative costs and the length of the time lag between PGP performance and incentive payments for that performance. Further analyses are needed to identify the advantages and disadvantages of these alternatively defined benchmarks.

Performance Payments and Savings

Rebasing

As in any model that uses a provider specific baseline, there is a potential for inefficient providers to unjustly receive performance payments. Historically inefficient providers can receive a performance payment for reducing expenditures to a more efficient level, while historically efficient providers may have little opportunity for further cost reductions. A provider-specific baseline encourages voluntary participation, reduces risk to providers, and focuses incentives where the greatest potential for improving cost control exists. However, it may become easier to achieve performance targets over time. Consequently, schedules for rebasing the financial reconciliation algorithm and the quality performance targets would need to be established for any longer term operation of the model. Rebasing will create incentives for further cost reductions and quality enhancements, and allow Medicare to capture more of the already achieved efficiency improvements.

Change in Medicare Payment Formulas or Policies

A design feature of the PGP savings methodology is that changes to Medicare payment formulas or policy can be ignored because of the geographic matching of the PGP and their comparison groups. This assumption is reasonable as long as the PGP and comparison groups are similarly affected by such changes. In support of this assumption, it was reasoned that "for any differential effect to arise, Assigned Beneficiaries would have to receive a very different set of services than comparison beneficiaries on average, which would be highly unusual for two large groups (15,000 or more) of beneficiaries in the same market area" (Pope, 2002). While Direct Medical Education (DME) payments were removed from Demonstration savings calculations since they are not paid through the claims system, adjustments for other changes in Medicare payments were not made.

Perhaps the most likely source of a differential payment effect is the Indirect Medical Education (IME) and Disproportionate Share (DSH) payments to hospitals made under the Inpatient Prospective Payment System (IPPS). Inpatient expenditures typically represent 40 percent of the services provided to Medicare beneficiaries, and IME and DSH payments are a significant proportion of IPPS payments, which are concentrated among certain hospitals, primarily large teaching hospitals. In contrast, IPPS payment adjustments for differences in geographic wage costs are unlikely to result in a differential payment effect since PGPs and their comparison groups are in the same geographic areas.

Differential changes in payments attributable to IME and DSH could occur in a number of ways. For example, a change in IPPS payment policy could either increase or decrease IME and/or DSH payments for all hospitals qualifying for these payments. Also, individual hospitals could experience a change due to changes in the hospital-specific statistics used to calculate IME and DSH payments. Finally, differential changes in the distribution of PGP and comparison group patients between hospitals receiving IME and/or DSH versus other hospitals would result in differential changes in payments and would affect measured Demonstration performance.

By including IME and DSH payments in the performance calculation, the Demonstration may give PGPs an incentive to admit patients to hospitals that receive lower IME and/or DSH payments. This incentive could disadvantage teaching and DSH hospitals. Excluding IME and DSH from the payment calculation removes payment differences between teaching and non-teaching hospitals, so that payment differences measured under the Demonstration only reflect utilization differences. Alternatively, including IME and DSH in the payment calculations could allow some groups, particularly those associated with academic medical centers, to earn a portion of the foregone IME and DSH from reduced admissions at their core teaching hospital.

To test the assumption that payment system changes do not disproportionately affect the Demonstration, savings and the resulting performance payments were re-estimated by removing IME and DSH payments from expenditures. These re-estimations employed the same algorithm used in the Demonstration's financial reconciliation, but IME and DSH were excluded from the calculations, for PGPs and CGs, and in all years (BY, PY1, PY2). The re-estimation did not attempt to determine changes in the distribution of hospitals (teaching vs. non-teaching) utilized

by assigned and comparison group beneficiaries during the base and performance years which may also influence results.

The re-estimation showed total earned performance payments across all PGPs would have been reduced by $1.3 million in PY1, while increasing by $1.5 million in PY2. Earned performance payments would have been redistributed among PGPs with some potentially earning more and some potentially earning less. At the individual PGP site level, the difference between Target Expenditures and Actual Expenditures changed, ranging from a 1.0 percentage point reduction to a 0.8 percentage point increase. To earn a performance payment, a PGP must reduce its Medicare expenditure growth rate by more than 2 percentage points relative to the local market area comparison group growth rate. Consequently, relative to the 2 percent threshold, these small changes between the Target Expenditures and Actual Expenditures can have a significant impact on the likelihood of a site earning a performance payment under the Demonstration. Thus, the assumption that geographic matching obviates the need to adjust for Medicare payment system changes may need to be reassessed in future Demonstration designs depending upon the desired incentives.

Quality Measures

The majority of the Demonstration quality indicators require medical record abstraction. While the costs of extracting medical records for these quality indicators have been lowered by using sampling techniques, the process is still costly. It is anticipated that in the future, the clinical information necessary to measure performance will be more easily accessed and submitted via electronic reporting methods. Developments in quality measurement should be

incorporated into the Demonstration as appropriate. The Demonstration quality indicators can be refined by adding, deleting, and grouping process measures over time as appropriate. Additional outcome measures could be included, and changes in scoring algorithms could be implemented to attach different weights to achieved quality levels versus changes in levels.

REFERENCES

GAO, February 2008. "Care Coordination Programs Used in Demonstration Show Promise, but Wider Use of Payment Approach May Be Limited." Report Number GAO-08-65.

John Kautter, Ph.D., Gregory C. Pope, M.S., Michael Trisolini, Ph.D., M.B.A., and Sherry Grund, R.N, "Medicare Physician Group Practice Demonstration Design: Quality and Efficiency Pay-for-Performance," Health Care Financing Review/Fall 2007/Volume 29, Number 1 PP. 15 -29.

Gregory C. Pope, M.S., Michael Trisolini, Ph.D., M.B.A., John Kautter, Ph.D., and Walter Adamache, Ph.D., "Physician Group Practice (PGP) Demonstration Design Report, 2002.

Report to Congress: "PHYSICIAN GROUP PRACTICE DEMONSTRATION FIRST EVALUATION REPORT," Michael O. Leavitt, Secretary of Health and Human Services 2006. Available at: http://www.cms.hhs.gov/reports/downloads/Leavitt1.pdf

Trisolini M, Aggarwal J, Leung M, Pope G, and J Kautter: *The Medicare Physician Group Practice Demonstration: Lessons Learned on Improving Quality and Efficiency.* New York: The Commonwealth Fund, 2008. Available at: http://www.cms.hhs.gov/DemoProjectsEvalRpts/downloads/PGP_Conference_Report.pdf

Michael Trisolini, Ph.D., M.B.A., Gregory Pope, M.S., John Kautter, Ph.D. et al.: *Medicare Physician Group Practices: Innovations in Quality and Efficiency.* New York: The Commonwealth Fund, 2006. Available at: http://www.commonwealthfund.org/Content/Publications/Fund-Reports/2006/Dec/Medicare-Physician-Group-Practices--Innovations-in-Quality-and-Efficiency.aspx

APPENDIX

Table A-1 Summary of Programs Implemented at the PGP Groups

Program category/name	Brief description
Disease Specific Programs	
Congestive Heart Failure (CHF) Care Management (no tele-management)	Program that assists patients diagnosed with CHF to ensure they receive appropriate care for their condition and education for self management techniques.
Tele-Management for Heart Failure or Other Conditions	An interactive voice response system that assists with care management of patients with congestive heart failure.
Diabetes Care Management	Program that assists patients living with diabetes to ensure proper self-care techniques and appropriate physician follow-up.
Anticoagulation Program/Clinic	Program that works with patients receiving anticoagulation therapy to ensure patient is receiving appropriate medications and to avoid any potential hospitalization.
Chronic Obstructive Pulmonary Disease (COPD) Management	COPD patients are provided with education and tools to assist with preventing functional decline. Preventive services are also provided to avoid future acute episodes.
Psychiatric Conditions	Programs that assist high risk patients with chronic psychiatric conditions.
Coronary Artery Disease (CAD) Management	Program that alerts physicians of required activities or services for CAD patients.
Cancer Care Management	Cancer patients are provided with coordinated cancer care.
Hypertension Disease Management Program	Reminder system for physicians to ensure that blood pressure is measured at each visit for hypertension patients.
General Care Management/Care Coordination Programs	
Moderate/High-Risk Case Management	Care managers work with all patients to reduce any risk factors that can be associated with increased risk of hospitalization.
Homecare/Post Acute Care Services	Care to patients who have been discharged and are receiving some form of post acute-care services.

(continued)

Table A-1 Summary of Programs Implemented at the PGP Groups (continued)

Program category/name	Brief description
Health Coaching	Care managers assist patients post-discharge with their care management activities.
Gold Star Population	Group identifies patients for further management based on diagnoses/co-morbidities, costs and hospitalization.
Complex Care Coordination	Care managers monitor patients with multiple chronic conditions and provide general care as required. Care managers also educate patients on self-management techniques.
Transitional Care Program	Programs that assist patients at hospital discharge to improve care transitions and avoid recurrent hospitalizations.
Palliative Care Program	End of life care planning for terminally ill patients.
Patient Registries	Development of patient registries as an administrative tool to improve care management processes.
Re-assignment of Non-Physician Staff	Examples include use of physician assistants for medication reconciliation and chart review at start of visit or moving physician assistants to nursing homes for improved patient care in nursing home.
Medication Reconciliation Program	A system either electronic or not that assists with tracking patient medications to avoid adverse events related to prescription medications.
Medication Access Program	Assists low income patients with obtaining low cost prescription medications from pharmaceutical companies.

SOURCE: Summary of data collected by CMS/RTI from groups participating in the PGP Demonstration and GAO (2008) Report: "Care Coordination Programs Used in Demonstration Show Promise, But Wider Use of Payment Approach May be Limited."

Table A-2 Quality Measures, Weights, and Quality Points by Module

Diabetes mellitus		Congestive heart failure		Coronary artery disease		Hypertension / preventive care	
Measure	**Weight**	**Measure**	**Weight**	**Measure**	**Weight**	**Measure**	**Weight**
DM-1 HbA1c management	4	HF-1 left ventricular function assessment	1	CAD-1 antiplatelet therapy	1	HTN-1 blood pressure screening	1
DM-2 HbA1c control	1	HF-2 left ventricular ejection fraction testing	4	CAD-2 drug therapy for lowering LDL cholesterol	1	HTN-2 blood pressure control	1
DM-3 blood pressure management	1	HF-3 weight measurement	1	CAD-3 beta-blocker therapy—prior MI	1	HTN-3 plan of care	1
DM-4 lipid measurement	4	HF-4 blood pressure screening	1	CAD-4 blood pressure	1	PC-5 breast cancer screening	4
DM-5 LDL cholesterol level	1	HF-5 patient education	1	CAD-5 lipid profile	4	PC-6 colorectal cancer screening	1
DM-6 urine protein testing	4	HF-6 beta-blocker therapy	1	CAD-6 LDL cholesterol level	1		
DM-7 eye exam	4	HF-7 ace inhibitor therapy	1	CAD-7 ace inhibitor therapy	1		
DM-8 foot exam	1	HF-8 Warfarin therapy for patients HF	1				
DM-9 influenza vaccination	1	HF-9 influenza vaccination	1				
DM-10 pneumonia vaccination	1	HF-10 pneumonia vaccination	1				
Total points	22		13		10		8

SOURCE**:** RTI International

Table A-3 Excerpts of Conference Proceedings

***Medicare Physician Group Practices: Innovations in Quality and Efficiency*.**
The Commonwealth Fund, 2006

Dartmouth-Hitchcock Clinic

Under the PGP Demonstration, the Dartmouth-Hitchcock Clinic (DHC) has implemented one set of interventions aimed at reducing costs, and another set aimed at improving quality. Cost reduction interventions include analysis of risk scores, predictive modeling, and strategies to reduce readmissions. Analysis of diagnostic risk categories showed that 23 percent of DHC's assigned beneficiaries represent 73 percent of Medicare payments for assigned beneficiaries overall. As a result, one of their goals is to find ways to target interventions to that high cost group.

Analysis of the cost effect of patients with readmissions showed that the 5,928 assigned beneficiaries with readmissions had annual Medicare costs of $30,052, while the 22,176 assigned beneficiaries without readmissions had annual costs of only $2,629. Closer study of those with readmissions indicated that they were more frequently dual eligibles who had both psychiatric and medical conditions. As a result, another goal is to tailor interventions to their unique set of issues.

DHC's quality improvement strategies include development of disease registries, cohort reports, and a health coaching program. A health coach is a specially trained professional (such as a nurse or dietitian) who instructs or directs patients in aspects of personal health care. Health coaches are charged with providing evidence-based health information to patients by telephone, during office visits, through educational materials, and through group classes. They are integrated into primary care practices within DHC divisions and target their interventions on chronic disease, high risk patients (e.g., diabetes, HF, CAD). Hospital-based sites target interventions on post-discharge follow-up and readmissions. The health coach model was developed in collaboration with Health Dialog who trained DHC clinicians in health coaching techniques.

DHC's disease registries include lists of patients who have a specific clinical condition, such as diabetes. They are used to proactively manage patients, order pre-work such as lab tests, and identify gaps in care. Each patient is tracked on multiple measures related to care for that clinical condition, including the disease-specific quality measures applied under the PGP Demonstration.

University of Michigan

The University of Michigan Health System (UMHS) identified the primary goal of its care coordination interventions to be quality improvement. UHMS is working to improve communication among providers and to improve patient compliance, self-management, and access to necessary services. Care coordination is viewed as having potential to affect a broad range of quality of care issues, including overuse, misuse and underuse of health care services.

UMHS has focused on two service delivery interventions to improve coordination of care: 1) transitional care; and 2) the "medical home." Transitional care is aimed at reducing readmissions, while the medical home is aimed at reducing avoidable first admissions, such as for ambulatory care sensitive conditions. UMHS has implemented a range of transition care interventions, and is working on broader implementation of the medical home approach.

The strategies for transition management are to assist with timely appointment scheduling, improve the availability of patient contact information, provide appropriate patient discharge counseling, reduce social barriers to care (e.g., transportation to appointments, affordability of medications), and provide home care. Transitional care interventions include post-discharge calls to follow up with patients within 24 hours of hospital discharge. The majority of these calls have been to medical patients, although some have been made to surgical patients. UMHS has also piloted a pharmacy discharge program, to ensure that patients discharged with medication changes understand the changes and receive the correct medications. They have found this to be a major issue for many patients, who are often discharged with five or more complex medications that may need ongoing monitoring. Moreover, increases in the number of medications prescribed for chronic disease patients in recent years, and the increased complexity of medication regimens, have made it more difficult for social workers or nurses to assist patients with managing those issues.

Medical home interventions have focused on patients needing complex care coordination services (see *Figure 3-4)*. These interventions are provided either at individual clinic sites or another central location for particular clinical groups. The patients targeted for medical home services include the vulnerable elderly and dual eligibles with mental health and social problems. A similar program is being considered for end-stage renal disease patients.

Marshfield Clinic

An Intervention List has been implemented under the PGP demonstration to stratify patients by risk level. It focuses attention on patients ranked at the top of an electronic list prepared for physicians, with rankings based on patients with multiple conditions and those needing interventions to satisfy quality measures. It also enables medical assistants to review high risk patients and, based on written protocols, order routine tests needed for some interventions without the need for physician involvement.

The Electronic Medical Record (EMR) at Marshfield Clinic has been operational since 1985, and has become increasingly sophisticated over the years. It is accessible at all Marshfield Clinic sites. The EMR enables physicians to generate graphs and other reports presenting a specific patient's health care improvement (or decline) over time. It includes a Dashboard that presents a patient's active medications, problem list, laboratory test results, medications, previous appointment dates, vital statistics, immunizations, and other data. The EMR also includes a Medications Manager, Document Manager, and a physician reminder system.

In addition to improving patient care, Marshfield Clinic has utilized informatics to provide feedback on quality metrics to individual providers. The EMR facilitates the collection of quality

data, and allows for timely distribution of feedback to physicians. Individual physicians can examine their quality performance and compare it to their department overall.

The Everett Clinic

Everett Clinic has been promoting palliative care through the presence of hospice nurses in primary care offices. They have also been providing intensive case management and end of life planning education. Everett Clinic has funded palliative care programs and educational information through partnership with a hospital-based hospice program. The palliative care promotion program is currently available at two Everett Clinic sites and is expected to expand to all four satellite sites in 2006.

Evidence has shown that proper use of palliative care programs can reduce hospital admissions. Everett Clinic staff studied 140 patients over age 65 who passed away between August 2004 and January 2006. They found that patients who had received palliative care were more likely to have zero hospital admissions prior to death (53 percent versus 28 percent). In addition, the total number of admissions per patient was lower for patients receiving palliative care (1.9) compared to those not receiving palliative care (2.4). Palliative care programs also increased use of hospice services. Everett Clinic found that the median hospice length of stay was 47 days for those receiving palliative care versus just 6 days for others.

Billings Clinic

Billings Clinic implemented an integrated EMR in July 2004. Its EMR provides a common data repository for information from laboratories, pharmacies, and radiologists, as well as provider documentation. It also allows for online medication ordering and prescribing, with full implementation of this module expected in 2006.

The EMR has been Billings Clinic's main vehicle for quality of care and process improvement under the PGP Demonstration. It has supported chronic disease management programs by facilitating the identification of eligible patients through registries, enabling development of disease management modules, generating quality and care performance reports for organizations and providers, and generating score cards for individual patients that highlight specific patient needs (e.g., laboratory tests). The EMR, through health maintenance modules, alerts providers regarding gaps in preventive services such as tests and screenings or immunizations.

The EMR has also improved patient safety through medication reconciliation applied during transitions in care. Medication reconciliation is made possible through online prescribing, use of the electronic medication record, and development of patient-friendly medication lists.

The EMR has been applied to generate both cost savings and quality improvement for diabetes and HF patients since they are a focus of the PGP Demonstration. For diabetes, the cost savings focus has been on preventing avoidable admissions, frequent readmissions, and readmission complications. Quality improvement efforts include a diabetes patient registry, a

disease management module, provider reports and benchmarking, and a patient score card focused on the PGP Demonstration quality measures. Quality improvement efforts under the demonstration for diabetes patients have resulted in improved quality measurement reporting and documentation at Billings Clinic. For example, foot exam documentation rates increased substantially from May 2005 (20 percent) to April 2006 (> 50 percent).

For HF, the cost savings goal is to reduce all cause admissions by 20 percent to 50 percent. Quality improvement is supported by HF clinic redesign, to increase the roles for nurse practitioners, an HF patient registry, disease management, enhanced provider education on new treatment guidelines, and improved patient education. Heart failure patients are monitored in between office visits by Billings Clinic nurses using an interactive voice response system that prompts patients to respond to questions about their weight, medications, and symptoms on a daily basis. Currently, over 700 patients are enrolled in this service, with a goal of 1,000 patients.

Future efforts are planned for other diseases and interventions targeted by the PGP Demonstration. Disease management modules will be expanded to include CAD and hypertension. Health maintenance modules will be developed for cancer screening, including mammography and colonoscopy.

Geisinger Health System

Geisinger Health System (GHS) is utilizing its electronic health record (EHR) as a key element in its response to the PGP Demonstration incentives. GHS has a longstanding commitment to health information technology. Its primary goals in developing an EHR were to develop an efficient, adaptable system that would reduce administrative burden and be scalable and exportable. It was also intended to be user-friendly for patients, so they would be able to access information from it regarding their health status and care.

The EHR now serves GHS providers, referring physicians, and patients. It collects data on over one million visits provided by GHS providers each year, and can accommodate more than 5,000 concurrent users. The EHR connects GHS to over 500 non-GHS physicians and 10,000 patient records. Patients may also access portions of the EHR, called "MyGeisinger," for viewing test and lab results, scheduling appointments, dialoguing with their physician, and renewing medications. The patient portion is expanding rapidly, currently adding over 2,000 new users per month.

Patient registries are used in conjunction with the EHR to provide analysis and intervention reminders to physicians on a number of topics. These "operational registries" include pre-defined lists of patients deficient in various aspects of standards-based care. They are used to initiate interventions such as letters, referrals, laboratory test orders, and secure e-mails to ensure that patients receive needed care.

The registries are focused on a range of PGP Demonstration and related diseases and interventions, including chronic disease return visits (for patients with HF, COPD, or diabetes), pneumococcal vaccination, and diabetes management. Figure 3-8 illustrates the standards set for diabetes care.

Geisinger has focused on setting high standards of care for treatment of diabetes, including LDL < 100 versus 130, blood pressure < 130/80 versus 140/90, and evaluation of smoking status versus smoking assessment or education. Quality of care analysis also includes evaluation of the number of patients who achieved standards for all of the measures.

The registries are updated automatically and reviewed on a monthly basis. The use of registries has increased the number of patients receiving clinical services. One particularly successful application has been informing patients of their need for pneumococcal vaccinations. The registries permit Geisinger Clinic to track and target patients who have not yet received the recommended vaccinations.

The Geisinger EHR also provides best practice alerts to providers at the point of care. They allow physicians to view a summary of the patient's care, receive reminders about tests and other interventions, and ensure that they have not missed anything regarding needed care.

Forsyth Medical Group

Forsyth Medical Group (FMG) introduced the Comprehensive Organized Medicine Provided Across a Seamless System (COMPASS) disease management program under the PGP Demonstration. The goals of COMPASS are to provide practice level tools to promote efforts to meet the PGP Demonstration quality measure targets, educational packets to address disease self-management with patients, disease-specific population-based interventions, and case management for high risk patients.

As part of COMPASS, FMG developed color-coded disease management worksheets. They are tools to remind physicians and other clinical staff about patients needing particular tests or interventions. They also serve to increase the available documentation for services provided to Medicare beneficiaries at FMG. Providers receive pocket cards explaining the PGP Demonstration quality measures and their components. The cards correlate with the disease management worksheets, so that providers can be reminded about why particular tests are required.

Park Nicollet Health Services

Park Nicollet Health Services (PNHS) implemented two major innovations under the PGP demonstration to improve care management for Medicare patients. They involve re-design of health care delivery for diabetes and heart failure patients. PNHS plans to implement similar health care delivery innovations for coronary artery disease, hypertension, and preventive care.

The diabetes program involves a disease registry, a nurse population manager, point of care testing, and a Certified Diabetes Educator (CDE). The disease registry identifies patients with diabetes and provides medical histories and laboratory test results. The nurse population manager is responsible for reviewing the registry to identify patients who may be overdue for tests and patients who have not yet met the standard of care. The population manager provides lists of targeted patients to receptionists who call and schedule necessary appointments. The population

manager works with physicians to plan next steps in treatment for those patients, and also works directly with patients to enhance their disease self-management skills.

For point of care testing, patients requiring laboratory testing are asked to arrive for their next appointment thirty minutes prior to the appointment time. They receive laboratory papers for required tests at check-in, have the tests administered on-site, and then the results can be made available to physicians prior to the patient's appointment through the EMR. This allows physicians to treat patients based on today's laboratory test results.

St. John's Health System

St. John's Health System (SJHS) developed a comprehensive patient registry to respond to the PGP Demonstration's quality improvement incentives. Its development was viewed as critical for the success of the PGP Demonstration at SJHS. It was developed internally and required significant senior staff commitment. An Advisory Board was formed to develop and implement the registry, including physicians, nurses, case management staff, office management staff, and IT staff. Development and implementation occurred over a period of about eight months.

A key element of the patient registry is the Visit Planner. It is designed to complement the established clinical work-flow process at SJHS. It provides a "to do" list for physicians prior to each patient visit, with reminders for needed tests or interventions. The Visit Planner consists of a one page summary for each patient showing key demographic and clinical data, including test dates and results. It highlights tests for which the patient is due, including those for the PGP Demonstration quality measures.

Physicians have responded positively to the Visit Planner, indicating that it assists them in preparing for patient encounters. For example, they do not need to look through the medical record to see if mammograms or colonoscopies have been done. As a result, physicians have assisted with the effort to keep the patient registry database up to date. Some enter data from their patient visits directly into the system, and others write notes that their staff enter. This has helped to keep the patient registry information current and complete for reports generated for management staff. SJHS views this as a lesson learned from the PGP Demonstration and the patient registry effort; when data systems are integrated into the physicians' workflow, and viewed as supportive by physicians, data are more easily obtained for management reports and monitoring efforts.

The patient registry also provides reports on areas where patient care can be improved. An Exception List includes patients that are due for tests or other interventions. This is viewed as a "clean-up" process, to identify gaps in care that were missed in the regular clinical work flow. Patients can then be called, or letters sent, indicating the need for a visit or test.

The patient registry generates quality measure and outcome summary reports at both the individual provider and clinic levels. They are unblinded to encourage competition among physicians for quality improvement.

Middlesex Health System

Middlesex Health System (MHS) aims to generate cost savings under the PGP demonstration through improved quality of care, enhanced patient safety, and appropriate coordination of care. MHS emphasizes two major strategies to achieve these goals: 1) participation in national quality and safety initiatives; and 2) transition management.

One of MHS's reasons for participating in the PGP demonstration was its interest in implementing a number of national hospital-based quality and safety initiatives. MHS is currently participating in several of these initiatives, including the Institute for Healthcare Improvement 100K Lives Campaign, the Surgical Care Improvement Project, the National Surgery Quality Improvement Program, the Leapfrog Group for Patient Safety, and the National Quality Form quality measures. MHS staff believe that by leveraging the knowledge and techniques provided as part of these initiatives they will be better able to respond to the PGP Demonstration incentives.