

September 2012

Evaluation of the Medicare Physician Group Practice Demonstration

Final Report

Prepared for

Fred Thomas, PhD
Heather Grimsley, MS
Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
Mail Stop C3-21-28
7500 Security Boulevard
Baltimore, MD 21244-1850

Prepared by

John Kautter, PhD
Gregory C. Pope, MS
Musetta Leung, PhD
Michael Trisolini, PhD
Walter Adamache, PhD
Kevin Smith, MS
Diana Trebino, BA
Jenya Kaganova, MS
Lindsey Patterson, BS
Olivia Berzin, BA
Margot Schwartz, MS
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0209853.007.005

Evaluation of the Medicare Physician Group Practice Demonstration

Final Report

Authors:

John Kautter, PhD
Gregory C. Pope, MS
Musetta Leung, PhD
Michael Trisolini, PhD
Walter Adamache, PhD
Kevin Smith, MS
Diana Trebino, BA
Jenya Kaganova, MS
Lindsey Patterson, BS
Olivia Berzin, BA
Margot Schwartz, MS

Federal Project Officers:

Fred Thomas
Heather Grimsley

RTI International

CMS Contract No. HHSM-500-2005-00029I

September 2012

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2005-00029I. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

CONTENTS

EXECUTIVE SUMMARY	ES-1
CHAPTER 1 INTRODUCTION	1
1.1 Overview of the Demonstration.....	1
1.2 Purpose of the Report and Overview	2
1.3 Sources of Information	2
CHAPTER 2 PGP DEMONSTRATION DESIGN	3
2.1 Demonstration Goals and Objectives	3
2.2 Key Design Elements.....	3
2.2.1 Assignment of Beneficiaries to PGPs	3
2.2.2 Demonstration Comparison Group	4
2.2.3 Measurement of Demonstration Savings	4
2.3 Performance Payments and Additional Design Features.....	5
2.4 Quality Measurement and Reporting.....	8
CHAPTER 3 PGP DEMONSTRATION PARTICIPANTS	11
3.1 Site Selection	11
3.2 PGP Participant Characteristics.....	12
3.2.1 Geographic Characteristics	12
3.2.2 Organizational Characteristics.....	12
CHAPTER 4 DEMONSTRATION INTERVENTIONS	17
4.1 Overview of Demonstration Interventions	17
4.1.1 Process Improvement Interventions.....	17
4.1.2 Patient Care Management Programs.....	21
4.2 Beneficiary Involvement in Site Programs	21
4.2.1 Disease Specific Care Management Programs	23
4.2.2 Complex Care Management Programs	26
4.2.3 Beneficiary Turnover	26
4.3 Factors Expected to Influence Financial Results.....	26
4.4 Factors Expected to Influence Quality Results.....	28
CHAPTER 5 DEMONSTRATION FINANCIAL RESULTS	31
5.1 Assigned Beneficiary and Comparison Group Profiles.....	31
5.2 Trends During the Demonstration Period in Expenditures, Risk Scores, Risk-Adjusted Expenditures, and Assigned Beneficiaries	36
5.2.1 Expenditures, Risk Scores, and Risk-Adjusted Expenditures	36
5.2.2 Assigned Beneficiary Actual Versus Target Expenditures.....	42
5.3 Financial Reconciliation by Performance Year	42

5.4	Summary of Demonstration Financial Results and Medicare Program Savings.....	58
5.4.1	Demonstration Savings and Performance Payments	59
5.4.2	Demonstration Savings and Performance Payments Per Person and as a Percentage of Target Expenditures	64
CHAPTER 6	QUALITY MEASUREMENT & REPORTING.....	69
6.1	Overview of PGP Demonstration Quality Measurement Methodology and Results.....	69
6.2	Diabetes (DM) Module Quality Performance	70
6.2.1	Description of DM Quality Measures.....	70
6.2.2	DM Results for BY and PY1 - PY5 by PGP	71
6.2.3	Description of Trends by PGP	71
6.2.4	Description of Trends by DM Measure	75
6.2.5	Number of Targets Hit per PGP for DM in PY1 through PY5.....	75
6.2.6	Number of PGPs Hitting a Target by Measure.....	76
6.3	Heart Failure (HF) Module Quality Performance.....	77
6.3.1	Description of HF Quality Measures	77
6.3.2	HF Results for BY and PY2 - PY5 by PGP.....	77
6.3.3	Description of Trends by PGP	81
6.3.4	Description of Trends by HF Measure.....	81
6.3.5	Number of Targets Hit per PGP for HF in PY2-PY5	81
6.3.6	Number of PGPs Hitting a Target by HF Measure in PY2 – PY5	81
6.4	Coronary Artery Disease (CAD) Module Quality Performance	83
6.4.1	Description of CAD Quality Measures.....	83
6.4.2	CAD Results for BY and PY2 – PY5 by PGP.....	83
6.4.3	Description of Trends by PGP	83
6.4.4	Description of Trends by CAD Measure	87
6.4.5	Number of Targets Hit per PGP for CAD in PY2-PY5.....	87
6.4.6	Number of PGPs Hitting a Target by CAD Measure in PY2-PY5.....	88
6.5	Hypertension (HTN) Module Quality Performance	88
6.5.1	Description of HTN Measures	88
6.5.2	HTN Results for BY and PY3 – PY5 by PGP.....	89
6.5.3	Description of Trends by PGP	89
6.5.4	Description of Trends by HTN Measure	89
6.5.5	Number of Targets Hit per PGP for HTN in PY3-PY5.....	89
6.5.6	Number of PGPs Hitting a Target by HTN Measure in PY3-PY5	92
6.6	Preventative Care (PC) Modules Quality Performance.....	93

6.6.1	Description of PC Quality Measures	93
6.6.2	PC Results for the BY and PY3 – PY5 by PGP	93
6.6.3	Description of Trends by PGP	93
6.6.4	Description of Trends by HTN and PC Measure.....	93
6.6.5	Number of Targets Hit per PGP for PC in PY3-PY5	93
6.6.6	Number of PGPs Hitting a Target by PC Measure in PY3-PY5	96
6.7	PQRI Incentive Payments in PY2-PY5	97
6.7.1	Description of Methodology Used for Calculating PY2 – PY5 PQRI Payments for PGP Sites	97
6.7.2	PQRI Incentive Payments Earned by PGP for PY2 – PY5	97
CHAPTER 7 DEMONSTRATION IMPACTS ON MEDICARE PROGRAM		
EXPENDITURES: SIMULATION ANALYSIS		99
7.1	Medicare Program Expenditures for Assigned Beneficiaries.....	100
7.1.1	Trends in Assigned Beneficiary Versus Target Expenditures in the Periods Prior to and During the Demonstration.....	100
7.1.2	Statistical Significance of Demonstration Expenditure Impacts	108
7.2	Medicare Program Expenditure Impacts for Subgroups of Assigned Beneficiaries	112
7.2.1	Methods and Presentation	113
7.2.2	Results.....	116
7.3	Components of Medicare Program Expenditures.....	117
CHAPTER 8 DEMONSTRATION IMPACTS ON QUALITY OF CARE		121
8.1	Trends in Assigned Beneficiary versus Comparison Group Quality of Care.....	122
8.1.1	Performance Results from Base Year to PY5.....	122
8.1.2	Change in Scores for all Seven Claims-based Quality Measures from Base Year to PY5.....	130
8.2	Statistical Estimates of Demonstration Quality of Care Impacts	131
8.2.1	Assessment of Quality Results from Base Year to PY5--Overall	131
8.2.2	Assessment of Quality Results from Base Year to PY5—By Site	132
8.2.3	Impact of Demonstration on Quality	141
CHAPTER 9 UTILIZATION ANALYSES		143
9.1	Introduction.....	143
9.2	Utilization Measures	144
9.2.1	Hospital Discharges	144
9.2.2	Emergency Department Visits	144
9.2.3	Hospital Readmissions.....	145
9.3	Descriptive Analyses	145

9.3.1	Hospital Discharges	145
9.3.2	Emergency Department Visits	145
9.3.3	Hospital Readmissions.....	146
9.4	Trends in Assigned Beneficiaries versus Comparison Groups.....	147
9.4.1	Hospital Discharges	147
9.4.2	Emergency Department Visits	151
9.4.3	All Cause Hospital Readmissions.....	151
9.5	Statistical Estimates of Demonstration Impacts	158
9.5.1	Hospital Discharges	158
9.5.2	Emergency Department Visits	162
9.5.3	All Cause Hospital Readmissions.....	165
9.6	Summary	168
CHAPTER 10 MULTIVARIATE ANALYSES OF DEMONSTRATION IMPACTS		169
10.1	Revised Comparison Group.....	169
10.2	Propensity Score Methodology.....	173
10.3	Propensity Model Evaluation	174
10.4	Repeated Cross Sections Difference-in-Differences Model.....	182
10.5	Demonstration Medicare Expenditure Effects Based on Repeated Cross Sections Difference-in-Differences Model.....	183
10.6	Possible Refinement to the County-Balancing Method.....	191
10.7	Comparison to Previous Demonstration Effect Estimates.....	191
10.8	Demonstration Effects on Medicare Expenditure Components	192
10.9	Subgroup Analyses of Demonstration Effects.....	196
10.10	Quality Outcomes	203
10.11	Demonstration Impacts on Utilization of Medical Services.....	208
CHAPTER 11 SENSITIVITY ANALYSES FOR REFINEMENTS IN DEMONSTRATION DESIGN		219
11.1	Patient Attribution.....	219
11.1.1	Type of Provider	219
11.1.2	Quantity of Services.....	221
11.2	Alternative Performance Benchmarks.....	224
11.3	Alternative Calculation of Savings and Sharing Rates.....	228
11.3.1	Design Assumptions of the Alternative Scenarios.....	228
11.3.2	Net Savings, Simulated Performance Payments, and Medicare Program Savings	232
11.4	Target Expenditures Without Risk Adjustment.....	234
11.5	Diagnostic Coding	237
11.5.1	Importance of Coding in the Demonstration	237
11.5.2	Reasons for Coding Differences	237

11.5.3	Observed Coding Changes.....	238
11.5.4	Changes in Risk Scores in Pre-Demonstration and Demonstration Periods.....	242
11.5.5	Concurrent versus Prospective Risk Scores.....	246
11.5.6	Stayer Analysis	250
11.6	Medicare Payment Rates	252
11.6.1	Inpatient Prospective Payment System.....	253
11.6.2	Other Medicare Payment Policies: Site of Service.....	255
11.7	Alternative Quality Performance Methodology	257
11.7.1	Setting Yearly Quality Improvement Targets.....	258
11.7.2	Setting Benchmark to be Highest of All Possible Targets.....	260
11.7.3	Composite Scoring.....	262
 CHAPTER 12 DISCUSSION OF DEMONSTRATION MODEL AND GENERALIZABILITY		
		265
12.1	Discussion of Demonstration Model: Strengths and Limitations.....	265
12.2	Generalizability of Demonstration Model.....	267
12.2.1	Large Physician Groups.....	267
12.2.2	Additional Issues for Small Physician Groups	268
 REFERENCES		
		269

This page intentionally left blank.

List of Figures

Figure 2-1	Process for calculating performance payments in the PGP demonstration	6
Figure 3-1	Service areas of PGP demonstration participants	15
Figure 5-1	Assigned beneficiary versus demonstration target expenditures, 2004 base year to PY5, all 10 PGPs.....	43
Figure 5-2	Assigned beneficiary versus demonstration target expenditures, by PGP, 2004 base year to PY5	44
Figure 5-3	PY1 target minus actual expenditures as a percentage of target expenditures, by PGP	47
Figure 5-4	PY2 target minus actual expenditures as a percentage of target expenditures, by PGP	50
Figure 5-5	PY3 target minus actual expenditures as a percentage of target expenditures, by PGP	52
Figure 5-6	PY4 target minus actual expenditures as a percentage of target expenditures, by PGP	54
Figure 5-7	PY5 target minus actual expenditures as a percentage of target expenditures, by PGP	57
Figure 5-8	Distribution of PGP demonstration gross savings (in millions of dollars) amounts within $\pm 2\%$ corridor not counted as savings or losses	63
Figure 7-1	Assigned beneficiary versus risk-adjusted target expenditures, 2001 to performance year five	101
Figure 7-1a	Assigned beneficiary versus risk-adjusted target expenditures for 2 PGPs sharing in savings all five years, 2001 to performance year five.....	102
Figure 7-1b	Assigned beneficiary versus risk-adjusted target expenditures for 3 PGPs not sharing in savings any of the five years, 2001 to performance year five	103
Figure 7-1c	Assigned beneficiary versus risk-adjusted target expenditures for 5 PGPs sharing in savings some of the five years, 2001 to performance year five.....	104
Figure 7-2	Assigned beneficiary versus risk-adjusted target expenditures by PGP, 2001 to performance year five	105
Figure 7-3	Assigned beneficiary versus non-risk-adjusted target expenditures, 2001 to performance year five	106
Figure 7-4	Assigned beneficiary versus non-risk-adjusted target expenditures by individual PGP, 2001 to performance year five.....	107
Figure 8-1	Trend in quality performance for DM-1 (HbA1c testing): PGPs vs. comparison groups.....	123
Figure 8-2	Trend in quality performance for DM-4 (LDL-C testing): PGPs vs. comparison groups	124
Figure 8-3	Trend in quality performance for DM-6 (Urine protein testing): PGPs vs. comparison groups.....	125
Figure 8-4	Trend in quality performance for DM-7 (Eye exam): PGPs vs. comparison groups.....	126
Figure 8-5	Trend in quality performance for HF-2 (LVF testing): PGPs vs. comparison groups.....	127
Figure 8-6	Trend in quality performance for CAD-5 (Lipid profile): PGPs vs. comparison groups.....	128

Figure 8-7	Trend in quality performance for PC-5 (Breast cancer screening): PGP vs. comparison groups	129
Figure 8-8	Range of change in performance scores from base year to PY5 across PGPs and CGs.....	131
Figure 9-1	Hospital discharge rates per 10,000 beneficiaries over all 10 PGPs by year.....	149
Figure 9-2	Hospital discharge rates per 10,000 beneficiaries by PGP and year.....	150
Figure 9-3	Emergency department visits per 10,000 beneficiaries over all 10 PGPs by year.....	153
Figure 9-4	Emergency department visits per 10,000 beneficiaries by PGP and year	154
Figure 9-5	Readmission rates over all 10 PGPs by year	156
Figure 9-6	Readmission rates by PGP and year	157
Figure 9-7	Hospital discharge rates: Overall difference in differences between the base year and each performance year	160
Figure 9-8	Hospital discharge rates: Difference in differences between the base year and each performance year by PGP	161
Figure 9-9	Emergency department visits: Difference in differences between the base year and each performance year	163
Figure 9-10	Emergency department visits: Difference in differences between the base year and each performance year by PGP	164
Figure 9-11	All Cause Readmission rates: Difference in differences between the base year and each performance year	166
Figure 9-12	Readmission rates: Difference in differences between the base year and each performance year by PGP.....	167
Figure 10-1	Distributions of predicted propensity scores for the assigned and comparison beneficiaries for each of the 10 PGP sites, aggregated over 2001 to PY5	179
Figure 10-2	Predicted Expenditures by Year, Assigned and Comparison Beneficiaries, 10 PGPs Combined, Assigned Beneficiary Effects by Year	190
Figure 10-3	Components for Difference-in-Differences Calculations	214
Figure 10-4	Components for Difference-in-Differences for the Combined Hurdle Calculations.....	215
Figure 11-1	Assigned beneficiary versus non-risk-adjusted target expenditures, 2004 base year to PY5, all 10 PGPs	235
Figure 11-2	Assigned beneficiary versus non-risk-adjusted target expenditures, by PGP, 2004 base year to PY5	236
Figure 11-3	Risk ratio of assigned beneficiaries vs. risk ratio of comparison group, 2004 base year to PY5 ²	244
Figure 11-4	Risk ratio of assigned beneficiaries vs. risk ratio of comparison group, 2001 base year to PY5 ²	245
Figure 11-5	Number of quality targets met in PY4, original PGP method vs. rebasing QI targets.....	259
Figure 11-6	Percent of total points earned in PY4, original PGP method vs. rebasing QI targets.....	260
Figure 11-7	Number of quality targets met in PY4, original PGP method vs. setting to highest target.....	261
Figure 11-8	Percent of total points earned in PY4, original PGP method vs. setting to highest target.....	262

List of Tables

Table 2-1	Quality measures, weights and total quality points by module for the PGP demonstration.....	9
Table 3-1	PGP demonstration participants: Organizational characteristics.....	13
Table 3-2	PGP demonstration participants: Geographic characteristics.....	14
Table 4-1	Summary of programs implemented or enhanced at the groups participating in the PGP demonstration.....	18
Table 4-2	Assigned beneficiaries enrolled in care management programs, performance year one.....	22
Table 4-3	Beneficiaries enrolled in selected ¹ disease or condition specific care management programs, performance year one.....	24
Table 4-4	Beneficiaries enrolled in selected ¹ general care management programs, performance year one.....	25
Table 4-5	Assigned beneficiary enrollment in care management programs from base year and performance year one.....	27
Table 5-1	Number of assigned and comparison group beneficiaries.....	32
Table 5-2	Beneficiaries excluded from PGP assignment due to Medicare advantage enrollment.....	34
Table 5-3	Assigned beneficiary and comparison group characteristics, performance year five.....	35
Table 5-4a	Expenditures per beneficiary.....	37
Table 5-4b	Risk scores per beneficiary.....	37
Table 5-4c	Risk-adjusted mean expenditures ¹ per beneficiary and risk scores.....	38
Table 5-5a	Selected risk-adjusted ¹ expenditure components per beneficiary—Inpatient expenditures (facility only).....	39
Table 5-5b	Selected risk-adjusted ¹ expenditure components per beneficiary—Outpatient ² expenditures.....	39
Table 5-5c	Selected risk-adjusted ¹ expenditure components per beneficiary—Part B Physician/Supplier.....	40
Table 5-6	Physician group practice demonstration performance payments calculation summary, performance year one.....	46
Table 5-7	Physician group practice demonstration performance payments calculation summary, performance year two.....	49
Table 5-8	Physician group practice demonstration performance payments calculation summary, performance year three.....	51
Table 5-9	Physician group practice demonstration performance payments calculation summary, performance year four.....	53
Table 5-10	Physician group practice demonstration performance payments calculation summary, performance year five.....	56
Table 5-11	Physician group practice demonstration shared savings per member per month, performance years one through five.....	58
Table 5-12	Savings, performance payments, and Medicare program savings, PY1-PY5 (amounts in thousands) amounts within \pm 2% corridor not counted as savings or losses.....	60

Table 5-13	Savings and performance payments as a proportion of Medicare allowable charges	65
Table 5-14	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, PY1-PY5 combined	65
Table 5-15a	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 1	66
Table 5-15b	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 2	67
Table 5-15c	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 3	67
Table 5-15d	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 4	68
Table 5-15e	Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 5	68
Table 6-1	Percent of quality targets ¹ met and total points ² earned from PY1 to PY5	70
Table 6-2a	Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 1 and 2	72
Table 6-2b	Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 3 and 4	72
Table 6-2c	Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 5 and 6	73
Table 6-2d	Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 7 and 8	73
Table 6-2e	Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 9 and 10	74
Table 6-3	Number of quality targets hit per PGP for Diabetes Mellitus (DM) in PY1 - PY5	75
Table 6-4	Number of PGPs hitting a quality target in PY1 – PY5 by Diabetes Mellitus (DM) measure	76
Table 6-5a	Heart Failure (HF) results for BY and PY2 - PY5, by PGP 1 and 2	78
Table 6-5b	Heart Failure (HF) results for BY and PY2 - PY5, by PGP 3 and 4	78
Table 6-5c	Heart Failure (HF) results for BY and PY2 - PY5, by PGP 5 and 6	79
Table 6-5d	Heart Failure (HF) results for BY and PY2 - PY5, by PGP 7 and 8	79
Table 6-5e	Heart Failure (HF) results for BY and PY2 - PY5, by PGP 9 and 10	80
Table 6-6	Number of quality targets hit per PGP for Heart Failure (HF) in PY2-PY5	82
Table 6-7	Number of PGPs hitting a quality target in PY2-PY5 by Heart Failure (HF) measure	82
Table 6-8a	Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 1 and 2	84
Table 6-8b	Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 3 and 4	84
Table 6-8c	Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 5 and 6	85
Table 6-8d	Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 7 and 8	85

Table 6-8e	Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 9 and 10.....	86
Table 6-9	Number of quality targets hit per PGP for Coronary Artery Disease (CAD) in PY2-PY5.....	87
Table 6-10	Number of PGPs hitting a quality target in PY2-PY5 by CAD measure	88
Table 6-11a	Hypertension (HTN) results for BY and PY3 – PY5, by PGP 1 and 2	90
Table 6-11b	Hypertension (HTN) results for BY and PY3 – PY5, by PGP 3 and 4	90
Table 6-11c	Hypertension (HTN) results for BY and PY3 – PY5, by PGP 5 and 6	90
Table 6-11d	Hypertension (HTN) results for BY and PY3 – PY5, by PGP 7 and 8	91
Table 6-11e	Hypertension (HTN) results for BY and PY3 – PY5, by PGP 9 and 10	91
Table 6-12	Number of quality targets hit per PGP for Hypertension (HTN) PY3-PY5	92
Table 6-13	Number of PGPs hitting a quality target in PY3-PY5 by HTN measures	92
Table 6-14a	Preventative Care (PC) results for BY and PY3 – PY5, by PGP 1 and 2.....	94
Table 6-14b	Preventative Care (PC) results for BY and PY3 – PY5, by PGP 3 and 4.....	94
Table 6-14c	Preventative Care (PC) results for BY and PY3 – PY5, by PGP 5 and 6.....	94
Table 6-14d	Preventative Care (PC) results for BY and PY3 – PY5, by PGP 7 and 8.....	94
Table 6-14e	Preventative Care (PC) results for BY and PY3 – PY5, by PGP 9 and 10.....	95
Table 6-15	Number of quality targets hit per PGP for Preventative Care (PC) in PY3-PY5	96
Table 6-16	Number of PGPs hitting a quality target in PY2-PY5 by PC measures	96
Table 6-17a	Physician Quality Reporting Initiative (PQRI) incentive payments earned by PGPs during PY2 and PY3	98
Table 6-17b	Physician Quality Reporting Initiative (PQRI) incentive payments earned by PGPs during PY4 and PY5	98
Table 7-1	Target minus assigned beneficiary expenditures in demonstration performance year five, compared to simulated pre-demonstration trend	109
Table 7-2	Target minus assigned beneficiary expenditures in demonstration performance year five, compared to simulated pre-demonstration trend by individual PGP	111
Table 7-3	Target minus assigned beneficiary expenditures in demonstration performance years one to five, compared to the simulated pre-demonstration trend	112
Table 7-4	Target minus assigned beneficiaries expenditures per person year by subgroup, performance year five	114
Table 7-5	Target minus assigned beneficiary expenditures per person year by expenditure component, performance year five.....	118
Table 8-1	Average differences in the claims-based quality performance scores between base year and performance year five across all ten PGPs	133
Table 8-2	Differences in DM1: HbA1c management quality performance score between base year and performance year five	134
Table 8-3	Differences in DM4: Lipid measurement quality performance score between base year and performance year five	135
Table 8-4	Differences in DM6: Urine protein testing quality performance score between base year and performance year five	136
Table 8-5	Differences in DM7: Eye exam quality performance score between base year and performance year five.....	137

Table 8-6	Differences in HF2: LVG testing quality performance score between base year and performance year five.....	138
Table 8-7	Differences in CAD5: Lipid profile quality performance score between base year and performance year five	139
Table 8-8	Differences in PC5: Breast cancer screening quality performance score between base year and performance year five	140
Table 9-1	Distribution of beneficiaries by number of hospital discharges, BY and PY5 ...	146
Table 9-2	Distribution of beneficiaries by number of emergency department visits, BY and PY5.....	146
Table 9-3	Readmission rates, BY and PY5.....	147
Table 9-4	Hospital discharge rates per 10,000 beneficiaries by PGP and year.....	148
Table 9-5	Emergency department visits per 10,000 beneficiaries by PGP and year	152
Table 9-6	Readmission rates by PGP and year	155
Table 9-7	Hospital discharge rates: Difference in differences between the base year and each performance year	159
Table 9-8	Emergency department visits: Difference in differences between the base year and each performance year	162
Table 9-9	All Cause Readmission rates: Difference in differences between the base year and each performance year	165
Table 10-1a	Number of demonstration assigned beneficiary years used in the regression analysis, total, 2001-PY5	171
Table 10-1b	Number of demonstration comparison group beneficiary years used in the regression analysis, total, 2001-PY5.....	172
Table 10-1c	Final sample size (number of beneficiary years) used in the regressions (assigned and comparison).....	172
Table 10-2	Propensity score models for regression weighting and sample selection, by PGP, 2004 (Table entries are coefficient estimates with p-values for statistical significance beneath)	175
Table 10-3	Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures (standard errors and p-values for statistical significance are shown below coefficient estimates in dollars) (regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)	184
Table 10-4	Demonstration impacts on components of per capita expenditures (standard errors and p-values are below coefficients)	193
Table 10-5	Subgroup percentages of assigned beneficiaries, total 2001-PY5, regression sample	197
Table 10-6	Demonstration impacts on assigned beneficiary subpopulations (standard errors and p-values are below coefficients)	199
Table 10-7	Demonstration impact on the seven claims-based quality indicators (Logistic regression results) (standard errors and p-values are below the odds ratio coefficients).....	204
Table 10-8	Demonstration impact on hospital admissions and emergency department visits, All PGPs (Hurdle regression results)	211
Table 10-9	PGP demonstration impacts on hospital admissions: Results from Hurdle regression models.....	216

Table 10-10	PGP demonstration impacts on emergency department visits: Results from Hurdle regression models	217
Table 11-1	Statistics on Alternative Beneficiary Assignment Algorithms, Performance Year 4.....	220
Table 11-2a	Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs One or More E&M Visits	222
Table 11-2b	Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs Plurality (PGP model)	222
Table 11-2c	Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs Majority	223
Table 11-3	Alternative performance benchmarks/expenditure target.....	225
Table 11-4	Actual performance year two per-capita expenditures and percent difference of target over actual expenditures	226
Table 11-5	Actual performance year two shared savings as percent of target expenditures	227
Table 11-6	Treatment of 2% corridor, MSR, sharing rate, and earned performance payment caps under the PGP demonstration methodology and the four alternative scenarios.....	230
Table 11-7	Net savings, net amounts within the 2% corridor, total simulated performance payments, and Medicare program savings under PGP Demonstration methodology and Scenarios One-Four, PY1-PY5 combined	233
Table 11-8	Percent change in the percentage of beneficiaries with NOHCC, base year to performance year two	239
Table 11-9a	Percentage change in rate of coding of four HCC groups, base year to performance year two, PGP 1 – PGP 5.....	240
Table 11-9b	Percentage change in rate of coding of four HCC groups, base year to performance year two, PGP 7 – PGP 10.....	240
Table 11-10a	Percentage change in rate of coding of four HCC groups, base year to performance year two	242
Table 11-10b	Percentage change in rate of coding of four HCC groups, base year to performance year two	242
Table 11-11a	Risk scores base year to performance year five—Assigned beneficiaries	243
Table 11-11b	Risk scores base year to performance year five—Comparison group.....	243
Table 11-12	Average Annual Percent Change in PGP risk score, pre-demonstration and demonstration periods.....	247
Table 11-13a	Concurrent (Demo model) risk scores	249
Table 11-13b	Prospective (MA model) risk scores.....	249
Table 11-14	PGP vs. FFS differential in prospective risk score growth for stayer cohorts....	251
Table 11-15	Changes in relative IPPS prices and relative Medicare payments per beneficiary.....	254
Table 11-16	Comparison of quality performance, using ‘All-or-None’ composite scoring ...	264

This page intentionally left blank.

EXECUTIVE SUMMARY

ES.1 Introduction

The Medicare Physician Group Practice (PGP) demonstration, which was Medicare's first physician pay-for-performance initiative, established incentives for quality improvement and cost efficiency at the level of the physician group practice. A legislative mandate for the PGP demonstration was included in the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act of 2000. The PGP Demonstration sought better coordination of health care furnished under fee for service Medicare through increased investment in administrative structures and processes for more efficient service delivery and rewards for physicians who improved health care processes and outcomes.

The Centers for Medicare & Medicaid Services (CMS) implemented the 5-year PGP demonstration on April 1, 2005. The demonstration's "base year" (BY) for measuring quality and efficiency improvements was calendar year 2004, and the five "performance years" (PY1-PY5) ran consecutively from April to March starting in 2005. This report summarizes the key findings and lessons learned from the PGP demonstration.

ES.2 PGP Demonstration Design

The primary purpose of the PGP demonstration was to effect changes in cost efficiency and quality through physician incentive payments. The key elements of the demonstration design involved identifying the PGP's patients, determining whether there were any changes in efficiency and quality of care, and assessing whether those changes were due to the incentive payments.

Assignment of Beneficiaries to PGPs. A PGP's ability to coordinate and manage the health care of a beneficiary depends on the type of services the PGP provides the beneficiary and the overall control the PGP has over the beneficiary's utilization of services. Because the PGP demonstration operated within fee for service Medicare, there was no enrollment process whereby beneficiaries accepted or rejected participation. Therefore, beneficiaries were annually "assigned" to PGP participants based on utilization of Medicare-covered services. If the beneficiary received more office or other outpatient Evaluation and Management (E&M) services from the PGP than from any other physician practice, then the beneficiary was assigned to that PGP. In PY5, the proportion of total E&M allowed charges received by beneficiaries assigned to the PGPs was 84 percent and the average number of E&M visits for assigned beneficiaries was 5.6.

Demonstration Comparison Group. The effect of demonstration incentive payments was assessed by comparing each PGP's annual growth in per capita costs to the annual growth in per capita costs of a local comparison group. The PGP's comparison group was drawn from the PGP's service area. The PGP's local service area consisted of all counties in which at least 1 percent of the PGP's assigned beneficiaries resided. These counties were combined to form the service area for the PGP.

Measurement of Demonstration Savings. A financial reconciliation algorithm was used to determine savings and performance payments to PGPs in the demonstration. Financial performance is evaluated by comparing each PGP's actual per capita Medicare expenditures to the PGP's Target Expenditures. Target Expenditures are defined for each PGP using the PGP's base year expenditures (2004), adjusted for expenditure growth of the local comparison group (CG) and changes in patient characteristics (risk adjustment) in a performance year.

A 2-percent band or corridor around Target Expenditures was used to account for normal variation in Medicare expenditures. This threshold was believed to be a reasonable balance between paying deserved performance payments and not paying undeserved performance payments. The portion of annual Medicare savings greater than the 2 percent performance payment threshold was then used to compute the incentive payment.

The "sharing rate" was the proportion of savings that CMS shared with a participating PGP. The sharing rate (up to 80 percent of gross savings) was set high enough to give PGPs sufficient incentive to participate in the demonstration, and yet allow for significant savings for the Medicare program. The incentive payment included both performance payments for cost efficiency and quality improvement.

If a PGP's Actual Expenditures were less than 98 percent of its Target Expenditures ("gross savings") the PGP was deemed to have saved Medicare expenditures and earned a performance payment (gross savings multiplied by the applicable sharing rate). If actual expenditures fell within 98 percent and 102 percent of the target, the PGP was not eligible for performance payments, but was considered to have met its target. Finally, if Actual Expenditures exceeded 102 percent of Target Expenditures ("loss"), the PGP received no performance payments and was considered to have achieved negative Medicare savings.

Quality Measurement and Reporting. There were 32 specific quality measures used in the PGP demonstration. They included measures from different condition modules, including diabetes (DM), congestive heart failure (CHF), coronary artery disease (CAD), hypertension (HTN) and preventive care (PC). As a result, they covered a broad range of conditions and indicated treatments. Two types of measurement processes were used to calculate quality performance in the PGP demonstration: claims-based (seven quality measures), and medical records-based (25 measures). Both threshold and improvement targets were available for PGPs to demonstrate they met the quality performance goals of the PGP demonstration. For a given performance year, the percentage of quality targets achieved by a PGP was used to determine performance payments.

ES.3 PGP Demonstration Participants

There were 10 PGP participants in the demonstration: 1) Billings Clinic, 2) Dartmouth-Hitchcock Clinic, 3) Everett Clinic, 4) Forsyth Medical Group, 5) Geisinger Clinic, 6) Marshfield Clinic, 7) Middlesex Health System, 8) Park Nicollet Health Services, 9) St. John's Health System, and 10) University of Michigan. The participating organizations were all large, ranging from 232 to 1,291 affiliated physicians. The variation in practice size allowed testing of the demonstration model's applicability across a range of large practice sizes. Together, the 10 PGP participants spanned four Census regions. Four were located in the Midwest, three in the

Northeast, two in the West, and one in the South. Half of the PGP participants were located in predominantly rural areas, which include scattered small cities or towns. Three PGP participants were located in small city suburban areas, one was located in a smaller urban area, and one was located in a suburb adjacent to a large city. No participant was located in the core of a large city.

Nine of the 10 PGP participants were integrated physician group practices. One participant was a physician network supported by a management services organization and hospital partner. The management services organization provided quality improvement, 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. Inclusion of a network model provided an opportunity to test the generalizability of the demonstration to the majority of physicians who do not practice in large integrated medical groups.

Two participants were faculty group practices within academic medical centers; five belonged to an integrated delivery system consisting of at least one hospital in addition to the physician group (and may include other health care providers such as home health agencies or nursing homes); two were freestanding physician group practices; and one was a hospital sponsored physician network of 60 small groups and individual physician practices.

Seven of the 10 participants had currently or previously owned a health maintenance organization (HMO). This experience promoted an interest in care management and pre-demonstration development of managed care infrastructure. Eight of the participants were not-for-profit organizations, and one was a for-profit subsidiary of a not-for-profit health system. All of the participants planned to initially reinvest any demonstration performance payments in their care management infrastructure, as opposed to sharing it with individual providers.

ES.4 Demonstration Interventions

The demonstration was not designed to test specific interventions; therefore, participating sites had complete autonomy in determining strategies that would provide higher quality care and expenditure savings. Since these strategies could be implemented prior to, or during the demonstration, and were not uniformly designed, defined, or implemented across the 10 PGPs, evaluations of interventions could not be done. However, information collected during site visits was used to compile a list of programs that were used at the PGP demonstration sites. Based on this information, interventions could be classified as either process interventions, which were implemented throughout a larger system, or program interventions, which often target a specific population and require patient or beneficiary enrollment.

Process interventions were widespread and included patient registries and electronic medical records, information system interventions (e.g., automated alert systems in medical records), medication reconciliation programs, and reporting/feedback and educational interventions for physicians and staff. In addition to process re-design and administrative interventions, groups implemented several clinically based care management programs that target specific patient populations. These programs can be classified as programs that target specific diseases or conditions or programs that target a subset of beneficiaries based on cost or patient complexity.

Some groups implemented disease related programs that were expected to generate cost savings, such as congestive heart failure care management programs. Other programs addressed anticoagulation therapy, diabetes, chronic obstructive pulmonary disease, cancer, psychiatric conditions, coronary artery disease, and hypertension. Care management programs for these conditions most often involved education for patient self-management techniques and periodic patient follow-up and assistance with scheduling of appointments and coordination of care. Each of the patient care management programs targeted a specific population either based on condition/disease or based on patient complexity defined by number of diagnoses or patient cost.

Additional programs existed at several groups that were not disease-based. Several of these programs focused on patients with multiple conditions or patients that were high risk.

ES.5 Demonstration Financial Results

Over the five years of the demonstration, gross savings of \$152.9 million were realized (see definition of gross savings above). Offsetting this amount was \$15.1 million in PGP losses (assigned beneficiary expenditures above target), leaving net savings of \$137.8 million. Of this amount, \$107.6 million was distributed to the PGPs as performance payments, leaving Medicare with program savings of \$30.2 million. On a per person basis, annual gross savings of \$143.18 were generated in each demonstration year, \$100.74 was paid to the PGPs as performance bonuses, \$14.15 was a loss offset, and \$28.29 comprised Medicare program savings.

Both actual expenditures and target expenditures increased steadily from 2004 through PY4 with PGP-average actual expenditures below PGP-average target expenditures in each performance year by an increasing amount over time. This upward trend leveled off in PY5 and is partially explained by the PY5 risk score cap. CMS implemented a risk score cap in PY5 due to concerns that the higher growth of assigned beneficiary risk scores was a response to financial incentives of the demonstration rather than a reflection of changes in the health status of the assigned population relative to the comparison population. The risk score cap became a limiting factor for two PGPs and constrained the earned performance payments in PY5. On average, the five year risk score growth from the BY to PY5 was greater for the PGP assigned beneficiaries than for comparison group beneficiaries, averaging 17.4 percent across PGPs compared to 10.0 percent for comparison group beneficiaries. PGPs with target minus actual expenditures in excess of the 2 percent minimum savings threshold were eligible to share in savings up to 80 percent of this amount. A portion of the performance payment was earned for efficiency, while another portion was contingent on quality performance. The relative sizes of these shares began in PY1 at 70 percent for efficiency and 30 percent for quality, and ended in PY5 at 50 percent for efficiency and 50 percent for quality.

The PGPs that earned performance payments varied in each performance year. In PY1 two PGPs earned performance payments totaling \$7,323,697. In PY2 four PGPs earned performance payments totaling \$13,840,014. In PY3 five PGPs earned performance payment of \$25,278,792. In PY4, the sum of performance payments peaked at \$31,679,844 with five PGPs earning performance payments. In PY5, four PGPs earned performance payments totaling \$29,434,607. Only two PGPs earned payments in all five demonstration years and two earned bonuses in only one year. Of the three sites that never earned a bonus in any year, one site had losses in all five years.

ES.6 Quality Measurement & Reporting for Performance Years One Through Five

The PGP demonstration included performance assessments for 32 quality measures, which were gradually introduced. The demonstration began in PY1 with 10 Diabetes Mellitus measures. In PY2, 10 Heart Failure and 7 Coronary Artery Disease measures were also included. By PY3, all 32 measures, including the three Hypertension and two Preventative Care measures, were active. PGPs were able to demonstrate positive performance for each quality measure through meeting either threshold targets or improvement over baseline targets.

While each PGP started at different levels of performance and thus had different improvement over baseline targets, all of the PGPs achieved improvements in their scores for most of the quality measures over time, even when there were 22 more active measures in PY5 than in PY1. For example, while only two PGPs met all 10 measure targets active in PY1 (i.e., 100 percent), by PY5, seven PGPs met all 32, or 100 percent of their targets, and the remaining 3 PGPs still met over 90 percent of the targets. This translated to an average of 90 percent of the possible quality points earned in PY1 across all 10 PGPs, compared to an average of 99 percent points earned in PY5 across all ten PGPs.

Diabetes Mellitus (DM): PGPs increased their scores on the 10 DM quality measures over the course of the demonstration. By the final year of the demonstration, seven of the 10 PGPs had demonstrated improvement in seven or more DM quality measures. Improvements were greatest for DM-8, complete foot exam. By PY5, six of the PGPs had improved more than 30 percent on this measure. For measures with minimal or no improvement, including DM-1 (HbA1c management) and DM-2 (HbA1c management control), PGPs began the demonstration with performance levels of 90 percent or higher, suggesting that improvements in performance would be difficult to achieve. During the demonstration, PGPs met the majority of their quality targets and the number of targets met improved annually. By PY5, only one site missed a DM quality target.

Heart Failure (HF): PGPs increased their scores on most of the 10 HF quality measures over the course of the demonstration. By PY5, all sites had improved on at least 8 measures. For the few quality measures where improvements were not made, most PGPs began the demonstration at performance levels of 90 percent or higher, suggesting that improvements in performance would be difficult to achieve. Some of the largest improvements were seen in HF-5 Patient Education, HF-9 Influenza Vaccination, and HF-10 Pneumonia Vaccination. These were also the measures with the lowest baseline scores. Overall, all 10 PGPs met all of their HF quality targets every year.

Coronary Artery Disease (CAD): PGPs increased their scores on most of the seven CAD quality measures over the course of the demonstration. By PY5, all sites had improved on four or more measures and four PGPs improved on all seven measures. PGPs struggled most to improve on CAD-6 LDL Cholesterol Level; however all sites maintained performance levels between 77 and 99 percent and consistently met targets for this measure. Only one quality measure target was missed during the demonstration (PY2-PY5), for a total of 289 out of 290 targets met across the 10 PGPs over those performance years.

Hypertension (HTN): The HTN measures were not introduced until PY3, giving PGPs less time to improve. However, the majority of sites were able to improve on at least two of the three HTN measures. All sites improved in HTN-1, Blood Pressure Screening. The most challenging measure for the PGPs was HTN-2 Blood Pressure Control, although the results for this measure did improve somewhat over the last three years of the demonstration. A total of 62 of the 90 possible quality targets were met across the 10 PGPs in PY3 through PY5. The number of targets hit each year increased over the course of the demonstration, with 27 out of 30 targets hit across the 10 PGPs in PY5.

Preventative Care (PC): Performance improved on the two PC measures over the course of the demonstration. In PY5 seven of the 10 PGPs had improved on both of the PC measures. PGPs hit 60 of the 60 possible PC quality targets in PY3 through PY5.

PQRI Incentive Payments: The PGPs satisfied their Physician Quality Reporting Initiative (PQRI) quality reporting requirement as a result of reporting quality measures under the demonstration. As such, the demonstration quality performance measures were used to determine the PQRI incentive payments. The PQRI incentive payment was calculated by multiplying the maximum PQRI bonus amount (based on the PQRI formula) by the PGP's overall percentage quality performance score. The PQRI payments for all PGPs totaled \$8,986,697 in PY5 with an average incentive payment of \$898,670.

ES.7 Demonstration Impacts on Medicare Program Expenditures: Simulation Analysis

Under the demonstration savings methodology, the calculations of expenditure growth and Medicare savings did not separate the impact attributable directly to the demonstration from impacts that might have occurred due to trends existing prior to and carrying forward into the demonstration. We analyzed the impact of the PGP demonstration on total Medicare program expenditures for the participating PGPs' overall assigned beneficiary populations, taking into account 2001–2004 pre-demonstration trends. We also simulated the actual demonstration payment methodology, and used the statistical “bootstrapping” technique to determine statistical significance. We analyzed the statistical significance of differences of assigned beneficiary from target expenditures in the demonstration period (2004 to PY5), and when the pre-demonstration (2001–2004) trend in expenditures was removed. These simulations included the 2 percent corridor around target expenditures. We also considered the impact of the demonstration on the expenditures for subgroups of assigned beneficiaries relative to analogously defined subgroups of comparison beneficiaries, as well as impacts by components of Medicare program expenditures.

In order to identify the effect of the demonstration on costs, we assessed cost trends that existed prior to the demonstration. On a risk adjusted basis, on average, expenditures of assigned beneficiaries were trending slightly below simulated target expenditures prior to start of the demonstration in 2005. Starting from equality with the target in 2001, by 2004 assigned beneficiary per capita expenditures were 1.4 percent lower than the simulated target. In the demonstration period, assigned beneficiary expenditures continued to trend lower than target. By PY3, assigned beneficiary expenditures were 3 percent lower than target and by PY5, assigned beneficiary expenditures were 3.9 percent lower than target. Therefore, there is evidence of a small reduction in the rate of assigned beneficiary expenditure growth in the

demonstration's five years, relative to contemporaneous comparison group expenditure growth. However, the slight pre-demonstration trend toward lower-than-target assigned beneficiary expenditures could have continued into the demonstration period even in the absence of the demonstration.

We analyzed the statistical significance of differences of assigned beneficiary actual expenditures from target expenditures in the demonstration period (2004 to PY5) with and without removing the pre-demonstration (2001–2004) trend in expenditures.¹ In PY1, on average across the 10 participating PGPs, assigned beneficiary expenditures were 0.7 percent lower than target expenditures. This increased every year to 2.1 percent lower than target expenditures in PY4, and decreased to 1.5 percent lower in PY5 (all statistically significant differences).² When adjusted for the pre-demonstration trend in expenditures, actual expenditures were lower than target expenditures for all years except PY5 (which was partially a result of the risk score cap put in place in PY5 which only affected two PGPs), but by a lower amount than when not adjusted. Moreover, adjusted for the pre-demonstration trend, assigned beneficiary expenditures are no longer statistically significantly lower than their expenditure target in any of the performance years.

We considered the impact of the demonstration on the expenditures for subgroups of assigned beneficiaries relative to analogously defined subgroups of comparison beneficiaries. Adjusted for prevalence, the largest subgroup contribution to the overall demonstration impact per beneficiary was from beneficiaries with any of 70 high-cost diagnoses, who accounted for all of the overall impact. Beneficiaries with inpatient expenditures, any of the 8 chronic diagnoses, as well as high risk score beneficiaries and beneficiaries with diabetes also accounted for a large portion of the demonstration expenditure impact. Averaged across the 10 demonstration PGPs, the demonstration had statistically significant expenditure reduction impacts on 13 of 18 selected subgroups between the BY and PY5. The largest statistically significant per beneficiary impacts were on the symptomatic coronary artery disease, inpatient utilization, and high risk score subgroups. Other subgroups showing statistically significant impacts were congestive heart failure, diabetes, chronic obstructive pulmonary disease, heart arrhythmia, having at least one of 8 prevalent chronic medical conditions, having one of 70 high-cost diagnoses, and dual Medicare-Medicaid eligibles.

We studied the demonstration impacts by type of Medicare program expenditure. The analysis was done with combined PY 5 data from the ten PGPs (pooled) and then done for each subset of PGPs: those that shared in savings, and those that did not. Overall, inpatient expenditures were \$95 per person year less than target; this difference was statistically significant ($p < .05$). In contrast, total outpatient expenditures were \$37 per person year less than target; this difference was not statistically significant. Home health expenditures were \$43 less than target and statistically significant ($p < .01$). The participating PGPs on average were not

¹ The 2001–2004 trend was multiplied by 5/3 to make the length of the trend comparable to the 5 year Demonstration period trend.

² If the PY5 risk score growth cap was not in place in PY5, the target minus actual expenditures would have increased to a difference of 2.1 percent instead of 1.5 percent. Therefore, the PY5 downturn in average target minus assigned beneficiary expenditures was a result of the PY5 risk score cap.

successful in controlling skilled nursing facility costs. Thus, demonstration cost savings were most strongly related to controlling inpatient and home health expenditures.

The four PGPs sharing in savings controlled both inpatient and outpatient costs, whereas these costs exceeded target in the six that did not share in savings. The four PGPs sharing in savings had lower home health and hospital outpatient costs than target, while those costs exceeded target in the six PGPs that did not share in savings.

ES.8 Demonstration Impacts on Quality of Care

Although the PGPs, in general, improved their quality performance over time, comparison group information is also necessary to discern any demonstration effects on quality of care. Performance assessment was conducted using Medicare FFS Inpatient, Outpatient and Part B Physician/Supplier claims. Descriptive and statistical analysis demonstrated notable differences over time for individual PGPs compared to their CGs, both in the magnitude and direction of quality performance. Nevertheless, the PGPs as a group had higher performance scores by PY5, had larger and more positive changes in performance scores, and met more quality performance targets compared to the CGs in four of the seven claims-based quality measures.

When comparing assigned beneficiaries to CG quality of care prior to and during the demonstration, several trends emerged. PGPs had mostly higher performance scores compared to their CGs. However, performance between the PGPs and the CGs did not differ widely for a number of measures, such as in DM-1 (HbA1c Testing), HF-2 (LVF Testing) and CAD-5 (Lipid Profile). There were also a number of instances (for both PGPs and CGs) where we saw improvements between the base year and PY2, but this improvement either leveled off or even reversed by PY5. Finally, when there were large changes in both PGP and CG scores from one year to the next, PGPs had steeper increases in their performance.

For all seven measures, performance among the PGPs showed more improvements over time and higher average change between base year and PY5 when compared to the CGs. In terms of the range of change, there were wider distributions among the PGPs compared to the CGs, with notable results found in DM-6 (Urine Protein Testing), DM-7 (Eye Exam) and HF-2 (LVEF Testing).

To ascertain whether the differences in quality improvement are statistically attributable to the PGP demonstration, we conducted a difference-in-differences (DD) assessment of each of the seven claims-based quality measures. The DD technique calculates the difference between the base year and PY5 quality scores for the PGPs, the difference between the base year and PY5 quality scores for the CGs, and finally the difference in the PGP and CG differences.

When evaluating the overall performance of each measure, all seven quality scores improved between the base year and PY5 for the PGPs. This was also true for the CGs for 6 of the 7 measures, although DM-1 performance declined among the CGs by an average of 1.9 percentage points. The DD analysis showed that the assigned beneficiaries in the PGPs had statistically significantly larger improvements over time in four of the seven measures (DM-1, DM-7, PC-5 and HF-2) compared to the changes in the quality measures for the CGs. For the

remaining three measures (DM-4, DM-6, and CAD-5), the DD analysis showed that the differences in the improvements between the PGPs and the CGs were not statistically different.

ES.9 Utilization Analyses

Ideally, savings would be achieved by reducing per beneficiary service utilization, and substituting lower cost services for higher cost services, particularly for expensive treatments. PGPs also had financial incentives to reduce “avoidable” hospitalizations for conditions for which viable alternative treatments were available.

To assess changes in utilization, changes in three types of utilization measures were reviewed: general acute-care hospitalizations per beneficiary (per 10,000), emergency department visits per beneficiary (rate per 10,000), and all cause hospital readmission rates (within 30 days of discharge). Three types of utilization analyses were performed for each utilization measure: descriptive, time trends, and statistical testing of differences-in-differences (DD: the change in PGP beneficiary utilization rates between the base year and performance years minus the comparable changes in the comparison group rates). This last type of analysis was conducted to ascertain how participating PGPs performed relative to their comparison groups.

Overall, on a non-risk adjusted basis, the PGPs had higher inpatient utilization than the local comparison groups. During the course of the demonstration, hospital discharge rates for both the assigned beneficiaries and the comparison groups first rose and then started falling in PY3 to rates, ultimately, below the base year rates.

The rate of emergency department visits grew for both the assigned beneficiaries and comparison groups between the base year and PY4 before falling in PY5. For both groups, emergency department visits in PY5 were above the base year rates. The DD analyses of emergency department visits indicate most participating PGPs did worse than their comparison groups. Only two PGPs did consistently better.

With one possible exception, all cause hospital readmission rates did not exhibit any particular pattern or trend. Seven of the DDs improved between PY4 and PY5.

ES.10 Multivariate Analyses of Demonstration Impacts

A multivariate analysis of demonstration impacts on cost, utilization, and quality was conducted. The methodology involved several alterations to the descriptive and univariate approaches presented elsewhere. These analyses included the use of 1) a revised comparison group to more closely match the county distribution in the PGP, 2) propensity score weights that were applied to further balance the beneficiary characteristics in each group, and 3) a multivariate Repeated Cross Sections difference-in-differences regression model to estimate demonstration impacts, which adjusts for beneficiary characteristics as well as pre- and post-demonstration time trends. The major findings from these analyses are:

- **Medicare Expenditures.** The overall impact of the demonstration across all PGP sites was a savings of \$171 per assigned beneficiary per person year during the demonstration performance period (standard error = \$22, 95% confidence interval =

\$127 to \$215, $p < 0.01$). Thus on average, for each year in the demonstration period Medicare expenditures for PGP assigned beneficiaries were \$171 less than for comparison beneficiaries, other things equal. Three PGPs showed statistically significant expenditure savings ($p < 0.01$) ranging from \$188 to \$818 per assigned beneficiary per year for three PGPs. Two PGPs had statistically significant savings ($p < 0.05$) of \$229 and \$142 per person year, respectively. A statistically significant dis-savings of \$323 per year was found for one PGP ($p < 0.01$), while no significant effects were found for four PGPs ($p > 0.10$).

- ***Comparison with Other Methods.*** Our best estimate of the impact of the demonstration on expenditures is \$171 per assigned beneficiary per year, based on the multivariate methodology. Impacts on expenditures estimated by the multivariate methodology were broadly consistent with those based on the accounting or descriptive analytical methods presented elsewhere. The various savings estimates are all in the range of \$50 to \$210 per assigned beneficiary per performance year when all ten sites are combined (savings estimates are before performance payments are made). This represents a small, but not negligible reduction in the level of medical expenditures.
- ***Medicare Cost Components.*** Of six major components of Medicare costs, demonstration savings arose almost exclusively from the reduced costs of inpatient care (\$228 when combining all sites). Among the individual PGPs, 6 of 10 achieved significant inpatient savings, and none showed dis-savings. Five PGPs achieved statistically significant savings on home health costs. In contrast, only two PGPs realized statistically significant savings in total outpatient expenditures, and three PGPs incurred significant outpatient dis-savings.
- ***Subgroup Analyses.*** Overall savings were greatest for those diagnosed with cardiovascular conditions (CHF, stroke, and vascular disease) as well as with diabetes and COPD. These results were driven largely by the findings in only three sites. The entire demonstration effect occurred in the 51 percent of beneficiaries with at least one of 7 high-cost conditions and in beneficiaries with the highest 10 percent or 25 percent of risk scores. All demonstration savings were generated among chronically ill beneficiaries and beneficiaries with high expected expenditures, and the majority of overall savings (about two-thirds) were generated among beneficiaries who were hospitalized at least once during a year. Not surprisingly given the PGPs' interventions, demonstration savings were achieved among sick, high-cost beneficiaries. There were no major effects for cancer, ESRD, Medicaid patients, or those who were originally disabled. All of the demonstration savings were achieved among the elderly. Expenditures were somewhat higher for assigned beneficiaries who were disabled.
- ***Quality Outcomes.*** Both demonstration and comparison beneficiaries tended to receive higher levels of care after the 2004 base year on all 7 quality measures. Re-analyses based on the multivariate method showed positive demonstration effects for

6 of the 7 quality of care indicators that were independent of time trends. Statistically significant quality impacts were found in half of the PGP sites.

- **Utilization Outcomes.** Across all the PGP sites combined, there was a 2.3 percent reduction in the number of hospital admissions attributed to the demonstration. Four PGPs had significantly lower admission rates. The demonstration also reduced ED visits by an estimated 2.1 percent per year. Significantly lower ED visit rates were found in 2 of the 10 sites.

ES.11 Sensitivity Analyses for Refinements in Demonstration Design

Our evaluation conducted sensitivity analyses of several demonstration design issues, including patient attribution, performance benchmarks, savings calculations, target expenditures, diagnostic coding, Medicare payment rates, and quality performance measurement.

Patient Attribution. PGPs were consulted during the pre-implementation phase to finalize the patient assignment algorithm that resulted in only using outpatient E&M services provided in physician offices to assign patients to the physician groups. In general, PGPs found the assignment methodology to be a reasonable approach that resulted in a set of assigned beneficiaries for whom they could be held accountable for cost and quality performance. The two PGPs that are academic medical centers had some reservations, however. They found that office or other outpatient 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.

The assignment algorithm under the PGP demonstration, which is based on outpatient E&M services provided by any specialty, was compared with an assignment algorithm based only on primary care E&M services. The empirical results, coupled with the PGP participants stated preferences, seem to indicate that a refinement of the PGP demonstration patient assignment algorithm that is based on E&M services provided by primary care physicians is worth considering for future shared savings demonstrations and programs.

Alternative Performance Benchmarks. The measured savings achieved by each participating PGP depended crucially on the benchmark used for measuring savings. Alternative performance benchmarks were simulated. These include comparison group growth rates based on statewide average, national average, and PGP pre-existing trend. Changing the benchmark for target expenditures can have dramatic impacts on measured savings.

Alternative Calculations of Savings. In the demonstration financial reconciliation methodology, savings or losses within the 2 percent corridor are considered normal variation in claims expenditures. Therefore, results within the corridor did not count toward net savings and could not be distributed as performance payments. However, including the net savings within the 2 percent corridor would have increased both earned performance payments and Medicare program savings to varying degrees based on the design assumptions.

As stated above, \$30.2 million is the net savings to Medicare of the demonstration. If the amount within the 2 percent corridor is considered savings, with no distribution of performance

payments to the PGPs, total Medicare savings would be \$113.0 million. If the demonstration shared savings rate of 80 percent is applied to the amount within the corridor, i.e., simulating a first dollar coverage distribution policy, total Medicare savings would be \$35.9 million. Reducing the shared saving rates from 80 percent to 50 percent and using a first dollar coverage distribution policy would result in net Medicare savings of \$100.2 million. Thus, changes in assumptions and design features of the demonstration can significantly affect program savings.

Impact of Risk Adjustment. The demonstration methodology adjusted for risk using a version of the CMS-HCC risk adjustment model implemented for Medicare Advantage risk adjustment. We analyzed the impact of risk adjustment on demonstration results by simulating a non-risk adjusted target. We find that without risk adjustment to remove the effects of changes in health status, the PGPs overall would have actual expenditures in excess of target expenditures. Six out of the ten PGPs had actual expenditures in excess of non-risk adjusted target expenditures for all five demonstration performance years, while only one PGP had actual expenditures less than non-risk adjusted target expenditures for all five performance years.

Diagnostic Coding. Changes in expenditures for the demonstration sites and their comparison groups were risk adjusted during the demonstration. A modification of the CMS-HCC risk adjustment model was developed for the demonstration. It was a concurrent risk adjustment model, which uses this year's diagnoses and demographics to predict this year's expenditures. The results of the demonstration thus depended on a) changes in expenditures for the demonstration sites and comparison groups and b) changes in risk of the populations in these groups as measured by risk score changes.

The PGPs during the demonstration had higher growth rates in risk scores than did their comparison groups. One reason for this could be that the demonstration provided the PGPs with an incentive for more complete diagnostic coding, which is important for effective care management and quality improvement. Another reason however could be that the PGPs experienced real casemix increases during the demonstration. This could have occurred because of structural changes at the group practice (e.g., the practice hired physicians or opened services treating a sicker patient mix), or perhaps because the demonstration provided incentives to attract beneficiaries with chronic conditions.

Several analyses were conducted to examine diagnostic coding growth. These analyses found that the PGPs generally had higher diagnostic coding growth during the demonstration than the comparison groups, and the overall difference in the risk score growth rates grew from the pre-demonstration to demonstration periods. However, these patterns were similar for PGPs sharing in savings versus those not sharing in savings, and the results of a "stayer" analysis was inconclusive.³

³ For the "stayer" analysis, we identified beneficiaries assigned to the PGP in each year 2006, 2007, and 2008, and then calculated the risk score growth from 2007 to 2008. We repeated this for the Medicare FFS population, which acted as the comparison population. The differential in the risk score growth rates was considered a measure of PGP diagnostic coding intensity. Finally, to check the robustness of our results, we conducted the entire "stayer" analysis again but using a different set of years (2007, 2008, and 2009).

Medicare Payment Rates and Factors. Improvement in efficiency through reduction of unnecessary utilization of Medicare-covered health services was an important goal of the PGP demonstration. However, changes in per beneficiary Medicare savings can also occur through changes in Medicare payment rates, payment factors, and rules that disproportionately affect the providers in the PGP and comparison groups. Additionally, a PGP might decide to admit a greater share of their patients to community hospitals instead of to a teaching hospital. Since community hospitals likely cost less than a teaching hospital for the same DRG, this substitution would likely result in savings to a PGP, savings that is not related to lower utilization. To assess the influence of payment rates and factors, inpatient hospital payments were analyzed because they account for the largest share of Medicare expenditures, 40 percent, and the IME and DSH factors were reduced during the demonstration period.

To examine the impact we compared the mean hospital per case payments (adjusting for case mix) of the assigned beneficiaries to those of the local comparison groups. The analysis found that one PGP would have had a modest change in its relative inpatient payments, which was almost 2 percent more than its local comparison group, while the nine other PGPs would have only minor changes of less than 1 percent. While changes in Medicare payment rates and factors would likely have some effect on the savings calculations, the overall effect for the ten PGPs would be small. Note that this analysis was performed for only the first two project years and it could not hold admissions constant by the type of hospital (teaching, community).

Alternative Quality Performance Methodology. We tested different quality performance benchmarking methodologies. Under the PGP demonstration methodology, performance targets for each topic (module) were set equal to the lowest of the following three values:

- (A) The higher of 75 percent compliance OR the Medicare HEDIS mean for the measure
- (B) The 70th percentile Medicare HEDIS level for the measure
- (C) The quality improvement target, which is defined as a 10 percent reduction in the gap between base year performance and 100 percent compliance.

We explored three additional quality performance methodologies: (1) using prior year results as quality improvement targets, (2) setting the benchmark to be the highest of all possible targets, and (3) using composite scores to assess quality achievements. Composite scores result from combining two or more quality measures into one global score. We compared the performance using these alternative methodologies with current results, and studied how changing the methodology would affect the PGPs' quality performance. Overall, the results show that the PGP demonstration performance targets were less difficult to achieve than targets constructed with alternative methodologies.

ES.12 Generalizability of PGP Demonstration Model

With the continuing strong policy interest in improving its quality and controlling its costs, Medicare is exploring alternative approaches to reform. Pay for performance is one approach that has attracted considerable attention. The PGP demonstration was Medicare's first pay for performance initiative for physicians. Unlike some other pay for performance initiatives, the PGP demonstration explicitly established incentives for efficiency as well as quality. It is a provider-based model that relies on the physician group as the organizational means to improve

the quality and efficiency of care. The Accountable Care Organization (ACO) Medicare Shared Savings Program (MSSP), which was implemented in 2012, followed in many respects the approach pioneered by the PGP demonstration.

The PGP demonstration model changes provider payment, not the insurance arrangements of Medicare beneficiaries, who remain enrolled in the traditional FFS program with complete freedom of provider choice. Disruptions to providers are minimized by the maintenance of standard FFS Medicare payments to them. The innovation of the model is that participating provider groups have the opportunity to earn an additional performance payment for providing high quality and efficient care. They share savings created in the care of their patients with the Medicare program, and retain more of the savings the higher their measured quality of care. The financial risk to providers is mitigated by the continuance of FFS payment, the use of provider-specific base costs as a starting point for measuring savings, and the lack of penalties for underperformance. Providers do face the business risk of investments to improve quality and efficiency without any upfront payments from Medicare, and the risk of foregone FFS revenues.

The PGP demonstration participants were large physician group practices or integrated delivery systems. However, most physicians are not affiliated with such organizations, but work in solo or small group practices. Therefore, the wide generalizability of the model depends on the formation of “network model” organizations that aggregate the experience of many small physician practices. Such physician organizations may also need to be integrated with hospitals and other institutional providers to optimize coordination of care.

CHAPTER 1 INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) initiated the Physician Group Practice (PGP) Demonstration in April 2005. CMS initially designed the PGP Demonstration for a 3-year period, but continued the Demonstration for an additional two years. This 5-year Demonstration offered PGP participants the opportunity to earn performance payments for improving the quality and efficiency of health care delivered to Medicare fee-for-service (FFS) beneficiaries. Ten large PGPs participated in the Demonstration.

1.1 Overview of the Demonstration

The Medicare Physician Group Practice (PGP) Demonstration, which was Medicare's first physician pay-for-performance initiative, established incentives for quality improvement and cost efficiency at the level of the physician group practice. A legislative mandate for the PGP Demonstration was included in Section 412 of the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act of 2000. It established several goals for the PGP Demonstration, including:

1. To encourage coordination of health care furnished under Medicare Parts A and B.
2. To encourage investment in administrative structures and processes for efficient service delivery.
3. To reward physicians for improving health care processes and outcomes.

The statute required "a bonus for each year under the demonstration equal to a portion of the Medicare savings realized for such year relative to the performance target," as well as an "additional bonus ... resulting from process improvements made by and patient outcome improvements attributable to activities of the group."

The premise of the PGP Demonstration was that physician group practices can achieve higher quality and greater cost efficiency by managing and coordinating patient care. The physician groups participating in the PGP Demonstration were engaged in a wide variety of care management interventions to improve the cost efficiency and quality of health care for Medicare FFS patients. These interventions included: chronic disease management programs, high risk/high cost care management, transitional care management, end-of-life/palliative care programs, practice standardization, and quality improvement programs. In addition, information technology, such as electronic medical records, patient disease registries, and patient monitoring systems, were being used by PGP participants to improve practice efficiency and quality of care delivered to patients, and to better understand the utilization of services by the Medicare FFS population. The PGP Demonstration tested whether care management initiatives generate cost savings by reducing avoidable hospital admissions, readmissions and emergency department visits, while at the same time improving the quality of care for Medicare beneficiaries.

The PGP Demonstration began on April 1, 2005 and ran for five years. The PGP Demonstration's "base year" (BY) for measuring quality and efficiency improvements was calendar year 2004, and the five "performance years" were April 2005 to March 2006 (performance year one, or PY1); April 2006 to March 2007 (performance year two, or PY2);

April 2007 to March 2008 (performance year three, or PY3); April 2008 to March 2009 (performance year four, or PY4); and April 2009 to March 2010 (performance year five, or PY5).

1.2 Purpose of the Report and Overview

The purpose of this report is to summarize the key findings and lessons learned from the PGP Demonstration. After this introductory chapter, Chapter 2 of the report summarizes the Demonstration design. Chapter 3 then describes the 10 PGP participants. Chapter 4 documents the care management and other initiatives used by the PGP participants to improve quality and efficiency during the Demonstration. Chapters 5 and 6 present the financial reconciliation results and the quality of care measurement and reporting results for the entirety of the Demonstration. Chapters 7 through 9 present evaluation findings on the impact of the Demonstration on Medicare expenditures, quality of care, and utilization. Chapter 10 then presents results from an alternative methodology for identifying Demonstration impacts, which involves the use of propensity score weighting techniques and multivariate regression modeling approaches. Chapter 11 examines refinement issues for the Demonstration, and finally, Chapter 12 provides a discussion of the Demonstration, including generalizability.

1.3 Sources of Information

To support and evaluate the PGP Demonstration, CMS contracted with RTI International, an independent, not-for-profit research organization. Working with CMS, RTI produced reports that specify the Demonstration's design and a plan for its evaluation. The primary source for the PGP Demonstration design is the *Physician Group Practice Demonstration Design Report* (Pope et al., 2002). Revisions, clarifications, and additional detailed specifications are contained in the *Physician Group Practice Demonstration Bonus Methodology Specifications* (Kautter et al., 2004) and *Physician Group Practice Demonstration Quality Measurement and Reporting Specifications, Version 2* (Trisolini et al., 2005). The primary source for the evaluation plan is the *Physician Group Practice Demonstration Evaluation Design* (Kautter et al., 2007a).

Key information sources for this report include:

- PGP Demonstration performance year one (PY1) through performance year five (PY5) results
 - Financial performance results for PY1 through PY5
 - Quality performance results for PY1 through PY5
- RTI analysis of Medicare claims and enrollment data for 2001–2010
- RTI-conducted interviews of PGPs during 2007–2008
- PGP-supplied information
 - PGP annual reports on Demonstration activities
 - PGP care management intervention enrollment lists
 - PGP-initiated Demonstration evaluation analyses
- Publicly available information (e.g., government reports)

CHAPTER 2

PGP DEMONSTRATION DESIGN

This chapter describes the design of the PGP Demonstration (Pope et al., 2002; Kautter et al., 2004; Trisolini et al., 2005; Kautter et al., 2007b). After an overview of Demonstration goals and objectives, several of the Demonstration's key design elements are explained, including beneficiary assignment, comparison population, and measurement of Demonstration savings. Next, the determination of performance payments in the PGP Demonstration is discussed and additional design elements are outlined. Quality measurement and reporting in the Demonstration is then described.

2.1 Demonstration Goals and Objectives

The PGP Demonstration was Medicare's first physician pay-for-performance initiative, and established incentives for quality improvement and cost efficiency at the level of the physician group practice. A legislative mandate for the PGP Demonstration was included in the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act (BIPA) of 2000. There were several goals of the PGP Demonstration: (1) testing the use of incentives for health care groups; (2) encouraging coordination of health care furnished under Medicare Parts A and B; (3) encouraging investment in care management infrastructure and processes for efficient service delivery; and (4) rewarding physicians for improving health care processes and outcomes.

2.2 Key Design Elements

The point of the PGP Demonstration was to effect changes in cost efficiency and quality through incentive payments. Identifying the patients whose Medicare program expenditures and quality of care will be used for evaluating the PGPs, determining whether there have been any changes in efficiency and quality of care, and assessing whether those changes are due to the incentive payments must be addressed by the Demonstration design.

2.2.1 Assignment of Beneficiaries to PGPs

A PGP's ability to coordinate and manage the health care of a beneficiary depends on the type of services the PGP provides the beneficiary and the overall control the PGP has over the beneficiary's utilization of services. Because the PGP Demonstration retained the structure of the Medicare FFS system, there was no enrollment process whereby beneficiaries accepted or rejected involvement. Therefore, beneficiaries were "assigned" to PGP participants based on utilization of Medicare-covered services. A beneficiary who received at least one "office or other outpatient" evaluation and management (E&M) service from a PGP during a year was eligible for assignment to the PGP that year. If the beneficiary received more office or other outpatient E&M services (as measured by Medicare allowed charges) from the PGP than from any other physician practice (group or solo)—in other words, if a plurality of E&M services were provided by the Demonstration PGP—then the beneficiary was assigned to that PGP.¹

¹ Assigned beneficiaries also had to meet additional criteria during a demonstration year, including enrollment in traditional Medicare FFS (no enrollment in Medicare Advantage), no MSP (Medicare secondary payer) status, US resident, and other more technical criteria (see Kautter et al., 2004).

Therefore, no beneficiary was assigned to multiple PGPs in a particular year. PGP assigned beneficiaries were identified retrospectively after the end of the year (whether base or performance) since only after the year ends can a plurality of a beneficiary's E&M services in that year be known.

2.2.2 Demonstration Comparison Group

The comparison group for each PGP, whose changes in Medicare expenditures were used to distinguish the effect of Demonstration incentive payments from unrelated trends among Medicare beneficiaries, was drawn from the PGP's service area. The PGP's service area consisted of all counties in which the PGP derived at least 1 percent of its assigned beneficiaries. These counties were combined to form the service area for the PGP. The service area was defined for the base year and redefined for each performance year, and may have differed between years to reflect changes in the geographic scope of the group's practice.

The comparison group assignment criteria were very similar to those for the PGP assigned beneficiaries. The goal of the comparison group assignment criteria was to ensure that beneficiaries assigned to the comparison group were similar to those assigned to the participating PGP. For example, to be assigned to the PGP's comparison group, a beneficiary must have received at least one "office or other outpatient" E&M service.² Finally, beneficiaries assigned to the PGP in the current or any prior performance year, or beneficiaries that have received any "office or other outpatient" E&M services at the PGP in the current performance year, were not eligible for assignment to the comparison group. This is because these beneficiaries' expenditures are likely to be affected by the Demonstration incentives. For a given year (base or performance), the PGP's comparison population was identified retrospectively after the end of the year.

2.2.3 Measurement of Demonstration Savings

Demonstration savings, termed "Medicare savings" in the PGP Demonstration, measured the cost savings impact of the PGP Demonstration, defining the pool of savings that the participating physician groups and the Medicare program shared. All components of Medicare Part A and B expenditures (except hospice expenditures) were used to measure Medicare savings (including hospital inpatient, hospital outpatient, physician/supplier, skilled nursing facility, home health agency, and durable medical equipment). To calculate Medicare savings in a performance year, first the participating PGP's annual per capita expenditure target, which had a PGP-specific expenditure base, was calculated as follows (all expenditures are on a per capita basis):

$$\text{Target Expenditures} = \text{PGP Base Year Expenditures} \times (1 + \text{Comparison Group Growth Rate})$$

PGP base year per capita expenditures were calculated for beneficiaries assigned to the PGP in the base year. The comparison group growth rate was defined as the growth in per capita expenditures in the PGP's comparison population between the base and performance years. Both

² As with assigned beneficiaries, comparison group beneficiaries had to meet additional criteria during a demonstration year (e.g., enrollment in traditional Medicare FFS, no MSP status, US resident, etc.—see Kautter et al., 2004).

the PGP base year expenditures and the comparison group expenditure growth rate were adjusted for casemix change between the base and performance years using a modification of the CMS-Hierarchical Condition Categories (HCC), or CMS-HCC, risk adjustment model (Pope, Kautter, Ellis, et al, 2004; Olmsted, Pope, Kautter, 2006).³ There was a growing concern by CMS that more complete diagnostic coding could result in higher risk score growth for the PGPs relative to their comparison groups. Because of this, a risk score cap was imposed in PY5. In effect, the difference in risk score growth rates for the PGP and comparison group between BY and PY5 was capped at 10 percentage points.

Medicare savings were computed as the difference between the per capita expenditure target and the PGP's per capita expenditures in the performance year (for beneficiaries assigned to the PGP in the performance year),⁴ multiplied by the number of beneficiaries (person years) assigned to the PGP in the performance year:⁵

$$\text{Medicare Savings} = (\text{Target Expenditures} - \text{PGP Performance Year} \\ \text{Actual Expenditures}) \times \text{Assigned Beneficiary Person Years}$$

This is a retrospective calculation, since neither actual nor target expenditures was known until after the end of the performance year.

2.3 Performance Payments and Additional Design Features

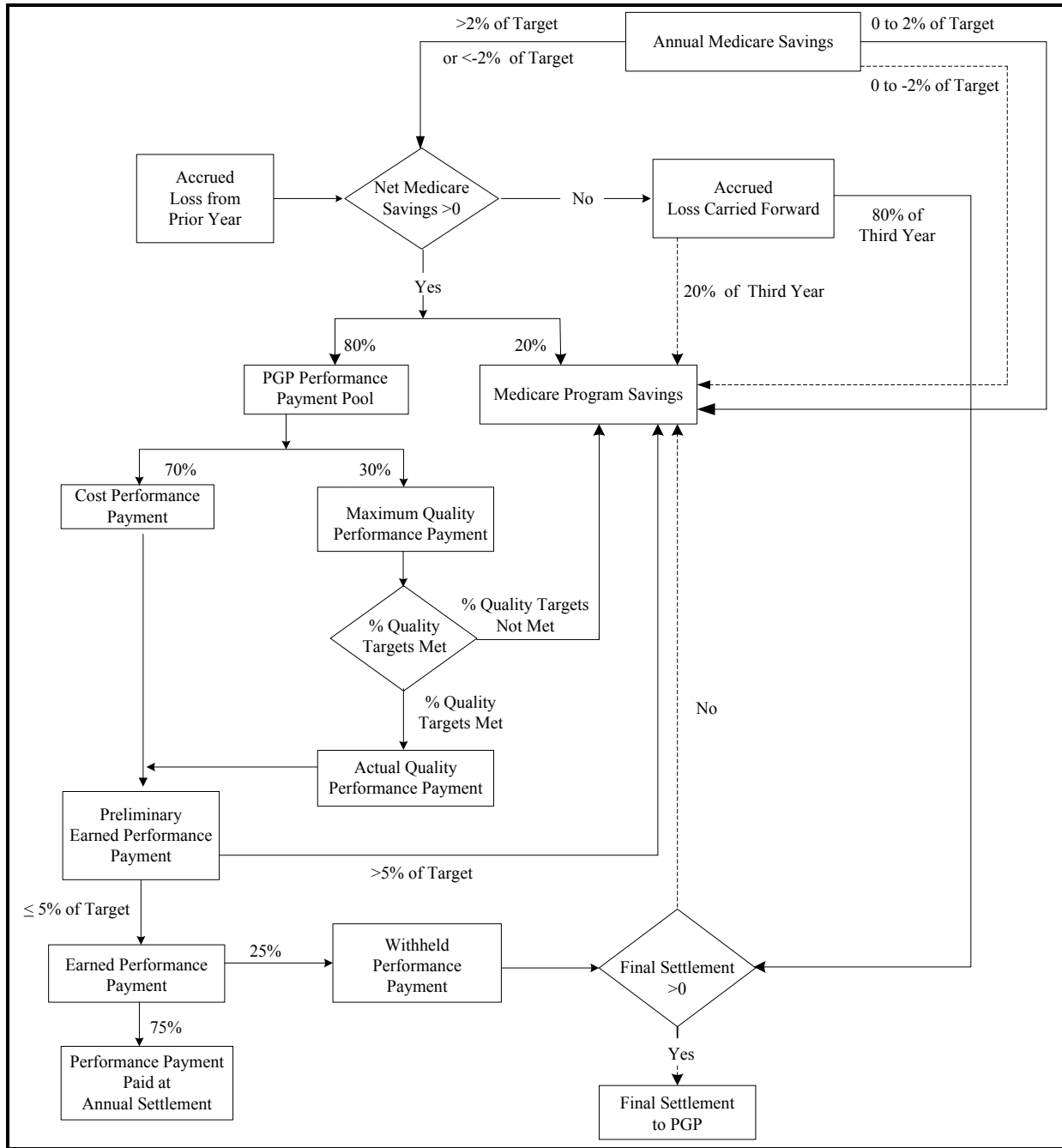
The flowchart in *Figure 2-1* shows the process of calculating performance payments in the PGP Demonstration. The first step involved calculating whether or not a PGP generated annual Medicare savings greater than two percent of its target expenditures. This was equivalent to “Net Medicare savings” being greater than \$0, where Net Medicare savings was defined as the portion of annual Medicare savings greater than two percent of its target expenditures or less than negative two percent of its 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. Actual observed expenditure growth rates are a combination of changes due to a PGP's efforts and those due to events specific to individual

³ To carry out their planned care management interventions during the demonstration, PGPs needed to engage in complete diagnostic coding of patients. While this was encouraged by Medicare, there was a growing concern that more complete diagnostic coding could result in higher risk score growth for the PGPs relative to their comparison groups. Because of this, in PY5 a risk score cap was imposed on the demonstration payment design. In effect, the difference in risk score growth rates for the PGP and comparison group between BY and PY5 was capped at 10 percentage points.

⁴ For each beneficiary assigned to the PGP in the performance year, performance-year expenditures were annualized by dividing expenditures by the fraction of the year they were alive and enrolled in Medicare. Performance year per capita expenditures were then weighted by this fraction.

⁵ Person-years were determined as follows. For each beneficiary assigned to the PGP in the performance year, the fraction of the year the beneficiary was alive and enrolled in Medicare was calculated. Person-years equaled the sum of these fractions.

Figure 2-1
Process for calculating performance payments in the PGP demonstration



NOTES:

1. Dotted lines represent negative contribution to Medicare program savings.
2. Annual Medicare Savings between -2% and 2% of target expenditures are not included in performance payment computations because they may result from random fluctuations. They are included in Medicare Program Savings.
3. In Performance Year 1, the cost performance payment and maximum quality performance payment shares of the PGP performance payment pool are 70% and 30%, respectively. In Performance Year 2, the shares are 60% and 40%, respectively, and in Performance Years 3 to 5, the shares are 50% and 50%, respectively.

SOURCE: RTI International

patients and entirely outside of a PGP's control. When observed expenditure growth rates are close to zero (both positive and negative), there is a large likelihood that the PGP-driven change is zero. As the observed expenditure growth rates move away from zero, the likelihood that the PGP-driven change is zero diminishes. The two-percent performance payment threshold represents a reasonable balance between paying deserved performance payments and not paying undeserved performance payments (Pope and Chromy, 1997). If the PGP held the expenditures for its assigned beneficiaries more than two percent below its target, it was eligible to earn a performance payment for that performance year (assuming there were no accrued losses from previous years). The portion of annual Medicare savings greater than the two percent performance payment threshold was then used to compute the incentive payment.

The "sharing rate" for Net Medicare savings was the proportion that CMS shared with a participating PGP. The sharing rate was set high enough (80 percent) to give PGPs sufficient incentive to participate in the Demonstration, and yet allow for significant savings for the Medicare program. The 80 percent shared Net Medicare savings contributed to the PGP's performance payment pool; the remaining 20 percent was retained by Medicare as program savings. The PGP performance payment pool was then divided between the cost performance payment and the maximum quality performance payment. The actual quality performance payment was then determined, based on the percentage of the PGP Demonstration's quality targets the PGP met in the performance year.⁶ If all of the quality targets were met, then the entire maximum quality performance payment was earned by the PGP. However, if some of the quality targets were not met, then a portion of the maximum quality performance payment was retained by the Medicare program. In performance year one, the cost performance payment and maximum quality performance payment shares of the PGP performance payment pool were 70 percent and 30 percent, respectively. In performance year two the respective shares were 60 percent and 40 percent, and in performance years three through five the shares were 50 percent and 50 percent. These percentages were set to gradually increase the importance of quality performance in the PGP Demonstration.

Once the actual quality performance payment had been determined, it was added to the cost performance payment to identify the preliminary earned performance payment. However, to avoid incentives for excessive cost cutting, a five percentage point performance payment cap was employed. The actual earned performance payment could not be greater than five percent of the PGP's target expenditures, which included both Part A and Part B expenditures; the final earned performance payment was capped at that five percent level if the preliminary earned performance payment was higher.

Finally, the performance payment paid to the PGP at the annual settlement equaled 75 percent of the earned performance payment amount. The remaining 25 percent of the earned performance payment was withheld until the end of the PGP Demonstration to protect Medicare against losses the PGP may generate in subsequent years. At the end of the PGP Demonstration, the cumulative amount of the withheld performance payment was paid to the PGP, net of any accrued losses.

⁶ See Section 2.4 for a description of the PGP Demonstration's quality measurement and reporting methodology.

In any given performance year, PGP participants may have underperformed their comparison group and generated “losses” to their performance payment pool. That is, assigned beneficiary expenditures may have exceeded target expenditures, in which case Medicare savings were negative. Losses were defined as Medicare *negative* savings beyond two percent of target expenditures. Participating PGPs were not at risk for reimbursing the Medicare program for either annual losses or an accrued net loss at final settlement. However, annual losses were carried forward to the subsequent performance year and were used to offset any positive Medicare savings generated in that year. No performance payments could be earned in a performance year unless Medicare savings were sufficient to offset accrued losses from earlier performance years.

As a result, annual Medicare savings within two percent of target expenditures generated neither losses to be carried forward nor performance payments to be paid (Figure 2-1). This portion of the annual Medicare savings (between negative and positive two percent) was assumed to be caused by random fluctuations in expenditure levels, not by the PGP’s performance.

Because of the relatively short period of the PGP Demonstration (five years), cost savings were measured cumulatively from the original Demonstration base year. Rebasings—updating the base year for setting targets for the annual performance payment computation—did not occur. By not rebasing, CMS gave participating PGPs the maximum incentive to generate savings during the Demonstration period. However, for an official Medicare program modeled after the Demonstration or for an extension of the Demonstration, CMS would want to rebase more often than 5 years, so as not to indefinitely pay for “past performance” and to capture more of the cost savings over time.

2.4 Quality Measurement and Reporting

Table 2-1 lists the 32 specific quality measures used in the PGP Demonstration. The quality measures for the PGP Demonstration are a subset of the measures developed by CMS’s Quality Measurement and Health Assessment Group for the Doctors Office Quality (DOQ) project. They include measures from different DOQ condition modules, including diabetes (DM), congestive heart failure (CHF), coronary artery disease (CAD), hypertension (HTN) and preventive care (PC). As a result, they cover a broad range of conditions and indicated treatments, and benefit from the extensive review and validation of measures conducted for the DOQ project.

Table 2-1
Quality measures, weights and total quality points by module for the PGP demonstration

Diabetes Mellitus		Weight	Congestive Heart Failure		Weight	Coronary Artery Disease		Weight	Hypertension / Preventive Care		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

The quality measures are phased-in under the following time frame:

Performance Year 1: Diabetes measures, including influenza and pneumonia vaccine measures for the diabetic population.

Performance Year 2: Year 1 measures plus the CHF and CAD measures, including influenza and pneumonia vaccine measures for the CHF population.

Performance Year 3: Year 2 measures plus the HTN and PC measures.

Performance Year 4: Year 3 measures.

Performance Year 5: Year 3 measures.

Two types of measurement processes were used to calculate quality performance in the PGP Demonstration: claims-based (seven quality measures), and medical records-based (25 measures). As shown in Table 2-1, the claims-based measures were weighted by a factor of four relative to the medical record-based measures. The reason for this weighting scheme was to reflect the administrative burden associated with reporting medical record based measures versus those measures that may be derived from claims data. Since claims-based measures data is more easily accessible and allows for larger denominators, they were given a higher weight compared with medical record based measures, which require additional resources to report and utilize a statistically valid sampling methodology.

Both threshold and improvement targets were available for PGPs to demonstrate they met the quality performance goals of the PGP Demonstration. For each quality measure, PGPs must have achieved at least one of the following three targets (the first two targets are threshold targets; the third target is an improvement target):

1. The higher of 75 percent compliance or the Medicare HEDIS mean for the measure (for those measures where HEDIS indicators are also available); or
2. Achieve the 70th percentile Medicare HEDIS level (for those measures where HEDIS indicators are also available); or
3. Demonstrate a 10 percent or greater reduction in the gap between the administrative baseline and 100 percent compliance.

Note that PGP and CG beneficiaries with less than a full year of data, who died during the year, or who entered hospice during the year were excluded from all quality analyses for that year.

As described above, for a given performance year, the percentage of quality targets achieved by a PGP was used to determine performance payments (Figure 2-1). Total quality measure points earned in a given performance year were calculated, with claims-based measures counting four points each and medical records-based measures counting one point each. Points earned were divided by total points possible to determine the percentage of quality targets achieved by the PGP for the given performance year. The ratio was applied to the maximum quality performance payment pool to derive the portion of the performance payment for quality.

CHAPTER 3 PGP DEMONSTRATION PARTICIPANTS

There were ten PGP participants in the Demonstration. This chapter describes CMS' Demonstration solicitation, application, and site selection process; and the characteristics of the 10 PGP participants.

3.1 Site Selection

In August 2003, the CMS Administrator selected 11 organizations to participate in the PGP Demonstration. The organizations were selected to achieve the following goals:

1. Geographic diversity through a mix of urban and rural settings and West Coast, Midwest, and East Coast variation.
2. Diversity of organizational models, including free-standing group practices, integrated delivery systems, faculty practices and a management service organization with an affiliated independent practice association.
3. Providing for large sample sizes for computing expenditure and quality measure changes.
4. Selecting Demonstration participants that are well organized and can “hit the ground running” with existing infrastructure and programs required to be successful under the Demonstration during the 3-year performance period.⁷
5. Including organizations that may benefit from the additional incentive the Demonstration offers to develop or expand infrastructure and programs.
6. A range of innovative models that provide insight on alternative approaches that could be used to improve the Medicare program.
7. Technically acceptable by the expert review panels.
8. Avoid geographic overlap of PGP participants to prevent Medicare beneficiaries from enrolling or being assigned to multiple PGP participants resulting in paying twice for the same or similar interventions and contamination of treatment and control groups.
9. Availability of CMS resources to administer and implement the Demonstration.

⁷ As discussed in Chapter 1, the Demonstration was extended to a fourth and fifth performance year.

In December 2004, CMS convened a two day pre-implementation meeting with the groups in Baltimore. The first day focused on educating and answering questions about the final financial model to be used to calculate savings, and PGP presentations about their implementation strategies. The second day of the meeting focused on reaching consensus on the 32 ambulatory care quality measures to use under the Demonstration, the phase-in plan, performance thresholds and quality improvement targets, and the relative weights to place on quality measures.

In early 2005, 10 organizations returned signed terms and conditions contracts to proceed with the Demonstration. The Demonstration began on April 1, 2005, and ran through March 31, 2010.

3.2 PGP Participant Characteristics

Table 3-1 identifies the 10 PGP participants and shows key characteristics of their organizations. *Table 3-2* characterizes PGP participants' geographic locations, and their service areas are shown on the *Figure 3-1* map. The order of the PGP participants may differ from the order in the rest of this report as well as in some prior work.

3.2.1 Geographic Characteristics

Together, the 10 PGP participants spanned four Census regions. Four were located in the Midwest, three in the Northeast, two in the West, and one in the South region. Half of the PGP participants were located in predominantly rural areas, which include scattered small cities or towns. Three PGP participants were located in small city suburban areas, one was located in a smaller urban area, and one was located in a suburb adjacent to a large city. No participant was located in a large urban core city.

3.2.2 Organizational Characteristics

Nine of the 10 PGP participants were integrated physician group practices. One participant was a physician network supported by a management services organization and hospital partner. The management services organization provided quality improvement, 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. Its participation as a network model provided an opportunity to test the generalizability of the Demonstration model to the majority of physicians who do not practice in large, traditional integrated medical groups.

Two participants were faculty group practices within academic medical centers; five belonged 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 were freestanding physician group practices; and one was a physician network that was sponsored by a hospital affiliate and comprised of 60 small and individual physician practices.

Table 3-1
PGP demonstration participants: Organizational characteristics

Participant	Organizational Structure	Part Of Integrated Delivery System?	Includes Academic Medical Center?	Owns Or Owned An HMO? ¹	Not For Profit?	Number Of Providers
Dartmouth-Hitchcock Clinic	Faculty/Community Group Practice	Yes	Yes	Yes	Yes	907
Billings Clinic	Group Practice	Yes	No	No	Yes	232
Geisinger Clinic	Group Practice	Yes	No	Yes	Yes	833
Middlesex Health System	Network Model	Yes	No	No	No ²	293
Marshfield Clinic	Group Practice	No	No	Yes	Yes	1039
Forsyth Medical Group	Group Practice	Yes	No	Yes	Yes	250
Park Nicollet Health Services	Group Practice	Yes	No	Yes	Yes	648
St. John's Health System	Group Practice	Yes	No	Yes	Yes	522
The Everett Clinic	Group Practice	No	No	No	No	250
University of Michigan Faculty Group Practice	Faculty Practice	Yes	Yes	Yes	Yes	1,291

NOTES:

¹ HMO may be owned by associated health system, not Demonstration participant per se.

² For profit subsidiary of not for profit health system.

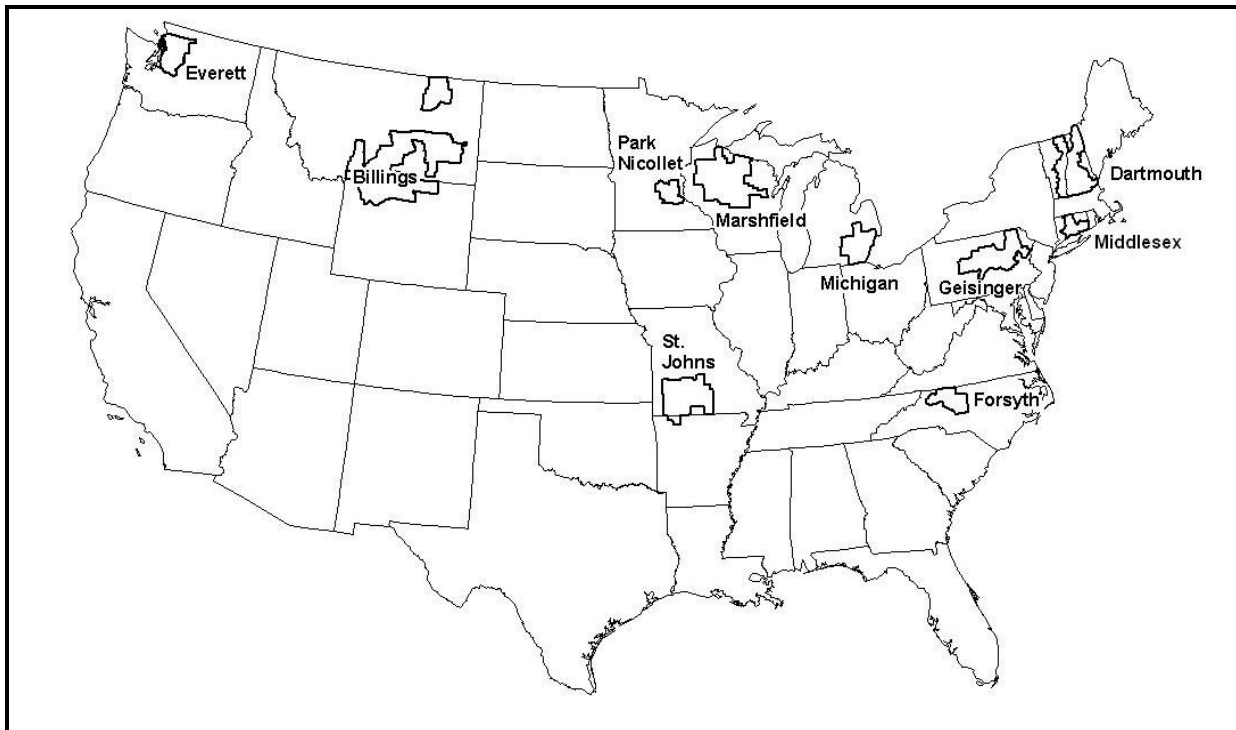
SOURCE: Multiple sources of information including 1) RTI International site visits and interviews with the PGPs, 2) Demonstration applications provided to RTI by CMS, and 3) publicly available information (e.g., PGP websites).

Table 3-2
PGP demonstration participants: Geographic characteristics

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

SOURCE: Multiple sources of information including 1) RTI International site visits and interviews with the PGPs, 2) Demonstration applications provided to RTI by CMS, and 3) publicly available information (e.g., PGP websites).

Figure 3-1
Service areas of PGP demonstration participants



SOURCE: RTI International

It is noteworthy that the presence of a hospital within the same organization facilitates care management and coordination, for example, discharge planning and coordination of inpatient and outpatient care. But, hospitals may be concerned about the loss of inpatient admissions and associated revenues from outpatient care management activities that are encouraged by the Demonstration, which could result in intra-organization tensions that may make continued participation more challenging. Alternatively, high occupancy hospitals may have an opportunity to free up capacity by working with physician groups to provide proactive care management interventions to patients in ambulatory settings thereby reducing avoidable admissions for patients and improving the quality of care and services they receive in alternative settings.

Seven of the 10 participants had currently or previously owned a health maintenance organization (HMO), a managed care insurer. This experience promoted an interest in care management and pre-Demonstration development of managed care infrastructure among the participants. Eight of the participants were not-for-profit organizations, and one was a for-profit subsidiary of a not-for-profit health system. All of the participants had planned to initially reinvest any Demonstration performance payments in their care management infrastructure, as opposed to sharing it with individual providers. The participating organizations were all large, ranging from 232 to 1,291 affiliated physicians, but their size dispersion did allow testing of the Demonstration model's applicability across a range of large practice sizes.

This page intentionally left blank.

CHAPTER 4 DEMONSTRATION INTERVENTIONS

The incentives of the PGP Demonstration triggered the development and implementation or enhancement of several different interventions. These interventions included process re-design or more administrative interventions, as well as interventions or programs that were clinically based, targeted specific populations and involved an enrollment process. Discussions and meetings with the groups participating in the PGP Demonstration throughout the course of the Demonstration revealed specific interventions that may have contributed to the outcomes seen under the Demonstration. Information presented in this chapter draws strongly from the February 2008 GAO Report, RTI site visits and telephone interviews with the PGPs, and annual reports submitted by the PGP groups. The feedback from the groups on which programs appeared to result in positive outcomes under the Demonstration are instrumental to understanding what works at the physician group practice level to generate savings and improve quality of care.

This chapter will (1) provide a brief overview of the types of interventions implemented and or enhanced at the groups throughout the Demonstration, (2) provide some description of the types of beneficiaries that were enrolled in the programs at the groups and (3) will discuss the factors that the groups believe were contributing to or hindering the goals of increased savings and improved quality under the PGP Demonstration.

4.1 Overview of Demonstration Interventions

The groups participating in the Demonstration implemented or enhanced several different interventions that could contribute to the goals of the PGP Demonstration. These interventions can be divided into two distinct groups. The first group is the process intervention, which includes interventions that are slightly more administrative in focus and are implemented throughout a larger system. The second group is the program intervention. These programs often target a specific population, such as a population with a specific disease, and require patient or beneficiary enrollment. Here we divide program interventions into disease management interventions and care coordination programs, as does the February 2008 GAO Report. Each of these types of interventions and specific examples of interventions are provided below. It is important to note that the interventions are not strictly defined. For that reason, we chose to identify program interventions by targeted disease or population of focus. A summary of the interventions and at which groups they were implemented can be found in *Table 4-1*.

4.1.1 Process Improvement Interventions

Process interventions at the groups participating in the PGP Demonstration were widespread and included interventions such as the patient registries, information system interventions (e.g., automated alert systems in medical records), medication reconciliation programs and reporting/feedback and educational interventions for physicians and staff. Some of these process improvement interventions are listed in Table 4-1.

Patient Registries and EMRs. The design and use of patient registries varied across the groups participating in the Demonstration. Some groups had disease specific registries that assist

Table 4-1
Summary of programs implemented or enhanced at the groups participating in the PGP demonstration

Program Category/Name	Brief Description	# of Groups with Program	Groups ¹
<i>Process Improvement Interventions</i>			
Transitional Care Program	Programs that assist patients at hospital discharge to improve care transitions and avoid recurrent hospitalizations.	7	PGP 1; PGP 3; PGP 10; PGP 5; PGP 7; PGP 4; PGP 9
Palliative Care Program	Examples include end of life care planning information, quality of life guidance, and alternative living options for terminally ill patients and their families.	5	PGP 2; PGP 3; PGP 1; PGP 4; PGP 10
Patient Registries	Development of patient registries as an administrative tool to improve care management processes.	2	PGP 1; PGP 9
Medication Reconciliation Program	A system either electronic or not that assists with tracking patient medications to avoid adverse events related to prescription medications.	1	PGP 1
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.	1	PGP 1
Medication Access Program ²	Assists low income patients with obtaining low cost prescription medications from pharmaceutical companies.	2	PGP 9; PGP 1
<i>Program Interventions: Disease Specific</i>			
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.	5	PGP 5, PGP 7; PGP 6; PGP 9; PGP 10
Tele-Management for Heart Failure or Other Conditions	An interactive voice response system that assists with care management of patients with congestive heart failure.	4	PGP 1; PGP 8; PGP 5; PGP 7
Diabetes Care Management	Program that assists patients living with diabetes to ensure proper self-care techniques and appropriate physician follow-up.	5	PGP 5; PGP 7; PGP 8; PGP 1; PGP 3
Anticoagulation Program/Clinic	Program that works with patients receiving anticoagulation therapy to ensure patient is receiving appropriate medications and to avoid any potential hospitalization.	4	PGP 7; PGP 1; PGP 3; PGP 6
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.	3	PGP 1; PGP 5; PGP 9

(continued)

Table 4-1 (continued)
Summary of programs implemented or enhanced at the groups participating in the PGP demonstration

Program Category/Name	Brief Description	# of Groups with Program	Groups ¹
Psychiatric Conditions	Programs that assist high risk patients with chronic psychiatric conditions.	2	PGP 1; PGP 10
Coronary Artery Disease (CAD) Management	Program that alerts physicians of required activities or services for CAD patients.	2	PGP 3; PGP 1
Cancer Care Management	Cancer patients are provided with coordinated cancer care.	2	PGP 1; PGP 7
Hypertension Disease Management Program	Reminder system for physicians to ensure that blood pressure is measured at each visit for hypertension patients.	1	PGP 3
<i>Program Interventions: General Care Management/Care Coordination</i> 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.	4	PGP 5; PGP 4; PGP 1; PGP 10
Homecare/Post Acute Care Services	Care to patients who have been discharged and are receiving some form of post acute care services.	2	PGP 7; PGP 10
Health Coaching	Care managers assist patients post-discharge with their care management activities.	1	PGP 2
Gold Star Population	Group identifies patients for further management based on diagnoses/comorbidities, costs and hospitalization.	1	PGP 2
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	1	PGP 10

NOTES:

¹ Interventions are not strictly defined. While groups may target the same disease in the case of program interventions, only the disease is the common target. Groups need not share the same intervention.

² The Medicare Part D program implemented in 2006 provides subsidies for prescription drug coverage for low income beneficiaries. Beneficiaries who do not qualify for Part D subsidies may still benefit from Medication Access Programs.

SOURCE: Summary of data collected by CMS/RTI from groups participating in the PGP Demonstration and the 2008 GAO (GAO-08-65) Report: "Care Coordination Programs Used in Demonstration Show Promise, But Wider Use of Payment Approach May be Limited." The GAO data was collected in 2007 using a questionnaire.

with the identification of patients with specific conditions and generate lists of patients that should be followed for some form of care management. Other groups utilized the patient registry as a central management system for patient care. The most comprehensive patient registry that was seen under the Demonstration was developed by PGP 9. At PGP 9 the Demonstration acted as a catalyst for the development of their registry, which was populated through automatic feeds from several different databases maintained throughout the system. The registry was then used to track patients and identify any gaps in care, ensuring that appropriate care is provided in a timely fashion.

The PGPs were at various stages of development on their electronic medical records (EMRs). A few PGPs had fully developed EMRs, whereas one PGP did not have any EMR. The remaining PGPs had some form of EMR, but still in development. Most PGPs said the EMR was used for facilitating their Demonstration interventions.

Information System Interventions. These types of interventions relate to the use of the registry or EMR to improve care provided to the beneficiary. Examples of the interventions referred to here are the visit planner at PGP 9, the iList (intervention list) at PGP 6 and alert systems built in to the electronic medical record at groups such as PGP 5 and PGP 6. The core role of these types of interventions is to recognize gaps in care and ensure that they are provided during the next patient encounter. They also prevent the occurrence of prolonged gaps in the future. The visit planner at PGP 9 was printed for the provider to review and use at the point of care. It listed the services required for the visit and reminded the physician and other providers to supply specific services during the encounter.

The iList is a software application at PGP 6 that was introduced as a mechanism for improving patient level care. The software allowed for the creation of lists of patients who may not be meeting specific quality metrics and that require additional follow-up. Finally, several groups, including PGP 5, PGP 1 and PGP 6 used an electronic medical record with automated alert systems to ensure that appropriate care is provided.

Medication Reconciliation. PGP 1 found that medication reconciliation activities need to be better integrated into the patient visit process. The process essentially consists of generating and maintaining an accurate list of medications taken by each individual patient. The definition of medication reconciliation and activities believed to be required for successful medication reconciliation varied across groups participating in the PGP Demonstration. However, there was general consensus that medication reconciliation is important for avoiding adverse outcomes, particularly in the Medicare population, which includes individuals that may be on several different medications.

Reporting/Feedback Mechanisms. Several groups have found that providing periodic feedback to physicians and staff on quality and performance measures has improved metrics throughout the systems or organizations. Specific examples of feedback mechanisms existed at several groups including PGP 2 where intranet feedback systems were in place for individual physicians to view their quality metrics online. Any issues with quality metrics or physician performance were reviewed carefully by management teams.

Physician and Staff Education. Education of physicians and other staff is important for the success of any project, including the PGP Demonstration. Informing all staff of what is happening within the organization increases and improves staff engagement. Groups found that open and clear communication was a key element for success under the Demonstration.

4.1.2 Patient Care Management Programs

In addition to process re-design and administrative type interventions, groups participating in the PGP Demonstration implemented several different clinically based care management programs that target specific patient populations. These programs can be classified as programs that target specific diseases or conditions or programs that target a subset of beneficiaries based on cost or patient complexity. The programs implemented by the groups participating in the PGP Demonstration are summarized in Table 4-1 and are discussed briefly below.

Disease Specific Programs. Groups participating in the PGP Demonstration focused somewhat on implementing disease related programs that would generate the most savings. Congestive heart failure care management programs fall in to this group. There are two types of congestive heart failure programs that were implemented by the groups. The first is a program that uses an interactive voice response system (IVRS) and allows for congestive heart failure patients to call in daily for monitoring of vitals (tele-management). The second program does not involve the IVRS but does involve patient follow-up and education services to help sustain self-management techniques. With both of these programs a total of eight of the ten groups participating in the PGP Demonstration had formal congestive heart failure related care management programs. Other conditions for which programs existed at the groups included: patients receiving anticoagulation therapy, diabetes, chronic obstructive pulmonary disease, cancer, psychiatric conditions, coronary artery disease and hypertension. Care management programs for these conditions most often involves education for self-management techniques and periodic patient follow-up and assistance with scheduling of appointments and coordination of care.

General Care Coordination Programs. Additional programs existed at several groups that casted a wider net for enrollment criteria, meaning that enrollment is not disease based. Several of these programs focused on patients with multiple conditions or patients that are high cost or high risk. The Gold Star Population at PGP 2 for example, was defined as a population that has three or more select comorbidities, seven or more evaluation and management visits or has been hospitalized with charges at \$10,000 or more. Once identified this population receives either complex care coordination or a more formal Health Coaching intervention. Groups such as the PGP 7 and PGP 10 may also have continued providing these types of care coordination services once a patient is discharged and is receiving home care services or other post acute care services.

4.2 Beneficiary Involvement in Site Programs

Each of the patient care management programs at the groups targeted a specific population either based on condition/disease or based on patient complexity defined by number or diagnoses or patient cost. Data studied in this section differ from information provided in previous sections. The prior sections discuss programs and processes implemented across PGPs.

Here we use enrollment data through PY1 for beneficiaries participating in these Demonstration care management programs provided by a subset of the PGPs. We investigated the Demonstration programs in more detail by tracking these enrolled beneficiaries. In this analysis, we blinded the names of the PGP sites because we provide counts of assigned beneficiaries that are also provided in later chapters. Our analysis focuses primarily on descriptive statistics.

For the seven groups that provided care management program enrollment data, approximately 10 percent of the assigned beneficiaries were in at least one care management program. **Table 4-2** shows the number of assigned beneficiaries for the seven PGPs and also shows counts for beneficiaries enrolled in one program, two programs, three programs, four programs, or five or more programs. While a small percentage of the assigned beneficiary population enrolled in at least one program, it is important to keep in mind that this portion of the population is a very costly group. Additionally, the enrollment numbers for programs may appear low; however, it is important to remember that data were only available to RTI through the end of PY1. Most groups have indicated that PY1 was more of a ramp up year for several of their programs and that enrollment numbers increased significantly in PY2 after processes for enrollment and more educational resources about programs were put in place.

Table 4-2
Assigned beneficiaries enrolled in care management programs, performance year one

PGP	Number of Assigned Beneficiaries	# of Assigned Beneficiaries in First Program	# of Assigned Beneficiaries in Second Program	# of Assigned Beneficiaries in Third Program	# of Assigned Beneficiaries in Fourth Program	# of Assigned Beneficiaries in Five + Program
PGP 6	42,017	1,920	—	—	—	—
PGP 9	32,646	1,041	139	46	9	1
PGP 2	29,211	3,451	289	23	—	—
PGP 7	17,822	422	81	5	—	—
PGP 4	14,018	248	14	—	—	—
PGP 1	13,809	876	119	—	—	—
PGP 3	9,313	3,482	2,152	850	206	38
Total	158,836	11,440	2,794	924	215	39
Percent	—	7.2%	1.8%	0.6%	0.1%	0.0%
Total (without PGP 3)	149,523	7,958	642	74	9	1
Percent	—	5.5%	.4%	.1%	0	0

SOURCE: Data from groups participating in the PGP Demonstration and RTI calculations using Medicare administrative data from 2004–2007.

From Table 4-2, it is also interesting to note that several assigned beneficiaries were enrolled in more than one care management program. For example, 46 beneficiaries assigned to PGP 9 in PY1 were enrolled in three of the care management programs offered by the group. This was most likely because the beneficiary had several comorbidities that required specific attention. It is also possible that a program referred to a specific level of care, for example, an inpatient care management program may be considered separately from an outpatient IVRS program. The prevalence of comorbidities increases the need for care coordination and more complex care management processes and procedures.

The care management programs enrolled different types of beneficiaries and could range in size from less than 10 Medicare FFS enrollees to greater than 2,000 enrollees. *Tables 4-3 and 4-4* provide an idea of the size of some of these programs and also shows how many beneficiaries assigned to the groups for the purposes of the Demonstration were actually participating in one of the disease specific care management programs. Data in these tables are limited to those groups/programs for which data was submitted to RTI/CMS on the beneficiaries enrolled in each program. Programs are also limited to those with over 100 Medicare beneficiaries actively enrolled.

4.2.1 Disease Specific Care Management Programs

There have been three major disease/condition focuses across the groups under the Demonstration that can be seen clearly in Table 4-3. The focus included programs related to congestive heart failure/coronary artery disease, diabetes and anticoagulation. Congestive heart failure was most likely a focus of care management programs due to the high cost of the condition and therefore the potential for cost reductions. The mean expenditures for the beneficiaries in these heart failure programs were over \$15,000 annually. The percent of beneficiaries that were hospitalized reached 47%. Therefore the key was to reduce costs by improving patient self-management and avoiding unnecessary hospitalization.

Reasoning for the focus on diabetes was slightly different. Several groups, in fact, reported through site visits and interviews that cost savings from diabetes care management would not be seen as immediately due to the nature of the condition. Groups felt that these savings would be witnessed after the PGP Demonstration period had ended. Nevertheless, there was a specific focus on diabetes in PY1. This was most likely due to the early emphasis on diabetes related quality measures. Characteristics of the typical diabetes population that were enrolled in the diabetes care management programs at various groups participating in the PGP Demonstration are provided in Table 4-3. It is interesting to note that on average the diabetes population is younger and less costly than other disease specific care management programs.

Comprehensive anticoagulation programs also existed at four of the groups participating in the Demonstration. While these programs existed prior to the start of the Demonstration, there was still significant increase in the number of beneficiaries enrolled in these programs during the Demonstration.⁸ The core idea behind the anticoagulation program is again to treat and monitor patients carefully in an outpatient setting and thus avoid costly hospitalizations.

⁸ See Table 4-5 below.

Table 4-3
Beneficiaries enrolled in selected¹ disease or condition specific care management programs, performance year one

Program Name/Description	Number of Assigned Beneficiaries Enrolled in Program	Mean Age	% Male	% Aged ²	% Disabled	% ESRD	Mean Expenditures	Mean Risk Score	% At Least 1 Discharge
<i>Hypertension Care Management</i>									
PGP 3	6,099	73	42%	86%	13%	1%	\$6,888	0.966	20.0%
<i>Anticoagulation Clinic</i>									
PGP 6	2,143	75	59%	93%	6%	1%	\$12,289	1.631	36%
PGP 1	846	75	54	92	8	1	11,461	1.313	35
PGP 3	819	75	46	89	11	0	12,803	1.743	37
Total	3,808	75	53%	91%	8%	1%	\$12,184	1.562	36%
<i>CHF/CAD Programs</i>									
PGP 3	1,788	74	57%	88%	12%	1%	\$8,154	1.183	25%
PGP 9	391	74	67	87	13	0	19,906	2.041	64
PGP 1	261	75	58	92	8	0	13,842	1.603	46
PGP 7	194	80	50	96	1	3	19,240	2.139	55
Total	2,634	76	58%	90%	9%	1%	\$15,285	1.742	47%
<i>Diabetes Care Management</i>									
PGP 3	1,635	70	46%	81%	19%	1%	\$6,948	1.072	19%
PGP 2	136	65	47	68	31	1	7,105	1.177	20
PGP 9	117	68	47	59	35	6	29,649	2.961	70
Total	1,888	68	47%	69%	28%	2%	\$14,567	1.737	36%
<i>Cancer Care Management</i>									
PGP 7	151	74	29%	95%	5%	0%	\$11,853	1.404	27%

NOTES:

1. Data in this table is limited to those groups/programs for which data was submitted to RTI/CMS on the beneficiaries enrolled in each program. Programs are also limited to those with over 100 Medicare beneficiaries actively enrolled.
2. Aged beneficiaries are those aged 65 or older.

SOURCE: Data from groups participating in the PGP Demonstration and RTI calculations using Medicare administrative data from 2004–2007.

Table 4-4
Beneficiaries enrolled in selected¹ general care management programs, performance year one

Program Name/Description	Group	Number of Assigned Beneficiaries Enrolled in Program	Mean Age	% Male	% Aged ²	% Disabled	% ESRD	Mean Expenditures	Mean Risk Score	% At Least 1 Discharge
<i>Health Coaching</i>	PGP 2	2,368	71	43%	81%	18%	1%	\$12,124	1.407	34%
<i>Gold Star Population - Location 1</i>	PGP 2	1,155	75	51	91	8	1	11,603	1.479	30
<i>Palliative Care Program</i>	PGP 3	535	81	35	95	4	1	15,071	1.922	40
<i>Hospital Transitional Care Program</i>	PGP 3	479	79	45	97	2	1	16,432	2.211	55
<i>Gold Star Population - Location 2</i>	PGP 2	439	73	49	82	15	3	22,746	2.312	42
<i>Homecare Program</i>	PGP 7	176	81	47	95	3	2	21,586	2.359	61
<i>COMPASS</i>	PGP 4	146	73	49	80	19	1	21,134	2.217	66
<i>SafeMed</i>	PGP 4	130	76	47	97	4	0	9,134	1.114	37
<i>Emergency Room Case Management</i>	PGP 9	109	73	41	74	24	2	14,257	1.627	46
<i>Total</i>	—	5,537	76	45%	88%	11%	1%	\$16,010	1.850	46%

NOTES:

1. Data in this table is limited to those groups/programs for which data was submitted to RTI/CMS on the beneficiaries enrolled in each program. Programs are also limited to those with over 100 Medicare beneficiaries actively enrolled.
2. Aged beneficiaries are those aged 65 or older.

SOURCE: Data from groups participating in the PGP Demonstration and RTI calculations using Medicare administrative data from 2004–2007.

4.2.2 Complex Care Management Programs

Additional programs that focus less on a specific disease and more on a population such as a high cost or complex population were also implemented at several groups or enhanced as part of the PGP Demonstration. Beneficiary enrollment characteristics for these programs can be found in Table 4-4. Because these programs typically focused on patients with multiple comorbidities the mean expenditures for the beneficiaries enrolled in these programs were very high (over \$16,000 on average) and the risk scores for this population were also very high (1.850 on average). The most costly subpopulation was found to be the Gold Star Population from the PGP 2 – Location 2 (over \$22,000). This is likely because the population there that was being enrolled in the program was a post-hospitalization, high cost/high risk population.

4.2.3 Beneficiary Turnover

An interesting design element of the PGP Demonstration was that beneficiaries were assigned to the groups retrospectively. Therefore, there were no real means of identifying beneficiaries assigned to the group and targeting specifically these beneficiaries for care management interventions during the Demonstration year. In general, it is estimated that for the PGP Demonstration approximately 60 percent of the beneficiaries assigned in one year were assigned in the next. This posed some concern because a beneficiary enrolled in a care management program one year may not be assigned the next and the full benefits of the care management program cannot be seen under the Demonstration.

In order to test the effect of turnover, we examined changes in the number of beneficiaries enrolled in care management programs from the base year to the first performance year. **Table 4-5** presents some information on the assigned stability of the beneficiaries that were enrolled in the care management programs. We present the number of assigned beneficiaries enrolled in each program only during the base year, only during performance year one, and during both years. On average, we see that about 70 percent of the beneficiaries enrolled in the care management programs were assigned in both the base year and in PY1. These results are not definitive as more data is required to fully understand the stability of the care management population as a group of assigned beneficiaries; however, based on these results it does appear that a significant portion of the care management population were assigned to groups throughout the PGP Demonstration years.

4.3 Factors Expected to Influence Financial Results

In addition to specific programs or process interventions, groups provided some information on what factors in general they expected would influence the financial results seen under the Demonstration. The factors expected to influence financial results can be classified as either clinical interventions or structural impacts.

Clinical Interventions. One of the key cost drivers for groups in the PGP Demonstration was avoidable hospitalizations (including avoidable readmissions) of assigned beneficiaries. A focus of clinical or care management interventions under the Demonstration was to improve self-management techniques to ensure medication or treatment compliance and reduce the likelihood

Table 4-5
Assigned beneficiary enrollment in care management programs from base year and performance year one

Program Name/Description	Group	Number of Beneficiaries in Program ¹	Assigned Beneficiaries in Program BY	Assigned Beneficiaries in Program PY1	Assigned Beneficiaries in Program BY and PY1
Hypertension Care Management	PGP 3	6,933	834	1,760	4,339
Health Coaching	PGP 2	2,368	0	418	1,950
Anticoagulation Clinic	PGP 6	2,143	223	295	1,625
CAD Disease Management	PGP 3	1,975	187	454	1,334
Diabetes Disease Management	PGP 3	1,800	165	459	1,176
Gold Star Population – Keene	PGP 2	1,155	0	123	1,032
Anticoagulation Clinic	PGP 3	935	116	227	592
Anticoagulation Clinic	PGP 1	846	93	135	618
Palliative Care Program	PGP 3	609	74	93	442
Hospital Transitional Care Program	PGP 3	522	43	106	373
Gold Star Population – Lebanon	PGP 2	475	36	108	331
CHF Management	PGP 9	430	39	122	269
CHF Tele-Assurance Management	PGP 1	408	47	73	288
CHF Inpatient Care Management	PGP 7	197	3	28	166
COMPASS	PGP 4	186	40	46	100
Homecare	PGP 7	182	6	25	151
Cancer Care Management	PGP 7	155	4	31	120
SafeMed	PGP 4	151	21	34	96
Diabetes Disease Management	PGP 2	136	0	26	110
Diabetes Care Management	PGP 9	127	10	28	89
Emergency Room Case Management	PGP 9	119	10	26	83
Diabetes Disease Management	PGP 7	42	4	8	30
Total	—	21,894	1,955	4,625	15,314
Percent	—	—	9%	21%	70%

NOTES:

¹ Total number of assigned beneficiaries with assignment in any year. Numbers will not match tables 4.2 or 4.3 which focus on PY1.

SOURCE: Data from groups participating in the PGP Demonstration and RTI calculations using Medicare administrative data from 2004–2007.

of avoidable hospitalization. Therefore, patient engagement also gains a central role to care management and ultimate cost savings.

Groups found that care management programs, particularly complex care management and transitional care management programs, assist with the reduction of re-hospitalization rates. Reduced hospitalization rates were intended to help considerably with meeting the cost savings goal of the PGP Demonstration. Internal analysis (not peer reviewed) at PGP 8 showed that with their congestive heart failure program, at least one admission is averted per beneficiary, per year.

Additionally, increased use of palliative care and hospice programs has the potential for reducing high cost care utilization during end stages of life. Groups increased their educational efforts around end of life issues to increase use and improve understanding of these kinds of programs.

Structural Impacts. One of the major structural or process changes that affected expenditures under the Demonstration involved coding initiatives. Several groups shared their complete coding initiatives under the PGP Demonstration. Groups worked with information technology systems and staff to ensure accurate and complete coding of diagnoses.

Structural impacts such as the opening of new clinics or addition of beds may also have impacted Demonstration financial results. Another factor, which came up at one of the groups, PGP 10, was the potential for transferring assigned beneficiary patients to less costly PGP 10 facilities to continue providing care to the assigned beneficiaries at a lower cost.

4.4 Factors Expected to Influence Quality Results

While factors the groups expected to influence financial results overlap considerably with factors that the groups expected to influence quality results, there are some additional interventions/processes implemented under the Demonstration that the groups expected would primarily impact quality. Some of the factors expected to influence quality results under the PGP Demonstration included physician behavior modification and performance feedback, information systems and measure specific strategies.

Physician Behavior Modification. Physician and staff behavior is central to improving quality of care. This is particularly true of physicians who have the largest influence on patient treatment and resource utilization. Methods for modifying physician's practice patterns included changes in work processes, encouraging physicians to consider the health of a panel of patients rather than individual patients and most importantly, feedback reports to improve care coordination and quality of care.

Another important behavior modification for physicians was to encourage them to delegate specific quality related tasks to health care staff such as the nurses and physician assistants to ensure that all required services are conducted in a timely manner during the visit and so that the physician can focus on the interaction with the patient regarding their condition. If each staff person is engaged in the patient's care and has responsibility for a quality related task, then quality measures will most likely improve throughout the system.

Enhancing Information Technology. Many groups mentioned enhancements to information technology systems as a key strategy for successfully meeting quality metrics under the Demonstration. Applications may have included disease or patient registries that identify and track specific beneficiary groups and provide physicians and staff with detailed reports on patient status. Automated reminders, when used sparingly, in an electronic medical record, are also useful information technology enhancements to avoid gaps in care and meet quality and performance measurements.

Measure Specific Strategies. Specific strategies were also implemented for some of the more difficult to obtain measures, such as the diabetic foot exam quality measure. Education about the diabetic foot exam and the components of the exam can be important for ensuring its completion. Some groups also created a dummy code in their system for the foot exam to avoid having to do a detailed chart review for the foot exam components during chart abstraction period for the PGP Demonstration. Another example is the creation of a dilated eye exam form to distribute to diabetes patients as well as eye care providers in the area, which would be subsequently returned to the group.

This page intentionally left blank.

CHAPTER 5 DEMONSTRATION FINANCIAL RESULTS

This chapter presents Demonstration financial results, including PGP savings and losses, performance payments to participating PGPs, and Medicare program savings. The chapter begins in Section 5.1 with a brief profile of the Demonstration PGP assigned beneficiary and comparison group populations including numbers over time, utilization at the participating PGP, and Medicare eligibility status. Then in Section 5.2 we discuss trends during the Demonstration period in expenditures, risk scores, and risk-adjusted expenditures, and in risk-adjusted assigned beneficiary versus target expenditures. Next in Section 5.3 the Demonstration financial reconciliation for performance years one (PY1) through five (PY5) is presented. The chapter concludes in Section 5.4 with a summary of Demonstration financial results: gross savings, performance payments and losses, and Medicare program savings. In this chapter's tables and figures, PGPs are sorted, unless otherwise specified, in numerical order of their designated blinded number.

5.1 Assigned Beneficiary and Comparison Group Profiles

This section is a brief profile of the characteristics of the participating PGPs and their comparison groups. These data are drawn from annual profile reports each PGP received to assist with gauging Demonstration progress. All of the data included in the profile reports are from the same data sources used to calculate Demonstration performance payments. While some data are presented for BY, PY1, PY2, PY3, and PY4, this section will give greater attention to PY5 data, which are the most recent data available and are representative of the other years.

Table 5-1 shows the number of assigned beneficiaries and comparison group beneficiaries at each PGP from BY through PY5. The total number of assigned beneficiaries in the Demonstration was stable in BY and PY1 at 223,203 and 223,893 respectively, slightly decreased to 219,577 in PY2, 219,725 in PY3, and 216,977 in PY4, and increased again to 220,779 in PY5. The number of assigned beneficiaries per PGP ranged from 10,257 to 35,081 in PY5. The total number of comparison group beneficiaries in the Demonstration decreased each year from 1,618,235 in BY to 1,375,158 in PY4, and then increased slightly to 1,384,347 in PY5. The number of comparison group beneficiaries per PGP ranged from 16,635 to 481,534 in PY5. In financial calculations, comparison group beneficiaries were weighted according to the proportion of assigned beneficiaries residing in each service area county. This weighting reduces the effective sample size of comparison group beneficiaries in terms of the variance of per capita expenditures. For example, a populous county with many comparison group beneficiaries might receive a low assigned beneficiary proportion weight, and its large sample size might not contribute proportionally to the comparison group per capita expenditures. Hence, the variance of the comparison group per capita expenditures would be greater than indicated by the nominal comparison group sample size. The implication is that the effective sample size of comparison group beneficiaries is lower than the counts shown in Table 5-1.

Table 5-1
Number of assigned and comparison group beneficiaries

PGP Name	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5
Total	223,203	223,893	219,577	219,725	216,977	220,779	1,618,235	1,605,052	1,524,031	1,456,216	1,375,158	1,384,347
PGP 1	14,347	13,809	13,424	13,760	13,825	14,397	17,969	19,096	17,572	17,142	16,614	16,635
PGP 2	28,107	29,211	30,646	31,284	32,127	33,429	101,911	108,614	115,784	114,070	107,099	109,052
PGP 3	8,383	9,313	9,715	10,035	10,184	10,257	157,640	163,736	162,714	162,275	161,680	163,708
PGP 4	14,688	14,018	13,997	15,533	15,285	15,733	107,270	98,474	93,233	87,943	72,652	85,224
PGP 5	25,767	25,709	25,412	26,707	25,908	26,979	172,283	169,419	162,641	157,635	139,877	135,702
PGP 6	44,609	42,017	38,743	37,026	35,148	35,081	51,595	48,048	45,370	39,211	36,474	39,605
PGP 7	17,551	17,822	17,668	16,991	16,489	16,342	231,434	229,964	220,539	210,204	203,063	200,432
PGP 8	19,034	18,843	19,027	18,950	18,562	17,908	137,393	126,411	120,596	117,730	111,935	107,491
PGP 9	31,233	32,646	31,706	31,435	31,408	31,237	50,927	49,949	46,603	45,813	44,977	44,964
PGP 10	19,484	20,505	19,239	18,004	18,041	19,864	589,813	591,341	538,979	504,193	480,787	481,534

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

As explained in Chapter 2, potential comparison group beneficiaries were excluded from the final comparison group for several reasons, including Medicare Advantage enrollment. **Table 5-2** presents the number of beneficiaries excluded from assignment due to Medicare Advantage (MA) partial year enrollment. The table also shows the number and percent of beneficiaries with at least one office or other outpatient E&M visit at a PGP who were excluded from the Demonstration due to MA enrollment. This ranged from 52 to 1,594 (0.1 percent to 3.8 percent) in BY and ranged from 527 to 4,582 (1.4 percent to 9.9 percent) in PY5.

Table 5-3 show PY5 demographic and eligibility characteristics of the assigned and comparison group beneficiaries. Further data regarding expenditures and risk scores are presented in the financial reconciliation section of this chapter.

PGPs provide a high level of E&M services to their assigned beneficiaries. The proportion of total E&M allowed charges received by assigned beneficiaries that were provided by participating PGPs ranged from 0.72 to 0.89 across beneficiaries served by the PGPs in PY5. The average number of E&M visits provided to beneficiaries ranged from 5.0 to 6.8. These data indicate that the participating PGPs had significant opportunities to manage and coordinate the care provided to their assigned beneficiaries. Comparison group beneficiaries had no utilization at the participating PGP (if potential comparison group beneficiaries did, they were excluded from the comparison group).

The percent of beneficiaries by Medicare and Medicaid eligibility status, by hospice status, and by count of CMS-HCCs shows a high degree of comparability between the assigned and comparison populations, although the higher counts of CMS-HCCs indicate that the assigned population may be in somewhat poorer health than the comparison population. The majority of assigned and comparison group beneficiaries are eligible for Medicare by reason of age. This percent of beneficiaries eligible by age ranged from 69.2 percent to 86.3 percent for the assigned beneficiaries and 72.8 percent to 84.2 percent for comparison group beneficiaries in PY5. Only a small percentage of beneficiaries are eligible for Medicare by reason of end stage renal disease (ESRD). In PY5 the ESRD population at each PGP ranged from 0.3 percent to 3.5 percent for assigned beneficiaries and 0.2 percent to 1.7 percent for comparison group beneficiaries.

Approximately two-thirds of beneficiaries in the Demonstration are diagnosed with at least one of the 71 Hierarchical Condition Categories (HCCs) used to calculate risks scores for the PGPs and the comparison groups. In PY5 the percent of beneficiaries with at least one HCC ranged from 66.6 percent to 78.4 percent for assigned beneficiaries and 62.9 percent to 72.4 percent for comparison group beneficiaries. The five most frequent HCCs for PGP-assigned beneficiaries during PY5 included the following: HCC19 Diabetes without Complications, HCC92 Specified Heart Arrhythmias, HCC108 Chronic Obstructive Pulmonary Disease, HCC80 Congestive Heart Failure, and HCC105 Vascular Disease.

Table 5-2
Beneficiaries excluded from PGP assignment due to Medicare advantage enrollment

PGP Name	BY*	PY1*	PY2*	PY3*	PY4*	PY5*	BY**	PY1**	PY2**	PY3**	PY4**	PY5**
PGP 1	52	652	684	491	616	607	0.2	3.1	3.4	2.4	3.0	2.9
PGP 2	73	156	483	777	833	780	0.2	0.3	1.0	1.5	1.6	1.4
PGP 3	372	626	784	796	853	832	3.2	4.6	5.4	5.3	5.6	5.2
PGP 4	995	2,108	1,387	1,413	1,999	1,122	3.8	7.9	5.4	5.1	7.1	4.0
PGP 5	417	2,165	2,510	1,665	2,887	1,752	1.0	5.0	5.8	3.7	6.4	3.6
PGP 6	1,594	3,165	3,414	2,751	3,434	2,600	2.7	5.5	6.4	5.4	6.9	5.3
PGP 7	53	194	676	932	805	527	0.2	0.8	2.7	3.8	3.3	2.2
PGP 8	667	1,960	1,850	1,759	1,897	2,301	2.6	7.2	6.8	6.4	6.9	8.4
PGP 9	749	1,790	2,050	2,032	1,828	1,860	1.7	3.8	4.4	4.4	4.0	4.1
PGP 10	56	602	4,360	3,614	1,521	4,582	0.1	1.4	9.7	8.7	3.9	9.9

NOTES:

* Number of Beneficiaries Excluded

** Percent of Excluded Beneficiaries with at Least One Evaluation and Management (E&M) Visit at the PGP

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Table 5-3
Assigned beneficiary and comparison group characteristics, performance year five

Beneficiary Characteristics	Assigned Beneficiaries Min ¹	Assigned Beneficiaries Max ¹	Assigned Beneficiaries Mean ¹	Comparison Group Min ¹	Comparison Group Max ¹	Comparison Group Mean ¹
Office or Other Outpatient E&M Services						
Mean Proportion of Allowed Charges for Office or Other Outpatient E&M Visits Provided at the PGP	0.72	0.89	0.84	—	—	—
Mean Number of Office or Other Outpatient E&M Visits at the PGP	5.0	6.8	5.6	—	—	—
Percent of Beneficiaries by Medicare Eligibility						
Age	69.2	86.3	78.9	72.8	84.3	79.5
Disability	13.4	27.3	20.2	15.2	25.5	19.8
ESRD	0.3	3.5	0.9	0.2	1.7	0.7
Percent of Beneficiaries by Medicaid Eligibility						
Medicaid Eligible for At Least One Month During the Year	12.0	24.7	17.1	12.5	26.0	17.3
Percent of Beneficiaries with Hospice Status						
Beneficiaries Enrolled in Hospice During the Year	1.5	2.7	1.9	1.2	2.2	1.7
Percent of Beneficiaries by Count of CMS-HCCs²						
5 or More CMS-HCCs	9.1	14.0	11.2	6.4	12.8	9.3
1 or More CMS-HCCs	66.6	78.4	72.0	62.9	72.4	67.9
0 CMS-HCCs	21.6	33.4	28.0	27.6	37.1	32.1

NOTES:

¹ Min, Max and Mean are across the 10 participating PGPs. Mean is unweighted.

² “Centers for Medicare and Medicaid Services Hierarchical Condition Categories”.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

5.2 Trends During the Demonstration Period in Expenditures, Risk Scores, Risk-Adjusted Expenditures, and Assigned Beneficiaries

This section examines trends in key Demonstration quantities—expenditures, risk scores, and their derivatives, risk-adjusted and target expenditures—during the Demonstration base and performance years. Demonstration financial reconciliation, which we discuss in Section 5.3, is based on these quantities. We first examine in Section 5.2.1 expenditures, risk scores and risk-adjusted expenditures for assigned and comparison group beneficiaries, then we consider in Section 5.2.2 a different combination of these factors, assigned beneficiary versus target expenditures.

5.2.1 Expenditures, Risk Scores, and Risk-Adjusted Expenditures

To adjust for health risk, the Demonstration uses a version of the CMS-HCC model implemented for Medicare Advantage risk adjustment. The PGP Demonstration risk adjustment model is concurrent—using diagnoses from the same year as expenditures—rather than prospective—using diagnoses from the year prior to expenditures, as is the Medicare Advantage model (Olmsted, Pope, and Kautter, 2006). The PGP Demonstration risk model was calibrated only once, using base year 2004 data, and so did not reflect any changes in medical treatment, expenditure, or diagnostic coding patterns over the course of the Demonstration. The model was, however, updated for new diagnosis and procedure codes for each year of the Demonstration. The risk adjustment model produces a “risk score” for each beneficiary. A risk score is an expenditure-weighted index of a beneficiary’s diagnoses. A higher risk score denotes poorer health status. The final average risk scores for the assigned and comparison group populations at each PGP are applied to the observed per capita expenditure growth rate to remove the effects of changes in health status in the performance payment calculations.

Tables 5-4a, 5-4b, 5-4c and *Tables 5-5a, 5-5b, and 5-5c* use these risk scores to present risk-adjusted expenditures where the average risk scores from the assigned and comparison group beneficiaries at each PGP have been applied to observed expenditures. All risk adjusted data are the observed expenditures divided by the average corresponding assigned or comparison group beneficiary mean risk score. *Tables 5-4a, 5-4b, and 5-4c* focus on overall expenditures while *Tables 5-5a, 5-5b, and 5-5c* present results for selected expenditure components.

Tables 5-4a, 5-4b, and 5-4c show the actual mean expenditures per beneficiary, risk scores, and risk-adjusted mean expenditures per beneficiary for the assigned and comparison group populations at each PGP from BY to PY5. In PY5 the mean risk score at the PGPs ranged from 0.958 to 1.491 for assigned beneficiaries and ranged from 0.781 to 1.147 for comparison group beneficiaries. On average the risk score growth rate from BY to PY5 was greater for assigned beneficiaries than comparison group beneficiaries. The five year risk score growth rate for assigned beneficiaries from BY to PY5 ranged from 11.7 percent to 25.5 percent with a simple average of 17.4 percent across all the PGPs. The five year risk score growth rate for comparison group beneficiaries ranged from 4.2 percent to 16.5 percent with a simple average of 10.0 percent across all ten PGPs locales.

Table 5-4a
Expenditures per beneficiary

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Assigned Beneficiaries Growth BY to PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5	Comparison Group Growth BY to PY5	CG Minus AB Growth
PGP 1	\$6,640	\$7,282	\$7,642	\$8,106	\$8,502	\$8,496	28.0%	\$5,673	\$6,212	\$6,498	\$6,609	\$6,944	\$6,949	22.5%	-5.5%
PGP 2	7,959	8,799	9,274	9,604	10,132	10,010	25.8	5,536	6,069	6,433	6,566	6,802	6,567	18.6	-7.1
PGP 3	6,843	7,052	7,393	7,845	8,344	8,398	22.7	6,969	7,175	7,417	7,693	7,938	7,999	14.8	-7.9
PGP 4	6,067	6,590	7,141	7,143	7,456	7,619	25.6	7,125	7,785	8,026	8,294	8,693	8,449	18.6	-7.0
PGP 5	6,840	7,602	7,958	8,618	9,198	9,327	36.3	6,431	6,844	7,197	7,494	7,658	7,621	18.5	-17.8
PGP 6	6,576	7,096	7,506	7,774	8,440	8,296	26.1	5,437	5,999	6,454	6,713	7,256	7,016	29.0	2.9
PGP 7	7,602	8,142	8,832	9,555	9,899	10,106	32.9	7,785	7,992	8,426	8,789	9,184	9,366	20.3	-12.6
PGP 8	6,320	6,938	7,077	7,609	7,945	7,907	25.1	6,836	7,274	7,652	8,048	8,516	8,685	27.0	1.9
PGP 9	6,347	6,960	7,355	7,672	7,691	7,882	24.2	6,098	6,573	6,940	7,257	7,579	7,410	21.5	-2.7
PGP 10	11,341	11,916	12,853	13,584	13,662	13,770	21.4	7,875	8,587	8,951	9,372	9,842	9,893	25.6	4.2

Table 5-4b
Risk scores per beneficiary

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Assigned Beneficiaries Growth BY to PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5	Comparison Group Growth BY to PY5	CG Minus AB Growth
PGP 1	0.871	0.876	0.902	0.966	0.998	0.985	13.1%	0.772	0.801	0.827	0.849	0.864	0.862	11.7%	-1.4%
PGP 2	0.991	1.042	1.086	1.095	1.118	1.109	11.9	0.735	0.758	0.764	0.770	0.783	0.781	6.3	-5.6
PGP 3	0.972	0.972	1.011	1.039	1.071	1.086	11.7	0.936	0.930	0.938	0.975	0.981	0.976	4.2	-7.5
PGP 4	0.850	0.900	0.937	0.930	0.938	0.958	12.8	0.990	1.047	1.051	1.076	1.086	1.067	7.7	-5.1
PGP 5	1.026	1.081	1.103	1.205	1.265	1.258	22.6	0.958	0.961	0.996	1.008	1.015	1.008	5.2	-17.3
PGP 6	0.916	0.960	1.004	1.055	1.141	1.149	25.5	0.757	0.780	0.825	0.841	0.874	0.867	14.4	-11.1
PGP 7	0.904	0.964	1.023	1.049	1.053	1.111	22.9	0.940	0.941	0.975	1.012	1.004	1.043	10.9	-12.0
PGP 8	0.821	0.869	0.872	0.923	0.940	1.011	23.1	0.888	0.906	0.932	0.971	1.011	1.035	16.5	-6.6
PGP 9	0.897	0.930	0.988	1.025	1.027	1.044	16.4	0.898	0.934	0.968	0.972	0.990	0.989	10.1	-6.3
PGP 10	1.308	1.328	1.415	1.500	1.483	1.491	14.0	1.018	1.036	1.075	1.120	1.141	1.147	12.7	-1.3

Table 5-4c
Risk-adjusted mean expenditures¹ per beneficiary and risk scores

PGP	Assigned	Assigned	Assigned	Assigned	Assigned	Assigned	Assigned	Comparison	Comparison	Comparison	Comparison	Comparison	Comparison	Comparison	CG
	Benefi- ciaries BY	Benefi- ciaries PY1	Benefi- ciaries PY2	Benefi- ciaries PY3	Benefi- ciaries PY4	Benefi- ciaries PY5	Benefi- ciaries Growth BY to PY5								
PGP 1	\$7,623	\$8,315	\$8,476	\$8,389	\$8,522	\$8,623	13.13%	\$7,350	\$7,754	\$7,854	\$7,781	\$8,034	\$8,059	9.66	-3.47%
PGP 2	8,036	8,446	8,536	8,769	9,060	9,027	12.34	7,538	8,010	8,420	8,527	8,684	8,410	11.58	-0.75
PGP 3	7,040	7,253	7,310	7,550	7,794	7,735	9.88	7,442	7,717	7,904	7,894	8,088	8,197	10.14	0.26
PGP 4	7,140	7,323	7,624	7,684	7,949	7,950	11.33	7,195	7,432	7,639	7,704	8,007	7,918	10.05	-1.28
PGP 5	6,667	7,031	7,212	7,152	7,269	7,416	11.24	6,716	7,118	7,224	7,432	7,544	7,563	12.61	1.37
PGP 6	7,182	7,391	7,479	7,371	7,399	7,217	0.48	7,178	7,694	7,823	7,986	8,298	8,095	12.78	12.30
PGP 7	8,406	8,451	8,635	9,110	9,401	9,094	8.19	8,278	8,489	8,642	8,689	9,146	8,983	8.52	0.32
PGP 8	7,699	7,981	8,114	8,241	8,451	7,822	1.60	7,701	8,024	8,207	8,284	8,420	8,394	9.00	7.41
PGP 9	7,072	7,484	7,446	7,484	7,485	7,547	6.72	6,791	7,039	7,173	7,465	7,658	7,496	10.38	3.66
PGP 10	8,673	8,971	9,086	9,055	9,212	9,237	6.51	7,738	8,288	8,326	8,370	8,622	8,625	11.47	4.96

NOTES:

¹ Risk adjusted expenditures are defined as expenditures divided by risk score.

Does not reflect PY5 Demonstration risk score cap.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Table 5-5a
Selected risk-adjusted¹ expenditure components per beneficiary—Inpatient expenditures (facility only)

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Assigned Beneficiaries Growth BY to PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5	Comparison Group Growth BY to PY5	CG Minus AB Growth
PGP 1	\$3,102	\$3,269	\$3,219	\$3,142	\$3,262	\$3,241	4.49%	\$3,099	\$3,279	\$3,248	\$3,211	\$3,304	\$3,322	7.21%	2.72%
PGP 2	3,602	3,704	3,749	3,835	3,946	3,673	2.00	3,167	3,153	3,294	3,366	3,396	3,043	-3.90	-5.89
PGP 3	3,095	2,955	3,069	3,262	3,266	2,991	-3.36	3,126	3,191	3,167	3,278	3,271	3,201	2.43	5.79
PGP 4	3,348	3,342	3,495	3,333	3,429	3,356	0.23	3,366	3,513	3,442	3,431	3,556	3,619	7.51	7.29
PGP 5	3,157	3,404	3,529	3,291	3,273	3,018	-4.42	3,005	3,168	3,178	3,266	3,128	3,000	-0.18	4.24
PGP 6	3,155	3,208	3,197	3,044	3,046	2,803	-11.19	3,151	3,324	3,176	3,261	3,333	3,081	-2.23	8.95
PGP 7	3,296	3,283	3,346	3,733	3,759	3,564	8.11	3,586	3,493	3,589	3,835	3,885	3,661	2.10	-6.01
PGP 8	3,498	3,526	3,528	3,560	3,626	3,133	-10.45	3,719	3,873	3,976	4,005	4,089	3,900	4.87	15.32
PGP 9	3,165	3,343	3,199	3,112	3,003	2,914	-7.93	3,029	3,078	3,034	3,102	3,134	2,958	-2.33	5.59
PGP 10	4,744	4,628	4,634	4,780	4,763	4,691	-1.12	3,472	3,656	3,607	3,608	3,677	3,512	1.16	2.27

39

Table 5-5b
Selected risk-adjusted¹ expenditure components per beneficiary—Outpatient² expenditures

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Assigned Beneficiaries Growth BY to PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5	Comparison Group Growth BY to PY5	CG Minus AB Growth
PGP 1	\$1,868	\$2,249	\$2,409	\$2,513	\$2,654	\$2,823	51.07%	\$1,213	\$1,362	\$1,397	\$1,445	\$1,564	\$1,679	38.43%	-12.64%
PGP 2	1,875	2,106	2,254	2,411	2,632	2,851	52.06	1,291	1,567	1,740	1,825	1,969	2,125	64.59	12.53
PGP 3	606	672	752	810	886	1,042	71.88	1,206	1,284	1,372	1,421	1,521	1,647	36.49	-35.39
PGP 4	1,029	1,043	1,108	1,224	1,424	1,393	35.43	1,008	1,092	1,145	1,231	1,353	1,377	36.66	1.23
PGP 5	1,222	1,321	1,358	1,490	1,602	1,761	44.17	823	904	953	998	1,087	1,140	38.43	-5.73
PGP 6	858	1,001	1,052	1,136	1,177	1,263	47.24	1,162	1,396	1,683	1,816	1,951	2,131	83.30	36.06
PGP 7	1,071	1,102	1,138	1,153	1,231	1,265	18.12	1,013	1,095	1,102	1,129	1,272	1,334	31.65	13.52
PGP 8	1,098	1,238	1,345	1,382	1,517	1,496	36.33	1,023	1,148	1,207	1,243	1,363	1,523	48.96	12.63
PGP 9	1,124	1,305	1,410	1,570	1,751	1,951	73.59	1,097	1,267	1,353	1,503	1,637	1,663	51.67	-21.92
PGP 10	2,134	2,273	2,310	2,311	2,498	2,624	22.96	1,087	1,211	1,205	1,226	1,265	1,358	24.93	1.97

Table 5-5c
Selected risk-adjusted¹ expenditure components per beneficiary—Part B Physician/Supplier

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Assigned Beneficiaries Growth BY to PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5	Comparison Group Growth BY to PY5	CG Minus AB Growth
PGP 1	\$1,701	\$1,788	\$1,725	\$1,691	\$1,670	\$1,688	-0.77%	\$2,141	\$2,108	\$2,103	\$2,073	\$2,124	\$2,188	2.20%	2.97%
PGP 2	1,649	1,623	1,598	1,596	1,594	1,613	-2.18	1,857	1,918	1,933	1,898	1,959	1,948	4.90	7.09
PGP 3	2,483	2,589	2,515	2,468	2,533	2,710	9.16	2,342	2,425	2,436	2,366	2,388	2,434	3.93	-5.23
PGP 4	1,888	1,980	2,015	2,030	2,064	2,020	7.02	1,999	2,036	2,138	2,134	2,217	2,256	12.89	5.87
PGP 5	1,606	1,653	1,670	1,638	1,671	1,789	11.39	2,012	2,107	2,117	2,143	2,200	2,245	11.55	0.16
PGP 6	2,369	2,315	2,328	2,254	2,305	2,415	1.93	1,893	1,842	1,792	1,760	1,845	1,756	-7.27	-9.20
PGP 7	2,596	2,654	2,685	2,708	2,789	2,790	7.47	2,454	2,611	2,647	2,605	2,735	2,794	13.83	6.35
PGP 8	2,135	2,191	2,235	2,197	2,267	2,156	0.98	2,175	2,276	2,304	2,263	2,243	2,257	3.76	2.77
PGP 9	1,936	2,006	1,924	1,863	1,814	1,787	-7.69	1,731	1,737	1,787	1,784	1,828	1,884	8.86	16.55
PGP 10	1,621	1,711	1,719	1,747	1,760	1,871	15.43	2,312	2,486	2,496	2,470	2,518	2,547	10.15	-5.28

NOTES:

¹ Risk adjusted expenditures are defined as expenditures divided by risk score.

² Hospital and other institutional outpatient.

Does not reflect PY5 Demonstration risk score cap.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

CMS was concerned that the greater growth of assigned beneficiary than comparison group risk scores reflected more complete coding of diagnoses by the participating PGPs for their patients in response to the financial incentives of the Demonstration, rather than reflecting changes in the health status of the assigned relative to the comparison population as intended. The Demonstration financial results could be affected by differential trends in assigned beneficiary versus comparison group diagnostic coding, in the form of higher performance payments to participating PGPs. Therefore, in PY5, CMS implemented a risk score cap. The risk score cap restricted the PY5 “relative risk ratio” for each PGP to lie between 0.90 and 1.10. The risk ratio for assigned beneficiaries is the ratio of their PY5 to BY risk scores. The risk ratio for comparison beneficiaries is defined analogously. The relative risk ratio is the ratio of the risk ratios for assigned and comparison beneficiaries. Two PGPs were affected by the PY5 risk score cap. As a result of the cap, PGP 5’s assigned beneficiary risk ratio was reduced from 1.226 to 1.158, and PGP 7’s assigned beneficiary risk ratio was reduced from 1.229 to 1.220. These reductions are equivalent to a reduction in PGP 5’s PY5 assigned beneficiary risk score from 1.258 to 1.188, and a reduction in PGP 7’s PY5 assigned beneficiary risk score from 1.111 to 1.103. The Demonstration PY5 risk score cap is not reflected in Tables 5-4 or 5-5; the actual rather than reduced PY5 risk scores are presented and utilized in the calculations for these tables. But the risk score cap does affect (for PY5 only) the assigned beneficiary versus target expenditure trends, and the financial reconciliation, discussed in the next sections.

Risk-adjusted mean expenditures per assigned beneficiary increased by a smaller percentage than those of comparison group beneficiaries at seven of the ten PGPs from BY to PY5. From BY to PY5 risk-adjusted mean expenditures increased by a range of 0.48 percent to 13.13 percent for assigned beneficiaries and 8.52 percent to 12.78 percent for comparison group beneficiaries. On average assigned beneficiary risk-adjusted expenditures increased by 2.48 percentage points less than the comparison group beneficiaries. For PY5, risk-adjusted mean expenditures ranged from \$7,217 to \$9,237 for assigned beneficiaries and from \$7,496 to \$8,983 for comparison group beneficiaries.

Table 5-5 presents selected components of risk-adjusted mean expenditures per beneficiary.⁹ Eight PGPs have a smaller percent increase in risk-adjusted inpatient expenditures per assigned beneficiary than their comparison group beneficiaries from BY to PY5. On average, mean risk-adjusted inpatient expenditures per assigned beneficiary decreased by 4.0 percentage points compared to those of the comparison group beneficiaries. PGP 8’s inpatient expenditures decreased by 15.32 percentage points compared to its comparison group.

Hospital outpatient mean risk-adjusted expenditures increased greatly from BY to PY5, with growth rates ranging from 18.12 percent to 73.59 percent for assigned beneficiaries, and 24.93 percent to 83.30 percent for comparison group beneficiaries. Six PGPs had a smaller percent increase in hospital outpatient expenditures for assigned beneficiaries than their comparison group beneficiaries. PGP 6’s outpatient expenditures grew 36.06 percent less than its comparison group’s, contributing substantially to its overall financial performance.

⁹ The risk adjustment model is calibrated on total expenditures. The total expenditure risk scores are applied here to risk adjust component expenditures for comparability of results with risk-adjusted total expenditures (the same risk score is used), and because risk scores specific to component expenditures are not available.

Conversely, PGP 3's outpatient expenditures grew by 35.39 percentage points more than its comparison group's. Risk-adjusted Part B Physician/Supplier expenditures per beneficiary grew less quickly for assigned beneficiaries than comparison group beneficiaries at seven PGPs. PGP 9's physician/supplier expenditures declined, and grew by 16.55 percentage points less than its comparison group's.

5.2.2 Assigned Beneficiary Actual Versus Target Expenditures

Demonstration target expenditures are determined by base year assigned beneficiary expenditures, comparison group expenditure growth rates from base to performance years, and the change in assigned beneficiary versus comparison group risk scores from base to performance year. In this section we examine trends in assigned beneficiary actual versus target expenditures during the Demonstration period. The Demonstration financial reconciliation, performance payment, and savings calculations presented in Sections 5.3 and 5.4 are based on the difference between assigned beneficiary and target expenditures.

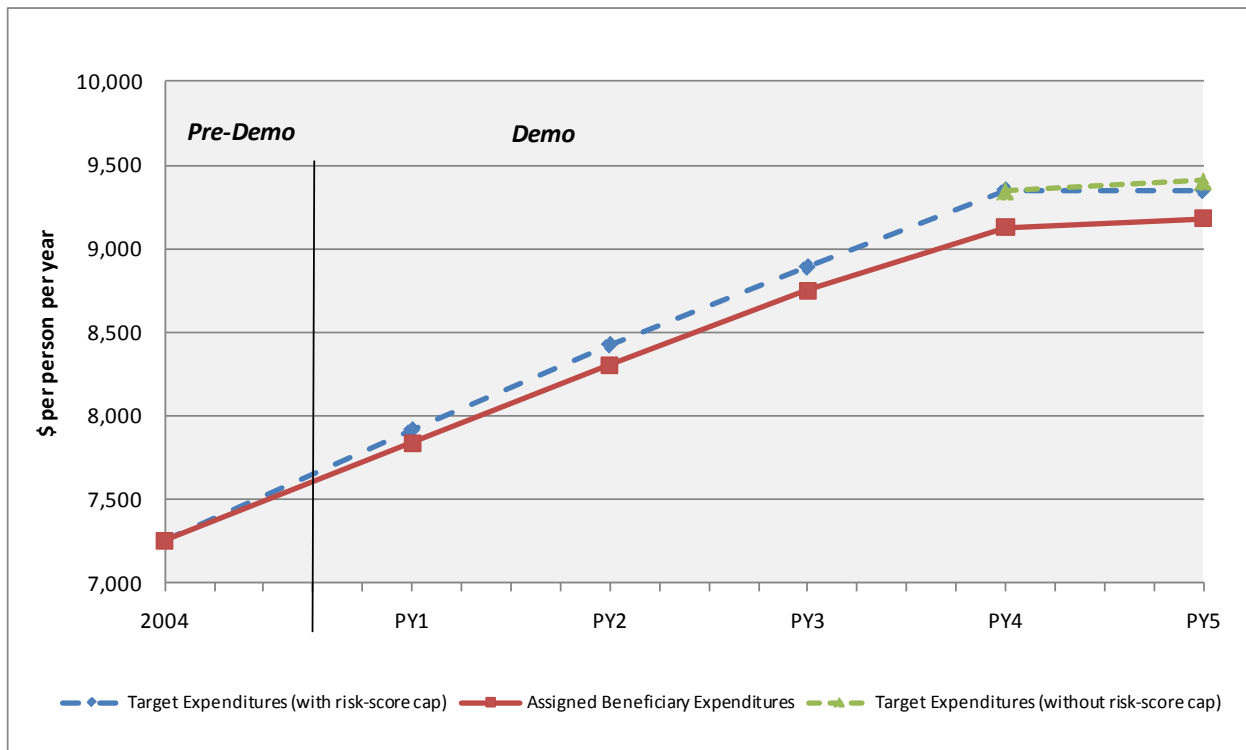
Figure 5-1 shows assigned beneficiary actual versus Demonstration target expenditures from 2004 through PY5. The data are unweighted averages of the 10 participating PGPs. In PY5, both the Demonstration target reflecting the Demonstration PY5 risk score cap and a simulated target not reflecting the risk score cap are shown. Both actual expenditures and target expenditures increase steadily from 2004 through PY4 with actual expenditures below target expenditures in each performance year and by an increasing amount over time. It is not known why this upward trend flattens in PY5 and the gap between assigned beneficiary and Demonstration target expenditures narrows. Demonstration target expenditures per person per year, reflecting the risk score cap, decreased from \$9,432 in PY4 to \$9,399 in PY5. Without the risk score cap, which affected only two PGPs, simulated target expenditures increased to \$9,468 in PY5.

Figure 5-2 shows assigned beneficiary versus Demonstration target expenditures from 2004 through PY5, by PGP. With the exception of PGP 6 and PGP 10, the individual graphs show assigned beneficiary expenditures tracking target expenditures fairly closely, with fluctuations for some years. For PGP 6 and PGP 10, assigned beneficiary expenditures are less than target expenditures in each of the five Demonstration performance years. These two PGPs earned performance payments in all five performance years of the Demonstration.

5.3 Financial Reconciliation by Performance Year

Using the expenditures and risk scores presented in the preceding section of this chapter and the quality measure results shown in Chapter 6, RTI calculated PY1 through PY5 performance payments. PY1 performance payments are shown in *Table 5-6*, PY2 results are shown in *Table 5-7*, PY3 results are shown in *Table 5-8*, PY4 results are shown in *Table 5-9*, and PY5 results are shown in *Table 5-10*. Each PGP that earned a performance payment is eligible to earn up to 80 percent of their target expenditures minus actual expenditures in excess of the 2 percent corridor. This amount is shown in Column D, "Shared Savings" of Tables 5-6 through 5-10. A detailed description of the PGP Demonstration performance payment methodology is given in Chapter 2.

Figure 5-1
Assigned beneficiary versus demonstration target expenditures, 2004 base year to PY5,
all 10 PGPs



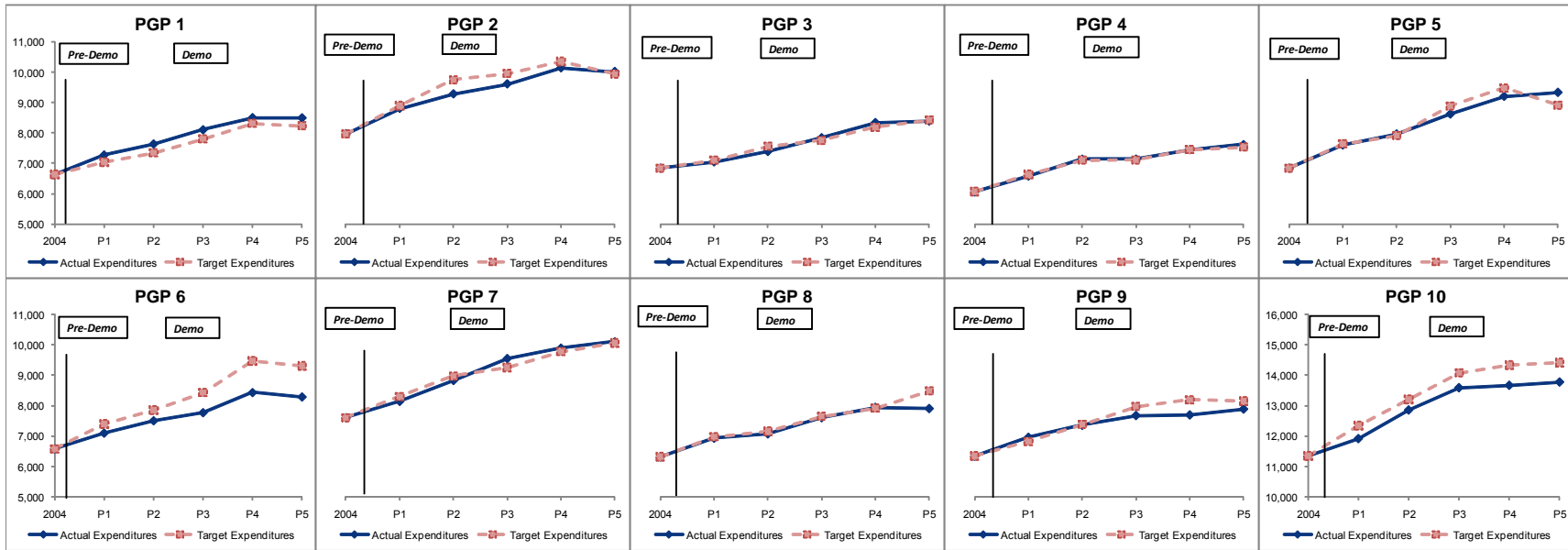
NOTES:

Unweighted average across the 10 PGPs.

The Demonstration risk score cap was in place for PY5 only.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 5-2
Assigned beneficiary versus demonstration target expenditures, by PGP, 2004 base year to PY5



NOTES:

Target expenditures reflect the risk score cap put in place for PY5.

The vertical scale is the same for all PGPs except PGP 10, for which it is higher.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Performance Year One

The PY1 through PY5 Demonstration financial results are presented graphically as a vertical bar chart in *Figures 5-3* through *5-7*. The height of the vertical bar for each PGP in the Figures is gross savings or loss (target minus actual expenditures) as a percentage of target expenditures. The minimum savings and loss corridor is shown in the Figures as two dashed horizontal lines at plus 2 percent and minus 2 percent. The amount of the PGP performance payment/bonus or loss, if any, is shown as a solid segment of the vertical bar for each PGP. Only amounts outside the 2 percent corridor were eligible for performance payments or accrued losses¹⁰. The PGP bonus did not equal the gross savings minus the 2 percent minimum savings threshold because a portion of savings were shared with the Medicare program and because if the PGP's quality score was less than the maximum score, the PGP's bonus was reduced. The PGP loss, on the other hand, was simply the gross loss minus the negative 2 percent minimum loss threshold. PGPs were not at risk for losses in the Demonstration, but accrued losses were used to offset future savings. In Figures 5-3 through 5-7, PGPs are not presented in numerical order, but are sorted from left to right horizontally by the size of their savings in the performance year. The horizontal order of the PGPs changes from year to year according to their savings/loss performance in particular years.

In PY1, two PGPs earned a performance payment by having target expenditures minus actual expenditures in excess of the 2 percent corridor. *Table 5-6* summarizes the performance payments calculations for PY1. Six PGPs had actual expenditures less than their target expenditures, but not over the 2 percent corridor. The remaining two PGPs had actual expenditure in excess of target expenditures and below the 2 percent corridor, therefore accruing losses carried forward. In these calculations target expenditures are the total risk-adjusted base year assigned beneficiary expenditures trended forward by the risk-adjusted expenditure growth rate of the comparison group from BY to PY1. Actual expenditures are the observed non-risk-adjusted expenditures at the PGP during PY1. The 2 percent corridor is 2 percent of target expenditures. In total the ten PGPs' actual expenditures were \$21 million less than target expenditures. *Figure 5-3* shows gross savings (height of the bars) for the 10 PGP sites. PGP 6 and PGP 10 each had target minus actual expenditures in excess of the 2 percent corridor and are shown earning a bonus in PY1. PGPs 1 and 9 each had target minus actual expenditures less than the minus 2 percent corridor, and are shown accruing a loss in PY1.

In PY1 each PGP kept 70 percent of shared savings as their performance payment for efficiency. The other 30 percent of shared savings represents the performance payment for quality. The amount of the performance payment for quality received was based on the percentage of quality targets met by the PGP. In PY1, PGP 6 hit 82 percent of the quality targets and therefore only received 82 percent of their maximum performance payment for quality. PGP 10 met 95 percent of the quality targets.¹¹ This resulted in final earned performance payments of

¹⁰ In chapter 2, amounts outside the 2 percent corridor are termed "Net Medicare savings".

¹¹ Further quality measure results are in chapter 6.

Table 5-6
Physician group practice demonstration performance payments calculation summary, performance year one

PGP Name	[A] Total Target Minus Actual Expenditures	[B] 2% Corridor (+/-)	[C] Expenditures Net 2% Corridor ¹	[D] Shared Savings	[E] Performance Payment for Efficiency	[F] Maximum Performance Payment for Quality	[G] Earned Performance Payment	[H] Payment at Annual Settlement
Total	\$20,955,837	—	\$8,001,192	\$7,625,215	\$5,337,651	\$2,287,565	\$7,323,697	\$5,492,773
PGP 1	-3,205,230	1,888,819	-1,316,410	0	0	0	0	0
PGP 2	2,763,374	5,038,554	0	0	0	0	0	0
PGP 3	407,850	1,274,756	0	0	0	0	0	0
PGP 4	649,289	1,809,372	0	0	0	0	0	0
PGP 5	958,992	3,798,915	0	0	0	0	0	0
PGP 6	12,062,878	6,026,989	6,035,889	4,828,711	3,380,098	1,448,613	4,565,327	3,423,995
PGP 7	2,826,924	2,882,206	0	0	0	0	0	0
PGP 8	656,929	2,559,899	0	0	0	0	0	0
PGP 9	-4,533,209	4,319,292	-213,917	0	0	0	0	0
PGP 10	8,368,038	4,872,408	3,495,630	2,796,504	1,957,553	838,951	2,758,370	2,068,777

NOTES:

¹ Accrued loss carried forward for PGP 9 is -\$213,917. Accrued loss carried forward for PGP 1 is -\$1,316,410.

[A] = [Total Target Expenditures] - [Total Actual Expenditures]. Target Expenditures and Actual Expenditures are determined using the PGP Demonstration Bonus Methodology.

[B] = [Total Target Expenditures] x [2%]. Total Target Expenditures are determined using the PGP Demonstration Bonus Methodology. The 2% Corridor is also known as the Savings Threshold.

[C] = Amount of Total Target Minus Actual Expenditures [A] that is outside the 2% corridor.

[D] = [C] x [80%]. The PGP Performance Payment Pool Sharing rate is set at 80%. Shared Savings are also referred to as the Bonus Pool.

[E] = [D] x [70%]. The share of the performance payment for efficiency is set at 70% for performance year one.

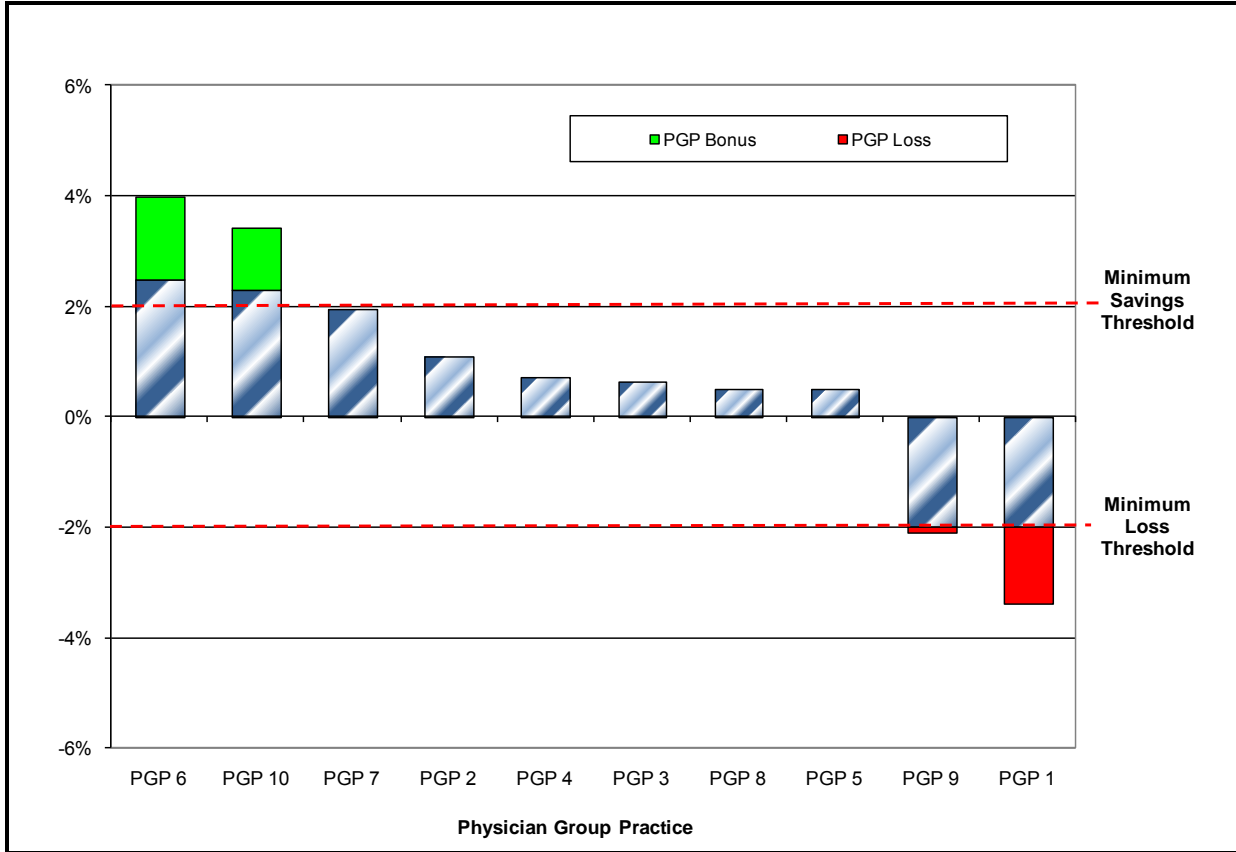
[F] = [D] x [30%]. The share of the performance payment for quality is set at 30% for performance year one.

[G] = [E] + [F]. This Earned Performance Payment takes into account actual percent quality targets achieved.

[H] = [G] x [75%]. The annual withhold is set at 25%.

SOURCE: RTI calculations with 2004–2006 Medicare claims and enrollment data.

Figure 5-3
PY1 target minus actual expenditures as a percentage of target expenditures, by PGP



NOTES:

1. The total height of the bar represents gross savings or loss (target minus actual expenditures).
2. The PGP bonus does not equal the gross savings minus the minimum savings threshold because of savings shared with the Medicare program and quality performance incentives.
3. PGPs were not at risk for “losses” during the Demonstration.

SOURCE: RTI calculations with 2004-PY1 Medicare claims and enrollment data.

\$4,565,327 for PGP 6 and \$2,758,370 for PGP 10. Each PGP was paid 75 percent of the final PY1 earned performance payments when the PY1 financial reconciliation was calculated. The remaining 25 percent was withheld until the end of the Demonstration to protect Medicare against losses the PGP may have generated in subsequent years. In PY1 \$5.5 million was paid to the two PGPs that earned performance payments.

Performance Year Two

As shown in *Table 5-7*, four PGPs earned a performance payment by having target expenditures minus actual expenditures in excess of the 2 percent corridor in PY2. Three PGPs had actual expenditures less than target expenditures, but did not exceed the 2 percent corridor.

Two PGPs had actual expenditures greater than target expenditures, but not below the 2 percent corridor. The remaining PGP had actual expenditures in excess of target expenditures and below payments are illustrated in **Figure 5-4** showing the gross savings (height of the bars), PGP bonus, and PGP loss of each site, as applicable. In these calculations target expenditures are the total risk-adjusted base year assigned beneficiary expenditures trended forward by the risk-adjusted expenditure growth rate of the comparison group from BY to PY2. In total the 10 PGPs' actual expenditures were \$34.5 million less than target expenditures.

The four PGPs that earned PY2 performance payments generated a total of \$13.9 million in shared savings. Following the Demonstration design, in PY2 the portion of shared savings kept by the PGP as the performance payment for efficiency decreased to 60 percent, down from 70 percent in PY1. The performance payment for quality was correspondingly increased to 40 percent in PY2. In PY2, both PGP 6 and PGP 10 hit 100 percent of their quality targets therefore earning the maximum performance payment for quality. PGP 2 and PGP 3 hit 98 percent and 96 percent of quality targets respectively. This resulted in final earned performance payments of \$6,689,879 for PGP 2, \$5,781,573 for PGP 6, \$1,239,294 for PGP 10, and \$129,268 for PGP 3. Each PGP was paid 75 percent of the final PY2 earned performance payments when the PY2 financial reconciliation was calculated. The remaining 25 percent was withheld along with the withholds from PY1 until the end of the Demonstration. In PY2 \$10.4 million was paid out to the four PGPs that earned a performance payment.

Performance Year Three

In PY3 five PGPs earned a performance payment by having target expenditures minus actual expenditures in excess of the 2 percent corridor (**Table 5-8**). One PGP had actual expenditures less than target expenditures, but not over the 2 percent corridor. Two PGPs had actual expenditures greater than target expenditures, but not below the 2 percent corridor. The remaining two PGPs had actual expenditures in excess of target expenditures and below the 2 percent corridor, therefore accruing losses carried forward. In these calculations target expenditures are the total risk-adjusted base year assigned beneficiary expenditures trended forward by the risk-adjusted expenditure growth rate of the comparison group from BY to PY3. In total the 10 PGPs' actual expenditures were \$49.0 million less than target expenditures. **Figure 5-5** shows the gross savings of all 10 PGP sites (the height of the bars); the bonuses of PGP 6, PGP 9, PGP 2, PGP 10, and PGP 5; and the losses of PGP 7 and PGP 1.

The five PGPs that earned performance payments generated a total of \$25.7 million in shared savings. Per the Demonstration design, in PY3 the portion of the shared savings kept by the PGP as performance payment for efficiency decreased to 50 percent, down from 60 percent in PY2. The performance payment for quality was correspondingly increased to 50 percent in PY3. PGP 5 hit 100 percent of the quality targets therefore earning the maximum performance payment for quality. PGP 6, PGP 2, PGP 9, and PGP 10 hit 98 percent, 92 percent, 96 percent, and 94 percent of quality targets respectively. This resulted in final earned performance payment of \$13,816,922 for PGP 6, \$3,570,173 for PGP 2, \$3,143,044 for PGP 9, \$2,798,005 for PGP 10, and \$1,950,649 for PGP 5. Each PGP was paid 75 percent of the final PY3 earned performance payments when the PY3 financial reconciliation was calculated. The remaining 25 percent was withheld along with the withholds from PY1 and PY2 until the end of the Demonstration. In PY3 \$19.0 million was paid out to the five PGPs that earned a performance payment.

Table 5-7
Physician group practice demonstration performance payments calculation summary, performance year two

PGP Name	[A] Total Target Minus Actual Expenditures	[B] 2% Corridor (+/-)	[C] Expenditures Net 2% Corridor [#]	[D] Shared Savings	[E] Performance Payment for Efficiency	[F] Maximum Performance Payment for Quality	[G] Earned Performance Payment	[H] Payment at Annual Settlement
Total	\$34,478,988	—	\$15,424,386	\$13,902,353	\$8,341,412	\$5,560,941	\$13,840,014	\$10,380,011
PGP 1	-3,864,259	1,910,704	-1,953,554	0	0	0	0	0
PGP 2	14,245,806	5,808,459	8,437,347	6,749,878	4,049,927	2,699,951	6,689,879	5,017,409
PGP 3	1,584,801	1,420,291	164,510	131,608	78,965	52,643	129,268	96,951
PGP 4	-550,320	1,929,294	0	0	0	0	0	0
PGP 5	-1,123,249	3,891,817	0	0	0	0	0	0
PGP 6	13,124,654	5,897,687	7,226,966	5,781,573	3,468,944	2,312,629	5,781,573	4,336,180
PGP 7	2,460,215	3,089,935	0	0	0	0	0	0
PGP 8	1,468,439	2,654,538	0	0	0	0	0	0
PGP 9	705,317	4,540,583	0	0	0	0	0	0
PGP 10	6,427,585	4,878,468	1,549,118	1,239,294	743,576	495,718	1,239,294	929,471

NOTES:

[A] = [Total Target Expenditures] - [Total Actual Expenditures]. Target Expenditures and Actual Expenditures are determined using the PGP Demonstration Bonus Methodology.

[B] = [Total Target Expenditures] x [2%]. Total Target Expenditures are determined using the PGP Demonstration Bonus Methodology. The 2% Corridor is also known as the Savings Threshold.

[C] = Amount of Total Target Minus Actual Expenditures [A] that is outside the 2% corridor.

[D] = [C] x [80%]. The PGP Performance Payment Pool Sharing rate is set at 80%. Shared Savings are also referred to as the Bonus Pool.

[E] = [D] x [60%]. The share of the performance payment for efficiency is set at 60% for performance year two.

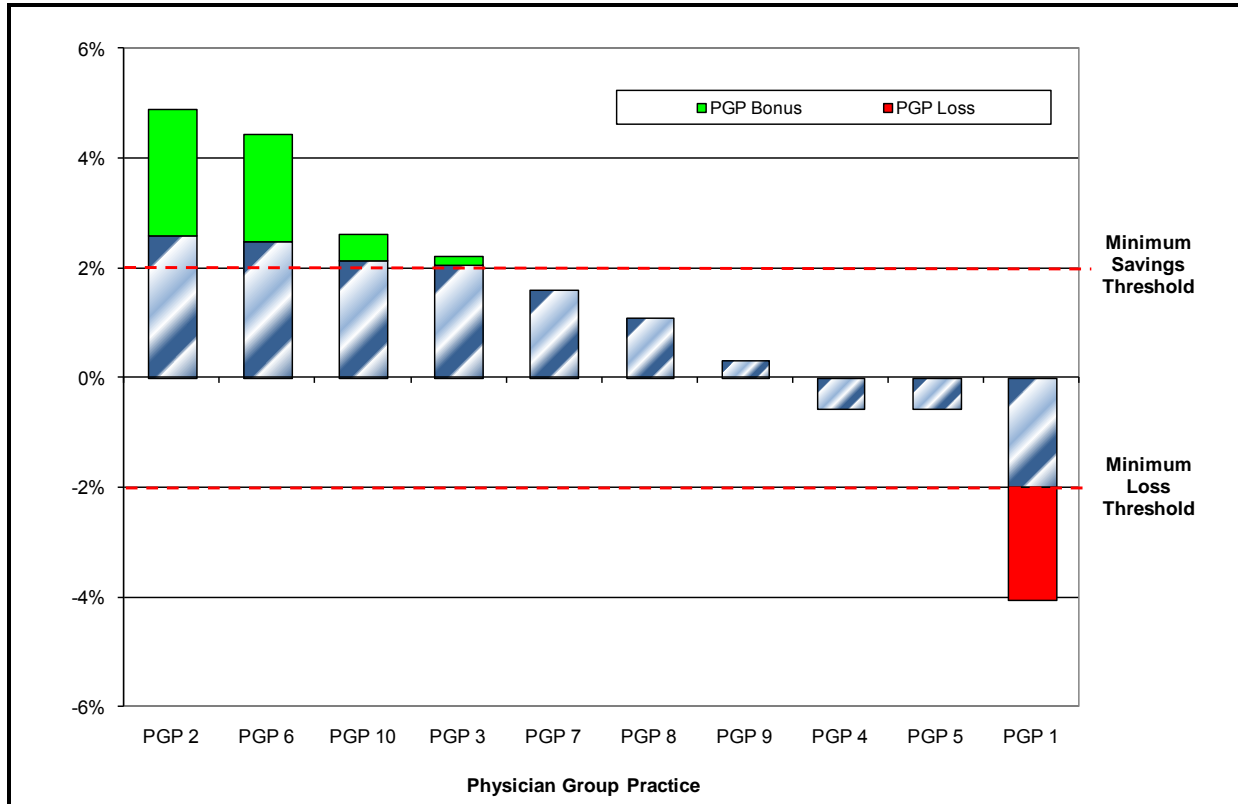
[F] = [D] x [40%]. The share of the performance payment for quality is set at 40% for performance year two.

[G] = [E] + [F]. This Earned Performance Payment takes into account actual percent quality targets achieved.

[H] = [G] x [75%]. The annual withhold is set at 25%.

SOURCE: RTI calculations with 2004–2007 Medicare claims and enrollment data.

Figure 5-4
PY2 target minus actual expenditures as a percentage of target expenditures, by PGP



NOTES:

1. The total height of the bar represents gross savings or loss (target minus actual expenditures).
2. The PGP bonus does not equal the gross savings minus the minimum savings threshold because of savings shared with the Medicare program and quality performance incentive.
3. PGPs were not at risk for “losses” during the Demonstration.

SOURCE: RTI calculations with 2004-PY2 Medicare claims and enrollment data.

**Table 5-8
Physician group practice demonstration performance payments calculation summary, performance year three**

PGP Name	[A] Total Target Minus Actual Expenditures	[B] 2% Corridor (+/-)	[C] Expenditures Net 2% Corridor [#]	[D] Shared Savings	[E] Performance Payment for Efficiency	[F] Maximum Performance Payment for Quality	[G] Earned Performance Payment	[H] Payment at Annual Settlement
Total	\$48,993,704	—	\$28,381,186	\$25,692,544	\$12,846,272	\$12,846,272	\$25,278,792	\$18,959,094
PGP 1	-4,104,817	-2,081,121	-2,023,696	0	0	0	0	0
PGP 2	10,681,304	6,043,511	4,637,793	3,710,234	1,855,117	1,855,117	3,570,173	2,677,630
PGP 3	-828,163	-1,503,025	0	0	0	0	0	0
PGP 4	-534,438	-2,143,256	0	0	0	0	0	0
PGP 5	7,035,110	4,596,799	2,438,311	1,950,649	975,324	975,324	1,950,649	1,462,987
PGP 6	23,491,760	6,055,839	17,435,921	13,948,737	6,974,369	6,974,369	13,816,922	10,362,691
PGP 7	-4,985,387	-3,060,671	-1,924,716	0	0	0	0	0
PGP 8	695,207	2,819,621	0	0	0	0	0	0
PGP 9	9,078,397	4,860,193	4,218,203	3,203,428	1,601,714	1,601,714	3,143,044	2,357,283
PGP 10	8,464,731	4,865,362	3,599,369	2,879,495	1,439,747	1,439,747	2,798,005	2,098,504

NOTES:

[A] = [Total Target Expenditures] - [Total Actual Expenditures]. Target Expenditures and Actual Expenditures are determined using the PGP Demonstration Bonus Methodology.

[B] = [Total Target Expenditures] x [2%]. Total Target Expenditures are determined using the PGP Demonstration Bonus Methodology. The 2% Corridor is also known as the Savings Threshold.

[C] = Amount of Total Target Minus Actual Expenditures [A] that is outside the 2% corridor.

[D] = [C] x [80%]. The PGP Performance Payment Pool Sharing rate is set at 80%. Shared Savings are also referred to as the Bonus Pool.

[E] = [D] x [50%]. The share of the performance payment for efficiency is set at 50% for performance year three.

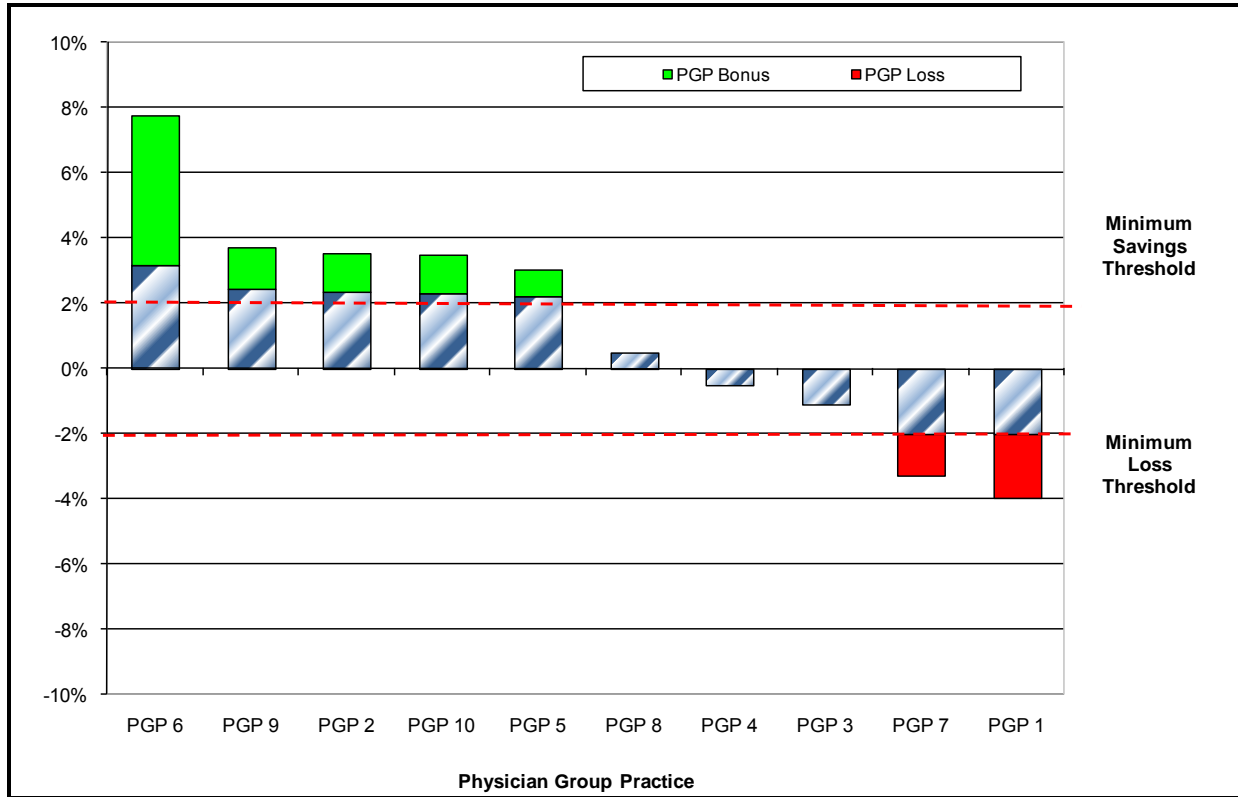
[F] = [D] x [50%]. The share of the performance payment for quality is set at 50% for performance year three.

[G] = [E] + [F]. This Earned Performance Payment takes into account actual percent quality targets achieved.

[H] = [G] x [75%]. The annual withhold is set at 25%.

SOURCE: RTI calculations with 2004–2008 Medicare claims and enrollment data.

Figure 5-5
PY3 target minus actual expenditures as a percentage of target expenditures, by PGP



NOTES:

1. The total height of the bar represents gross savings or loss (target minus actual expenditures).
2. The PGP bonus does not equal the gross savings minus the minimum savings threshold because of savings shared with the Medicare program quality performance incentive.
3. PGPs were not at risk for “losses” during the Demonstration.

SOURCE: RTI calculations with 2004-PY3 Medicare claims and enrollment data.

Performance Year Four

In PY4 five PGPs earned a performance payment by having target expenditures minus actual expenditures in excess of the 2 percent corridor (*Table 5-9*). Four PGPs had actual expenditures greater than target expenditures, but not below the 2 percent corridor. The remaining PGP had actual expenditures in excess of target expenditures and below the 2 percent corridor and therefore accrued losses carried forward. In these calculations target expenditures are the total risk-adjusted base year assigned beneficiary expenditures trended forward by the risk-adjusted expenditure growth rate of the comparison group from BY to PY4. *Figure 5-6* illustrates the gross savings of all 10 PGP sites (the height of the bars); the bonuses of PGP 6, PGP 9, PGP 10, PGP 5, and PGP 2; and the loss of PGP 1. In total the 10 PGPs’ actual expenditures were \$69.4 million less than target expenditures.

**Table 5-9
Physician group practice demonstration performance payments calculation summary, performance year four**

PGP Name	[A] Total Target Minus Actual Expenditures	[B] 2% Corridor (+/-)	[C] Expenditures Net 2% Corridor	[D] Shared Savings	[E] Performance Payment for Efficiency	[F] Maximum Performance Payment for Quality	[G] Earned Performance Payment	[H] Payment at Annual Settlement
Total	\$69,427,208	—	\$48,076,105	\$38,697,635	\$19,348,818	\$19,348,818	\$31,679,844	\$23,759,883
PGP 1	-2,523,613	-2,227,674	-295,939	0	0	0	0	0
PGP 2	6,878,100	6,455,132	422,968	338,375	169,187	169,187	328,798	246,599
PGP 3	-1,513,813	-1,615,249	0	0	0	0	0	0
PGP 4	-36,330	-2,216,506	0	0	0	0	0	0
PGP 5	6,977,096	4,741,851	2,235,245	1,788,196	894,098	894,098	1,788,196	1,341,147
PGP 6	35,191,769	6,461,697	28,730,072	22,984,057	11,492,029	11,492,029	16,154,242 ¹	12,115,682
PGP 7	-1,932,138	-3,135,723	0	0	0	0	0	0
PGP 8	-562,752	-2,859,335	0	0	0	0	0	0
PGP 9	15,322,653	4,993,008	10,329,645	8,263,716	4,131,858	4,131,858	8,185,757	6,139,317
PGP 10	11,626,235	4,972,121	6,654,114	5,323,291	2,661,646	2,661,646	5,222,852	3,917,139

NOTES:

¹ Earned Performance Payment is capped at 5% of Target Expenditures. The uncapped Earned Performance Payment for PGP 6 is \$22,984,057.

[A] = [Total Target Expenditures] - [Total Actual Expenditures]. Target Expenditures and Actual Expenditures are determined using the PGP Demonstration Bonus Methodology.

[B] = [Total Target Expenditures] x [2%]. Total Target Expenditures are determined using the PGP Demonstration Bonus Methodology. The 2% Corridor is also known as the Savings Threshold.

[C] = Amount of Total Target Minus Actual Expenditures [A] that is outside the 2% corridor.

[D] = [C] x [80%]. The PGP Performance Payment Pool Sharing rate is set at 80%. Shared Savings are also referred to as the Bonus Pool.

[E] = [D] x [50%]. The share of the performance payment for efficiency is set at 50% for performance year four.

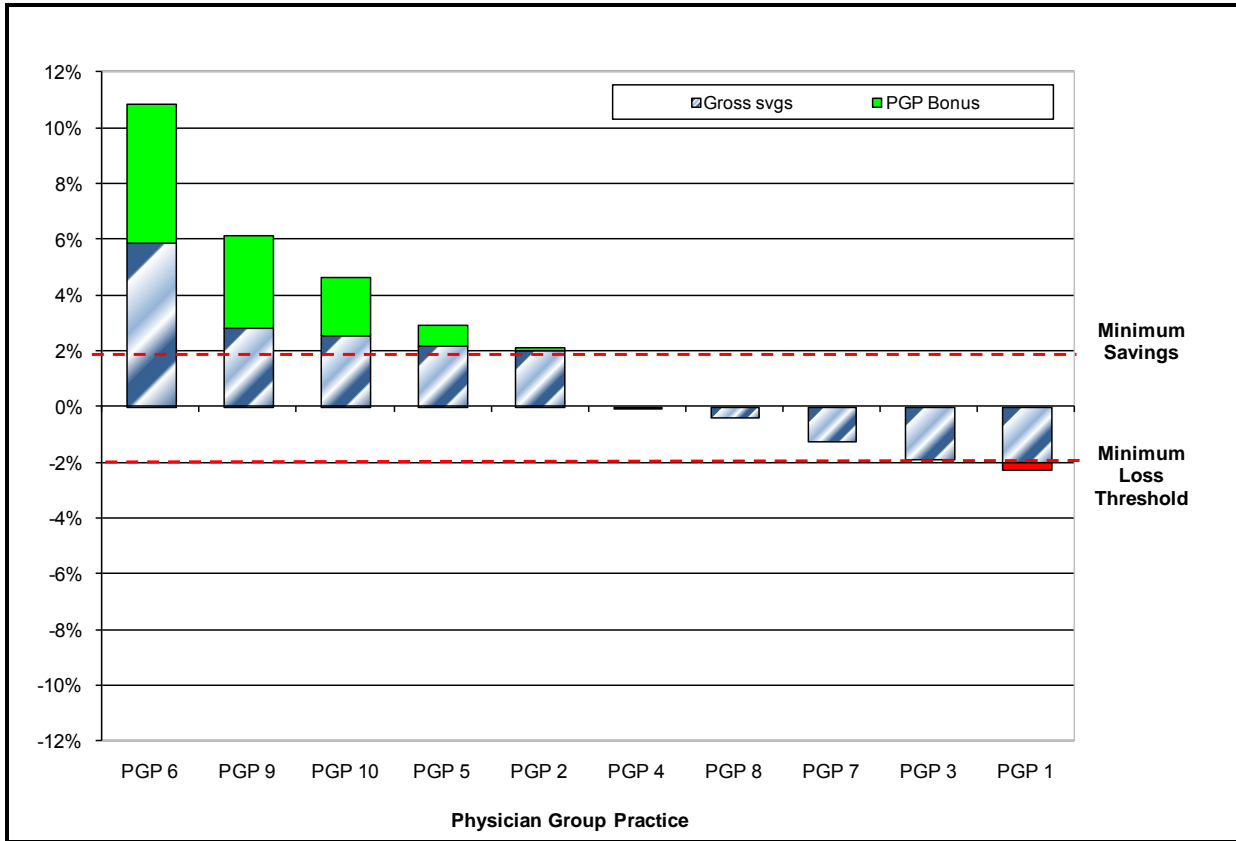
[F] = [D] x [50%]. The share of the performance payment for quality is set at 50% for performance year four.

[G] = [E] + [F]. This Earned Performance Payment takes into account actual percent quality targets achieved.

[H] = [G] x [75%]. The annual withhold is set at 25%.

SOURCE: RTI calculations with 2003–2009 Medicare claims and enrollment data.

Figure 5-6
PY4 target minus actual expenditures as a percentage of target expenditures, by PGP



NOTES:

1. The total height of the bar represents gross savings or loss (target minus actual expenditures).
2. The PGP bonus does not equal the gross savings minus the minimum savings threshold because of savings shared with the Medicare program quality performance incentive.
3. The PGP bonus for PGP 6 reflects the final earned performance payment capped at 5% of target expenditures.
4. PGPs were not at risk for “losses” during the Demonstration.

SOURCE: RTI calculations with 2004-PY4 Medicare claims and enrollment data.

The five PGPs that earned PY4 performance payments generated a total of \$38.7 million in shared savings. In PY4 the portion of the shared savings kept by the PGP as the performance payment for efficiency and the performance payment for quality remained at 50 percent and 50 percent, respectively. In PY4 both PGP 6 and PGP 5 hit 100 percent of their quality targets, therefore earning the maximum performance payments for quality. PGP 9, PGP 10, and PGP 2 hit 98 percent, 96 percent, and 94 percent of quality targets respectively. This resulted in final earned performance payment of \$8,185,757 for PGP 9, \$5,222,852 for PGP 10, \$1,788,196 for PGP 5, and \$328,798 for PGP 2. Final earned performance payments were capped at 5 percent

of target expenditures, therefore resulting in a final earned performance payment of \$16,154,242 for PGP 6.¹² Each PGP was paid 75 percent of the final PY4 earned performance payments when the PY4 financial reconciliation was calculated. The remaining 25 percent was withheld along with the withholds from PY1, PY2, and PY3 until the end of the Demonstration. In PY4 \$23.8 million was paid out to the five PGPs that earned a performance payment.

Performance Year Five

In PY5 four PGPs earned a performance payment by having target expenditures less than actual expenditures in excess of the 2 percent corridor (*Table 5-10*). One PGP had actual expenditures less than target expenditures, but not over the 2 percent corridor. Three PGPs had actual expenditures greater than target expenditures, but not below the 2 percent corridor. The remaining two PGPs had actual expenditures in excess of target expenditures and below the 2 percent corridor, therefore accruing losses carried forward. In these calculations target expenditures are the total risk-adjusted base year assigned beneficiary expenditures trended forward by the risk-adjusted expenditure growth rate of the comparison group from BY to PY5. *Figure 5-7* graphically illustrates the gross savings of all 10 PGP sites (the height of the bars; the bonuses of PGP 6, PGP 8 PGP 10, and PGP 9; and the losses of PGP 1 and PGP 5). In total the 10 PGPs' actual expenditures were \$46.6 million less than target expenditures.

The four PGPs that earned PY5 performance payment generated a total of \$36.2 million in shared savings. In PY5 the portion of shared savings kept by the PGP as the performance payment for efficiency and the performance payment for quality remained at 50 percent and 50 percent, respectively. In PY5 PGP 8 and PGP 9 hit 100 percent of their quality targets, therefore earning the maximum performance payment for quality. PGP 6 and PGP 10 both hit 98 percent of quality targets. This resulted in final earned performance payment of \$5,673,177 for PGP 8, \$5,329,967 for PGP 10, and \$2,598,859 for PGP 9. Final earned performance payments were capped at 5% of target expenditures, therefore resulting in a final earned performance payment of \$15,832,603 for PGP 6.¹³ In PY5 \$22.1 million was paid to the four PGPs that earned a performance payment.

At the end of the Demonstration, the 25% withholds from PY1 through PY4 less any accrued losses carried forward were paid to each PGP earning a performance payment. This final settlement to each PGP, including the PY5 annual settlement, resulted in final settlement payments of \$25,912,119 to PGP 6, \$8,334,597 to PGP 10, \$5,673,177 to PGP 8, and \$5,431,059 to PGP 9. In total, \$45.4 million was paid in the final settlement plus PY5 annual settlement.

¹² PGP 6's uncapped Earned Performance Payment was \$22,984,057.

¹³ PGP 6's uncapped Earned Performance Payment was \$22,338,610.

**Table 5-10
Physician group practice demonstration performance payments calculation summary, performance year five**

PGP Name	[A] Total Target Minus Actual Expenditures	[B] 2% Corridor (+/-)	[C] Expenditures Net 2% Corridor	[D] Shared Savings	[E] Performance Payment for Efficiency	[F] Performance Payment for Quality	[G] Earned Performance Payment	[H] Payment at Annual Settlement
Total	\$46,635,970	—	\$37,879,238	\$36,204,124	\$18,102,062	\$17,838,522	\$29,434,607	\$22,075,955
PGP 1	-3,632,820	-2,297,269	-1,335,551	0	0	0	0	0
PGP 2	-2,181,523	-6,458,120	0	0	0	0	0	0
PGP 3	201,021	1,673,731	0	0	0	0	0	0
PGP 4	-1,335,406	-2,297,612	0	0	0	0	0	0
PGP 5 ²	-10,685,905	-4,645,538	-6,040,367	0	0	0	0	0
PGP 6	34,522,240	6,333,041	28,189,199	22,551,359	11,275,680	11,062,931	15,832,603 ¹	11,874,452
PGP 7 ²	-708,319	-3,197,688	0	0	0	0	0	0
PGP 8	10,049,998	2,958,526	7,091,472	5,673,177	2,836,589	2,836,589	5,673,177	4,254,883
PGP 9	8,187,372	4,938,798	3,248,574	2,598,859	1,299,430	1,299,430	2,598,859	1,949,145
PGP 10	12,219,312	5,493,401	6,725,911	5,380,729	2,690,364	2,639,603	5,329,967	3,997,475

NOTES:

¹ Earned Performance Payment is capped at 5% of Target Expenditures. The uncapped Earned Performance Payment for PGP 6 is \$22,338,610.

² Reflects the PY5 risk score cap. Without the risk score cap, PGP 5's Total Target Minus Actual Expenditures = \$2,981,555, Expenditures Net 2% Corridor = \$0. Without the risk score cap, PGP 7's Total Target Minus Actual Expenditures = \$479,435, Expenditures Net 2% Corridor = \$0.

[A] = [Total Target Expenditures] - [Total Actual Expenditures]. Target Expenditures and Actual Expenditures are determined using the PGP Demonstration Bonus Methodology.

[B] = [Total Target Expenditures] x [2%]. Total Target Expenditures are determined using the PGP Demonstration Bonus Methodology. The 2% Corridor is also known as the Savings Threshold.

[C] = Amount of Total Target Minus Actual Expenditures [A] that is outside the 2% corridor.

[D] = [C] x [80%]. The PGP Performance Payment Pool Sharing rate is set at 80%. Shared Savings are also referred to as the Bonus Pool.

[E] = [D] x [50%]. The share of the performance payment for efficiency is set at 50% for performance year five.

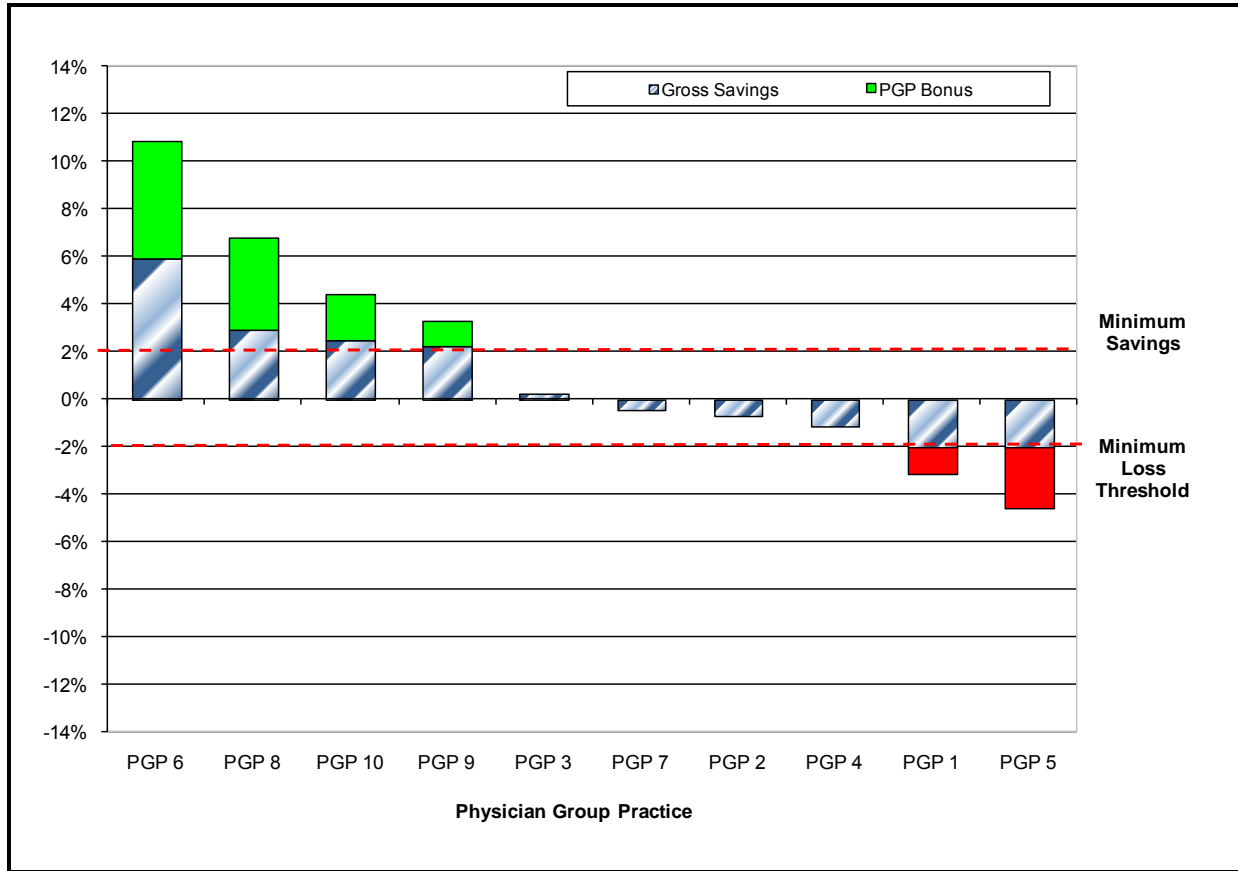
[F] = [D] x [50%] x Weighted Percentage of Quality Targets Met. The share of the performance payment for quality is set at 50% for performance year five.

[G] = [E] + [F]. This Earned Performance Payment assumes that all quality targets were achieved.

[H] = [G] x [75%]. The annual withhold is set at 25%. Assumes all quality targets were achieved.

SOURCE: RTI calculations with 2003–2010 Medicare claims and enrollment data.

Figure 5-7
PY5 target minus actual expenditures as a percentage of target expenditures, by PGP



NOTES:

1. The total height of the bar represents gross savings or loss (Target minus Actual Expenditures).
2. The PGP bonus does not equal the gross savings minus the minimum savings threshold because of savings shared with the Medicare program quality performance incentive.
3. The PGP bonus for PGP 6 reflects the final earned performance cap at 5% of Target Expenditures.
4. Reflects the PY5 risk score cap.
5. PGPs were not at risk for “losses” during the Demonstration.

SOURCE: RTI calculations with 2004-PY5 Medicare claims and enrollment data.

Finally, **Table 5-11** presents the shared savings of the PGP on an assigned beneficiary “per member per month” (PMPM) basis for PY1 through PY5. (Shared savings was the maximum PGP bonus, if all quality targets were met.) In PY1, PGP 6 and PGP 10, the two PGPs that earned performance payments, had shared savings PMPM of \$9.87 and \$11.80 respectively. In PY2 shared savings PMPM ranged from \$1.17 to \$18.89 at the four PGPs that

earned performance payments. In PY3 shared savings PMPM ranged from \$6.29 to \$32.35. In PY4 shared savings PMPM ranged from \$0.90 to \$39.47.¹⁴ In PY5 shared savings PMPM ranged from \$7.15 to \$38.79.¹⁵ PGP 6 and PGP 10 earned payments in all five Demonstration years.

Table 5-11
Physician group practice demonstration shared savings per member per month, performance years one through five

PGP Name	PY1 Shared Savings Per Member Per Month ¹ (\$)	PY2 Shared Savings Per Member Per Month (\$)	PY3 Shared Savings Per Member Per Month (\$)	PY4 Shared Savings Per Member Per Month (\$)	PY5 Shared Savings Per Member Per Month (\$)
PGP 1	—	—	—	—	—
PGP 2	—	18.89	10.19	0.90	—
PGP 3	—	1.17	—	—	—
PGP 4	—	—	—	—	—
PGP 5	—	—	6.29	5.96	—
PGP 6	9.87	12.84	32.35	39.47 ²	38.79 ²
PGP 7	—	—	—	—	—
PGP 8	—	—	—	—	27.11
PGP 9	—	—	8.75	22.60	7.15
PGP 10	11.80	5.59	13.88	25.57	23.53

NOTES:

¹ Shared Savings Per Member Per Month = Shared Savings / Person Years / 12.

² Shared Savings were capped at 5% of Target Expenditures. The uncapped Shared Savings PMPM was \$56.15 in PY4 and \$55.26 in PY5.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data

5.4 Summary of Demonstration Financial Results and Medicare Program Savings

This section summarizes the financial results of the Demonstration. The quantities that are analyzed are:

- gross savings: Medicare expenditures for Demonstration assigned beneficiaries that are less than expected, defined as positive target minus actual expenditures;
- PGP losses: Medicare expenditures for Demonstration assigned beneficiaries that are greater than expected, defined as negative target minus actual expenditures;
- performance payments to PGPs: PGPs' share of gross savings, defined as earned performance payouts from CMS to participating PGPs;

¹⁴ Shared savings are capped at 5% of target expenditures. The uncapped maximum shared savings PMPM in PY4 (for PGP 6) is \$56.15.

¹⁵ The uncapped maximum shared savings PMPM in PY5 (for PGP 6) is \$55.26.

- Medicare program savings: estimated Medicare Trust Fund savings from the Demonstration, the estimated amount by which program expenditures for assigned beneficiaries were lower under the Demonstration than in the absence of the Demonstration. Defined as gross savings minus PGP losses minus performance payments to PGPs.

We summarize Demonstration financial results under the Demonstration methodology that savings or losses within the 2 percent corridor represent normal variation in claims expenditures and are therefore counted as neither (gross) savings nor losses. Chapter 11 presents an alternative calculation (sensitivity analysis) of savings under the assumption that savings or losses within the 2 percent corridor are Medicare program expenditures, or lack of expenditures, and are therefore counted as (gross) savings or losses. We first report aggregate savings and performance payments in Section 5.4.1. Then we examine savings and performance payments on a per person basis, and as a percentage of target expenditures, in Section 5.4.2.

5.4.1 Demonstration Savings and Performance Payments

In the Demonstration, savings are defined for each PGP as the difference between target expenditures and actual expenditures that exceed a two percent threshold of target expenditures, multiplied by the number of assigned beneficiary person years. Positive savings suggest 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 savings associated with the Demonstration.

Table 5-12 summarizes the Demonstration savings and losses above or below the 2% corridor, performance payments, and Medicare program savings for all five performance years. In PY1, the savings to the Medicare Trust Fund were \$0.7 million. In total, the PGPs saved \$8.0 million (consisting of gross savings of \$9.5 million minus losses of \$1.5 million) and earned \$7.3 million. Two PGPs earned performance payments, two PGPs accrued losses below the negative two percent corridor, and six PGPs had actual expenditures lower than their targets within the two percent corridor (which are shown as '0' in Table 5-12).

Table 5-12
Savings, performance payments, and Medicare program savings, PY1-PY5 (amounts in thousands)
amounts within ± 2% corridor not counted as savings or losses

PGP Site	PY 1 Total Target Minus Actual Expenditures ²	PY 1 Earned Performance Payments	PY 2 Total Target Minus Actual Expenditures ²	PY 2 Earned Performance Payments	PY 3 Total Target Minus Actual Expenditures ²	PY 3 Earned Performance Payments
PGP 1	-\$1,316	\$0	-\$1,954	\$0	-\$2,024	\$0
PGP 2	0	0	8,437	6,690	4,638	3,570
PGP 3	0	0	165	129	0	0
PGP 4	0	0	0	0	0	0
PGP 5	0	0	0	0	2,438	1,951
PGP 6	6,036	4,565	7,227	5,782	17,436	13,817
PGP 7	0	0	0	0	-1,925	0
PGP 8	0	0	0	0	0	0
PGP 9	-214	0	0	0	4,218	3,143
PGP 10	3,496	2,758	1,549	1,239	3,599	2,798
Total	8,001	7,324	15,424	13,840	28,381	25,279
Gross Savings⁴	9,532	—	17,378	—	32,330	—
Less: Losses⁵	-1,530	—	-1,954	—	-3,948	—
Savings	8,001	—	15,424	—	28,381	—
Less: Performance Payments	-7,324	—	-13,840	—	-25,279	—
Medicare Program Savings	\$677	—	\$1,584	—	\$3,102	—

(continued)

Table 5-12 (continued)
Savings, performance payments, and Medicare program savings, PY1-PY5 (amounts in thousands)
amounts within ± 2% corridor not counted as savings or losses

PGP Site	PY 4 Total Target Minus Actual Expenditures ²	PY 4 Earned Performance Payments	PY 5 ¹ Total Target Minus Actual Expenditures ²	PY 5 ¹ Earned Performance Payments	Combined Years Total Target Minus Actual Expenditures ²	Combined Years Earned Performance Payments
PGP 1	-\$296	\$0	-\$1,336	\$0	-\$6,925	\$0
PGP 2	423	329	0	0	13,498	10,589
PGP 3	0	0	0	0	165	129
PGP 4	0	0	0	0	0	0
PGP 5	2,235	1,788	-6,040	0	-1,367	3,739
PGP 6	28,730	16,154 ³	28,189	15,833 ³	87,618	56,151
PGP 7	0	0	0	0	-1,925	0
PGP 8	0	0	7,091	5,673	7,091	5,673
PGP 9	10,330	8,186	3,249	2,599	17,583	13,928
PGP 10	6,654	5,223	6,726	5,330	22,024	17,348
Total	48,076	31,680	37,879	29,435	137,762	107,557
Gross Savings⁴	48,372	—	45,255	—	152,866	—
Less: Losses⁵	-296	—	-7,376	—	-15,104	—
Savings	48,076	—	37,879	—	137,762	—
Less: Performance Payments	-31,680	—	-29,435	—	-107,557	—
Medicare Program Savings	\$16,396	—	\$8,445	—	\$30,205	—

NOTES:

¹ Reflects PY 5 risk score cap. Without the cap, PY 5 Gross Savings = \$45,255,156 (unchanged), Savings = \$43,920,605, and Medicare Program Savings = \$14,485,998.

² Amounts within ± 2% corridor are not counted as savings or losses.

³ Earned Performance Payment is capped at 5% of Target Expenditures.

⁴ Gross Savings represents the sum of Total Target Minus Actual Expenditures above the 2% minimum savings threshold.

⁵ Losses are Total Target Minus Actual Expenditures below the -2% minimum loss threshold.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data

These results are illustrated in *Figure 5-8*, which graphs summary Demonstration financial results for all five performance years. The height of the bar represents gross savings, the top portion of the bar represents PGP losses, the middle portion represents PGP bonuses (Demonstration performance payments or PGP earnings), and the bottom portion represents savings to Medicare. In other words, Figure 5-8 shows the distribution of gross Demonstration savings among PGP losses, PGP bonuses, and Medicare savings. In PY1, most savings were paid as bonuses to participating PGPs, resulting in limited Medicare savings.

In PY2, gross savings nearly doubled to \$17.4 million. PGP losses and performance payments also nearly doubled. Savings to the Medicare Trust Fund more than doubled to \$1.6 million, but most gross savings were still paid as bonuses to the PGPs. Four PGPs earned performance payments totaling \$13.8 million. PGP 1 had a negative savings of \$2.0 million.

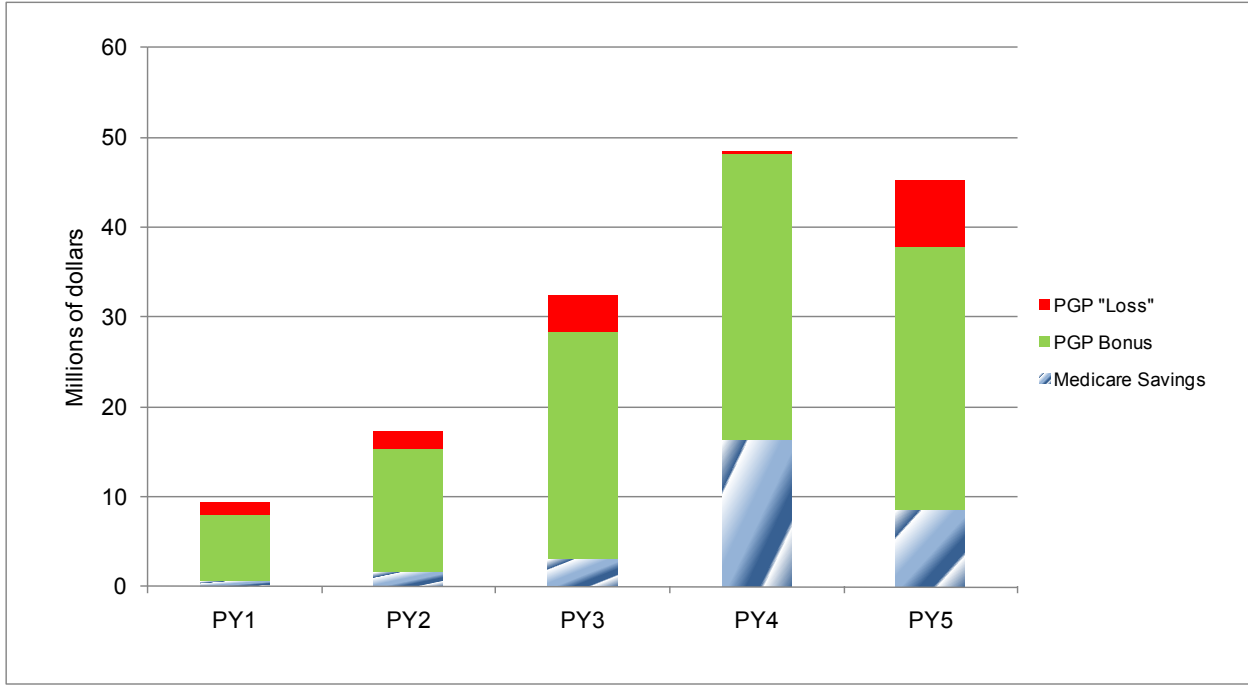
In PY3, gross savings, losses, and performance payments nearly doubled again. Medicare Program savings amounted to \$3.1 million with five PGPs earning performance payments. PGP's performance payments were \$25.3 million. Most gross savings continued to be distributed to the PGPs as performance payments.

In PY4, gross savings rose 50 percent to \$48.4 million, but Medicare savings increased more than fivefold to \$16.4 million. About two-thirds of gross savings were distributed to the PGPs as \$31.7 million in bonuses, but Medicare savings comprised a much larger share of gross savings—about one-third—than in prior years. In part this was the result of the 5 percent of target expenditures bonus cap, which was binding on PGP 6 in PY4 and PY5. Amounts in excess of the cap accrued to the Medicare program as savings. Savings to the Medicare Trust Fund peaked in PY4, with 5 PGPs earning performance payments. PGP losses were minimized in this performance year and PGP performance payments were maximized.

In PY5, gross savings fell slightly to \$45.3 million. Four PGPs earned performance payments, totaling \$29.4 million. PGP losses widened substantially to \$7.4 million, the most of any performance year, which was largely due to the PY5 risk score cap that affected two PGPs. These factors resulted in \$8.4 million in contributions to the Medicare Trust Fund, about one-half the level of PY4. Without the risk score cap, PY5 Medicare program savings would have been estimated at \$14.5 million, only a slight reduction from PY4.

Over the five years of the Demonstration combined, gross savings were \$152.9 million. Of this amount, \$15.1 million were absorbed by PGP losses. Of the remaining \$137.8 million in savings, \$107.6 million was distributed to the participating PGPs as performance payments and \$30.2 million comprised Medicare Program savings. PGP 6 accounted for 57 percent of total gross savings and 52 percent of total performance payments during the Demonstration. Six other PGPs generated overall savings (net of any losses) during the Demonstration. Three PGPs generated overall losses (net of any savings) during the performance period. (PGPs were not at risk for “losses” during the Demonstration.) PGP 4 was not credited with savings or losses in any of the five performance years. PGPs 6 and 10 earned performance payments in each of the Demonstration's five performance years. PGP 1 generated losses in all performance years.

Figure 5-8
Distribution of PGP demonstration gross savings
(in millions of dollars)
amounts within $\pm 2\%$ corridor not counted as savings or losses



NOTES:

1. The height of the bar reflects gross Demonstration savings, which are Target minus Actual Expenditures above the 2 percent minimum savings threshold. (Amounts within the plus or minus 2 percent minimum savings/loss corridor are not included in this graph.) PGP “losses” are Target minus Actual Expenditures below the negative 2 percent minimum loss threshold. Gross savings are divided between PGP bonuses, PGP losses, and Medicare savings.
2. Reflects PY5 risk score cap.
3. PGP 6’s bonus was capped in PY4 and PY5. This graph reflects the capped bonus. Savings above the cap are counted as Medicare savings.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Of the 7 PGPs, which earned performance payments during the Demonstration, five groups were paid between 78 and 80 percent of the aggregate savings they generated across the five years of the Demonstration (ratio of last column to the second-to-column of Table 5-12). PGP 6 was paid only 64 percent of its savings because of the 5 percent performance payment cap that limited its bonus in PY4 and PY5. PGP 5 generated dis-savings (losses) during the Demonstration overall, yet received performance payments. This anomaly was entirely due to the PY5 risk score cap, which turned a zero savings/loss for PGP 5 in PY5 into a loss.

BIPA required an initial demonstration period of three years, with renewal terms of up to three years; however, the legislation was silent on rebasing in the renewal terms. Consequently, the Demonstration had only one base year, 2004, and it was not updated in PY4 for the renewal term. 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, as the base year becomes outdated. Evidence of a stale base year is suggested by the distribution of savings across performance years. In the initial three year term of the demonstration, 18% of Medicare savings, \$5,363,000, and 38% of performance payments, \$51,806,000, were achieved. In the last two years of the demonstration, 82% of Medicare savings, \$24,841,000 and 62% of performance payments, \$85,955,000 were experienced. Thus, most of the Medicare savings and performance payments were achieved in the renewal term of the demonstration. The increased Medicare savings and performance payments in PY 4 and PY 5 could be due to an increase in the number of assigned beneficiaries, as this factor directly affects the performance calculation. During the 5 year demonstration period, the number of assigned beneficiaries remained almost constant, around 220,000 per year, although site specific numbers varied. The number of assigned beneficiaries changed by 5 percent (increase / decrease) from the base year at 5 sites, while 2 sites experienced changes of more than 20 percent (increase / decrease) from the base year.

During the 5 year demonstration, as shown in *Table 5-13*, Medicare allowable charges of the assigned beneficiaries were \$9,204 million. Medicare savings were \$30 million and performance payments were \$138 million. Expressed as a percentage, the demonstration saved Medicare .3% of the claims amounts, while performance payments were 1.5% of the claims amounts.

5.4.2 Demonstration Savings and Performance Payments Per Person and as a Percentage of Target Expenditures

For the five Demonstration performance years combined, *Table 5-14* shows gross savings, performance payments and losses, and Medicare program savings per person year and month, and as percentages of target expenditures and gross savings.

Table 5-13
Savings and performance payments as a proportion of Medicare allowable charges

Performance Year	Amounts in 000's: Medicare Savings	Amounts in 000's: Perf. Pay	Amounts in 000's: Gross Savings	Amounts in 000's: Total Medicare Costs	Percentages: Medicare Savings	Percentages: Performance Payments	Percentages: Gross Savings
PY1	677	8,001	9,532	1,702,605	0.040%	0.470%	0.560%
PY2	1,584	15,424	17,378	1,766,610	0.090%	0.873%	0.984%
PY3	3,102	28,381	32,330	1,852,476	0.167%	1.532%	1.745%
PY4	16,396	48,076	48,372	1,914,488	0.856%	2.511%	2.527%
PY5	8,445	37,879	45,255	1,968,050	0.429%	1.925%	2.299%
Total	30,204	137,761	152,867	9,204,229	0.328%	1.497%	1.661%

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Table 5-14
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, PY1-PY5 combined

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	143.18	11.93	1.62%	100.0%
Losses ³	14.15	1.18	0.16%	9.9%
Earned Performance Payments	100.74	8.39	1.14%	70.4%
Medicare Program Savings ⁴	28.29	2.36	0.32%	19.8%

NOTES:

This table reflects the PY5 Demonstration risk score cap.

¹ Person years is a count of assigned beneficiaries adjusted for the number of months of Demonstration eligibility per assigned beneficiary during a performance year.

² Gross Savings are Target minus Actual Expenditures above the 2 percent minimum savings threshold.

³ PGP losses are Target minus Actual Expenditures below the negative 2 percent minimum loss threshold.

⁴ Medicare Program Savings are Gross Savings minus Losses minus Earned Performance Payments.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data

Gross savings were \$143.18 per person year, or \$11.93 per person month. Of the annual gross savings, \$100.74 (70.4 percent) was paid to the participating PGPs as performance bonuses, \$14.15 (9.9 percent) was offset as losses, and \$28.29 (19.8 percent) comprised Medicare savings. Gross savings represented 1.62 percent of target expenditures, Medicare savings were 0.32 percent of target expenditures, and PGP performance payments were 1.14 percent of target expenditures.

Tables 5-15a through 5-15e segments Table 5-14 by individual performance year. Savings increase each year from PY1 to PY4, but then decline in PY5. The distribution of gross savings is consistent with the combined year results already discussed. Chapter 11 presents a sensitivity analysis of estimates of gross savings, performance payments and losses, and Medicare program savings per person year and month, and as percentages of target expenditures and gross savings under the alternative assumption that (gross) savings and losses within the 2 percent corridor are Medicare program savings, or lack of savings.

Table 5-15a
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 1

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	43.91	3.66	0.55%	100.0%
Losses ³	7.05	0.59	0.09%	16.1%
Earned Performance Payments	33.74	2.81	0.42%	76.8%
Medicare Program Savings ⁴	3.12	0.26	0.04%	7.1%

Table 5-15b
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 2

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	81.62	6.80	0.96%	100.0%
Losses ³	9.18	0.76	0.11%	11.2%
Earned Performance Payments	65.01	5.42	0.77%	79.6%
Medicare Program Savings ⁴	7.44	0.62	0.09%	9.1%

Table 5-15c
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 3

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	151.79	12.65	1.70%	100.0%
Losses ³	18.54	1.54	0.21%	12.2%
Earned Performance Payments	118.68	9.89	1.33%	78.2%
Medicare Program Savings ⁴	14.57	1.21	0.16%	9.6%

Table 5-15d
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 4

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	229.97	19.16	2.44%	100.0%
Losses ³	1.41	0.12	0.01%	0.6%
Earned Performance Payments	150.61	12.55	1.60%	65.5%
Medicare Program Savings ⁴	77.95	6.50	0.83%	33.9%

Table 5-15e
Demonstration gross savings, performance payments and losses, and Medicare program savings per person year and month, and as a percentage of target expenditures and gross savings, Performance Year 5

Savings, Losses, and Performance Payments	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Year ¹	Amounts within ± 2% Corridor Not Counted as Savings or Losses: \$ Per Assigned Beneficiary Person Month	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Target Expenditures	Amounts within ± 2% Corridor Not Counted as Savings or Losses: % of Gross Savings
Gross Savings ²	211.12	17.59	2.25%	100.0%
Losses ³	34.41	2.87	0.37%	16.3%
Earned Performance Payments	137.31	11.44	1.46%	65.0%
Medicare Program Savings ⁴	39.39	3.28	0.42%	18.7%

NOTES:

- ¹ Person years is a count of assigned beneficiaries adjusted for the number of months of Demonstration eligibility per assigned beneficiary during a performance year.
- ² In the first panel, Gross Savings are Target minus Actual Expenditures above the 2 percent minimum savings threshold.
- ³ In the first panel, PGP losses are Target minus Actual Expenditures below the negative 2 percent minimum loss threshold.
- ⁴ Medicare Program Savings are Gross Savings minus Losses minus Earned Performance Payments.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

CHAPTER 6 QUALITY MEASUREMENT & REPORTING

This chapter reviews the PGP Demonstration quality measurement results for PY1 through PY5. It includes seven sections. Section 6.1 provides an overview of the quality measurement methodology and results. Section 6.2 focuses on the quality performance of the participating PGPs for the diabetes mellitus (DM) module. Section 6.3 presents quality performance results for the heart failure (HF) module. Section 6.4 highlights results for the coronary artery disease (CAD) module. Section 6.5 includes results for the hypertension (HTN) module. Section 6.6 describes results for the preventative care (PC) module. Finally, Section 6.7 describes the methodology and results for the Medicare Physician Quality Reporting Initiative (PQRI) bonus payments made to the participating PGPs for PY2 through PY5.

6.1 Overview of PGP Demonstration Quality Measurement Methodology and Results

The PGP Demonstration included performance assessment for 10 DM quality measures in PY1. In PY2 the HF and CAD modules became active, with 10 and seven quality measures respectively. As a result, a total of 27 quality measures were active in PY2 across the three modules. In PY3 the three HTN and two PC quality measures became active, resulting in a total of 32 active quality measures across five modules in PY3, PY4 and PY5. Detailed descriptions of the quality measures and the performance targets are explained in Chapter 2 and also included in Trisolini et al. (2005). Performance targets in the PGP Demonstration included multiple options for demonstrating positive performance through meeting either threshold targets or improvement over time targets.

In PY1, the weighting method provided for the 4 claims-based diabetes measures to be worth 4 points each, while the rest of the 6 DM chart-based measures were worth 1 point each, resulting in 22 possible points earned. In PY2, the weighting method provided for 6 claims-based quality measures to be weighted at 4 points each and 21 chart-based quality measures to be weighted at 1 point each, with a possible 45 quality points earned. In PY3 through PY5, a total of 53 quality points were possible (7 claims-based quality measures worth 4 points each and 25 chart-based quality measures worth 1 point each).

The PGPs improved their quality performance for all five disease modules over the course of the Demonstration. Many PGPs started with and maintained high scores throughout the Demonstration, but over time, all of the PGPs had improvements in their scores for most of the quality measures. Measures of performance (rates and targets met) over time are presented in details in subsequent sections.

Table 6-1 provides a summary across all active measures for each performance year, which is represented by two columns: “Percent of targets met” and “Percent of points earned”. The first is based on the number of all active measures during that performance year, and the second is based on the total number of possible points for that year after accounting for the different measure weights, as described above. Comparing results over time, it is clear that every PGP had improvements in their quality results over time, even when there were 22 more active measures in PY5 than in PY1. For example, only two PGPs met all 10 measure targets in PY1

(i.e., 100 percent), but by PY5, seven PGPs met all 32, or 100 percent of their targets, and the remaining 3 PGPs still meeting the over 90 percent of the targets. This translated to an average of 90 percent quality points earned in PY1 compared to an average of 99 percent points earned in PY5 across all ten PGPs.

Table 6-1
Percent of quality targets¹ met and total points² earned from PY1 to PY5

PGP	PY1 % targets met	PY1 % of points earned	PY2 % targets met	PY2 % of points earned	PY3 % targets met	PY3 % of points earned	PY4 % targets met	PY4 % of points earned	PY5 % targets met	PY5 % of points earned
PGP 1	80%	90.9%	96%	97.8%	97%	98.1%	97%	92.5%	100%	100.0%
PGP 2	90%	95.5%	96%	97.8%	88%	92.5%	91%	94.3%	94%	96.2%
PGP 3	70%	86.4%	93%	95.6%	91%	94.3%	91%	94.3%	100%	100.0%
PGP 4	100%	100.0%	100%	100.0%	94%	96.2%	94%	96.2%	100%	100.0%
PGP 5	70%	72.7%	100%	100.0%	97%	100.0%	100%	100.0%	100%	100.0%
PGP 6	90%	81.8%	100%	100.0%	97%	98.1%	100%	100.0%	97%	98.1%
PGP 7	70%	86.4%	93%	95.6%	88%	92.5%	91%	94.3%	100%	100.0%
PGP 8	90%	95.5%	96%	97.8%	100%	100.0%	100%	100.0%	100%	100.0%
PGP 9	100%	100.0%	100%	100.0%	94%	96.2%	97%	98.1%	100%	100.0%
PGP 10	80%	95.5%	100%	100.0%	91%	94.3%	94%	96.2%	97%	98.1%

¹ 10 measures were in effect in PY1; 27 measures were in effect in PY2; and 32 measures were in effect in PY3, PY4 and PY5.

² Each claims-based measure was worth 4 points, and each chart-based measure was worth 1 point. Therefore, the total available points earned were 22 points in PY1, 45 points in PY2, and 53 points in PY3 through PY5.

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.2 Diabetes (DM) Module Quality Performance

6.2.1 Description of DM Quality Measures

The 10 DM quality measures included a range of testing, clinical results, and preventive care measures. Summary descriptions of them are as follows:

- DM-1: Percentage of diabetic patients with one or more A1c test(s).
- DM-2: Percentage of diabetic patients with most recent A1c level > 9.0% (poor control).
- DM-3: Percentage of diabetic patients with most recent BP < 140/90 mmHg.

- DM-4: Percentage of diabetic patients with at least one low-density lipoprotein (LDL) cholesterol test.
- DM-5: Percentage of diabetic patients with most recent LDL cholesterol < 130 mg/dl.
- DM-6: Percentage of diabetic patients with at least one test for microalbumin during the measurement year; or who had evidence of medical attention for existing nephropathy (diagnosis of nephropathy or documentation of microalbuminuria or albuminuria).
- DM-7: Percentage of diabetic patients who received a dilated eye exam or evaluation of retinal photographs by an optometrist or ophthalmologist during the measurement year, or during the prior year.
- DM-8: Percentage of eligible diabetic patients receiving at least one complete foot exam (including visual inspection, sensory exam with monofilament, and pulse exam).
- DM-9: Percentage of diabetic patients 50 years and older who received an influenza vaccination from September through February of the year prior to the measurement year.
- DM-10: Percentage of diabetic patients 65 years and older who ever received a pneumococcal vaccination.

6.2.2 DM Results for BY and PY1 - PY5 by PGP

Tables 6-2a through 6-2e present the DM quality measure performance results for the base year (BY), PY1, PY2, PY3, PY4 and PY5 for all 10 PGPs. As a result, this table includes a total of 600 quality measurement assessments made to date for the DM module for the PGP Demonstration (10 DM quality measures each for all 10 PGPs over six years). Quality results that met a PGP Demonstration performance target are highlighted in bold in Tables 6-2a through 6-2e for PY1 through PY5.

6.2.3 Description of Trends by PGP

Tables 6-2a through 6-2e show that all PGPs increased their scores on most of the DM quality measures over the course of the Demonstration. In the first measurement year (PY1), comparing scores in PY1 to the BY year, most PGPs improved on a majority of the quality measures. Only two PGPs improved on less than 5 DM measures. During years PY2 through PY5, all PGPs demonstrated continued improvements in DM measure scores, with each PGP improving in six or more (compared to BY) measures in years PY3-PY5. In the final year of the Demonstration, PY5, seven of the 10 PGPs improved in seven or more DM quality measures and all but one of the PGPs was meeting one or more of the performance targets for all 10 quality measures. This was accomplished in spite of the fact that the first two quality measures, DM-1 and DM-2, had BY results above 90 percent for most PGPs, indicating that improvements in performance results were difficult to achieve for those two measures given the high baseline levels.

Table 6-2a
Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 1 and 2

Quality Measure	PGP 1 BY	PGP 1 PY1	PGP 1 PY2	PGP 1 PY3	PGP 1 PY4	PGP 1 PY5	PGP 2 BY	PGP 2 PY1	PGP 2 PY2	PGP 2 PY3	PGP 2 PY4	PGP 2 PY5
DM-1 HbA1c management	91	90	91	91	90	95	91	91	96	96	98	96
DM-2 HbA1c Management Control (HbA1c ≤ 9.0%) ¹	94	94	96	97	95	87	93	94	91	95	85	85
DM-3 Blood Pressure Management (<140/90 mm Hg)	67	68	67	76	75	74	72	68	66	61	72	68
DM-4 Lipid Measurement	83	85	85	85	82	87	79	84	89	91	90	91
DM-5 LDL Cholesterol Level (<130md/dL)	81	87	89	90	84	86	87	90	88	81	82	77
DM-6 Urine Protein Testing	71	82	88	83	85	85	69	76	78	85	87	87
DM-7 Eye Exam	75	79	77	77	76	74	77	80	81	81	80	82
DM-8 Complete Foot Exam	32	51	58	59	59	56	35	44	43	55	64	69
DM-9 Influenza Vaccination	77	74	78	82	84	78	58	77	81	80	82	81
DM-10 Pneumonia Vaccination	49	75	87	87	95	88	59	81	79	88	87	88

Table 6-2b
Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 3 and 4

Quality Measure	PGP 3 BY	PGP 3 PY1	PGP 3 PY2	PGP 3 PY3	PGP 3 PY4	PGP 3 PY5	PGP 4 BY	PGP 4 PY1	PGP 4 PY2	PGP 4 PY3	PGP 4 PY4	PGP 4 PY5
DM-1 HbA1c management	95	95	94	98	97	99	97	95	96	95	98	99
DM-2 HbA1c Management Control (HbA1c ≤ 9.0%) ¹	96	92	88	91	88	88	91	93	93	94	92	91
DM-3 Blood Pressure Management (<140/90 mm Hg)	77	67	69	70	70	75	60	74	67	62	70	71
DM-4 Lipid Measurement	88	89	91	91	90	93	87	88	90	91	91	92
DM-5 LDL Cholesterol Level (<130md/dL)	85	81	85	83	84	80	88	90	88	87	87	86
DM-6 Urine Protein Testing	88	92	86	91	92	91	74	89	74	73	69	91
DM-7 Eye Exam	75	82	80	75	77	73	66	78	70	70	72	71
DM-8 Complete Foot Exam	15	13	38	64	71	72	19	44	54	46	48	54
DM-9 Influenza Vaccination	65	62	68	75	68	73	66	77	82	81	74	76
DM-10 Pneumonia Vaccination	30	53	71	82	84	91	66	73	74	79	77	73

Table 6-2c
Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 5 and 6

Quality Measure	PGP 5	PGP 5	PGP 5	PGP 5	PGP 5	PGP 5	PGP 6	PGP 6	PGP 6	PGP 6	PGP 6	PGP 6
	BY	PY1	PY2	PY3	PY4	PY5	BY	PY1	PY2	PY3	PY4	PY5
DM-1 HbA1c management	94	96	97	98	98	99	95	95	96	98	98	98
DM-2 HbA1c Management Control (HbA1c ≤ 9.0%) ¹	91	88	90	91	86	88	97	93	96	94	89	92
DM-3 Blood Pressure Management (<140/90 mm Hg)	74	64	76	76	75	78	58	67	74	72	72	73
DM-4 Lipid Measurement	81	86	93	92	95	93	84	86	88	90	93	95
DM-5 LDL Cholesterol Level (<130md/dL)	81	83	85	87	87	87	77	88	87	87	84	86
DM-6 Urine Protein Testing	69	70	97	97	95	89	80	85	89	90	93	94
DM-7 Eye Exam	69	69	76	73	74	74	79	82	81	80	77	77
DM-8 Complete Foot Exam	8	6	18	36	58	63	14	25	60	67	71	72
DM-9 Influenza Vaccination	68	77	90	89	91	90	59	79	78	86	89	80
DM-10 Pneumonia Vaccination	76	87	96	94	95	91	79	88	93	95	93	93

Table 6-2d
Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 7 and 8

Quality Measure	PGP 7	PGP 7	PGP 7	PGP 7	PGP 7	PGP 7	PGP 8	PGP 8	PGP 8	PGP 8	PGP 8	PGP 8
	BY	PY1	PY2	PY3	PY4	PY5	BY	PY1	PY2	PY3	PY4	PY5
DM-1 HbA1c management	96	94	97	98	97	96	94	98	98	99	99	99
DM-2 HbA1c Management Control (HbA1c ≤ 9.0%) ¹	95	94	94	92	91	93	96	95	94	91	92	93
DM-3 Blood Pressure Management (<140/90 mm Hg)	66	71	76	79	80	78	73	72	74	80	81	86
DM-4 Lipid Measurement	93	94	95	96	95	92	86	88	91	93	92	92
DM-5 LDL Cholesterol Level (<130md/dL)	87	88	91	86	89	89	87	90	94	89	93	89
DM-6 Urine Protein Testing	73	70	75	75	79	80	87	89	96	98	96	97
DM-7 Eye Exam	78	80	79	78	79	80	82	81	81	81	79	76
DM-8 Complete Foot Exam	48	29	46	50	62	56	16	27	35	31	27	35
DM-9 Influenza Vaccination	72	71	77	78	83	75	61	86	88	85	85	90
DM-10 Pneumonia Vaccination	69	63	67	72	78	84	78	87	91	92	93	94

Table 6-2e
Diabetes Mellitus (DM) results for BY and PY1 - PY5, by PGP 9 and 10

Quality Measure	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10
	BY	PY1	PY2	PY3	PY4	PY5	BY	PY1	PY2	PY3	PY4	PY5
DM-1 HbA1c management	88	92	90	96	96	96	85	87	91	96	97	97
DM-2 HbA1c Management Control (HbA1c ≤ 9.0%) ¹	93	93	93	95	87	86	91	91	92	93	86	82
DM-3 Blood Pressure Management (<140/90 mm Hg)	45	67	68	70	70	73	66	66	70	69	74	77
DM-4 Lipid Measurement	81	88	85	90	91	91	74	80	89	90	88	88
DM-5 LDL Cholesterol Level (<130md/dL)	75	88	86	79	83	80	91	90	89	80	84	84
DM-6 Urine Protein Testing	62	80	79	80	78	84	76	81	87	87	87	89
DM-7 Eye Exam	60	65	65	74	71	68	44	56	72	72	74	73
DM-8 Complete Foot Exam	11	65	62	64	59	63	35	42	59	56	54	52
DM-9 Influenza Vaccination	48	74	80	81	72	81	78	79	87	88	85	88
DM-10 Pneumonia Vaccination	63	74	85	88	84	85	73	76	81	86	87	89

74 ¹ Expressed in reverse to indicate ‘HbA1c in control’: as 100% minus percent of patients with poor control.

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.2.4 Description of Trends by DM Measure

Tables 6-2a through 6-2e also show that performance scores improved on most DM quality measures across PGPs. As noted, the scores on DM-1 and DM-2 started out very high, so improvements for those quality measures were less likely. Overall, there was a slight decrease in performance on DM-2 between BY and PY5 at nine of the sites. However, levels remained close to or above 90 percent and all sites continued to meet performance targets throughout the Demonstration. For DM-1 PGPs were able to maintain or improve their already high performance throughout the Demonstration.

In contrast, the last three DM quality measures, DM-8, DM-9, and DM-10, had the lowest baseline performance levels, which were all below 80 percent for all 10 PGPs. As a result, these quality measures showed some of the largest improvements between the BY and PY5. DM-8 (Complete Foot Exam) had very low BY scores, never above 50 percent for all of the 10 PGPs, and often below 20 percent. This is probably due to the strict clinical requirements in the specifications for this measure, which requires that three types of foot testing all be done in order to achieve a positive quality score for testing for this neuropathy quality measure on diabetic patients. Many PGPs reported that their clinical guidelines prior to joining the Demonstration recommended applying one or two of these testing procedures, but not all three. Nonetheless, after the five years of the Demonstration, six of the PGPs improved by over 30 percentage points on DM-8. DM-10 Pneumonia Vaccination was another quality measure with lower levels of BY performance. After joining the Demonstration, eight of the PGPs improved by over 20 percentage points on this quality measure.

6.2.5 Number of Targets Hit per PGP for DM in PY1 through PY5

Table 6-3 includes a summary of the number of quality performance targets hit for each of the PGPs for the DM quality measures in PY1 through PY5. It shows that a total of 84 performance targets out of 100 possible were met in PY1 and this number steadily increased to 99 out of 100 possible targets met in PY5. In PY5, nine of the 10 PGPs met all 10 performance targets and the remaining PGP met nine targets. The one unmet target was DM-3, for just one PGP.

Table 6-3
Number of quality targets hit per PGP for Diabetes Mellitus (DM) in PY1 -PY5

PGP	Number of DM Targets Hit PY1	Number of DM Targets Hit PY2	Number of DM Targets Hit PY3	Number of DM Targets Hit PY4	Number of DM Targets Hit PY5
PGP 1	8	9	10	9	10
PGP 2	9	9	9	9	9
PGP 3	7	8	9	8	10
PGP 4	10	10	9	10	10
PGP 5	7	10	9	10	10

(continued)

Table 6-3 (continued)
Number of quality targets hit per PGP for Diabetes Mellitus (DM) in PY1 -PY5

PGP	Number of DM Targets Hit PY1	Number of DM Targets Hit PY2	Number of DM Targets Hit PY3	Number of DM Targets Hit PY4	Number of DM Targets Hit PY5
PGP 6	9	10	10	10	10
PGP 7	7	8	9	10	10
PGP 8	9	9	10	10	10
PGP 9	10	10	10	10	10
PGP 10	8	10	9	10	10
TOTAL	84	93	94	96	99

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.2.6 Number of PGPs Hitting a Target by Measure

Table 6-4 includes a summary of the number of PGPs meeting a quality performance target for each of the DM quality measures in PY1 through PY5. Following the previous table, this also shows that in PY5, 99 out of a possible 100 quality targets were met, and DM-3 was the least frequently hit target. However it was also the measure with the greatest improvement in number of targets hit. In PY1 only four PGPs were able to hit this target, and this number increased every year, with nine PGPs hitting a target for DM-3 in PY5. All PGPs met their targets throughout all five years for DM-1, DM-2, DM-5 and DM-6.

Table 6-4
Number of PGPs hitting a quality target in PY1 – PY5 by Diabetes Mellitus (DM) measure

DM Quality Measure	Number of PGPs Hitting Target PY1	Number of PGPs Hitting Target PY2	Number of PGPs Hitting Target PY3	Number of PGPs Hitting Target PY4	Number of PGPs Hitting Target PY5
DM-1 HbA1c Testing	10	10	10	10	10
DM-2 HbA1c Control	10	10	10	10	10
DM-3 Blood Pressure Management	4	6	6	8	9
DM-4 Lipid Measurement	9	10	10	9	10
DM-5 LDL Cholesterol Level	10	10	10	10	10
DM-6 Urine Protein Testing	10	10	10	10	10
DM-7 Eye Exam	8	10	10	10	10
DM-8 Complete Foot Exam	7	9	9	10	10
DM-9 Influenza Vaccination	7	9	10	9	10
DM-10 Pneumonia Vaccination	9	9	10	10	10

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.3 Heart Failure (HF) Module Quality Performance

6.3.1 Description of HF Quality Measures

The 10 HF quality measures included a range of testing, clinical results, pharmaceutical prescription, and preventive care measures. Summary descriptions of them are as follows:

- HF-1: Percentage of HF patients who have quantitative or qualitative results of LVEF assessment recorded.
- HF-2: Percentage of patients hospitalized with a principal diagnosis of HF during the current year who had left ventricular ejection fraction testing during the current year.
- HF-3: Percentage of HF patient visits with weight measurement recorded.
- HF-4: Percentage of HF patient visits with blood pressure measurement recorded.
- HF-5: Percentage of HF patients who were provided with patient education on disease management and health behavior changes during one or more visit(s) within a six-month period.
- HF-6: Percentage of HF patients who also have LVSD who were prescribed beta-blocker therapy.
- HF-7: Percentage of HF patients who also have LVSD who were prescribed ACE inhibitor therapy.
- HF-8: Percentage of HF patients who also have paroxysmal or chronic atrial fibrillation who were prescribed warfarin therapy.
- HF-9: Percentage of HF patients 50 years and older who received an influenza vaccination from September through February of the year prior to the measurement year.
- HF-10: Percentage of HF patients 65 years and older who ever received a pneumococcal vaccination.

6.3.2 HF Results for BY and PY2 - PY5 by PGP

Tables 6-5a through 6-5e present the HF quality measure performance results for the BY and PY2 through PY5 for all 10 PGPs. As a result, this table includes a total of 500 quality measurement assessments made to date for the HF module for the PGP Demonstration. In contrast to the DM module, the HF module was not active in PY1, so there are no quality performance results available for HF in PY1. Quality results that met at least one of the PGP Demonstration quality performance targets are highlighted in bold for PY2 through PY5.

Table 6-5a
Heart Failure (HF) results for BY and PY2 - PY5, by PGP 1 and 2

Quality Measure	PGP 1 BY	PGP 1 PY1	PGP 1 PY2	PGP 1 PY3	PGP 1 PY4	PGP 1 PY5	PGP 2 BY	PGP 2 PY1	PGP 2 PY2	PGP 2 PY3	PGP 2 PY4	PGP 2 PY5
HF-1 Left Ventricular Function Assessment	86	N/A	93	92	96	97	89	N/A	98	98	98	99
HF-2 Left Ventricular Ejection Fraction Testing	84	N/A	83	84	86	100	92	N/A	94	93	94	91
HF-3 Weight Measurement	73	N/A	83	83	88	88	69	N/A	75	73	75	77
HF-4 Blood Pressure Screening	92	N/A	98	99	99	99	93	N/A	96	97	97	97
HF-5 Patient Education	62	N/A	96	91	95	81	83	N/A	83	83	82	84
HF-6 Beta-Blocker Therapy	89	N/A	97	97	99	100	93	N/A	96	96	96	94
HF-7 ACE Inhibitor or ARB Therapy	94	N/A	98	97	99	99	88	N/A	98	91	94	95
HF-8 Warfarin Therapy for Patients	88	N/A	95	92	94	97	94	N/A	89	88	92	89
HF-9 Influenza Vaccination	63	N/A	84	83	90	81	61	N/A	77	78	82	85
HF-10 Pneumonia Vaccination	44	N/A	84	91	94	88	53	N/A	73	80	84	87

Table 6-5b
Heart Failure (HF) results for BY and PY2 - PY5, by PGP 3 and 4

Quality Measure	PGP 3 BY	PGP 3 PY1	PGP 3 PY2	PGP 3 PY3	PGP 3 PY4	PGP 3 PY5	PGP 4 BY	PGP 4 PY1	PGP 4 PY2	PGP 4 PY3	PGP 4 PY4	PGP 4 PY5
HF-1 Left Ventricular Function Assessment	75	N/A	95	97	98	99	79	N/A	92	94	97	95
HF-2 Left Ventricular Ejection Fraction Testing	81	N/A	89	94	88	91	88	N/A	93	85	95	88
HF-3 Weight Measurement	87	N/A	81	91	92	90	87	N/A	94	94	93	93
HF-4 Blood Pressure Screening	96	N/A	95	99	99	99	95	N/A	98	98	98	98
HF-5 Patient Education	99	N/A	99	100	94	99	16	N/A	74	77	78	73
HF-6 Beta-Blocker Therapy	78	N/A	99	97	100	97	92	N/A	98	97	99	96
HF-7 ACE Inhibitor or ARB Therapy	80	N/A	97	98	98	95	89	N/A	93	93	95	90
HF-8 Warfarin Therapy for Patients	81	N/A	90	97	96	95	70	N/A	84	86	89	84
HF-9 Influenza Vaccination	61	N/A	73	72	77	76	62	N/A	79	78	76	73
HF-10 Pneumonia Vaccination	23	N/A	72	83	90	89	36	N/A	67	66	72	77

Table 6-5c
Heart Failure (HF) results for BY and PY2 - PY5, by PGP 5 and 6

Quality Measure	PGP 5 BY	PGP 5 PY1	PGP 5 PY2	PGP 5 PY3	PGP 5 PY4	PGP 5 PY5	PGP 6 BY	PGP 6 PY1	PGP 6 PY2	PGP 6 PY3	PGP 6 PY4	PGP 6 PY5
HF-1 Left Ventricular Function Assessment	79	N/A	89	97	90	94	91	N/A	96	98	98	99
HF-2 Left Ventricular Ejection Fraction Testing	92	N/A	92	91	93	95	87	N/A	95	89	92	90
HF-3 Weight Measurement	93	N/A	95	97	96	94	88	N/A	93	92	94	94
HF-4 Blood Pressure Screening	98	N/A	99	99	99	99	97	N/A	99	99	99	99
HF-5 Patient Education	72	N/A	83	92	90	91	40	N/A	91	82	92	92
HF-6 Beta-Blocker Therapy	83	N/A	96	94	94	95	95	N/A	97	99	96	97
HF-7 ACE Inhibitor or ARB Therapy	85	N/A	87	92	91	91	96	N/A	99	95	95	100
HF-8 Warfarin Therapy for Patients	85	N/A	97	86	83	84	95	N/A	94	98	91	99
HF-9 Influenza Vaccination	62	N/A	81	87	87	90	58	N/A	86	88	87	85
HF-10 Pneumonia Vaccination	77	N/A	87	91	88	90	83	N/A	89	92	90	92

Table 6-5d
Heart Failure (HF) results for BY and PY2 - PY5, by PGP 7 and 8

Quality Measure	PGP 7 BY	PGP 7 PY1	PGP 7 PY2	PGP 7 PY3	PGP 7 PY4	PGP 7 PY5	PGP 8 BY	PGP 8 PY1	PGP 8 PY2	PGP 8 PY3	PGP 8 PY4	PGP 8 PY5
HF-1 Left Ventricular Function Assessment	84	N/A	93	97	97	96	93	N/A	97	99	98	99
HF-2 Left Ventricular Ejection Fraction Testing	95	N/A	95	93	90	96	90	N/A	92	100	95	89
HF-3 Weight Measurement	N/A	N/A	84	81	85	91	80	N/A	87	88	88	88
HF-4 Blood Pressure Screening	N/A	N/A	92	89	92	96	94	N/A	99	99	100	100
HF-5 Patient Education	61	N/A	89	98	97	97	32	N/A	91	88	97	94
HF-6 Beta-Blocker Therapy	91	N/A	95	97	97	97	89	N/A	98	98	99	99
HF-7 ACE Inhibitor or ARB Therapy	90	N/A	91	91	98	94	92	N/A	94	100	97	99
HF-8 Warfarin Therapy for Patients	92	N/A	96	95	93	98	73	N/A	95	95	95	93
HF-9 Influenza Vaccination	67	N/A	72	72	88	76	78	N/A	88	87	87	93
HF-10 Pneumonia Vaccination	55	N/A	63	65	88	84	74	N/A	83	88	87	90

Table 6-5e
Heart Failure (HF) results for BY and PY2 - PY5, by PGP 9 and 10

Quality Measure	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10
	BY	PY1	PY2	PY3	PY4	PY5	BY	PY1	PY2	PY3	PY4	PY5
HF-1 Left Ventricular Function Assessment	81	N/A	93	93	96	96	87	N/A	91	99	99	99
HF-2 Left Ventricular Ejection Fraction Testing	93	N/A	90	94	92	95	91	N/A	92	96	92	95
HF-3 Weight Measurement	81	N/A	87	89	90	90	66	N/A	70	91	94	92
HF-4 Blood Pressure Screening	95	N/A	97	95	98	99	80	N/A	89	97	99	99
HF-5 Patient Education	91	N/A	93	93	95	96	90	N/A	89	98	94	94
HF-6 Beta-Blocker Therapy	76	N/A	93	92	91	97	81	N/A	87	99	99	98
HF-7 ACE Inhibitor or ARB Therapy	91	N/A	95	89	93	96	56	N/A	72	96	96	96
HF-8 Warfarin Therapy for Patients	75	N/A	79	86	92	88	42	N/A	59	92	92	96
HF-9 Influenza Vaccination	56	N/A	87	86	76	84	79	N/A	87	89	93	91
HF-10 Pneumonia Vaccination	69	N/A	90	92	90	89	76	N/A	81	91	96	93

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.3.3 Description of Trends by PGP

Tables 6-5a through 6-5e show that the PGPs increased their scores on most of the HF quality measures over the course of the Demonstration. Comparing scores between the BY and PY2, the first year of measurement, all of the PGPs improved on at least seven of the quality measures. Moreover, two PGPs improved on all 10 measures. In PY3 through PY5 performance continued to improve with PGPs consistently improving in seven or more measures. By PY5 all sites had improved on at least 8 measures. For the few quality measures where improvements were not made by PY5, most were already at a 90 percent or higher performance level, indicating that additional improvements in performance results were difficult to achieve given the already high levels.

6.3.4 Description of Trends by HF Measure

Tables 6-5a through 6-5e also show that performance scores improved on most HF quality measures across PGPs. Three HF quality measures had consistently low baseline quality performance scores across PGPs, including HF-5 Patient Education, HF-9 Influenza Vaccination, and HF-10 Pneumonia Vaccination. HF-5 baseline scores were above 80 percent for only half of the PGPs, and were below 50 percent for three PGPs. HF-9 baseline scores were below 70 percent for 8 PGPs. HF-10 baseline scores were below 70 percent for 6 PGPs. As a result, these HF quality measures showed some large improvements between the BY and PY5. HF-5 scores improved by over 25 percentage points for six of the PGPs, due to improved HF patient education protocols implemented by the PGPs after joining the PGP Demonstration. HF-9 scores improved by at least 10 percentage points for all 10 of the PGPs. HF-9 scores improved by at least 20 percentage points for nine of the PGPs.

6.3.5 Number of Targets Hit per PGP for HF in PY2-PY5

Table 6-6 includes a summary of the number of quality performance targets hit for each of the PGPs for the HF quality measures in PY2 through PY5. It shows that all of the 400 performance targets possible were met in PY2-PY5 (10 quality measure targets for each of the 10 PGPs over 4 years). All PGPs were able to meet one or more of the quality targets from the first year of measurement and maintain high levels of performance throughout the Demonstration.

6.3.6 Number of PGPs Hitting a Target by HF Measure in PY2 – PY5

Table 6-7 includes a summary of the number of PGPs hitting a quality performance target for the HF quality measures in PY2 through PY5. Following the previous table, this also shows that all of the 400 possible performance targets were met in PY2 through PY5 (10 PGPs meeting targets for each of 10 quality measures over 4 years).

Table 6-6
Number of quality targets hit per PGP for Heart Failure (HF) in PY2-PY5

PGP	Number of HF Targets Hit PY1	Number of HF Targets Hit PY2	Number of HF Targets Hit PY3	Number of HF Targets Hit PY4	Number of HF Targets Hit PY5
PGP 1	N/A	10	10	10	10
PGP 2	N/A	10	10	10	10
PGP 3	N/A	10	10	10	10
PGP 4	N/A	10	10	10	10
PGP 5	N/A	10	10	10	10
PGP 6	N/A	10	10	10	10
PGP 7	N/A	10	10	10	10
PGP 8	N/A	10	10	10	10
PGP 9	N/A	10	10	10	10
PGP 10	N/A	10	10	10	10
TOTAL	N/A	100	100	100	100

SOURCE: RTI analysis of 2005–2010 Medicare claims.

Table 6-7
Number of PGPs hitting a quality target in PY2-PY5 by Heart Failure (HF) measure

HF Quality Measure	Number of PGPs Hitting Target PY1	Number of PGPs Hitting Target PY2	Number of PGPs Hitting Target PY3	Number of PGPs Hitting Target PY4	Number of PGPs Hitting Target PY5
HF-1 LVEF Assessment	N/A	10	10	10	10
HF-2 LVEF Testing for Hospitalized Patients	N/A	10	10	10	10
HF-3 Weight Measurement	N/A	10	10	10	10
HF-4 Blood Pressure Screening	N/A	10	10	10	10
HF-5 Patient Education	N/A	10	10	10	10
HF-6 Beta Blocker Therapy	N/A	10	10	10	10
HF-7 ACE Inhibitor or ARB Therapy	N/A	10	10	10	10
HF-8 Warfarin Therapy	N/A	10	10	10	10
HF-9 Influenza Vaccination	N/A	10	10	10	10
HF-10 Pneumonia Vaccination	N/A	10	10	10	10

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.4 Coronary Artery Disease (CAD) Module Quality Performance

6.4.1 Description of CAD Quality Measures

The 7 CAD quality measures included a range of testing, clinical results, pharmaceutical prescription, and preventive care measures. Summary descriptions of them are as follows:

- CAD-1: Percentage of CAD patients who were prescribed antiplatelet therapy.
- CAD-2: Percentage of CAD patients who were prescribed a lipid-lowering therapy (based on current ATP III guidelines).
- CAD-3: Percentage of CAD patients with prior MI who were prescribed beta-blocker therapy.
- CAD-4: Percentage of CAD patients who had a blood pressure measurement during the last office visit.
- CAD-5: Percentage of CAD patients receiving at least one lipid profile during the reporting year.
- CAD-6: Percentage of CAD patients with most recent LDL cholesterol < 130 mg/dl.
- CAD-7: Percentage of CAD patients who also have diabetes and/or LVSD who were prescribed ACE inhibitor therapy.

6.4.2 CAD Results for BY and PY2 – PY5 by PGP

Tables 6-8a through 6-8e present the CAD quality measure performance results for the BY and PY2 through PY5 for all 10 PGPs. As a result, this table includes a total of 350 quality measurement assessments made for the CAD module for the PGP Demonstration (10 PGPs each with 7 CAD quality measurements for 5 years). As with the HF module, the CAD module was not active in PY1, so there are no quality performance results available for CAD in PY1. Quality results that met a PGP Demonstration quality performance target are highlighted in bold in Tables 6-8a through 6-8e for PY2.

6.4.3 Description of Trends by PGP

Tables 6-8a through 6-8e show that the PGPs have increased their scores on most of the CAD quality measures over the course of the Demonstration. Comparing scores between the BY and PY2 (the first Demonstration year of measurement for CAD), eight of the PGPs improved on at least five of the seven CAD quality measures. Moreover, three PGPs improved on all seven measures. Positive trends continued and by PY5 all sites had improved on four or more measures and four PGPs improved on all seven measures. When quality improvements were not found, the PGPs were already performing close to 90 percent or higher, indicating that additional improvement in performance results was difficult to achieve.

Table 6-8a
Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 1 and 2

Quality Measure	PGP 1 BY	PGP 1 PY1	PGP 1 PY2	PGP 1 PY3	PGP 1 PY4	PGP 1 PY5	PGP 2 BY	PGP 2 PY1	PGP 2 PY2	PGP 2 PY3	PGP 2 PY4	PGP 2 PY5
CAD-1 Antiplatelet Therapy	90	N/A	93	95	97	90	91	N/A	88	90	93	93
CAD-2 Drug Therapy for Lowering LDL Cholesterol	97	N/A	96	98	98	100	97	N/A	97	98	98	98
CAD-3 Beta-Blocker Therapy – Prior MI	72	N/A	93	92	95	91	92	N/A	93	92	93	95
CAD-4 Blood Pressure	96	N/A	97	99	99	99	93	N/A	96	98	100	100
CAD-5 Lipid Profile	73	N/A	79	77	81	80	73	N/A	81	77	76	77
CAD-6 LDL Cholesterol Level	92	N/A	92	92	87	83	93	N/A	86	90	77	80
CAD-7 ACE Inhibitor or ARB Therapy	80	N/A	91	87	93	88	85	N/A	84	82	84	82

Table 6-8b
Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 3 and 4

Quality Measure	PGP 3 BY	PGP 3 PY1	PGP 3 PY2	PGP 3 PY3	PGP 3 PY4	PGP 3 PY5	PGP 4 BY	PGP 4 PY1	PGP 4 PY2	PGP 4 PY3	PGP 4 PY4	PGP 4 PY5
CAD-1 Antiplatelet Therapy	68	N/A	83	84	89	93	87	N/A	92	93	95	93
CAD-2 Drug Therapy for Lowering LDL Cholesterol	91	N/A	96	97	97	97	88	N/A	96	96	96	98
CAD-3 Beta-Blocker Therapy – Prior MI	82	N/A	92	95	92	92	73	N/A	90	90	91	88
CAD-4 Blood Pressure	86	N/A	95	100	100	100	98	N/A	99	98	99	99
CAD-5 Lipid Profile	74	N/A	81	82	84	84	78	N/A	84	82	84	87
CAD-6 LDL Cholesterol Level	92	N/A	88	91	85	84	86	N/A	88	88	88	87
CAD-7 ACE Inhibitor or ARB Therapy	72	N/A	85	93	92	92	74	N/A	81	82	88	82

Table 6-8c
Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 5 and 6

Quality Measure	PGP 5 BY	PGP 5 PY1	PGP 5 PY2	PGP 5 PY3	PGP 5 PY4	PGP 5 PY5	PGP 6 BY	PGP 6 PY1	PGP 6 PY2	PGP 6 PY3	PGP 6 PY4	PGP 6 PY5
CAD-1 Antiplatelet Therapy	88	N/A	90	93	91	95	86	N/A	90	92	94	95
CAD-2 Drug Therapy for Lowering LDL Cholesterol	90	N/A	92	95	96	96	95	N/A	97	98	99	99
CAD-3 Beta-Blocker Therapy – Prior MI	87	N/A	95	93	91	92	91	N/A	93	93	93	94
CAD-4 Blood Pressure	98	N/A	98	100	99	100	97	N/A	99	100	99	98
CAD-5 Lipid Profile	67	N/A	83	78	77	78	78	N/A	81	89	91	87
CAD-6 LDL Cholesterol Level	88	N/A	91	87	85	86	88	N/A	93	90	89	85
CAD-7 ACE Inhibitor or ARB Therapy	72	N/A	80	82	84	80	68	N/A	85	87	92	91

Table 6-8d
Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 7 and 8

Quality Measure	PGP 7 BY	PGP 7 PY1	PGP 7 PY2	PGP 7 PY3	PGP 7 PY4	PGP 7 PY5	PGP 8 BY	PGP 8 PY1	PGP 8 PY2	PGP 8 PY3	PGP 8 PY4	PGP 8 PY5
CAD-1 Antiplatelet Therapy	89	N/A	95	95	92	91	95	N/A	94	94	93	97
CAD-2 Drug Therapy for Lowering LDL Cholesterol	93	N/A	98	98	98	98	94	N/A	96	99	98	99
CAD-3 Beta-Blocker Therapy – Prior MI	85	N/A	92	92	96	94	85	N/A	96	97	92	95
CAD-4 Blood Pressure	90	N/A	99	99	100	100	94	N/A	99	99	100	100
CAD-5 Lipid Profile	83	N/A	88	93	88	88	78	N/A	83	83	85	80
CAD-6 LDL Cholesterol Level	86	N/A	95	92	91	96	93	N/A	93	90	86	93
CAD-7 ACE Inhibitor or ARB Therapy	87	N/A	89	89	95	93	85	N/A	84	94	93	87

Table 6-8e
Coronary Artery Disease (CAD) results for BY and PY2-PY5, by PGP 9 and 10

Quality Measure	PGP 9 BY	PGP 9 PY1	PGP 9 PY2	PGP 9 PY3	PGP 9 PY4	PGP 9 PY5	PGP 10 BY	PGP 10 PY1	PGP 10 PY2	PGP 10 PY3	PGP 10 PY4	PGP 10 PY5
CAD-1 Antiplatelet Therapy	74	N/A	84	84	88	91	96	N/A	95	92	93	94
CAD-2 Drug Therapy for Lowering LDL Cholesterol	87	N/A	95	96	95	94	98	N/A	99	99	99	100
CAD-3 Beta-Blocker Therapy – Prior MI	59	N/A	76	78	78	84	93	N/A	96	94	93	94
CAD-4 Blood Pressure	97	N/A	97	99	99	100	94	N/A	98	97	98	99
CAD-5 Lipid Profile	75	N/A	83	84	82	86	66	N/A	84	84	89	87
CAD-6 LDL Cholesterol Level	84	N/A	87	90	83	85	93	N/A	89	90	85	85
CAD-7 ACE Inhibitor or ARB Therapy	77	N/A	82	81	81	91	86	N/A	89	90	85	91

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.4.4 Description of Trends by CAD Measure

Tables 6-8a through 6-8e show that performance scores improved on most CAD quality measures across the PGPs. However, one quality measure, CAD-6 LDL Cholesterol Level, failed to improve in PY5 for six of the 10 PGPs and many PGPs had fluctuating CAD-6 score over the years. Four of the PGPs that failed to show improvement in this measure by PY5 had baseline scores above 90 percent and the other two PGPs who did not improve in this measure had baseline scores above 80 percent, again indicating that additional improvements would be hard to achieve. All sites maintained high levels (between 77 and 99 percent in all years) and consistently met targets for this CAD measure.

None of the CAD quality measures had consistently low baseline scores across PGPs in the same way as was found for some measures in the DM and HF modules. Only one quality measure, CAD-5 Lipid Profile, had baseline scores below 70 percent for even as many as two PGPs.

6.4.5 Number of Targets Hit per PGP for CAD in PY2-PY5

Table 6-9 includes a summary of the number of quality performance targets hit for each of the PGPs for the CAD quality measures in PY2 through PY5. It shows that 279 of the possible 280 targets possible were met in PY2 through PY5 (7 CAD quality measure targets for each of the 10 PGPs across 4 years), with only one PGP missing only one target in PY3

Table 6-9
Number of quality targets hit per PGP for Coronary Artery Disease (CAD) in PY2-PY5

PGP	Number of CAD Targets Hit PY1	Number of CAD Targets Hit PY2	Number of CAD Targets Hit PY3	Number of CAD Targets Hit PY4	Number of CAD Targets Hit PY5
PGP 1	N/A	7	7	7	7
PGP 2	N/A	7	6	7	7
PGP 3	N/A	7	7	7	7
PGP 4	N/A	7	7	7	7
PGP 5	N/A	7	7	7	7
PGP 6	N/A	7	7	7	7
PGP 7	N/A	7	7	7	7
PGP 8	N/A	7	7	7	7
PGP 9	N/A	7	7	7	7
PGP 10	N/A	7	7	7	7
TOTAL	N/A	70	69	70	70

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.4.6 Number of PGPs Hitting a Target by CAD Measure in PY2-PY5

Table 6-10 includes a summary of the number of PGPs hitting a quality performance target for the CAD quality measures in PY2-PY5. Following the previous table, this also shows that 279 of the possible 280 targets were hit (10 PGPs hitting targets for each of the 7 CAD quality measures across 4 years with only one missed target for only one quality measure for only one PY over the course of the 4 years of measurement).

Table 6-10
Number of PGPs hitting a quality target in PY2-PY5 by CAD measure

CAD Quality Measure	Number of PGPs Hitting Target PY1	Number of PGPs Hitting Target PY2	Number of PGPs Hitting Target PY3	Number of PGPs Hitting Target PY4	Number of PGPs Hitting Target PY5
CAD-1 Antiplatelet Therapy	N/A	10	10	10	10
CAD-2 Drug Therapy for Lowering LDL Cholesterol	N/A	10	10	10	10
CAD-3 Beta Blocker Therapy – Prior MI	N/A	10	9	10	10
CAD-4 Blood Pressure	N/A	10	10	10	10
CAD-5 Lipid Profile	N/A	10	10	10	10
CAD-6 LDL Cholesterol Level	N/A	10	10	10	10
CAD-7 ACE Inhibitor or ARB Therapy	N/A	10	10	10	10

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.5 Hypertension (HTN) Module Quality Performance

6.5.1 Description of HTN Measures

The HTN module included three quality measures covering testing, clinical levels, and care planning. Summary descriptions of them are as follows:

- HTN-1: Percentage of hypertensive patients’ visits with blood pressure measurement recorded.
- HTN-2: Percentage of hypertensive patients with last blood pressure < 140/90 mmHg.
- HTN-3: Percentage of hypertensive patients’ visits with either systolic blood pressure > 140 mmHg or diastolic blood pressure > 90 mmHg with a documented plan of care for hypertension.

6.5.2 HTN Results for BY and PY3 – PY5 by PGP

Tables 6-11a through 6-11e present the HTN quality measure performance results for the BY and PY3 through PY5 for nine of the 10 PGPs, and for the BY and PY5 years for one PGP. PGP 7 did not submit HTN data for PY3 or PY4, and as such Table 6-11 only includes the BY and PY5 data for this PGP. As a result the table includes a total of 114 quality measurement assessments made for the HTN module for the PGP Demonstration (9 PGPs with 3 HTN quality measurements over 4 years and 1 PGP with 3 HTN quality measurements over 2 years). The HTN module was not active in PY1 or PY2 therefore there are no quality performance results available for those years. Quality results that met a PGP Demonstration quality performance target are highlighted in bold in Tables 6-11a through 6-11e for PY3-PY5.

6.5.3 Description of Trends by PGP

Comparing scores between BY and PY3 (the first year of measurement), two of the nine PGPs for which data was available improved on two of the three quality measures and three PGPs improved on all three quality measures. PGPs improved their performance in these measures throughout the Demonstration and by PY5 four (out of 10) PGPs had improved on two HTN measures and four more had improved on all three HTN measures. Moreover, the HTN quality measures were introduced later in the Demonstration giving PGPs less time to improve.

6.5.4 Description of Trends by HTN Measure

Tables 6-11a through 6-11e show that most sites were able to improve on at least two (out of three) HTN measures. All sites improved in HTN-1, blood pressure screening. The most challenging measure for PGPs was HTN-2 Blood Pressure Control, although the results for this measure did improve over the three years of study. Only three (out of nine) PGPs demonstrated improvement on HTN-2 in PY3, but five (out of nine) PGPs improved in PY4, and seven (out of 10) PGPs had improvements on this measure between PY5 and the BY.

6.5.5 Number of Targets Hit per PGP for HTN in PY3-PY5

Table 6-12 includes a summary of the number of quality performance targets hit for each of the PGPs for the HTN quality measures in PY3 through PY5. It shows that 62 of the possible 84 targets were met in PY3 through PY5 (three HTN quality measure targets each for nine PGPs across three years, and three HTN quality measure targets for PGP 7 in PY5). The number of targets hit each year increased over the course of the Demonstration. In PY3 four (out of nine) of the PGPs hit only one of their HTN targets and three PGPs hit two targets. By PY5 seven of the 10 PGPs hit all three of their HTN targets and the remaining three PGPs hit two out of three targets.

Table 6-11a
Hypertension (HTN) results for BY and PY3 – PY5, by PGP 1 and 2

Quality Measure	PGP 1 BY	PGP 1 PY1	PGP 1 PY2	PGP 1 PY3	PGP 1 PY4	PGP 1 PY5	PGP 2 BY	PGP 2 PY1	PGP 2 PY2	PGP 2 PY3	PGP 2 PY4	PGP 2 PY5
HTN-1 Blood Pressure Screening	95	N/A	N/A	99	99	99	94	N/A	N/A	98	98	98
HTN-2 Blood Pressure Control	71	N/A	N/A	66	71	72	93	N/A	N/A	61	64	67
HTN-3 Plan of Care	56	N/A	N/A	71	74	66	73	N/A	N/A	66	61	69

Table 6-11b
Hypertension (HTN) results for BY and PY3 – PY5, by PGP 3 and 4

Quality Measure	PGP 3 BY	PGP 3 PY1	PGP 3 PY2	PGP 3 PY3	PGP 3 PY4	PGP 3 PY5	PGP 4 BY	PGP 4 PY1	PGP 4 PY2	PGP 4 PY3	PGP 4 PY4	PGP 4 PY5
HTN-1 Blood Pressure Screening	78	N/A	N/A	99	100	99	95	N/A	N/A	97	99	98
HTN-2 Blood Pressure Control	66	N/A	N/A	62	62	69	69	N/A	N/A	65	63	67
HTN-3 Plan of Care	60	N/A	N/A	62	74	65	45	N/A	N/A	56	50	52

Table 6-11c
Hypertension (HTN) results for BY and PY3 – PY5, by PGP 5 and 6

Quality Measure	PGP 5 BY	PGP 5 PY1	PGP 5 PY2	PGP 5 PY3	PGP 5 PY4	PGP 5 PY5	PGP 6 BY	PGP 6 PY1	PGP 6 PY2	PGP 6 PY3	PGP 6 PY4	PGP 6 PY5
HTN-1 Blood Pressure Screening	97	N/A	N/A	99	100	100	98	N/A	N/A	99	99	99
HTN-2 Blood Pressure Control	69	N/A	N/A	70	71	78	97	N/A	N/A	70	71	75
HTN-3 Plan of Care	67	N/A	N/A	76	70	73	72	N/A	N/A	70	78	69

Table 6-11d
Hypertension (HTN) results for BY and PY3 – PY5, by PGP 7 and 8

Quality Measure	PGP 7 BY	PGP 7 PY1	PGP 7 PY2	PGP 7 PY3	PGP 7 PY4	PGP 7 PY5	PGP 8 BY	PGP 8 PY1	PGP 8 PY2	PGP 8 PY3	PGP 8 PY4	PGP 8 PY5
HTN-1 Blood Pressure Screening	—	N/A	N/A	N/A	N/A	98	93	N/A	N/A	99	100	100
HTN-2 Blood Pressure Control	—	N/A	N/A	N/A	N/A	72	69	N/A	N/A	73	80	80
HTN-3 Plan of Care	—	N/A	N/A	N/A	N/A	82	55	N/A	N/A	78	83	69

Table 6-11e
Hypertension (HTN) results for BY and PY3 – PY5, by PGP 9 and 10

Quality Measure	PGP 9 BY	PGP 9 PY1	PGP 9 PY2	PGP 9 PY3	PGP 9 PY4	PGP 9 PY5	PGP 10 BY	PGP 10 PY1	PGP 10 PY2	PGP 10 PY3	PGP 10 PY4	PGP 10 PY5
HTN-1 Blood Pressure Screening	97	N/A	N/A	95	95	99	96	N/A	N/A	98	99	98
HTN-2 Blood Pressure Control	63	N/A	N/A	63	66	69	66	N/A	N/A	64	69	67
HTN-3 Plan of Care	70	N/A	N/A	59	77	91	72	N/A	N/A	70	70	64

SOURCE: RTI analysis of 2006–2010 Medicare claims.

Table 6-12
Number of quality targets hit per PGP for Hypertension (HTN) PY3-PY5

PGP	Number of HTN Targets Hit PY1	Number of HTN Targets Hit PY2	Number of HTN Targets Hit PY3	Number of HTN Targets Hit PY4	Number of HTN Targets Hit PY5
PGP 1	N/A	N/A	2	3	3
PGP 2	N/A	N/A	1	1	2
PGP 3	N/A	N/A	1	2	3
PGP 4	N/A	N/A	2	1	3
PGP 5	N/A	N/A	3	3	3
PGP 6	N/A	N/A	2	3	2
PGP 7	N/A	N/A	NA	NA	3
PGP 8	N/A	N/A	3	3	3
PGP 9	N/A	N/A	1	2	3
PGP 10	N/A	N/A	1	1	2
TOTAL	NA	NA	16	19	27

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.5.6 Number of PGPs Hitting a Target by HTN Measure in PY3-PY5

Table 6-13 includes a summary of the number of PGPs hitting a quality performance target for the HTN quality measures in PY3-PY5. Following the previous table, this also shows that 62 of the possible 84 targets were hit. PGPs consistently met targets for HTN-1. All nine PGPs for which data was available met a target for this measure in PY3 and PY4, and all 10 PGPs met a target in PY5. PGPs had more difficulty meeting targets for HTN-2 and HTN-3. In PY3, only three (out of nine) PGPs met targets for HTN-2 and four PGPs met targets for HTN-3. However, improvements were made in both of these measures over time. By PY5 seven PGPs met a target for HTN-3 and all 10 met a target for HTN-2.

Table 6-13
Number of PGPs hitting a quality target in PY3-PY5 by HTN measures

HTN Quality Measure	Number of PGPs Hitting Target PY1	Number of PGPs Hitting Target PY2	Number of PGPs Hitting Target PY3	Number of PGPs Hitting Target PY4	Number of PGPs Hitting Target PY5
HTN-1 Blood Pressure Screening	N/A	N/A	9	9	10
HTN-2 Blood Pressure Control	N/A	N/A	3	4	10
HTN-3 Plan of Care	N/A	N/A	4	6	7

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.6 Preventative Care (PC) Modules Quality Performance

6.6.1 Description of PC Quality Measures

The PC module included two quality measures covering preventive care screening tests. Summary descriptions of them are as follows:

- PC-5: Percentage of female beneficiaries aged 50-69 years who had a mammogram during the measurement year or the year prior to the measurement year.
- PC-6: Percentage of beneficiaries aged 50 years or more who were screened for colorectal cancer during the one-year measurement period.

6.6.2 PC Results for the BY and PY3 – PY5 by PGP

Tables 6-14a through 6-14e present PC quality measure performance results for the BY and PY3 through PY5 for all 10 PGPs. As a result this table includes a total of 80 quality measurement assessments made for the PC module during the PGP Demonstration (10 PGPs with 2 PC quality measurements over 4 years). Like the HTN module, the PC module was not active in PY1 or PY2, and therefore there are no quality performance results available for those years. Quality results that met a PGP Demonstration quality performance target are highlighted in bold in Tables 6-14a through 6-14e for PY3-PY5.

6.6.3 Description of Trends by PGP

Performance improved on the two PC measures for most of the PGPs during the Demonstration. Comparing PY3 to the BY, six PGPs improved on both of their PC measures and the remaining four PGPs improved on one measure. In PY4, six PGPs again improved on both of their measures. Moreover, this number increased again in PY5, and seven of the PGPs improved on both of their PC measures and the other three PGPs improved in one. As in the HTN module, improvements were made over just three years of measurement, since these measurements were introduced later in the Demonstration, giving PGPs less time to improve.

6.6.4 Description of Trends by HTN and PC Measure

Because there were only two measures in this module there were a limited number of trends in the data. PGPs that failed to improve did not necessarily fail to improve in one measure more than the other. In PY5 two of the PGPs did not improve in breast cancer screening and one did not improve in colon cancer screening.

6.6.5 Number of Targets Hit per PGP for PC in PY3-PY5

Table 6-15 includes a summary of the number of quality performance targets met for each of the PGPs for the PC quality measures in PY3 through PY5. It shows that 60 of the possible 60 targets were met in PY3 through PY5 (two PC quality measure targets for each of the 10 PGPs across three years). All of the sites hit both of their PC targets in every year of the Demonstration.

Table 6-14a
Preventative Care (PC) results for BY and PY3 – PY5, by PGP 1 and 2

Quality Measure	PGP 1 BY	PGP 1 PY1	PGP 1 PY2	PGP 1 PY3	PGP 1 PY4	PGP 1 PY5	PGP 2 BY	PGP 2 PY1	PGP 2 PY2	PGP 2 PY3	PGP 2 PY4	PGP 2 PY5
PC-5 Breast Cancer Screening	81	N/A	N/A	78	79	78	75	N/A	N/A	76	76	76
PC-6 Colorectal Cancer Screening	63	N/A	N/A	77	67	67	58	N/A	N/A	63	64	64

Table 6-14b
Preventative Care (PC) results for BY and PY3 – PY5, by PGP 3 and 4

Quality Measure	PGP 3 BY	PGP 3 PY1	PGP 3 PY2	PGP 3 PY3	PGP 3 PY4	PGP 3 PY5	PGP 4 BY	PGP 4 PY1	PGP 4 PY2	PGP 4 PY3	PGP 4 PY4	PGP 4 PY5
PC-5 Breast Cancer Screening	81	N/A	N/A	81	81	81	62	N/A	N/A	67	67	68
PC-6 Colorectal Cancer Screening	27	N/A	N/A	73	70	66	50	N/A	N/A	66	66	64

Table 6-14c
Preventative Care (PC) results for BY and PY3 – PY5, by PGP 5 and 6

Quality Measure	PGP 5 BY	PGP 5 PY1	PGP 5 PY2	PGP 5 PY3	PGP 5 PY4	PGP 5 PY5	PGP 6 BY	PGP 6 PY1	PGP 6 PY2	PGP 6 PY3	PGP 6 PY4	PGP 6 PY5
PC-5 Breast Cancer Screening	78	N/A	N/A	81	87	90	80	N/A	N/A	83	84	83
PC-6 Colorectal Cancer Screening	47	N/A	N/A	61	59	66	72	N/A	N/A	80	73	76

Table 6-14d
Preventative Care (PC) results for BY and PY3 – PY5, by PGP 7 and 8

Quality Measure	PGP 7 BY	PGP 7 PY1	PGP 7 PY2	PGP 7 PY3	PGP 7 PY4	PGP 7 PY5	PGP 8 BY	PGP 8 PY1	PGP 8 PY2	PGP 8 PY3	PGP 8 PY4	PGP 8 PY5
PC-5 Breast Cancer Screening	78	N/A	N/A	80	80	79	82	N/A	N/A	88	89	85
PC-6 Colorectal Cancer Screening		N/A	N/A	77	77	79	83	N/A	N/A	75	72	74

Table 6-14e
Preventative Care (PC) results for BY and PY3 – PY5, by PGP 9 and 10

Quality Measure	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 9	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10	PGP 10
	BY	PY1	PY2	PY3	PY4	PY5	BY	PY1	PY2	PY3	PY4	PY5
PC-5 Breast Cancer Screening	68	N/A	N/A	72	78	72	75	N/A	N/A	80	81	79
PC-6 Colorectal Cancer Screening	47	N/A	N/A	66	60	69	62	N/A	N/A	69	68	70

SOURCE: RTI analysis of 2006–2010 Medicare claims.

Table 6-15
Number of quality targets hit per PGP for Preventative Care (PC) in PY3-PY5

PGP	Number of PC Targets Hit PY1	Number of PC Targets Hit PY2	Number of PC Targets Hit PY3	Number of PC Targets Hit PY4	Number of PC Targets Hit PY5
PGP 1	N/A	N/A	2	2	2
PGP 2	N/A	N/A	2	2	2
PGP 3	N/A	N/A	2	2	2
PGP 4	N/A	N/A	2	2	2
PGP 5	N/A	N/A	2	2	2
PGP 6	N/A	N/A	2	2	2
PGP 7	N/A	N/A	2	2	2
PGP 8	N/A	N/A	2	2	2
PGP 9	N/A	N/A	2	2	2
PGP 10	N/A	N/A	2	2	2
TOTAL	N/A	N/A	20	20	20

SOURCE: RTI analysis of 2006–2010 Medicare claims.

6.6.6 Number of PGPs Hitting a Target by PC Measure in PY3-PY5

Table 6-16 includes a summary of the number of PGPs hitting a quality performance target for the PC quality measures in PY3-PY5. Following the previous table, this also shows that 60 of a possible 60 targets were hit. All of the PGPs hit at least one of the performance targets for both of their PC measures in all three years of the Demonstration

Table 6-16
Number of PGPs hitting a quality target in PY2-PY5 by PC measures

PC Quality Measure	Number of PGPs Hitting Target PY1	Number of PGPs Hitting Target PY2	Number of PGPs Hitting Target PY3	Number of PGPs Hitting Target PY4	Number of PGPs Hitting Target PY5
PC-5 Breast Cancer Screening	N/A	N/A	10	10	10
PC-6 Colorectal Cancer Screening	N/A	N/A	10	10	10

SOURCE: RTI analysis of 2005–2010 Medicare claims.

6.7 PQRI Incentive Payments in PY2-PY5

6.7.1 Description of Methodology Used for Calculating PY2 – PY5 PQRI Payments for PGP Sites

Data collection that the participating PGPs conducted for quality measurement in the PGP Demonstration was accepted by CMS to qualify for its Physician Quality Reporting Initiative (PQRI) “pay-for-reporting” bonus payment program that is available to all physicians nationwide. Consequently, the PGPs were exempted from the PQRI reporting requirement as a result of participating in the Demonstration. That maximum bonus amount was calculated for each PY by another contractor working for CMS on the PQRI program. For the PGP Demonstration participants, that maximum bonus amount was multiplied by the PGP’s overall percentage quality performance score for each year to calculate each PGP’s actual PQRI incentive payments for PY2 through PY5. Since the PQRI bonus is paid outside the fee for service payment system, it is never considered in the savings and bonus calculations.

6.7.2 PQRI Incentive Payments Earned by PGP for PY2 – PY5

Tables 6-17a through 6-17e present the methodology and calculation used to determine the PQRI incentive payment amounts earned by the participating PGPs for PY2 through PY5. As noted earlier, *Table 6-1* shows the percentage of quality “points” earned by each site each performance year. For each PGP, the table first restates the number of quality “points” earned, based on weighting the results for the quality measures active in each PY. The “maximum possible PQRI Incentive Payment” for each year is calculated using relevant Part B claim allowed charges in the designated year¹⁶, subject to final PQRI rules. The last column, “PQRI incentive payments earned”, is simply the proportion of the maximum possible payment based on the percent of quality points earned for the year (i.e., percent points earned x PQRI maximum payment).

Over years PY2 through PY5, the range of PQRI incentive payments ranged from a low of \$98,407, earned by PGP 4 in PY2, to a high of \$1,582,357, earned by PGP 10 in PY5. The average PGP incentive payments were \$295,130 in PY2, \$575,277 in PY3, \$814,852 in PY4, and \$898,670 in PY5. This increase in PQRI payments over time is evident: while the PQRI incentive payments totaled \$2,951,299 in PY2, by PY5, these payments totaled three times higher, and reached \$8,986,697 across all ten PGPs.

¹⁶ PY2 PQRI incentive payments were based on 2007 Part B allowed charges; PY3 PQRI incentive payments were based on 2008 Part B allowed charges; PY4 PQRI incentive payments were based on 2009 Part B allowed charges; and PY5 PQRI incentive payments were based on 2010 Part B allowed charges.

Table 6-17a
Physician Quality Reporting Initiative (PQRI) incentive payments earned by PGPs during PY2 and PY3

PGP	Percentage of Quality Points Earned in PY2	Maximum Possible PQRI Incentive Payment for PY2	PQRI Incentive Payments Earned in PY2	Percentage of Quality Points Earned in PY3	Maximum Possible PQRI Incentive Payment for PY3	PQRI Incentive Payments Earned in PY3
PGP 1	44/45 = 97.8%	\$151,727	\$148,355	52/53 = 98.1 %	\$307,816	\$302,008
PGP 2	44/45 = 97.8%	\$372,847	\$364,561	49/53 = 92.5%	\$789,351	\$729,777
PGP 3	43/45 = 95.6%	\$132,929	\$127,021	50/53 = 94.3%	\$285,314	\$269,164
PGP 4	45/45 = 100%	\$98,407	\$98,407	51/52 = 96.2%	\$213,485	\$205,429
PGP 5	45/45 = 100%	\$376,734	\$376,734	53/53 = 100%	\$768,881	\$768,881
PGP 6	45/45 = 100%	\$500,107	\$500,107	52/53 = 98.1%	\$1,037,741	\$1,018,161
PGP 7	43/45 = 95.6%	\$234,693	\$224,262	49/53 = 92.5%	\$438,399	\$405,312
PGP 8	44/45 = 97.8%	\$253,223	\$247,596	53/53 = 100%	\$496,548	\$496,548
PGP 9	45/45 = 100%	\$404,802	\$404,802	51/53 = 96.2 %	\$782,346	\$738,062
PGP 10	45/45 = 100%	\$459,454	\$459,454	50/53 - 94.3%	\$857,998	\$809,432
Total	—	—	\$2,951,299	—	—	\$5,742,774

Table 6-17b
Physician Quality Reporting Initiative (PQRI) incentive payments earned by PGPs during PY4 and PY5

PGP	Percentage of Quality Points Earned in PY4	Maximum Possible PQRI Incentive Payment for PY4	PQRI Incentive Payments Earned in PY4	Percentage of Quality Points Earned in PY5	Maximum Possible PQRI Incentive Payment for PY5	PQRI Incentive Payments Earned in PY5
PGP 1	49/53 = 92.5%	\$425,523	\$394,332	53/53 = 100%	\$443,828	\$443,828
PGP 2	50/53 = 94.3%	\$1,102,906	\$1,040,477	51/53 = 96.2%	\$1,165,115	\$1,121,148
PGP 3	50/53 = 94.3%	\$414,962	\$391,474	53/53 = 100%	\$409,079	\$409,079
PGP 4	51/53 = 96.2%	\$299,298	\$288,004	53/53 = 100%	\$332,715	\$332,715
PGP 5	53/53 = 100%	\$1,113,400	\$1,113,400	53/53 = 100%	\$1,313,312	\$1,313,312
PGP 6	53/53 = 100%	\$1,400,245	\$1,400,245	52/53 = 98.1%	\$1,463,108	\$1,435,502
PGP 7	50/53 = 94.3%	\$591,993	\$558,484	53/53 = 100%	\$530,995	\$530,995
PGP 8	53/53 = 100%	\$672,006	\$672,006	53/53 = 100%	\$661,951	\$661,951
PGP 9	52/53 = 98.1%	\$1,103,097	\$1,082,284	53/53 = 100%	\$1,155,810	\$1,155,810
PGP 10	51/53 = 96.2%	\$1,255,177	\$1,207,811	52/53 = 98.1%	\$1,612,787	\$1,582,357
Total	—	—	\$8,148,517	—	—	\$8,986,697

SOURCE: RTI International

CHAPTER 7

DEMONSTRATION IMPACTS ON MEDICARE PROGRAM EXPENDITURES: SIMULATION ANALYSIS

The PGP Demonstration creates an incentive for participating PGPs to reduce the rate of expenditure growth of their assigned beneficiaries. If the expenditures of their assigned beneficiaries are more than 2 percent below their target expenditures in a performance year, then the participating PGP will share in Demonstration savings. This chapter analyzes Demonstration impacts on total Medicare program expenditures for the participating PGPs' overall assigned beneficiary populations, taking into account simulating expenditure trending prior to the Demonstration. We also analyze the statistical significance of differences of assigned beneficiary from target expenditures in the Demonstration period (2004 to PY5), and when the pre-Demonstration (2001–2004) trend in expenditures is removed. We also consider the impact of the Demonstration on the expenditures for subgroups of assigned beneficiaries relative to analogously defined subgroups of comparison beneficiaries, as well as impacts by components of Medicare program expenditures.

This chapter uses a simulation methodology. We simulate the actual Demonstration payment methodology, and use the statistical “bootstrapping” technique to determine statistical significance. The simulation methodology contrasts with the methods used in Chapters 5, 10, and 11. Chapter 5 presents an accounting analysis of the actual Demonstration financial results. This chapter is similar to Chapter 5 in that it simulates the actual Demonstration payment methodology, although it differs from Chapter 5 in that the 2% corridor between target and assigned beneficiary expenditures in determining performance payments is not reflected. In contrast to Chapter 5, this chapter considers the impact of pre-Demonstration expenditure trends in the 2001–2004 period on the interpretation of Demonstration 2004–2010 financial results.

Also, in this chapter, all 10 PGPs are weighted equally in combined PGP results. In Chapter 5, dollars are added across PGPs, so effectively PGPs are weighted by number of assigned beneficiaries and level of expenditures per PGP. PGPs are weighted equally in this chapter so that the larger or higher-expenditure PGPs do not have a disproportionate impact on results.¹⁷

Effectively in this chapter the unit of observation is each of the 10 PGPs, regardless of their number of assigned beneficiaries or level of expenditures (for percentage results). Chapter 10 on the other hand presents a multivariate regression statistical analysis of Demonstration financial results. This chapter is similar to Chapter 10 in that the 2% corridor is included in the estimation of Demonstration financial impacts, and also the statistical significance of results is considered and pre-Demonstration trends are incorporated. This chapter differs from Chapter 10 in that this chapter simulates the actual Demonstration payment methodology whereas Chapter 10 uses a regression specification that approximates, but is not identical to, the Demonstration payment methodology. Also, unlike Chapter 5's simulations, Chapter 10's regression model

¹⁷ As seen in Table 5-1, the number of assigned beneficiaries per PGP varies by a factor of more than 5 to 1 in some performance years (e.g., PY1). Also, the number of beneficiaries per PGP varies across performance years, meaning that the weight of individual PGPs in aggregate results would vary across performance years in a weighted analysis. Finally, PGP 10 has a much higher expenditure level than the other PGPs, meaning that it has disproportionate influence on dollar-weighted results.

“controls” for additional factors impacting expenditures. Chapter 11 presents an accounting analysis of the Demonstration financial results that includes the 2% corridor. The similarities/differences between this chapter and Chapter 11 are the same as the similarities/differences between this chapter and Chapter 5, except that this chapter and Chapter 11 include the 2% corridor, whereas Chapter 5 excludes the 2% corridor. For these reasons, the results of Chapters 5, 7, 10, and 11 are broadly similar, but differ in detail.

In Section 7.1, we analyze Demonstration impacts on total Medicare program expenditures¹⁸ for the participating PGPs’ overall assigned beneficiary populations. In Section 7.2, we consider the impact of the Demonstration on expenditures for subgroups of assigned beneficiaries. Section 7.3 studies Demonstration impacts by components of Medicare program expenditures.

7.1 Medicare Program Expenditures for Assigned Beneficiaries

To study the impact of the PGP Demonstration on Medicare expenditures, we compare trends in assigned beneficiary versus target expenditures in the periods prior to and during the Demonstration. Section 7.1.1 addresses these trends graphically. Section 7.1.2 considers the statistical significance of the observed trends and differences.

7.1.1 Trends in Assigned Beneficiary Versus Target Expenditures in the Periods Prior to and During the Demonstration

Figure 7-1 compares assigned beneficiary to risk adjusted target expenditures per person year from 2001 to Demonstration Performance Year 5 (PY5). The figure represents an unweighted composite average for the 10 PGPs participating in the Demonstration. It shows expenditure trends for both Demonstration (2004 through PY5) and pre-Demonstration (2001 through 2004) periods. Pre-Demonstration trends are simulated. In Figure 7-1, participating PGP cost control performance is measured relative to the first year of the pre-Demonstration period, 2001, rather than relative to the Demonstration base year of 2004. This is accomplished by setting target expenditures equal to assigned beneficiary expenditures in 2001. Target expenditures represent expenditure trends among comparison groups for the assigned beneficiaries. This does not reflect the 2% corridor (i.e., this does not exclude the 2% corridor). Figure 7-1 also does not reflect the PY5 risk score cap.

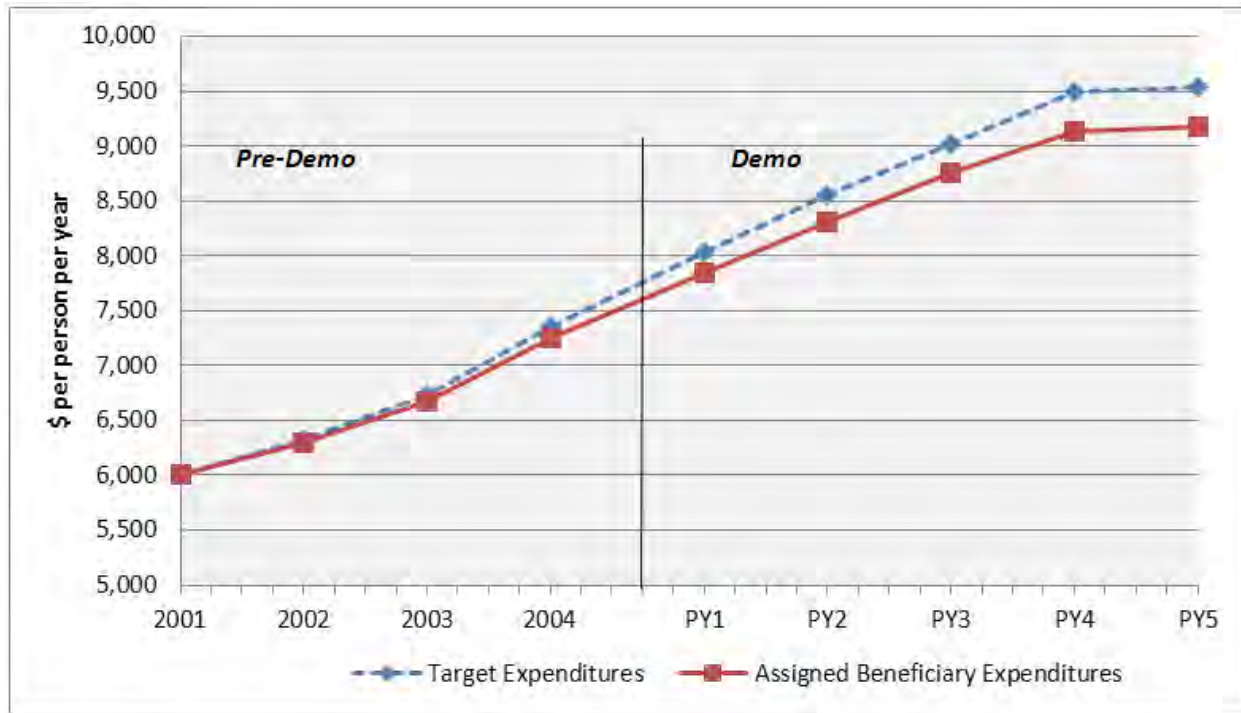
Figure 7-1 shows that, on a risk adjusted basis, on average, expenditures of assigned beneficiaries were trending very slightly below target expenditures prior to the Demonstration. Starting from equality with the target in 2001, by 2004 assigned beneficiary expenditures were \$7,253, 1.4 percent lower than the target of \$7,359. Even in the absence of the Demonstration financial incentive, assigned beneficiary expenditures tended to trend lower than target expenditures. This pre-Demonstration trend is an unexpected result under the Demonstration’s methodology.

In the Demonstration period, from 2004 to PY5, assigned beneficiary expenditures continued to trend lower than target. The difference between assigned beneficiary and target

¹⁸ Hospice expenditures were not included in the Demonstration (see Chapter 2).

expenditures increased during the Demonstration period, and the gap widened at a slightly greater rate than pre-Demonstration. By PY5, assigned beneficiary expenditures were \$9,181, 3.9 percent lower than the target of \$9,542.

Figure 7-1
Assigned beneficiary versus risk-adjusted target expenditures,
2001 to performance year five



NOTES: Unweighted average of all 10 participating physician group practices. Does not reflect PY5 risk score cap.

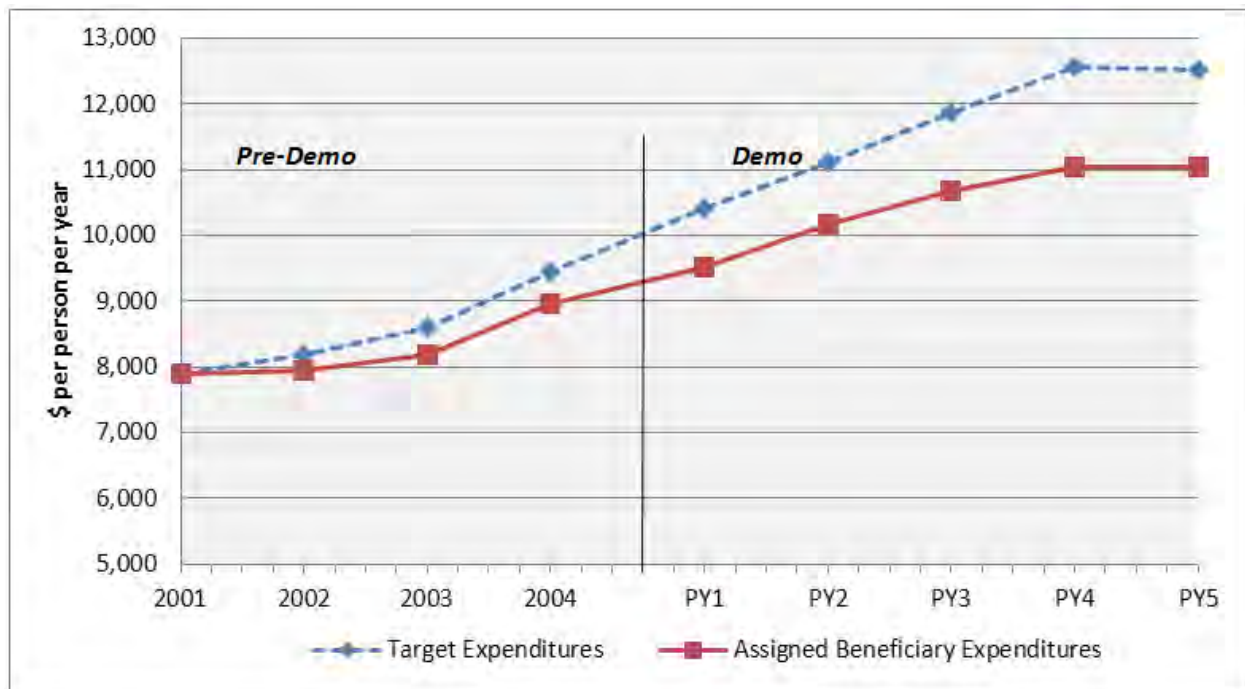
SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Figure 7-1 provides graphical evidence of a small reduction in the rate of assigned beneficiary expenditure growth in the Demonstration’s five years, relative to contemporaneous comparison group expenditure growth. However, the slight pre-Demonstration trend toward lower-than-target assigned beneficiary expenditures could have continued into the Demonstration period even in the absence of the Demonstration.

Figure 7-1a shows these results for the 2 PGPs that shared in savings in all of the 5 demonstration years. For these two groups, on a risk adjusted basis, on average, expenditures of assigned beneficiaries were trending slightly below target expenditures prior to the Demonstration. Starting from equality with the target in 2001, by 2004 assigned beneficiary expenditures were \$8,958, 5.5 percent lower than the target of \$9,450. Even in the absence of the Demonstration financial incentive, assigned beneficiary expenditures tended to trend lower than target expenditures. In the Demonstration period, from 2004 to PY5, assigned beneficiary expenditures continued to trend lower than target. The difference between assigned beneficiary

and target expenditures increased during the Demonstration period, and the gap widened at a greater rate than pre-Demonstration. By PY5, assigned beneficiary expenditures were \$11,033, 13.5 percent lower than the target of \$12,528.

Figure 7-1a
Assigned beneficiary versus risk-adjusted target expenditures for 2 PGPs sharing in savings all five years, 2001 to performance year five

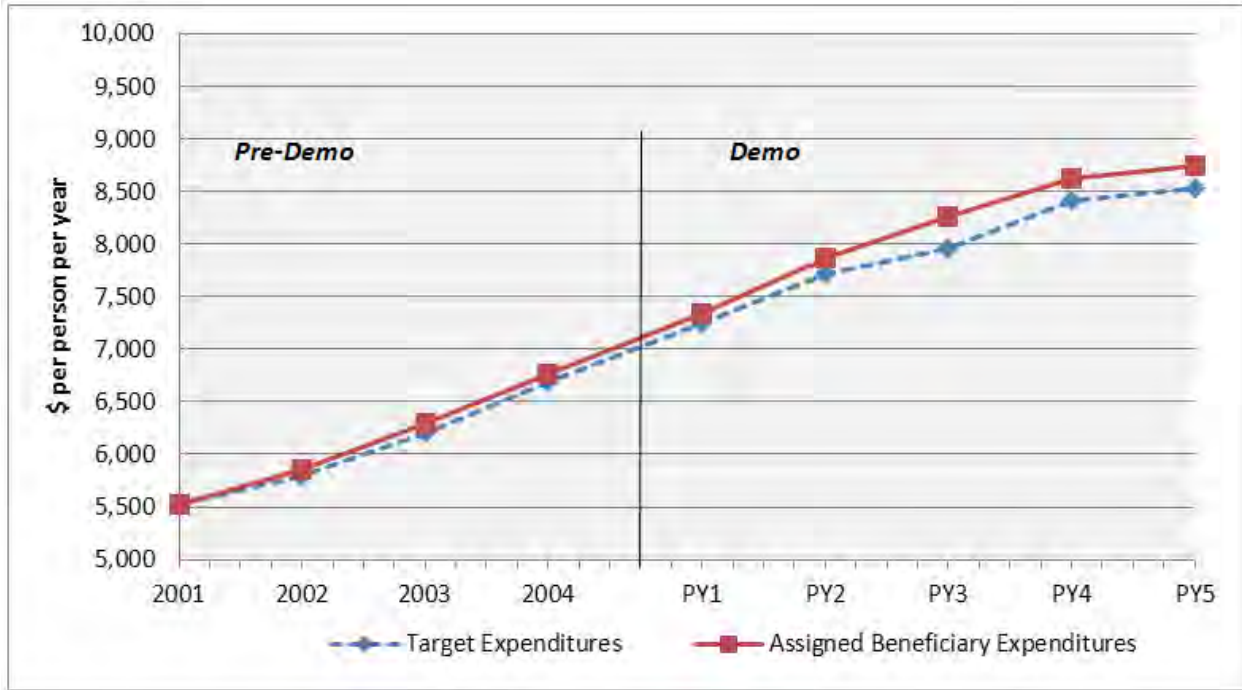


NOTES: Unweighted average of 2 participating physician group practices. Does not reflect PY5 risk score cap.

SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Figure 7-1b shows these results for the 3 PGPs that did not share in savings in any of the 5 demonstration years. For these three groups, on a risk adjusted basis, on average, expenditures of assigned beneficiaries were trending very slightly above target expenditures prior to the Demonstration. Starting from equality with the target in 2001, by 2004 assigned beneficiary expenditures were \$6,769, 1.1 percent higher than the target of \$6,693. Even in the absence of the Demonstration financial incentive, assigned beneficiary expenditures tended to trend slightly higher than target expenditures. In the Demonstration period, from 2004 to PY5, assigned beneficiary expenditures continued to trend higher than target. The difference between assigned beneficiary and target expenditures increased during the Demonstration period, and the gap widened at a greater rate than pre-Demonstration. By PY5, assigned beneficiary expenditures were \$8,740, 2.4 percent higher than the target of \$8,531.

Figure 7-1b
Assigned beneficiary versus risk-adjusted target expenditures for 3 PGPs not sharing in savings any of the five years, 2001 to performance year five

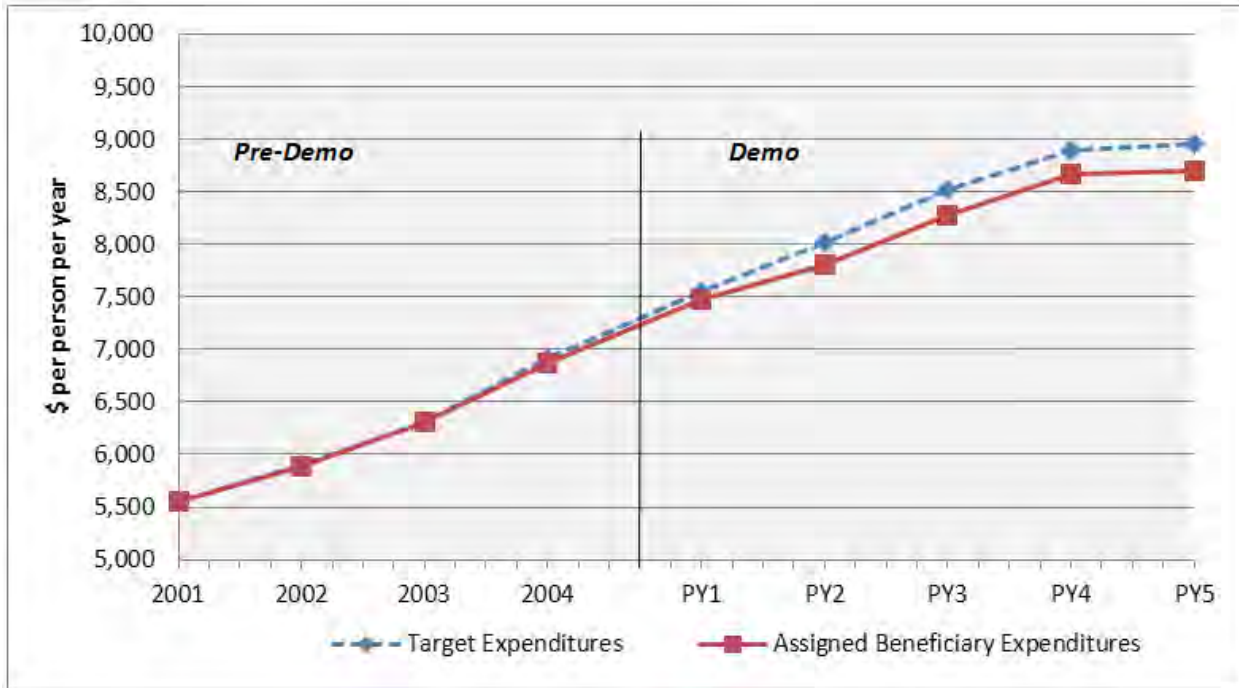


NOTES: Unweighted average of 3 participating physician group practices. Does not reflect PY5 risk score cap.

SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Finally, *Figure 7-1c* shows these results for the 5 PGPs that shared in savings some of the 5 demonstration years. For these five groups, on a risk adjusted basis, on average, expenditures of assigned beneficiaries were trending very similarly to target expenditures prior to the Demonstration. Starting from equality with the target in 2001, by 2004 assigned beneficiary expenditures were \$6,862, 0.85 percent lower than the target of \$6,921. In the Demonstration period, from 2004 to PY5, assigned beneficiary expenditures trended lower than target. The difference between assigned beneficiary and target expenditures increased during the Demonstration period, and the gap widened at a greater rate than pre-Demonstration. By PY5, assigned beneficiary expenditures were \$8,705, 2.9 percent lower than the target of \$8,954.

Figure 7-1c
Assigned beneficiary versus risk-adjusted target expenditures for 5 PGPs sharing in savings some of the five years, 2001 to performance year five

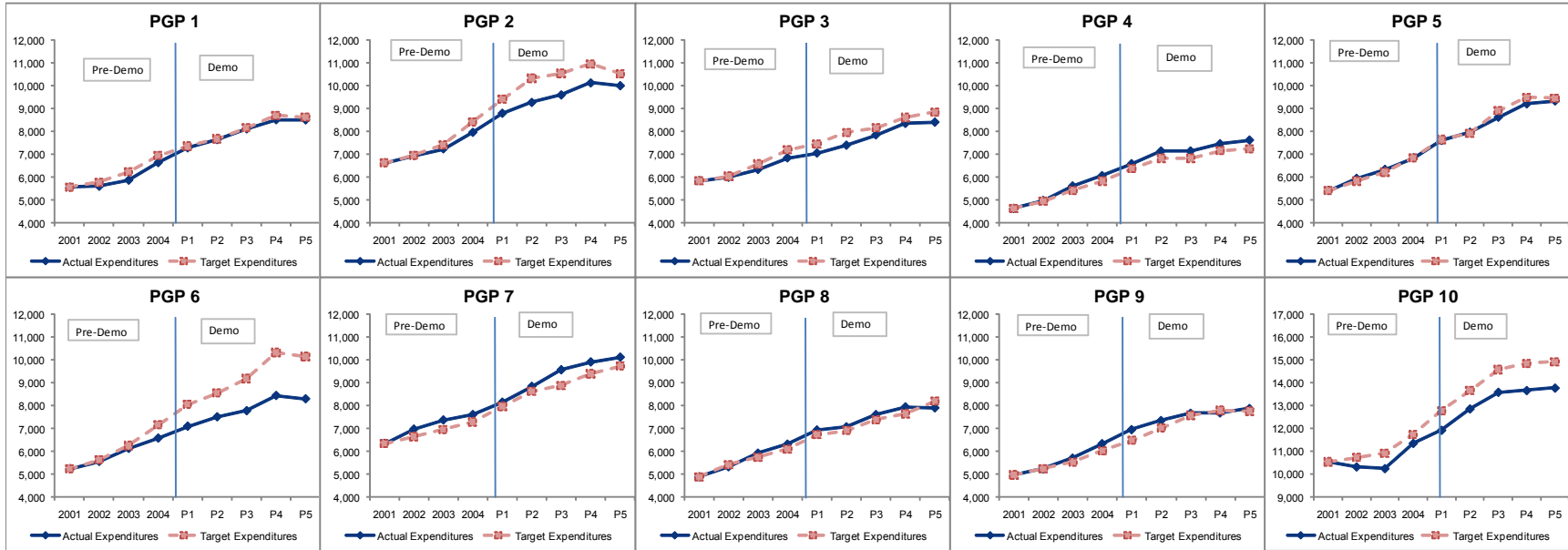


NOTES: Unweighted average of 5 participating physician group practices. Does not reflect PY5 risk score cap.

SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Figure 7-2 shows assigned beneficiary versus risk adjusted target expenditures from 2001 to PY5 for each PGP. PGPs' 1, 3, 4, 5, 7, 8, and 9 assigned beneficiary expenditures closely track their target expenditures throughout all or most of the Demonstration and pre-Demonstration periods. PGPs' 2, 6, and 10 assigned beneficiary expenditures grow less rapidly than their target expenditures. Although this slower growth is apparent in the pre-Demonstration period for all 3 PGPs, the slower growth appears to be greater in the Demonstration period, especially for PGP 10.

Figure 7-2
Assigned beneficiary versus risk-adjusted target expenditures by PGP, 2001 to performance year five



NOTE: Does not reflect PY5 risk score cap.

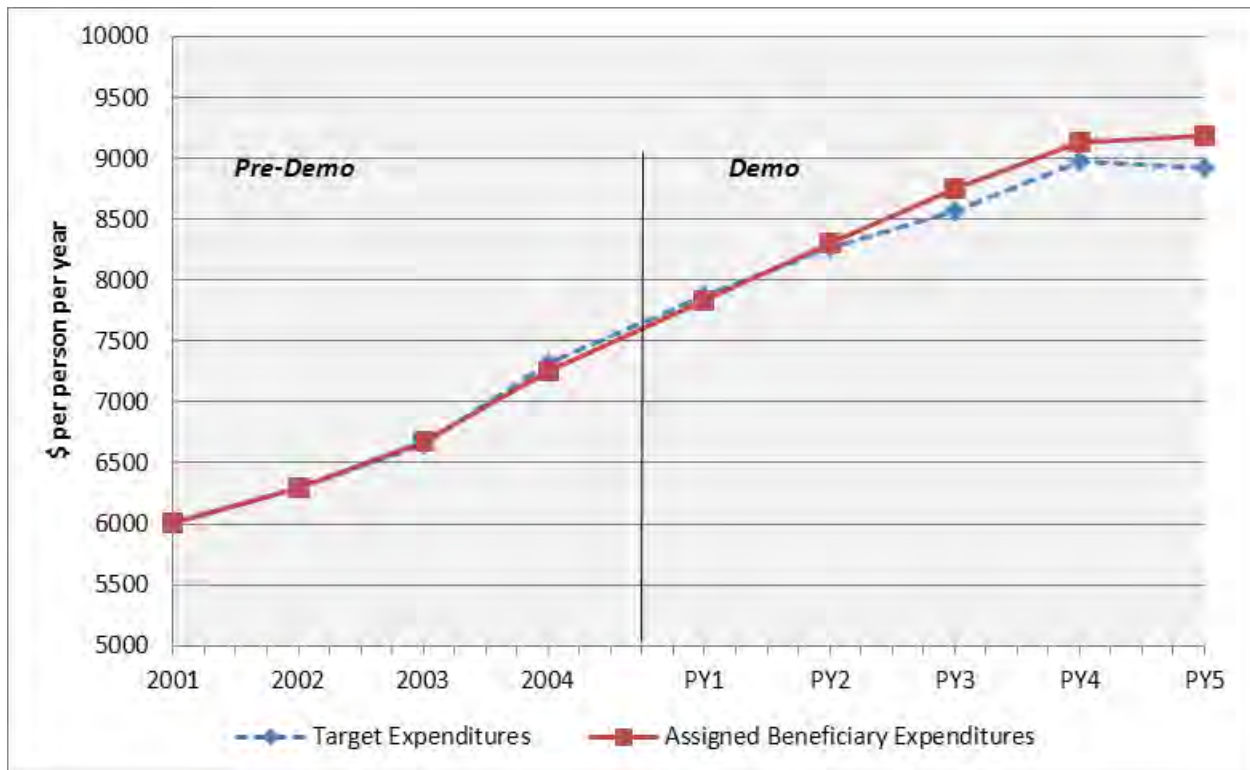
SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Figure 7-3 shows the 2001 to PY5 assigned beneficiary versus target expenditures trends on a non-risk-adjusted basis. On a non-risk adjusted basis, on average, expenditures of assigned beneficiaries were trending very close, but slightly below, target expenditures prior to the Demonstration. Starting from equality with the target in 2001, by 2004 assigned beneficiary expenditures were \$7,253, 0.9 percent lower than the target of \$7,320. During the Demonstration period on a non-risk-adjusted basis, from 2004 to PY5, assigned beneficiary expenditures actually trended higher than target expenditures. By PY5, assigned beneficiary expenditures were \$9,181, 2.9 percent higher than the target of \$8,918.

Figure 7-4 shows these results on a non-risk-adjusted basis by PGP. Four of the PGPs' (PGP 1, 2, 3, 9) assigned beneficiary expenditures track target expenditures closely for all or most of the pre-Demonstration and Demonstration periods. Two PGPs (6 and 10) show a slower assigned beneficiary than target expenditures trend, which may be more pronounced in the Demonstration period. Four of the PGPs' (PGPs 4, 5, 7, 8) assigned beneficiaries expenditures seem to trend higher than their target, especially in the Demonstration period for PGPs 5 and 7.

We now proceed to a more detailed numerical analysis of the Demonstration's expenditure impact, including examining the impact of adjusting for pre-Demonstration trends.

Figure 7-3
Assigned beneficiary versus non-risk-adjusted target expenditures,
2001 to performance year five

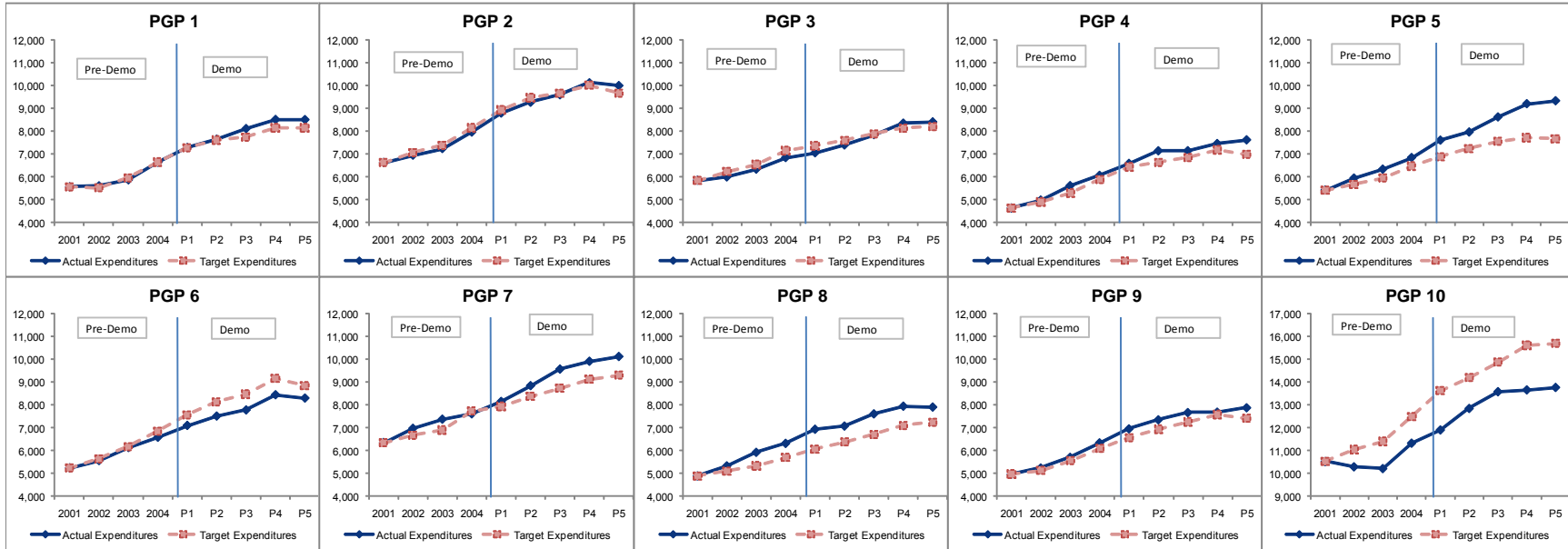


NOTE: Unweighted average of all 10 participating physician group practices.

SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

Figure 7-4

Assigned beneficiary versus non-risk-adjusted target expenditures by individual PGP, 2001 to performance year five



SOURCE: RTI analysis of Medicare claims and enrollment data 2001-PY5.

7.1.2 Statistical Significance of Demonstration Expenditure Impacts

We now turn to the question of the statistical significance of estimates of Demonstration expenditure impacts. Statistical significance refers to whether the estimated Demonstration expenditure impacts could have arisen from the normal year-to-year variations in Medicare expenditures (“chance”), or whether they are likely to represent actual behavioral impacts of the Demonstration. Statistical significance would represent actual behavioral impacts. To investigate this question, we begin by analyzing the PY5 difference of assigned beneficiary from target expenditures, in dollar as well as percentage terms, for all PGPs combined (later in this section we analyze results for individual PGPs and for the other 4 performance years). We continue to compare Demonstration results to pre-Demonstration trends. *Table 7-1* shows the Target minus Assigned Beneficiary expenditures in the Demonstration (2004 to PY5) and the Pre-Demonstration (2001–2004, multiplied by 5/3) periods. The pre-Demonstration period is multiplied by 5/3 in order to make its length analogous to the Demonstration period, i.e., 5 years in both cases. The table is stratified by PGPs earning and not earning performance payments in PY5. Statistical significance as reported in Table 7-1 was estimated by “bootstrapping” the standard errors. Table 7-1 reflects the Demonstration PY5 risk score cap, which effectively lowered the PY5 risk scores of 2 of the 10 PGPs. By lowering their risk scores, the cap reduced target expenditures and the difference of target and actual expenditures.

In PY5, on average across the 10 participating PGPs, assigned beneficiary expenditures were \$162 or 1.5 percent lower than target expenditures. These differences of assigned beneficiary from target expenditures were “statistically significant”, meaning that they were not likely to arise by chance. In the simulated pre-Demonstration PY5, assigned beneficiary expenditures were \$175 or 1.6 percent lower than target expenditures and these differences were statistically significant. When the Demonstration (PY5) results are adjusted for the pre-Demonstration trend (simulated PY5 results), the estimated difference of assigned beneficiary from target expenditures falls to -\$13 or -0.1 percent. Adjusted for the pre-Demonstration trend, the difference of assigned beneficiary from target expenditures is not statistically significantly different from zero. Without the PY5 risk score cap, target minus actual expenditures for the average of the PGPs for the Demonstration period was \$225 instead of \$162, and the Demonstration/pre-Demonstration period difference was \$50 instead of -\$13.

Table 7-1 also shows the results for the 4 PGPs who shared in PY5 savings. These 4 PGPs show a statistically significant lower-than-target assigned beneficiary expenditure per person year in PY5 of \$626 or 6.4 percent. Once the pre-Demonstration trend is adjusted for, the PY5 assigned beneficiary expenditures are estimated to be \$434 lower than target, or 5.1 percent, and are also statistically different from the target. Therefore, while these PGPs were trending favorably pre-Demonstration, they continued and widened this performance gap during the five performance years. This does reflect the 2% corridor (i.e., the difference between actual and target within the 2% corridor is included).

Table 7-1
Target minus assigned beneficiary expenditures in demonstration performance year five,
compared to simulated pre-demonstration trend

\$'s Per Person Year and % of Target Expenditures	Average of Physician Group Practices (\$): All 10	Average of Physician Group Practices (\$): 4 sharing PY5 savings	Average of Physician Group Practices (\$): 6 not sharing PY5 savings
Dollars per person year			
Demonstration performance year 5	162**	626***	-147**
Pre-demonstration 5-year trend	175**	192**	164*
Demonstration minus pre-demonstration	-13	434***	-311**
As a % of target expenditures			
Demonstration performance year 5	1.5%**	6.4%***	-1.7%**
Pre-demonstration 5-year trend	1.6%*	1.2%	1.9%
Demonstration minus pre-demonstration	-0.1%	5.1%***	-3.6%**

NOTES:

1. Target minus assigned beneficiary expenditures are calculated using the Demonstration financial reconciliation. The plus or minus 2 percent threshold between target and assigned beneficiary expenditures in calculating physician group practice performance payments is **not** reflected in this table.
2. Demonstration Performance Year 5 (PY5) is April 2009 to March 2010, with a base year of 2004.
3. The Pre-Demonstration trend is the change from 2001 to 2004, multiplied by 5/3. For the Pre-Demonstration trend, statistical testing is based on the 3-year (2001–2004) trend.
4. Physician Group Practice averages are unweighted averages of results for the applicable practices.
5. Reflects PY5 risk score cap. If the risk score cap was not in place in PY5, the target minus actual expenditures for the average of the PGPs for the Demonstration period would be \$225 instead of \$162.
6. Statistical significance is a two-tailed test of difference from zero.

*=statistically significant at 10% level

**=statistically significant at 5% level

***=statistically significant at 1% level

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to Performance Year 5.

For the 6 PGPs that did not share in PY5 savings, pre-Demonstration assigned beneficiary expenditures trended less than target expenditure growth on average for these PGPs by \$164. Once the pre-Demonstration trend is adjusted for, the PY5 assigned beneficiary expenditures are estimated to be \$311 greater than target expenditures, or 3.6 percent, and are also statistically different from the target. These PGPs showed unfavorable results in the Demonstration period relative to pre-Demonstration. The Demonstration financial reconciliation method has only a single base year and does not take into account pre-Demonstration trends.

Table 7-2 shows target minus assigned beneficiary expenditures in PY5 versus the pre-Demonstration 5-year trend, by PGP. The only two PGPs with a statistically significant favorable Demonstration period dollar and percentage trend once the pre-Demonstration trend is adjusted for are PGP 8 and PGP 9. Both of these PGPs had actual expenditures trending higher than target expenditures pre-Demonstration, but their actual expenditures trended lower than their targets in the Demonstration period. PY5 target minus actual expenditures once the pre-Demonstration trend was adjusted for were \$933 and \$783 for PGP 8 and PGP 9, respectively. In contrast, PGP 1, PGP 2, PGP 3, and PGP 5 all had statistically significant unfavorable Demonstration period trends once the pre-Demonstration trends were adjusted for. (PGP 5, along with PGP 7, had its risk score and target expenditures reduced by the PY5 risk score cap.) PGP 6, which had the largest target minus actual difference during the Demonstration (\$1,015), also had a favorable pre-Demonstration trend, and its Demonstration gross savings (target minus actual expenditures) was not statistically significant once its pre-Demonstration trend was removed.

Finally, **Table 7-3** shows target minus assigned beneficiary expenditures in the Demonstration and pre-Demonstration periods for the (unweighted) average of the ten PGPs in each of the five performance years. The pre-Demonstration trend is the 3-year 2001–2004 change, adjusted for the length of the Demonstration period in each performance year (i.e., the 2001–2004 trend is multiplied by 1/3 for PY1, 2/3 for PY2, 3/3 for PY3, 4/3 for PY4, and 5/3 for PY5). In Demonstration PY1, on average across the 10 participating PGPs, assigned beneficiary expenditures were \$77 (or 0.7 percent) lower than target expenditures, which was statistically significant. This increased every year to \$221 lower than target expenditures in PY4, and decreased again to \$162 lower than target expenditures in PY5 (all statistically significant). However, if the risk score growth cap was not in place in PY5, the target minus actual expenditures would have increased to a difference of \$225 instead of \$162. In the simulated pre-Demonstration trend for all performance years, actual expenditures were lower than target expenditures (all statistically significant), and the gap widened from PY1 to PY5. When the Demonstration results are adjusted for the pre-Demonstration trend, actual expenditures were lower than target expenditures for all years except PY5 (which as noted was a result of the risk score cap put in place in PY5). However, the estimated target minus assigned beneficiary expenditures, adjusted for the pre-Demonstration trend, are not statistically significant for any of the performance years.

Table 7-2
Target minus assigned beneficiary expenditures in demonstration performance year five, compared to simulated pre-demonstration trend by individual PGP

\$'s per person year and % of target expenditures	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
Dollars per person year										
Demonstration performance year 5	-260	-67	20	-88	-410**	1,015***	-78	576***	270**	641***
Pre-demonstration 5-year trend	515**	766**	583**	-390**	14	977***	-505**	-357**	-513***	661**
Demonstration minus pre-demonstration	-775**	-833***	-563**	302	-424**	38	427	933***	783***	-20
As a % of target expenditures										
Demonstration performance year 5	-3.2%	-0.7%	0.2%	-1.2%	-4.6%**	10.9%***	-0.8%	6.8%***	3.3%**	4.4%***
Pre-demonstration 5-year trend	7.4%**	9.1%***	8.1%**	-6.7%**	0.2%	13.6%***	-6.9%**	-5.8%**	-8.5%***	5.6%**
Demonstration minus pre-demonstration	-10.6%**	-9.8%***	-7.9%**	5.5%*	-4.8%*	-2.7%	6.1%*	12.6%***	11.8%***	-1.2%

NOTES:

1. Target minus assigned beneficiary expenditures are calculated using the Demonstration financial reconciliation. The plus or minus 2 percent threshold between target and assigned beneficiary expenditures in calculating physician group practice performance payments is **not** reflected in this table.
2. Performance Year 5 (PY5) is April 2009 to March 2010, with a base year of 2004.
3. The Pre-Demonstration trend is the change from 2001 to 2004, multiplied by 5/3. For the Pre-Demonstration trend, statistical testing is based on the 3-year (2001–2004) trend.
4. Reflects risk score cap in PY5.
5. Statistical significance is a two-tailed test of difference from zero.

*=statistically significant at 10% level

**=statistically significant at 5% level

***=statistically significant at 1% level

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to Performance Year 5.

Table 7-3
Target minus assigned beneficiary expenditures in demonstration performance years one to five, compared to the simulated pre-demonstration trend

Dollars per person year and % of target expenditures	Average of physician group practices: PY1	Average of physician group practices: PY2	Average of physician group practices: PY3	Average of physician group practices: PY4	Average of physician group practices: PY5
	Difference (\$)	Difference (\$)	Difference (\$)	Difference (\$)	Difference (\$)
Dollars per person year					
Demonstration performance year	76.5*	119.9**	137.2**	220.8***	162.0**
Pre-demonstration trend	35.0**	70.1**	105.1**	140.2**	175.2**
Demonstration minus pre-demonstration trend	41.5	49.8	32.1	80.6	-13.3
As a % of target expenditures					
Demonstration performance year	0.7%	1.2%**	1.3%**	2.1%***	1.5%**
Pre-demonstration trend	0.3%*	0.6%*	1.0%*	1.3%*	1.6%*
Demonstration minus pre-demonstration trend	0.4%	0.6%	0.4%	0.8%	-0.1%

NOTES:

1. Target minus assigned beneficiary expenditures are calculated using the Demonstration financial reconciliation. The plus or minus 2 percent threshold between target and assigned beneficiary expenditures in calculating physician group practice performance payments is **not** reflected in this table.
2. The Pre-Demonstration trend is the change from 2001 to 2004, multiplied by 1/3, 2/3, 3/3, 4/3, 5/3 for PYs 1-5, respectively.
3. Physician Group Practice averages are unweighted averages of results for the applicable practices.
4. Reflects PY5 risk score cap. If the risk score cap was not in place in PY5, the target minus actual expenditures for the average of the PGPs for the Demonstration period in PY5 would be \$225 instead of \$162.
5. Statistical significance is a two-tailed test of difference from zero.

*=statistically significant at 10% level

**=statistically significant at 5% level

***=statistically significant at 1% level

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to Performance Year 5.

7.2 Medicare Program Expenditure Impacts for Subgroups of Assigned Beneficiaries

To further understand Demonstration impacts on Medicare expenditures, we analyzed expenditure impacts for selected subgroups of beneficiaries assigned to PGPs participating in the Demonstration. This analysis helps us understand where savings were, and were not, achieved. Although this section is not an explicit analysis of the impact of the PGPs' care management programs, the participating PGPs targeted certain subgroups of their beneficiaries for interventions. We would not expect the effects of the Demonstration to be uniform across all subgroups.

7.2.1 Methods and Presentation

To analyze subgroup effects, we first defined subgroups of interest. We defined subgroups that were feasible to identify with administrative data, and that were high-cost or otherwise might have been targets for care management. Thus, the selected subgroups might be likely to show Demonstration financial impacts. Eight of these subgroups were beneficiaries with claims diagnoses of a major medical condition, for example, congestive heart failure (CHF) or cancer. A ninth subgroup consisted of beneficiaries with at least one of the 8 major medical conditions. A tenth and eleventh subgroup were an expanded version of the ninth, and consisted of beneficiaries with at least one of the 70 high-cost conditions used in the Demonstration risk adjustment (the same 70 conditions that are used to risk adjust Medicare Advantage capitation payments) and beneficiaries with none of the 70 high-cost conditions. A twelfth subgroup consisted of decedents, who are high-cost and may be the target of palliative care programs. A thirteenth subgroup consisted of assigned beneficiaries with inpatient utilization, i.e., at least one discharge, who are high-cost and may be the target of transitional care management programs as well as a fourteenth subgroup of beneficiaries with zero inpatient expenditures. The fifteenth and sixteenth subgroups consisted of beneficiaries with risk scores exceeding the 90th and 75th percentiles of the risk score distribution, respectively, two other high-cost groups.¹⁹ The last two subgroups were Medicaid dual eligible beneficiaries and beneficiaries entitled to Medicare by disability, both of which could be targets for care management. Note that beneficiaries may belong to more than one subgroup, i.e., a person with multiple chronic conditions will be included in the same disease subgroup as a person with only the one qualifying condition.

To analyze expenditure impacts on these subgroups, we limited the BY and PY5 samples of both assigned and comparison group beneficiaries for each PGP to beneficiaries qualifying for these subgroups. Then we calculated the PY5 difference between assigned beneficiary and target expenditures using the same Demonstration financial reconciliation model that we used for the overall groups of assigned and comparison beneficiaries. The PY5 assigned beneficiary/target expenditure differences were not adjusted for pre-Demonstration trends in the subgroup analysis. The expenditure differences reflect the PY5 risk score cap. Statistical significance was calculated in the same way as for the overall sample, by “bootstrapping” the standard errors. Results are reported in **Table 7-4**.

Average prevalence of each subgroup among assigned beneficiaries is also shown in Table 7-4. An approximate indication of the contribution of a subgroup to the overall PY5 assigned beneficiary and target expenditure difference is given by the product of the prevalence and the estimated difference for the subgroup, which is shown in Table 7-4 as the “prevalence-adjusted difference” column. For example, in Table 7-4, as an average of all 10 PGPs, the estimated target minus assigned beneficiary difference for CHF is \$583 and its prevalence is 13.3 percent. Hence, the estimated contribution of beneficiaries with CHF to the overall \$162 assigned beneficiary/target expenditure difference is \$78 (= \$583 multiplied by 13.3 percent).

¹⁹ The risk score distribution was based on all assigned and comparison group beneficiaries for all 10 participating PGPs, it was not PGP-specific.

Table 7-4
Target minus assigned beneficiaries expenditures per person year by subgroup, performance year five

Beneficiary Subpopulations	Average of Physician Group Practices All 10 Difference (\$)	Average of Physician Group Practices All 10 Prevalence (%)	Average of Physician Group Practices All 10 Prev-adj. Difference (\$)	Average of Physician Group Practices 4 Sharing Savings Difference (\$)	Average of Physician Group Practices 4 Sharing Savings Prevalence (%)	Average of Physician Group Practices 4 Sharing Savings Prev-adj. Difference (\$)	Average of Physician Group Practices 6 Not Sharing Savings Prevalence (%)	Average of Physician Group Practices 6 Not Sharing Savings Prev-adj. Difference (\$)	
All assigned beneficiaries	162**	100.0%	162	626***	100.0%	626	-147*	100.0%	-147
Only beneficiaries with:									
Congestive heart failure	583**	13.3%	78	1,648***	13.3%	219	-127	13.4%	-17
Diabetes	461***	25.3%	116	1,139***	25.7%	293	8	24.6%	2
Coronary artery disease	1,085***	8.4%	91	2,567***	8.8%	226	97	7.6%	7
Cancer	-2	15.9%	0	1,093***	16.2%	177	-732**	15.6%	-114
Chronic obstructive pulmonary disease	454**	13.9%	63	1,144***	13.6%	155	-6	14.5%	-1
Stroke	683	4.2%	29	2,193***	4.1%	90	-325	4.4%	-14
Vascular disease	229	15.9%	36	1,289***	16.2%	208	-478*	15.3%	-73
Heart arrhythmia	423**	15.0%	63	1,074***	14.3%	153	-11	16.2%	-2
Any of the above 8 diagnoses	266***	59.5%	158	937***	59.4%	557	-181*	59.5%	-108
Any of the 70 risk adjustment diagnoses	259***	72.5%	188	856***	73.3%	628	-139	71.0%	-99
None of the 70 risk adjustment diagnoses	-50*	27.5%	-14	-91**	26.7%	-24	-22	29.0%	-6
Decedents	716	3.5%	25	2,264**	3.6%	80	-317	3.4%	-11
Inpatient expenditures>0	742***	21.5%	159	1,342**	21.8%	293	342	20.8%	71

(continued)

Table 7-4 (continued)
Target minus assigned beneficiaries expenditures per person year by subgroup, performance year five

Beneficiary Subpopulations	Average of Physician Group Practices All 10 Difference (\$)	Average of Physician Group Practices All 10 Prevalence (%)	Average of Physician Group Practices All 10 Prev-adj. Difference (\$)	Average of Physician Group Practices 4 Sharing Savings Difference (\$)	Average of Physician Group Practices 4 Sharing Savings Prevalence (%)	Average of Physician Group Practices 4 Sharing Savings Prev-adj. Difference (\$)	Average of Physician Group Practices 6 Not Sharing Savings Difference (\$)	Average of Physician Group Practices 6 Not Sharing Savings Prevalence (%)	Average of Physician Group Practices 6 Not Sharing Savings Prev-adj. Difference (\$)
Inpatient expenditures=0	82**	78.5%	65	282***	78.2%	220	-51	79.2%	-40
Risk score in upper 10%	780**	12.5%	98	2,279***	13.0%	297	-220	11.6%	-26
Risk score in upper 25%	516**	30.2%	156	1,642***	31.4%	515	-235	28.1%	-66
Medicaid enrollee	84	17.4%	15	551**	18.4%	102	-227	15.5%	-35
Entitled to Medicare by disability	313**	19.3%	60	727***	21.0%	153	37	16.3%	6

NOTES:

1. Target minus assigned beneficiary expenditures are calculated using the Demonstration financial reconciliation. The plus or minus 2 percent threshold between target and assigned beneficiary expenditures in calculating physician group practice performance payments is **not** reflected in this table.
2. Prevalence-adjusted difference is target minus assigned beneficiary expenditures multiplied by prevalence. The sum of subgroup differences is greater than the difference for all beneficiaries because beneficiaries may belong to more than one subgroup.
3. Risk adjustment diagnoses are a set of 70 high-cost diagnoses used in Demonstration risk-adjustment.
4. Demonstration Performance Year 5 (PY5) is April 2009 to March 2010, with a base year of 2004.
5. Physician Group Practice averages are unweighted averages of results for the applicable practices.
6. Coronary artery disease is heart attack, unstable angina, and stable angina.
7. Reflects PY5 risk score cap.
8. Statistical significance is a two-tailed test of difference from zero.

*=statistically significant at 10% level

**=statistically significant at 5% level

***=statistically significant at 1% level

SOURCE: RTI analysis of Medicare claims and enrollment data for 2004 to Performance Year 5.

From Table 7-4, one can also determine the approximate percentage contribution of each subgroup to the total Demonstration expenditure impact for all assigned beneficiaries. This is defined as the ratio of the “prevalence-adjusted difference” amount for a subgroup to the “prevalence-adjusted difference” amount for all assigned beneficiaries. For example, in Table 7-4, for the average of all 10 PGPs, the contribution of CHF to the total Demonstration expenditure impact of \$162 for all assigned beneficiaries is \$78 divided by \$162, or 48 percent.

There is considerable overlap among the subgroups shown in Table 7-4, i.e., individual beneficiaries can be in more than one subgroup. The contributions of the subgroups to the overall difference sum to much more than the overall difference because of the large amount of overlap among the subgroups.

7.2.2 Results

Averaged across the 10 Demonstration PGPs, the Demonstration had statistically significant impacts on 13 subgroups between the BY and PY5. The largest statistically significant per beneficiary impacts were on the coronary artery disease (CAD)²⁰, inpatient utilization, and high risk score subgroups. Other subgroups showing statistically significant impacts were congestive heart failure, diabetes, chronic obstructive pulmonary disease, heart arrhythmia, having at least one of the 8 prevalent medical conditions, having one of the 70 high-cost risk adjustment diagnoses, and entitled to Medicare by disability.

Adjusted for prevalence, the largest subgroup contribution to the overall Demonstration impact per beneficiary was from beneficiaries with any of the 70 risk adjustment diagnoses, who accounted for all of the overall impact (\$188 subgroup impact versus \$162 overall impact). Beneficiaries with inpatient expenditures, any of the 8 diagnoses, as well as high risk score beneficiaries and beneficiaries with diabetes also accounted for a large amount of the Demonstration expenditure impact.²¹

The importance of the beneficiary subgroup with hospital stays is not surprising because this is a relatively large and high-cost group where expenditure impacts are likely to be seen, if anywhere. Inpatients were targeted by some participating PGPs--in particular for transition management programs. The importance of beneficiaries with the 70 high-cost diagnoses is also not surprising for similar reasons. The diabetes group as well as high risk patients are targeted by some PGPs.

CHF was one of the subgroups of beneficiaries showing a Demonstration impact, and this was statistically significant. A positive impact is estimated (\$583 per beneficiary, \$78 adjusted for prevalence). This is not surprising because reducing expenditures for CHF patients was a major focus of nearly all of the participating PGPs, and several participating PGPs implemented

²⁰ 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.

²¹ There is significant overlap between the inpatient-using subgroup and the subgroups with at least one of the 70 high-cost risk adjustment diagnoses, and much of the expenditure impact may occur among beneficiaries belonging to both groups.

significant CHF disease management programs. However, this was not found to be a statistically significant subgroup in the early Demonstration (PY2). No statistically significant impact was found for decedents despite palliative or end of life programs of some participating PGPs. Decedents are a small group with highly variable expenditures, however.

Among provider groups sharing PY5 savings, the Demonstration expenditure impact per beneficiary, \$626, was nearly quadruple the amount for all 10 PGPs, \$162. All subgroups showed statistically significant impacts for those PGPs sharing in savings. The 4 PGPs sharing PY5 savings were more successful in controlling expenditure growth for all subgroups than the 10 PGP average. Subgroups showing a large statistically significant Demonstration impact among PGPs sharing PY5 savings were coronary artery disease, stroke, beneficiaries with high risk scores, and decedents.

For all 10 PGPs, adjusted for prevalence, and the 4 PGPs sharing PY5 savings, adjusted for prevalence, assigned beneficiaries with any of the 70 risk adjustment diagnoses had the largest impact in accounting for the Demonstration overall net expenditure impact. For the 4 PGPs sharing PY5 savings, adjusted for prevalence, beneficiaries with any of the eight diagnoses as well as high risk score beneficiaries were also large contributors to the overall reductions.

On average as a group, the 6 PGPs that did not share in PY5 savings had higher expenditure growth than their target from the base year to PY5. Their total expenditure impact of -\$147 is statistically different from zero, meaning that during the Demo, their costs increased more than their local Comparison Groups. The shortcomings in their performance were largest in three subgroups, beneficiaries with cancer, with vascular disease, and with any of the 8 major chronic diagnoses. These three subgroups had negative expenditure impacts that were statistically different from zero.

7.3 Components of Medicare Program Expenditures

In this Section, we analyze the impact of the PGP Demonstration on components of Medicare program expenditures. Our analysis is shown in *Table 7-5*. The components are: hospital inpatient, skilled nursing facility, total outpatient, Part B physician/supplier, hospital outpatient, home health agency, and durable medical equipment. “Part B physician/supplier” expenditures include primarily professional services furnished by physicians and other clinicians, but also includes some other Part B services such as clinical laboratory tests furnished by independent clinical laboratories. “Hospital outpatient” primarily includes services furnished by hospital outpatient departments, but also includes services furnished by other institutional outpatient providers. In addition to Part B physician/supplier and hospital outpatient services, “total outpatient” expenditures include home health and durable medical equipment expenditures.

Table 7-5 shows target expenditures minus actual expenditures by service component for PY5. Over all 10 PGPs, inpatient expenditures were \$95 per person year less than target; this difference was statistically significant ($p < .05$). In contrast, total outpatient expenditures were only \$37 per person year less than target; this difference was not statistically significant. Thus, Demonstration cost savings seem to be more related to controlling inpatient than outpatient

Table 7-5
Target minus assigned beneficiary expenditures per person year by expenditure component, performance year five

Expenditure Component	Average of Physician Group Practices (\$): All 10	Average of Physician Group Practices (\$): 4 sharing savings	Average of Physician Group Practices (\$): 6 not sharing savings
Total	162.0**	625.7***	-147.2*
Hospital Inpatient	94.9**	313.3***	-50.8
Skilled nursing facility	-33.2**	20.5	-69.0**
Total outpatient	37.0	243.5***	-100.7**
Physician/supplier	2.6	5.7	0.5
Hospital outpatient	9.7	70.2**	-30.6
Home Health Agency	43.0***	84.2***	15.5
Durable medical equipment	2.9	9.8	-1.6

NOTES:

1. Target minus assigned beneficiary expenditures are calculated using the Demonstration financial reconciliation. The plus or minus 2 percent threshold between target and assigned beneficiary expenditures in calculating physician group practice performance payments is **not** reflected in this table.
2. Expenditure components do not sum to the total because the financial reconciliation algorithm is nonlinear.
3. Demonstration Performance Year 5 (PY5) is April 2009 to March 2010, with a base year of 2004.
4. Physician Group Practice averages are unweighted averages of results for the applicable practices.
5. Reflects the PY5 risk score cap.
6. Statistical significance is a two-tailed test of difference from zero.

*=statistically significant at 10% level.

**=statistically significant at 5% level.

***=statistically significant at 1% level.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2004 to Performance Year 5.

expenditures. This was a slight change from our earlier report²² in PY2 that found in the early Demonstration years, savings were mostly from outpatient expenditures. All or most PGP outpatient expenditure savings arose from lower PGP home health expenditures. The participating PGPs on average were not successful in controlling skilled nursing facility costs, as they were \$33 per person year higher among assigned beneficiaries than target.

Among the 4 PGPs earning performance payments in PY5, inpatient hospital expenditures were \$313 per person year lower than target, and total outpatient expenditures were \$244 per person year lower than target (both were statistically significant). This indicates that financially successful PGPs were able to control both inpatient and outpatient costs. Home health agency expenditures were \$84 per person year lower than target and hospital outpatient expenditures were \$70 per person year lower than target, making the largest contribution to lower total outpatient expenditures. For the 6 PGPs not earning performance payments in PY5, actual expenditures were statistically greater than target expenditures for skilled nursing facilities and total outpatient, but not significantly different than target for inpatient. This indicates that not-financially-successful PGPs had trouble controlling skilled nursing facility and total outpatient costs more than inpatient costs.

²² Sebelius, Kathleen. *Physician Group Practice Evaluation Report*. Report to Congress. 2009.

This page intentionally left blank.

CHAPTER 8 DEMONSTRATION IMPACTS ON QUALITY OF CARE

Chapter 6 discussed the quality performance of the ten physician group practices from the base year to the last performance year of the Demonstration. Although that chapter showed that the PGPs, in general, improved in their quality performance over time, comparison group information is necessary to discern any Demonstration effects on the groups' quality of care. Comparing the quality performance between the PGPs and their respective Comparison Groups (CGs) is required to evaluate any improvement in quality scores by the PGPs that were attributable to the interventions they applied in the Demonstration.

This chapter presents results of the seven claims-based quality measures as compared between the ten participating PGPs and their respective CGs. Each CG is comprised of beneficiaries who reside in the geographic areas served by the PGPs. The CG Medicare claims, and in particular, their claims-based quality measure performance, can be analyzed and compared to the PGPs. Thus, the impact of the Demonstration on quality can be examined by comparing the values of the seven claims-based quality measures for each PGP and its CG²³. As a result, evaluating the Demonstration's effect on quality of care depends on the following seven claims-based quality measures:

- Diabetes Mellitus
 - DM-1 HbA1c Management (Testing)
 - DM-4 LDL-C Screening
 - DM-6 Urine Protein Testing
 - DM-7 Eye Exam
- Heart Failure
 - HF-2 LVF-Testing
- Coronary Artery Disease
 - CAD-5 Lipid Profile
- Preventive Care
 - PC-5 Breast Cancer Screening

Differences in the quality results between the base year and PY5 were compared between the PGPs and their CGs. Performance assessment among the PGPs was conducted using

²³ Only the seven claims-based measures are reported (and not all 32 measures in the PGP Demonstration), because the 25 chart-based measures, including clinical values, could not be collected for the comparison groups.

Medicare FFS Inpatient, Outpatient and Part B Physician/Supplier claims for all beneficiaries assigned to the groups. Assessment of CG results was conducted using the same data source for assigned beneficiaries residing in service area counties, and weighted by county according to the method used in the financial analysis.

8.1 Trends in Assigned Beneficiary versus Comparison Group Quality of Care

8.1.1 Performance Results from Base Year to PY5

Figures 8-1 to 8-7 show the quality of care performance score results for each of the seven claims-based measures in the base year, and in selected performance years PY2 and PY5. Quality performance was defined as the rate of numerator hits to the denominator of eligible beneficiaries. The seven trend figures were consistent in a number of ways, albeit with a few exceptions. First, the PGPs (shown in blue lines) mostly had higher performance scores compared to their comparison group (shown in red dotted lines). Second, the performance between the PGPs and the CGs did not differ widely for a number of measures, as indicated by the proximity and overlaps of the two lines, such as in DM-1 (HbA1c testing), HF-2 (LVF testing) and CAD-5 (lipid profile). There were also a number of instances where there were improvements between the base year and PY2, but this improvement either leveled off or actually reversed trend by PY5 (see Figure 8-3 for DM-6: Urine protein testing), although these trends generally affected both the PGPs and CGs. Finally, when there were large changes in scores from one year to the next, these ‘jumps’ were limited to the blue solid lines, suggesting that PGPs had steeper increases in their performance. These jumps were the most evident in Figure 8-3 (multiple PGPs), Figure 8-4 (PGP 10), Figure 8-5 (PGP 8), and Figure 8-7 (PGP 5).

Figure 8-1

Trend in quality performance for DM-1 (HbA1c testing): PGPs vs. comparison groups

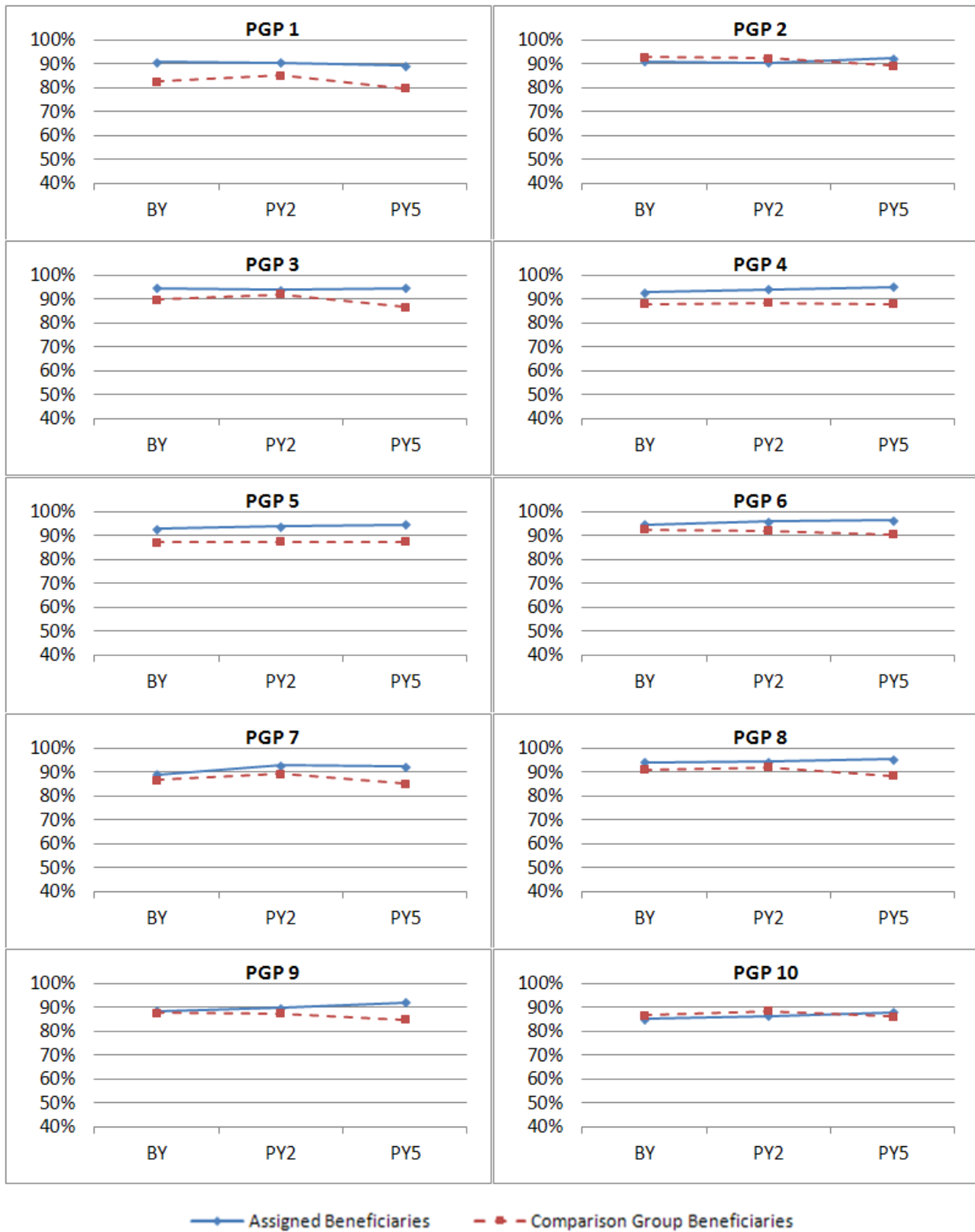


Figure 8-2

Trend in quality performance for DM-4 (LDL-C testing): PGPs vs. comparison groups

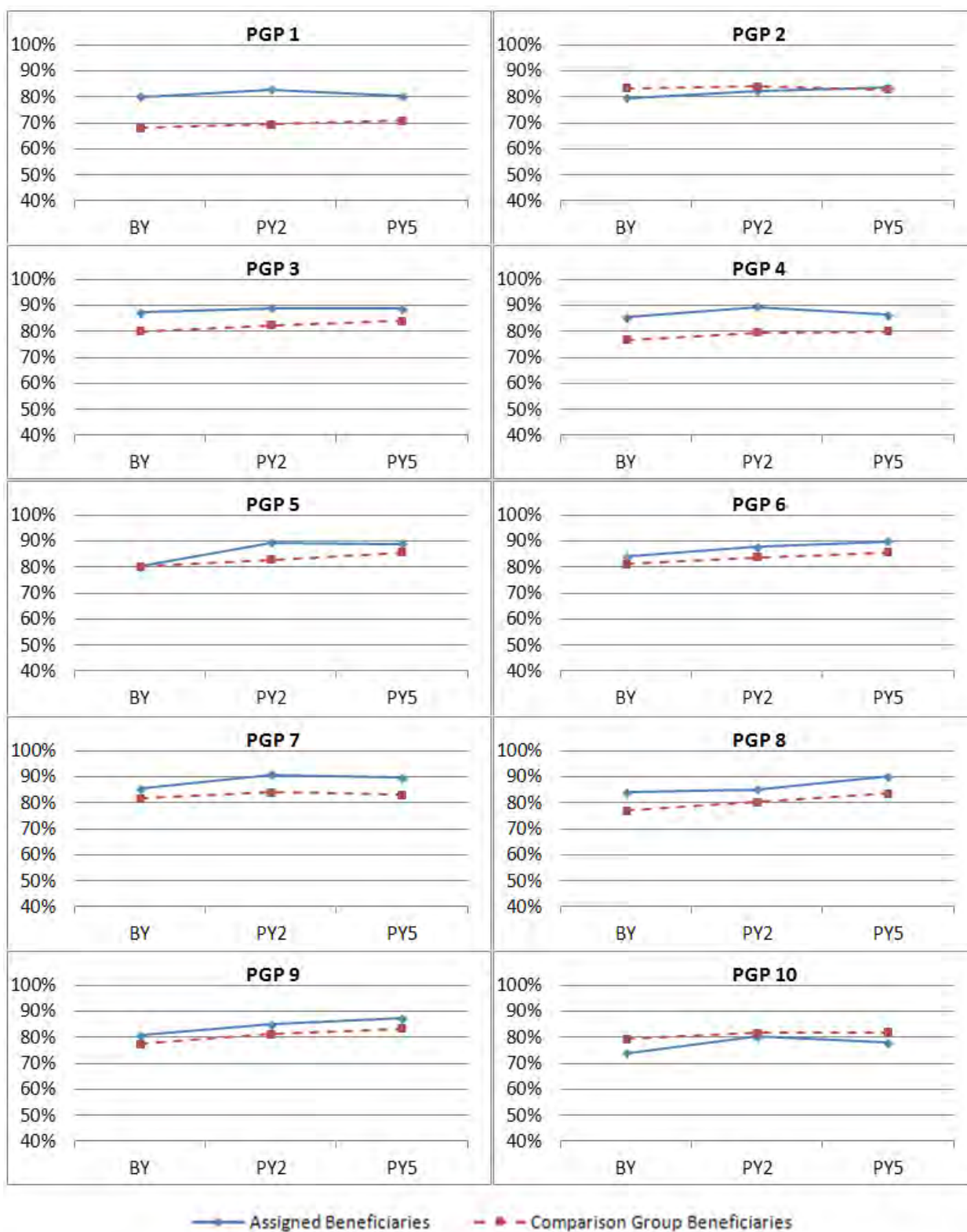


Figure 8-3

Trend in quality performance for DM-6 (Urine protein testing): PGPs vs. comparison groups

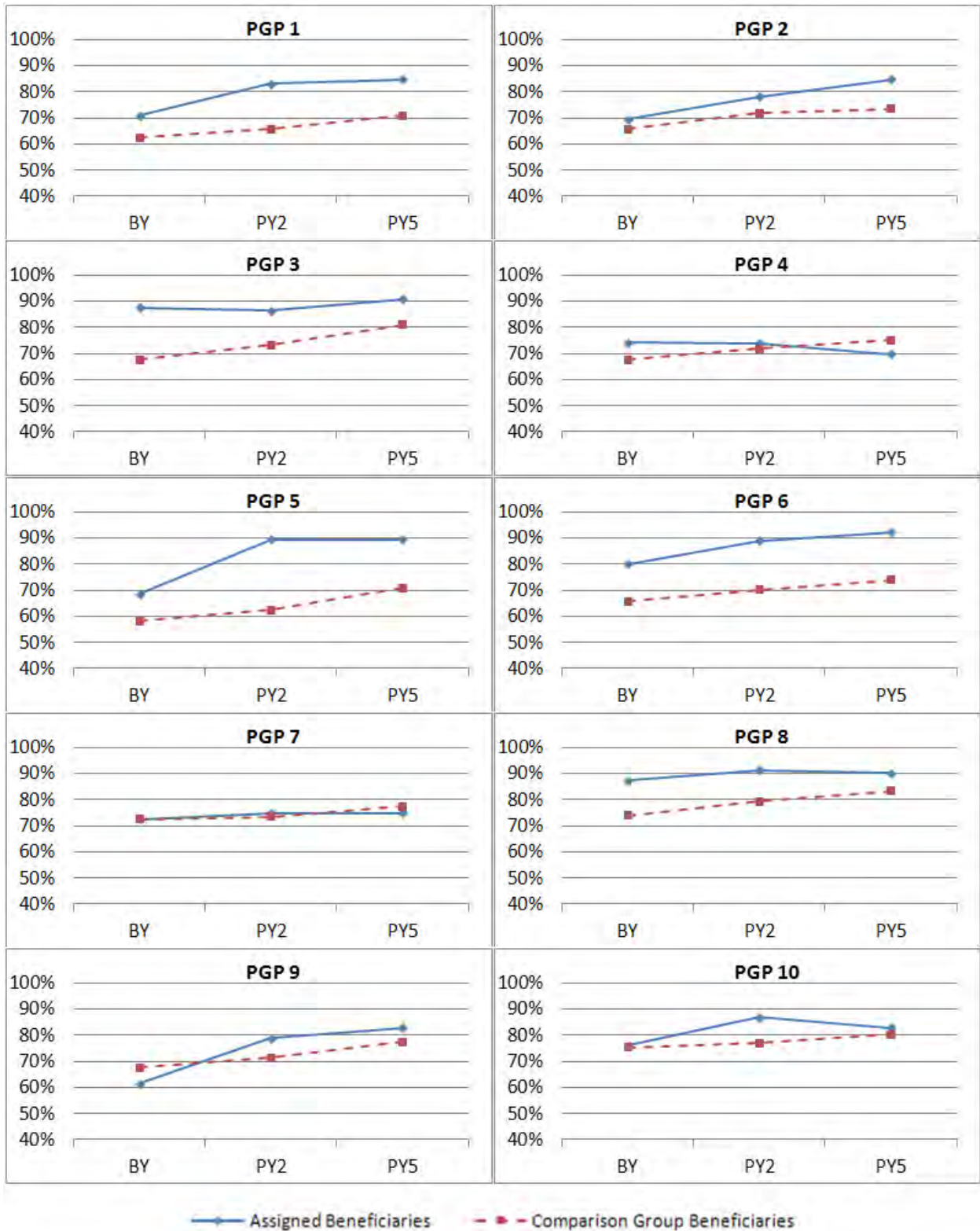


Figure 8-4
Trend in quality performance for DM-7 (Eye exam): PGPs vs. comparison groups

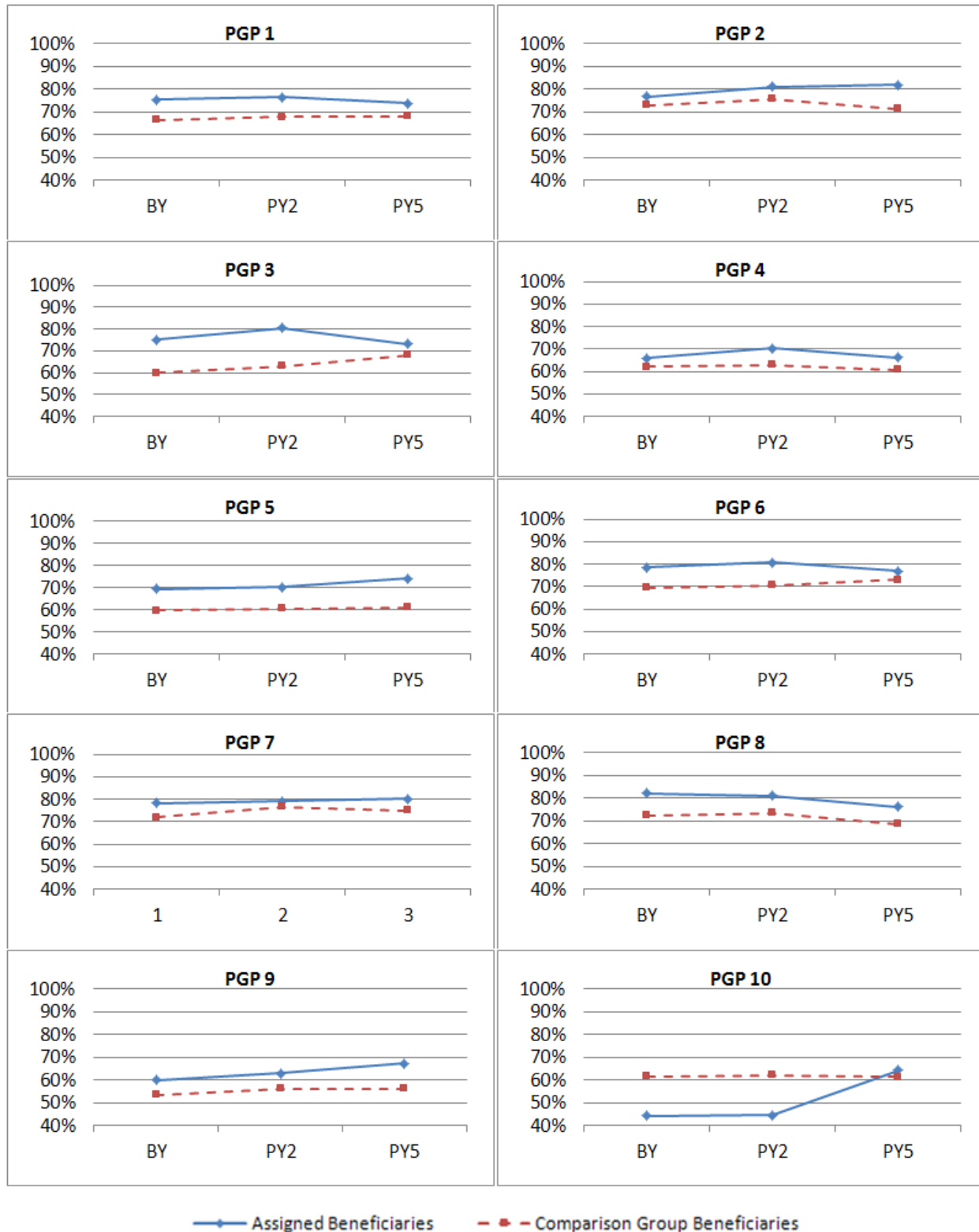


Figure 8-5
Trend in quality performance for HF-2 (LVF testing): PGPs vs. comparison groups

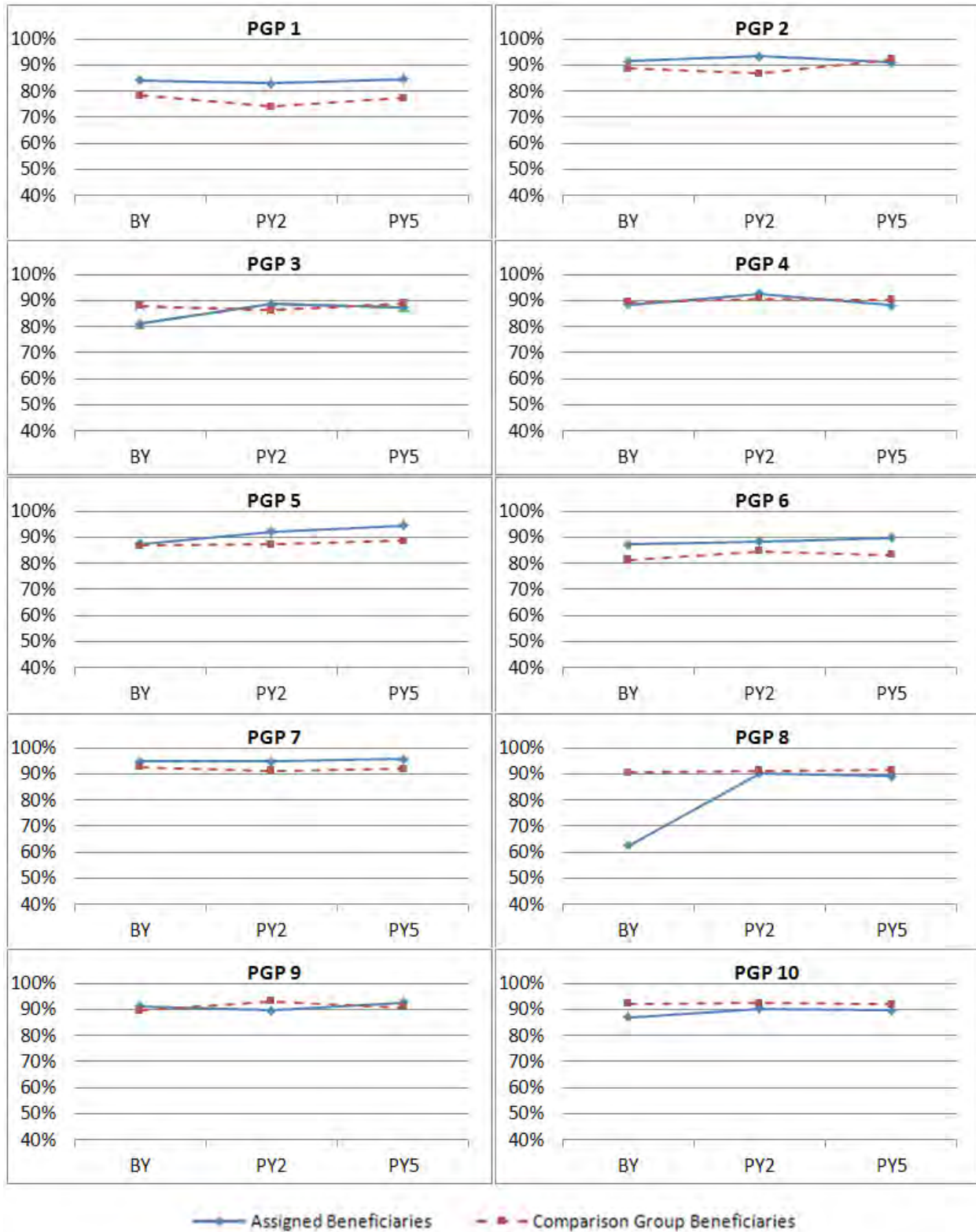


Figure 8-6
Trend in quality performance for CAD-5 (Lipid profile): PGP vs. comparison groups

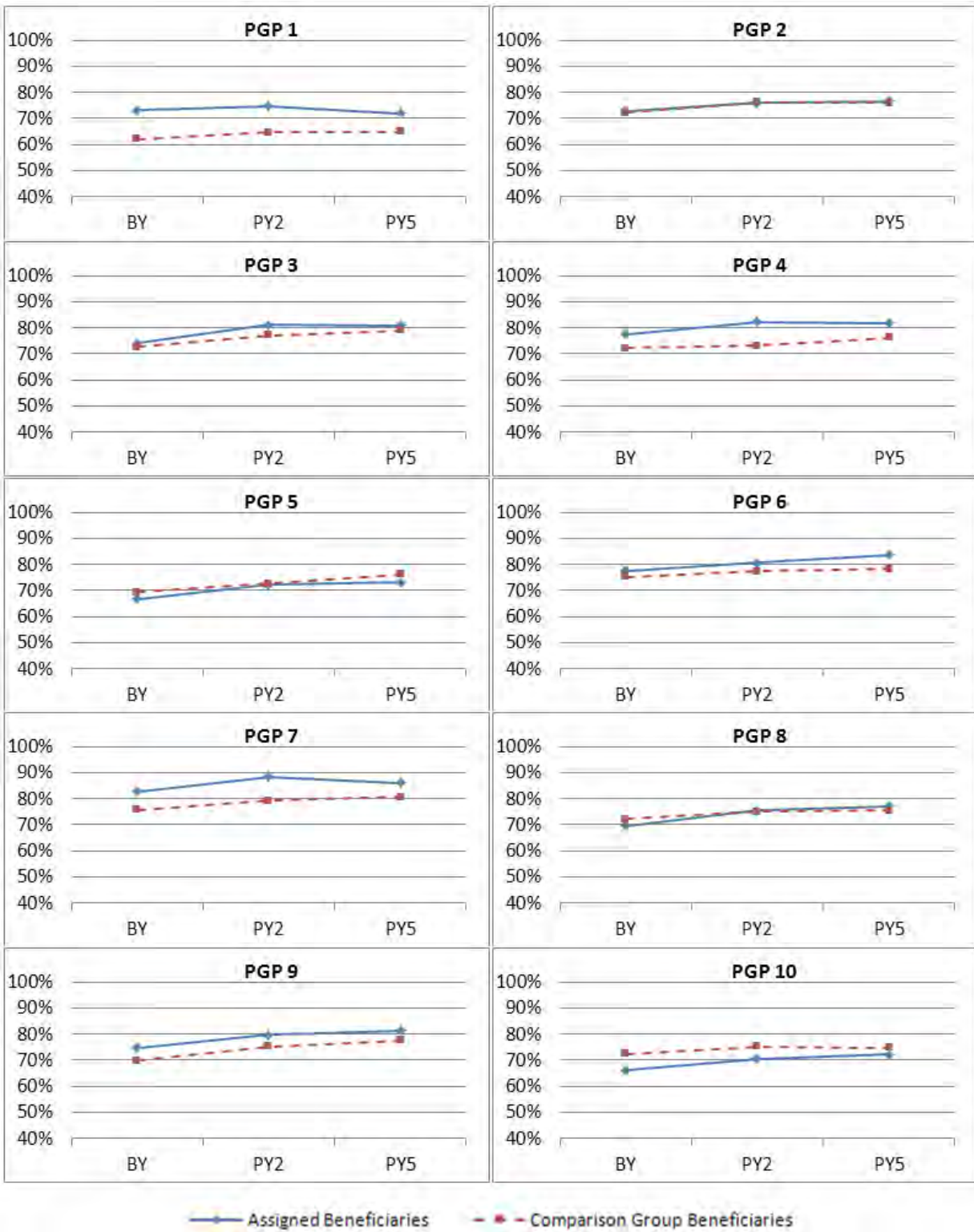
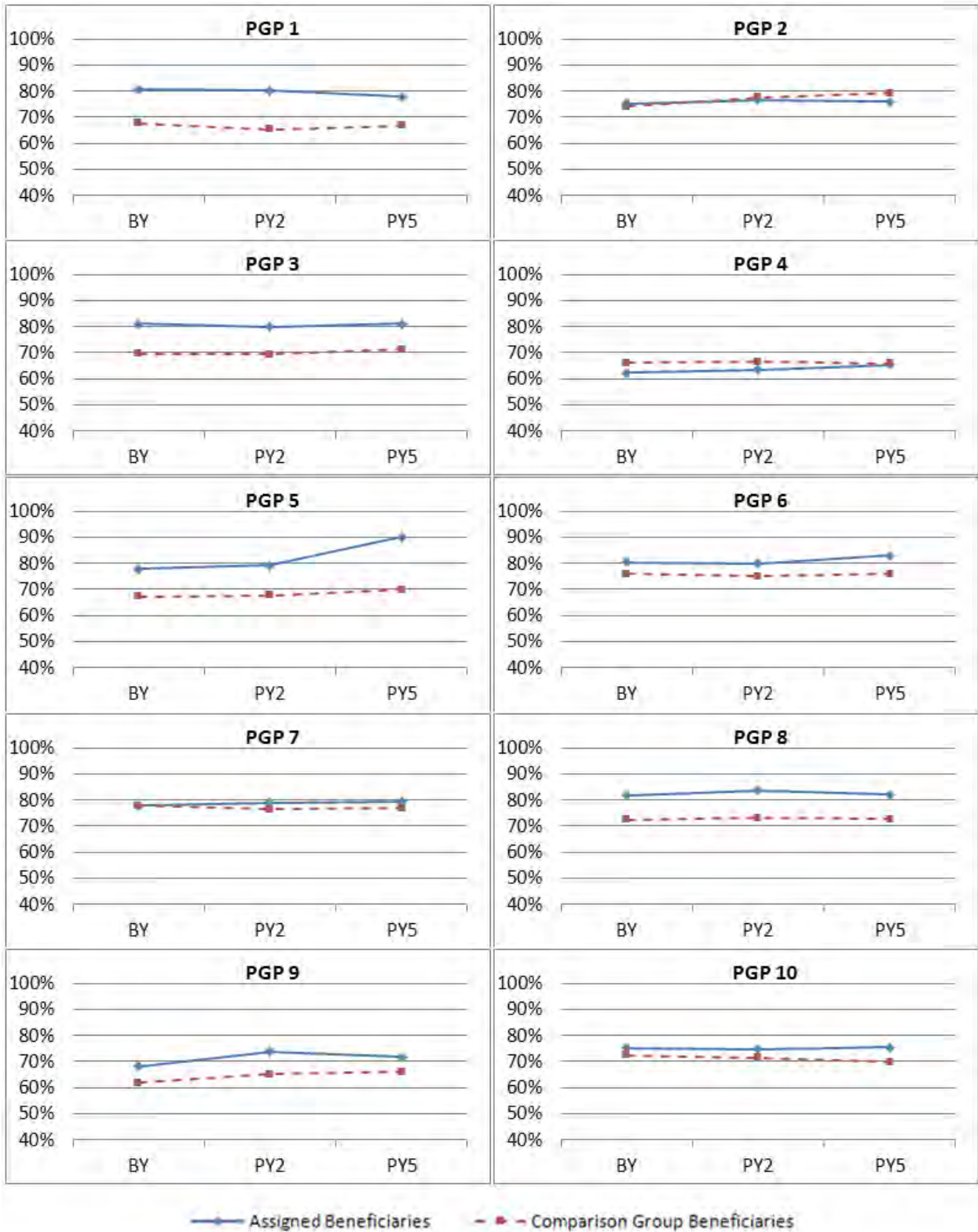


Figure 8-7

Trend in quality performance for PC-5 (Breast cancer screening): PGP vs. comparison groups



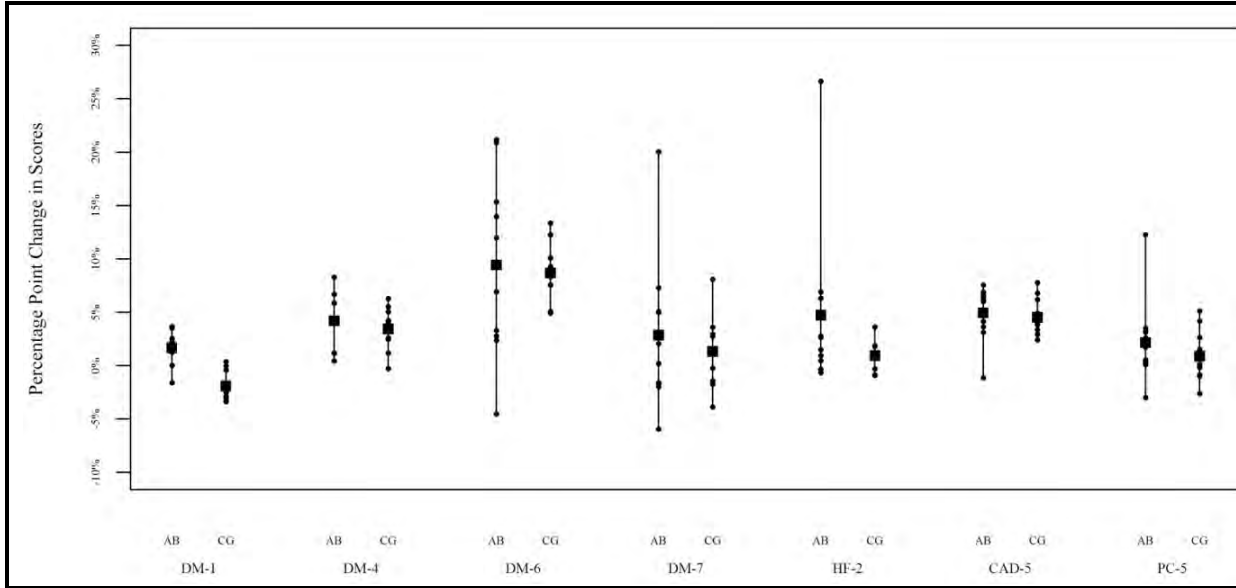
8.1.2 Change in Scores for all Seven Claims-based Quality Measures from Base Year to PY5

Improvements in quality during the Demonstration may be better captured by the change in scores among all ten PGPs compared to the ten CGs from the base year to PY5. *Figure 8-8* shows the patterns of changes between the two sets of groups, expressed as the difference in score within each group between PY5 and the Base Year (PY5 minus BY). Thus, positive differences signify improvement over time, while negative differences signify the opposite. The lines reflect the range of the differences in scores across the groups, and the square in the line shows the mean difference across the PGPs, or across the CGs.

For all seven measures, performance among the PGPs showed more positive changes or improvements, as shown by the higher average changes in all seven measures. For example, in DM-4, the average difference between BY and PY5 for the PGPs was 4.2 percent while the average difference for the CGs was 3.5 percent. Notably, for DM-1 (HbA1c Testing), the CGs performed worse in PY5 than in the base year, with an average 1.9 percent decrease across all CGs.

Looking at the range of change, there were wider distributions among the PGPs with the most notable differences found in DM-6 (urine protein testing), DM-7 (eye exam) and HF-2 (LVF testing following hospitalization). For DM-6 (urine protein testing), from base year to PY5, the PGPs varied from a decrease of 4.5 percentage points in score to an increase of 21.2 percentage points, compared to a positive range of change from 5 to 13.4 percentage points in the CGs. Similarly, the range of change in DM-7 (Eye Exam) was -5.9 to 20.0 percentage points across the PGPs, while the CGs had a range of change from -3.9 to 8.1 percentage points. For HF-2 (LVF testing following hospitalization), the difference was even more pronounced given the notable improvement in one PGP: the PGPs varied from a decrease of 0.7 percentage points in score to an increase of 26.3 percentage points, compared to changes from -0.8 to 3.6 percentage points for the CGs. The PGP with the largest change, PGP 8, improved from 62.5 percent to 89.1 percent in the rate of LVF testing following hospitalization for heart failure patients (HF-2). On the other hand, the range of change in scores for DM-1 (HbA1c testing) was more compact for both the PGPs and the CGs, which may be due to little room left for improvement since the achievement rate in the base year for both groups was already high.

Figure 8-8
Range of change in performance scores from base year to PY5 across PGPs and CGs



NOTES: DM-1=Diabetes Mellitus HbA1c Management; DM-4= Diabetes Mellitus Lipid Measurement; DM-7= Diabetes Mellitus Eye Exam; HF-2=Congestive Heart Failure Left Ventricular Ejection Fraction Testing; CAD-5=Coronary Artery Disease Lipid Profile; PC-5=Hypertension/Preventive Care Breast Cancer Screening.

SOURCE: RTI analysis of 2004–2010 Medicare claims.

8.2 Statistical Estimates of Demonstration Quality of Care Impacts

8.2.1 Assessment of Quality Results from Base Year to PY5--Overall

The previous sections provided descriptive changes in quality between the base year and PY5, and showed some quality improvements identified in the PGPs compared to the CGs. To ascertain whether the differences in quality improvement are statistically meaningful, or attributable to the PGP Demonstration, we conducted a difference-in-differences (DD) assessment of each of the seven claims-based quality measures. The DD technique calculates the difference between the base year and PY5 quality scores for the PGPs, the difference between the base year and PY5 quality scores for the CGs (weighted by county), and finally the difference in the PGP and CG differences. As before, a positive difference for the PGP or CG between the two years shows improvement in scores, and a positive difference in the DD shows that PGPs improved more compared to the CGs. To test for statistical significance, we calculated three sets of standard errors. First, the standard errors for each quality measure rate were calculated for both the baseline rates and the PY5 rates. Next, standard errors for each group's difference (PY5 minus BY) were calculated using the formula for the difference of two proportions. Third, the standard errors for the difference in difference rates were calculated using the formula for the difference of two means. Finally, we calculated the z-values as the

quotient of the DD rate and its standard error, from which we generated the final p-value assuming a normal distribution of results.

As shown in **Table 8-1**, all seven quality scores improved between the base year and PY5 for both groups, except for one measure, DM-1, which declined among the CGs by an average of 1.9 percentage points. The assigned beneficiaries in the PGPs had statistically significantly larger improvements over time in four of the seven measures compared to those in the CGs. For example, the 1.7 percentage point improvement among the PGPs in DM-1 was statistically different from the 1.9 percentage point decline in performance score among the CGs ($p < 0.001$). For DM-7 (eye exam) and PC-5 (Breast cancer screening), the DD results showed significant differences between the PGPs and the CGs (1.5 and 1.3 percentage points, respectively; $p < 0.01$). For HF-2 (LVF testing following hospitalization), the 4.7 percentage point increase over time among the PGPs was significantly larger than the 1 percentage point increase among the CGs ($p < 0.01$). For the remaining three measures (DM-4, DM-6, and CAD-5), the difference in the improvements between the PGPs and the CGs were not statistically different.

8.2.2 Assessment of Quality Results from Base Year to PY5—By Site

DD scores for each PGP and CG for each of the seven claims-based quality measures are provided in **Table 8-2** through **Table 8-8**. The measure with the largest number of groups with significant fluctuations over their respective CGs was DM-6 (urine protein testing), where five PGPs had statistically significant positive DD scores, but three groups had a statistically significant negative DD score, suggesting that for these three PGP sites the improvements by the CGs were larger than the improvements in the PGPs (Table 8-4). In fact, for PGP 4, there was a decline in PGP score over time (by 4.5 percentage points) compared to an improvement in score (by 7.6 percentage points) in their CG ($p < 0.001$). Similarly, in DM-7 (eye exam), four PGPs had significantly larger improvements than their comparison group, but this was not the case for PGP 3, where their score declined over time as their CG gained an extra 8.1 percentage points (Table 8-5). In fact, regardless of statistical significance, three PGPs had declines in their eye exam performance scores compared to their CGs.

In DM-1, interestingly, nine out of ten of the CGs had a decline in HbA1c testing, while the same number of PGPs trended the other (i.e., more appropriate) direction, which rendered most PGPs to perform statistically better compared to their CGs (Table 8-2). Notably, there were a number of large improvements in certain PGP scores from BY to PY5 that were associated with the statistically significant differences between the groups. For example, in DM-6, PGP 9 had a 21.2 percentage point improvement over the six years measured compared to the 10.1 percentage point improvement in their CG (Table 8-4). In DM-7, PGP 10 had a 20.0 percentage point increase while their CG had a slight decrease over time (Table 8-5). In HF-2, PGP 8 had a jump of 26.6 percentage points between the base year and PY5, while their CG showed a less than 1 percentage point increase (Table 8-6). Finally, unlike our 2008 interim report (Kautter et al., 2008), where no significant differences in DD scores were detected between the PGPs and CGs in DM-7 (eye exam) and PC-5 (breast cancer screening), by PY5, there were significant differences in at least one PGP for all of the seven quality measures.

Table 8-1
Average differences in the claims-based quality performance scores between base year and performance year five
across all ten PGPs

Quality Measure	Percentage Point Change in Quality Performance Scores Assigned Beneficiaries: BY	Percentage Point Change in Quality Performance Scores Assigned Beneficiaries: PY5	Percentage Point Change in Quality Performance Scores Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in Quality Performance Scores Comparison Group Beneficiaries: BY	Percentage Point Change in Quality Performance Scores Comparison Group Beneficiaries: PY5	Percentage Point Change in Quality Performance Scores Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
DM-1 HbA1c Management	91.32%	92.99%	1.66%	88.49%	86.60%	-1.89%	3.55%***
DM-4 Lipid Measurement	82.08%	86.28%	4.20%	78.58%	82.03%	3.45%	0.75%
DM-6 Urine Protein Testing	74.82%	84.25%	9.43%	67.71%	76.41%	8.70%	0.74%
DM-7 Eye Exam	70.58%	73.43%	2.85%	65.06%	66.40%	1.33%	1.52%**
HF-2 Left Ventricular Ejection Fraction Testing	85.53%	90.25%	4.73%	87.70%	88.66%	0.96%	3.77%**
CAD-5 Lipid Profile	73.55%	78.50%	4.95%	71.39%	75.92%	4.53%	0.42%
Breast Cancer Screening	76.04%	78.21%	2.17%	70.54%	71.45%	0.91%	1.26%**

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

NOTES: DM = Diabetes; HF = Heart Failure; CAD = Coronary Artery disease; PC = Preventive Care

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-2

Differences in DM1: HbA1c management quality performance score between base year and performance year five

PGP	Percentage Point Change in DM1 – HbA1c Management Quality Score Assigned Beneficiaries: BY	Percentage Point Change in DM1 – HbA1c Management Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in DM1 – HbA1c Management Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in DM1 – HbA1c Management Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in DM1 – HbA1c Management Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in DM1 – HbA1c Management Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	90.71%	89.14%	-1.57%	82.58%	79.71%	-2.87%	1.30%
PGP 2	91.10%	92.37%	1.27%	92.73%	89.48%	-3.25%	4.52%***
PGP 3	94.51%	94.56%	0.05%	89.91%	86.55%	-3.36%	3.41%**
PGP 4	92.79%	95.01%	2.22%	87.95%	87.93%	-0.02%	2.24%**
PGP 5	92.92%	94.43%	1.51%	87.11%	87.50%	0.39%	1.12%
PGP 6	94.52%	96.31%	1.79%	92.65%	90.42%	-2.23%	4.02%***
PGP 7	89.07%	92.60%	3.53%	86.75%	85.09%	-1.66%	5.19%**
PGP 8	93.96%	95.59%	1.63%	91.02%	88.51%	-2.51%	4.14%***
PGP 9	88.42%	92.08%	3.66%	87.57%	84.63%	-2.94%	6.60%***
PGP 10	85.22%	87.76%	2.54%	86.59%	86.18%	-0.41%	2.95%**
Average***	91.32%	92.99%	1.66%	88.49%	86.60%	-1.89%	3.55%

* Statistically significant at p<0.05

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-3

Differences in DM4: Lipid measurement quality performance score between base year and performance year five

PGP	Percentage Point Change in DM4 – Lipid Measurement Quality Score Assigned Beneficiaries: BY	Percentage Point Change in DM4 – Lipid Measurement Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in DM4 – Lipid Measurement Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in DM4 – Lipid Measurement Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in DM4 – Lipid Measurement Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in DM4 – Lipid Measurement Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	79.82%	80.26%	0.44%	68.30%	70.91%	2.61%	-2.17%
PGP 2	79.37%	83.56%	4.19%	83.07%	82.79%	-0.28%	4.47%**
PGP 3	87.49%	88.64%	1.15%	79.94%	84.06%	4.12%	-2.97%*
PGP 4	85.43%	86.63%	1.20%	76.81%	79.98%	3.17%	-1.97%
PGP 5	80.53%	88.84%	8.31%	80.49%	85.56%	5.07%	3.24%**
PGP 6	84.10%	89.97%	5.87%	81.28%	85.50%	4.22%	1.65%
PGP 7	85.32%	89.63%	4.31%	81.69%	82.97%	1.28%	3.02%*
PGP 8	84.15%	90.01%	5.86%	77.16%	83.45%	6.29%	-0.43%
PGP 9	80.66%	87.37%	6.71%	77.70%	83.26%	5.56%	1.15%
PGP 10	73.91%	77.92%	4.01%	79.32%	81.81%	2.49%	1.52%
Average	82.08%	86.28%	4.20%	78.58%	82.03%	3.45%	0.75%

* Statistically significant at p<0.05

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-4

Differences in DM6: Urine protein testing quality performance score between base year and performance year five

PGP	Percentage Point Change in DM6: Urine Protein Testing Quality Score Assigned Beneficiaries: BY	Percentage Point Change in DM6: Urine Protein Testing Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in DM6: Urine Protein Testing Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in DM6: Urine Protein Testing Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in DM6: Urine Protein Testing Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in DM6: Urine Protein Testing Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	70.84%	84.80%	13.96%	62.53%	70.86%	8.33%	5.63%**
PGP 2	69.43%	84.77%	15.34%	65.73%	73.32%	7.59%	7.75%***
PGP 3	87.60%	90.91%	3.31%	67.62%	80.99%	13.37%	-10.06%***
PGP 4	74.24%	69.74%	-4.50%	67.55%	75.10%	7.55%	-12.05%***
PGP 5	68.55%	89.45%	20.90%	58.52%	70.81%	12.29%	8.61%***
PGP 6	80.08%	92.10%	12.02%	65.66%	74.02%	8.36%	3.66%**
PGP 7	72.55%	74.95%	2.40%	72.61%	77.62%	5.02%	-2.62%
PGP 8	87.23%	90.01%	2.78%	73.97%	83.25%	9.28%	-6.50%***
PGP 9	61.73%	82.92%	21.19%	67.56%	77.67%	10.11%	11.08%***
PGP 10	75.91%	82.84%	6.93%	75.39%	80.45%	5.06%	1.87%
Average	74.82%	84.25%	9.43%	67.71%	76.41%	8.70%	0.74%

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-5
Differences in DM7: Eye exam quality performance score between base year and performance year five

PGP	Percentage Point Change in DM7: Eye Exam Quality Score Assigned Beneficiaries: BY	Percentage Point Change in DM7: Eye Exam Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in DM7: Eye Exam Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in DM7: Eye Exam Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in DM7: Eye Exam Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in DM7: Eye Exam Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	75.25%	73.66%	-1.59%	66.55%	68.16%	1.61%	-3.20%
PGP 2	76.73%	81.81%	5.08%	73.04%	71.31%	-1.73%	6.81%***
PGP 3	75.18%	73.23%	-1.95%	60.07%	68.15%	8.08%	-10.03%***
PGP 4	66.07%	66.29%	0.22%	62.13%	60.71%	-1.42%	1.64%
PGP 5	69.26%	74.25%	4.99%	59.64%	61.11%	1.47%	3.52%**
PGP 6	78.62%	76.91%	-1.71%	69.61%	73.23%	3.62%	-5.33%***
PGP 7	78.28%	80.35%	2.07%	72.09%	75.11%	3.02%	-0.95%
PGP 8	82.07%	76.14%	-5.93%	72.42%	68.56%	-3.86%	-2.07%
PGP 9	60.07%	67.39%	7.32%	53.47%	56.24%	2.77%	4.55%**
PGP 10	44.21%	64.24%	20.03%	61.63%	61.40%	-0.23%	20.26%***
Average**	70.58%	73.43%	2.85%	65.06%	66.40%	1.33%	1.52%**

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-6
Differences in HF2: LVG testing quality performance score between base year and performance year five

PGP	Percentage Point	Percentage Point	Percentage Point	Percentage Point	Percentage Point	Percentage Point	Difference in Difference (AB minus CG)
	Change in HF2: LVF Testing Quality Score Assigned Beneficiaries: BY	Change in HF2: LVF Testing Quality Score Assigned Beneficiaries: PY5	Change in HF2: LVF Testing Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Change in HF2: LVF Testing Quality Score Comparison Group Beneficiaries: BY	Change in HF2: LVF Testing Quality Score Comparison Group Beneficiaries: PY5	Change in HF2: LVF Testing Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	
PGP 1	84.17%	84.67%	0.50%	78.24%	77.41%	-0.83%	1.33%
PGP 2	91.56%	90.91%	-0.65%	88.61%	92.25%	3.64%	-4.29%
PGP 3	81.16%	87.50%	6.34%	87.85%	88.95%	1.10%	5.24%
PGP 4	88.42%	88.10%	-0.32%	89.38%	90.37%	0.99%	-1.31%
PGP 5	87.57%	94.49%	6.92%	86.85%	88.67%	1.82%	5.10%*
PGP 6	87.16%	89.83%	2.67%	81.28%	83.17%	1.89%	0.78%
PGP 7	94.67%	95.63%	0.96%	92.64%	91.84%	-0.80%	1.77%
PGP 8	62.50%	89.13%	26.63%	90.47%	91.39%	0.92%	25.71%***
PGP 9	91.12%	92.61%	1.49%	89.55%	90.71%	1.16%	0.33%
PGP 10	86.92%	89.67%	2.75%	92.17%	91.86%	-0.31%	3.06%
Average**	85.53%	90.25%	4.73%	87.70%	88.66%	0.96%	3.77%**

* Statistically significant at p<0.05

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-7
Differences in CAD5: Lipid profile quality performance score between base year and performance year five

PGP	Percentage Point Change in CAD5: Lipid Profile Quality Score Assigned Beneficiaries: BY	Percentage Point Change in CAD5: Lipid Profile Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in CAD5: Lipid Profile Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in CAD5: Lipid Profile Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in CAD5: Lipid Profile Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in CAD5: Lipid Profile Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	73.09%	71.97%	-1.12%	62.05%	64.98%	2.93%	-4.05%**
PGP 2	72.96%	76.60%	3.64%	72.21%	76.07%	3.86%	-0.22%
PGP 3	74.00%	80.91%	6.91%	72.76%	78.96%	6.20%	0.71%
PGP 4	77.68%	81.84%	4.16%	72.15%	76.15%	4.00%	0.16%
PGP 5	66.76%	73.21%	6.45%	69.37%	76.17%	6.80%	-0.35%
PGP 6	77.53%	83.70%	6.17%	75.36%	78.33%	2.97%	3.20%**
PGP 7	82.88%	86.02%	3.14%	75.65%	80.62%	4.97%	-1.83%
PGP 8	69.58%	77.14%	7.56%	72.15%	75.51%	3.36%	4.20%**
PGP 9	74.89%	81.48%	6.59%	69.90%	77.69%	7.79%	-1.20%
PGP 10	66.14%	72.17%	6.03%	72.32%	74.75%	2.43%	3.60%**
Average	73.55%	78.50%	4.95%	71.39%	75.92%	4.53%	0.42%

** Statistically significant at p<0.01

SOURCE: RTI analysis of Medicare claims and enrollment data.

Table 8-8

Differences in PC5: Breast cancer screening quality performance score between base year and performance year five

PGP	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Assigned Beneficiaries: BY	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Assigned Beneficiaries: PY5	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Assigned Beneficiaries: Difference (PY5-BY)	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Comparison Group Beneficiaries: BY	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Comparison Group Beneficiaries: PY5	Percentage Point Change in PC5: Breast Cancer Screening Quality Score Comparison Group Beneficiaries: Difference (PY5-BY)	Difference in Difference (AB minus CG)
PGP 1	80.74%	77.75%	-2.99%	67.81%	66.82%	-0.99%	-2.00%
PGP 2	75.32%	75.89%	0.57%	74.16%	79.30%	5.14%	-4.57%**
PGP 3	81.02%	81.14%	0.12%	69.69%	71.23%	1.54%	-1.42%
PGP 4	62.27%	65.44%	3.17%	65.97%	65.81%	-0.16%	3.33%
PGP 5	77.80%	90.08%	12.28%	67.24%	69.91%	2.67%	9.61%***
PGP 6	80.46%	83.09%	2.63%	75.89%	75.88%	-0.01%	2.64%**
PGP 7	77.74%	79.48%	1.74%	77.82%	77.01%	-0.81%	2.56%
PGP 8	81.74%	82.12%	0.38%	72.50%	72.65%	0.15%	0.23%
PGP 9	68.15%	71.65%	3.50%	61.87%	66.05%	4.18%	-0.68%
PGP 10	75.17%	75.46%	0.29%	72.48%	69.87%	-2.61%	2.90%*
Average**	76.04%	78.21%	2.17%	70.54%	71.45%	0.91%	1.26%**

** Statistically significant at p<0.01

*** Statistically significant at p<0.001

SOURCE: RTI analysis of Medicare claims and enrollment data.

8.2.3 Impact of Demonstration on Quality

This chapter compared quality performance results of the PGPs against their respective comparison groups' results from the base year to PY5 in order to demonstrate the impact of the PGP Demonstration on quality of care. Our descriptive and statistical analysis showed that there were notable differences over time for individual PGPs relative to their own comparison groups, both in the magnitude and direction of these differences. The PGPs as a group had higher performance scores by PY5, had larger and more positive changes in performance scores, and met more quality performance targets compared to their comparison groups in four of the seven measures. Given these findings, we believe the observed differences (i.e., larger improvements by the PGPs) were beyond random chance, and that the Demonstration had a positive effect on the quality of care delivered by the participating PGPs.

This page intentionally left blank.

CHAPTER 9 UTILIZATION ANALYSES

9.1 Introduction

A primary goal of the Physician Group Practice (PGP) Demonstration was to improve the efficiency of Medicare-provided services. Participating PGPs achieve savings by having lower expenditure growth rates per Medicare beneficiary than their comparison group. There are two basic ways to lower expenditure growth rates. One way is through lower prices or lower price increases while the other is through reduced utilization or utilization increases.

In the PGP Demonstration, providers continue to be paid standard Medicare fee-for-service (FFS) payment rates. Savings, then, cannot be achieved through lower prices or lower price increases. On the other hand, savings can be achieved by shifting the modality of care from more expensive places of service to less expensive places of service. Since the demonstration savings methodology does not adjust for site of service payment differentials or some hospital specific factors (IME and DSH), PGP physicians can admit their patients to community hospitals instead of teaching hospitals, and thereby realize savings.

Savings can also be achieved by reducing per assigned beneficiary service utilization or utilization growth, particularly for expensive treatments. Given how expensive they are, PGPs have strong financial incentives to reduce hospitalizations for conditions for which viable alternative treatments are available. For example, there is strong clinical evidence (Boden, et al., 2007) that most patients with stable coronary artery disease have similar outcomes when treated with medical, secondary prevention therapy compared to more costly percutaneous revascularization procedures or coronary bypass surgery (CABG).

High hospital readmission rates may indicate low quality of care. There are several channels through which low quality of care can lead to high readmission rates. One is through premature discharge. Others are through lack of discharge instructions and planning, insufficient post-discharge follow-up care, inadequate care processes (e.g., non-sterile conditions leading to iatrogenic conditions) during a hospitalization, and postponed and cancelled admissions that require more frequent subsequent hospitalizations. Aside from quality issues, high readmission rates also contribute to the high cost of medical care (Jencks, et al., 2009). Reducing readmissions is one way of lowering overall admissions.

Reduction of readmissions can be achieved through several methods. One way is to improve the quality of care rendered during initial hospitalizations. Another way is to improve discharge planning, including increased length of stay. And yet another way is to discharge patients to more appropriate care settings (e.g., skilled nursing facilities or home health agencies), when available, instead of directly to home. Participating PGPs might have been able to reduce readmission rates for assigned beneficiaries because they may have been more able to easily identify and intervene with these patients. Several of the participating PGPs had “transition care” or “discharge planning” initiatives that focused on the post-hospitalization period.

Another way participating PGPs might have been able to reduce the growth of expenditures is by reducing patients’ reliance on emergency departments (ED) and shifting

greater shares of care to primary care providers in physician offices and other non-institutional settings. This substitution in the place of service can generate savings because Medicare payment rates for EDs are typically higher than for equivalent services provided in physician offices.

Another long-standing concern is the apparent ED overuse. High ED use has been associated with higher patient costs beyond the simple differences between EDs and physician offices in the price per visit. In particular, it is widely believed that EDs are over-utilized by patients who have poor access to primary care and other lower-cost providers. It is also widely believed that use of EDs instead of primary care providers results in poorer patient health because patients who use EDs have done so because they delayed obtaining timely care.

In the next section we describe the utilization measures used for the analyses. In Section 9.3, we describe the number of discharges and ED visits per beneficiary during the base year (BY) and Performance Year 5 (PY5) and the readmission rates for both time periods. Time trends are the subject of Section 9.4. We then test the hypotheses that discharge, ED visit, and readmission rates for assigned patients dropped relative to those for comparison patients during the Demonstration period (Section 9.5). Since the comparison group is comprised of all the residual FFS beneficiaries in an area with at least one E&M visit (PGP assigned beneficiaries are excluded), comparisons between the PGPs and these populations do not constitute a rigorous evaluation model as in Chapter 10. Therefore, any interpretations of such comparisons are limited.

The results are summarized in a final section. We considered risk adjusting the utilization measures presented in this chapter, but did not do so. The literature does not show a satisfactory methodology of risk adjusting utilization. Developing such a method was beyond the scope of this project.

9.2 Utilization Measures

In this section we briefly define the utilization measures analyzed in this chapter, including hospital discharges, ED visits, and readmissions. The overall rates presented in the tables and figures later in the chapter are simple averages of the ten PGP-specific means for the participating PGPs and their respective comparison groups.

9.2.1 Hospital Discharges

All hospitalizations were counted. About 90 percent of the hospitalizations were short-term acute care (STAC) discharges paid under Medicare's Inpatient Prospective Payment System (IPPS) and discharges from critical access hospital (CAHs). Non-STAC hospitalizations were also included, from rehabilitation hospitals and units, psychiatric hospitals and units, and long-term care hospitals (LTCHs). Discharges from Skilled Nursing Facilities (SNFs) and Skilled Nursing Units (SNUs) are not included because SNFs and SNUs are not considered hospitals.

9.2.2 Emergency Department Visits

All ED visits were counted whether or not they led to a hospitalization. The standard Research Data Assistance Center (RESDAC) definition of ED visits was used (Merriman and

Caldwell, 2008). Revenue center (UB) codes 0450 through 0459 and 0981 on inpatient and outpatient Standard Analytic Files (SAFs) were used to identify to identify EDs.

9.2.3 Hospital Readmissions

The basic observational unit (INDEX) for the analysis was a live IPPS discharge or a live discharge from a CAH. This criterion excludes discharges from special units (e.g., certified psychiatric and rehabilitation units) within general STAC hospitals and other discharges that do not qualify for IPPS payments. The INDEX discharge date had to be within the relevant base or performance year. For a subsequent hospitalization to qualify as a readmission, it had to be an IPPS (acute) admission at either a general STAC hospital or a CAH. The readmission did not have to occur at the same hospital as the INDEX discharge. Readmissions could be for any reason, they were not restricted by diagnosis, procedure, or Diagnosis Related Group (DRG). Readmissions were not required to be clinically “related” to the index discharge, i.e., they are for any cause. A readmission itself becomes a new INDEX discharge if its discharge date is within the relevant year. Note that truncation was not an issue because if INDEX occurred in the last month of a period, we had access to Medicare claims data during the relevant post-period and thus were able to measure readmissions that occurred after the period was over.

In addition to the criteria set forth above, there are many ways to measure readmissions. Commonly used measures include 3 days, 14 days, 30 days, 60 days, and 90 days following discharge. The shorter periods are more likely to capture readmissions that are related to the original (index) admission, but may be only short-run effects. We use only the standard 30-day readmission rate for the analyses here to attempt to identify longer-run, more significant effects.

9.3 Descriptive Analyses

9.3.1 Hospital Discharges

The number of hospitalizations (discharges) per beneficiary for both the assigned beneficiaries and the comparison groups fell between the base year and PY5, which is consistent with national trends (MedPAC, 2011). The share of beneficiaries with no discharges, just over 77 percent in the base year for both participating PGPs and their comparison groups, increased at least one percentage point for both groups (*Table 9-1*). The share of beneficiaries with one, two, and three discharges fell between the base year and PY5. The share of beneficiaries with four or more discharges was roughly the same in both time periods.

9.3.2 Emergency Department Visits

The number of emergency department (ED) visits per beneficiary increased for the assigned beneficiaries and decreased for the comparison group beneficiaries between the base year and PY5. The share of beneficiaries with no ED visits, just over 68 percent in the base year, decreased more than two percentage points for assigned beneficiaries and more than one percentage point for comparison group beneficiaries (*Table 9-2*). The share of assigned beneficiaries with one, two, and three ED visits increased between the base year and PY5. The share of comparison group beneficiaries with one visit fell while those with two or three ED visits increased in both time periods. The share of assigned beneficiaries with four or more ED visits increased by about 1 percentage point between the base year and PY5. The share of

comparison group beneficiaries with four or more ED visits increased by half a percentage point over the same time period. As will be shown in Section 9.4, ED visits did not increase in each year between the base year and PY5.

Table 9-1
Distribution of beneficiaries by number of hospital discharges, BY and PY5

Number of Discharges	Assigned Beneficiaries:	Assigned Beneficiaries:	Assigned Beneficiaries:	Assigned Beneficiaries:	Comparison Group: Base Year Frequency	Comparison Group: Base Year Percent	Comparison Group: PY5 Frequency	Comparison Group: PY5 Percent
	Base Year Frequency	Base Year Percent	PY5 Frequency	PY5 Percent				
0	172,063	77.1%	173,762	78.5%	1,250,318	77.3%	1,085,272	78.4%
1	31,677	14.2%	30,244	13.7%	222,322	13.7%	186,468	13.5%
2	11,301	5.1%	9,808	4.4%	82,123	5.1%	62,476	4.5%
3	4,331	1.9%	3,896	1.8%	32,571	2.0%	25,590	1.8%
4	1,933	0.9%	1,774	0.8%	15,194	0.9%	11,785	0.9%
5	901	0.4%	846	0.4%	7,201	0.4%	5,717	0.4%
6+	997	0.4%	896	0.4%	8,506	0.5%	7,042	0.5%

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Table 9-2
Distribution of beneficiaries by number of emergency department visits, BY and PY5

Number of Visits	Assigned Beneficiaries:	Assigned Beneficiaries:	Assigned Beneficiaries:	Assigned Beneficiaries:	Comparison Group: Base Year Frequency	Comparison Group: Base Year Percent	Comparison Group: PY5 Frequency	Comparison Group: PY5 Percent
	Base Year Frequency	Base Year Percent	PY5 Frequency	PY5 Percent				
0	151,999	68.1%	146,022	66.0%	1,103,918	68.2%	929,103	67.1%
1	41,209	18.5%	41,363	18.7%	299,410	18.5%	254,762	18.4%
2	15,478	6.9%	16,328	7.4%	110,808	6.8%	98,082	7.1%
3	6,748	3.0%	7,657	3.5%	48,272	3.0%	45,061	3.3%
4	3,240	1.5%	3,769	1.7%	23,258	1.4%	22,925	1.7%
5	1,715	0.8%	2,197	1.0%	12,362	0.8%	12,448	0.9%
6+	2,813	1.3%	3,888	1.8%	20,205	1.2%	21,965	1.6%

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.3.3 Hospital Readmissions

Across all PGPs, the 30-day hospital readmission rate for assigned beneficiaries was 16.5 percent during the base year and increased slightly to 16.9 percent during PY5 (*Table 9-3*). As will be seen in Section 9.4.3, the readmission rates for the assigned beneficiaries did change from year to year, but not by very much. For the overall comparison group, the base year and PY5 readmission rates were both 16.0 percent. The overall rates for the comparison group were not completely constant between the base year and PY5. The assigned beneficiary readmission rate

rose for 6 PGPs, fell for 2 PGPs, and was the same for 2 PGPs. The comparison group readmission rate rose for 3 PGPs, fell for 5 PGPs, and was the same for 2 PGPs.

Table 9-3
Readmission rates, BY and PY5

PGP	Assigned Beneficiaries Base Year	Assigned Beneficiaries PY5	Comparison Group Base Year	Comparison Group PY5
Overall	16.5%	16.9%	16.0%	16.0%
PGP 1	15.5%	15.2%	13.0%	13.0%
PGP 2	16.3%	17.8%	13.4%	13.2%
PGP 3	15.2%	15.3%	16.0%	15.6%
PGP 4	16.0%	15.7%	18.2%	18.0%
PGP 5	17.0%	17.0%	16.3%	15.8%
PGP 6	16.3%	16.4%	14.9%	14.7%
PGP 7	13.8%	15.8%	17.1%	17.1%
PGP 8	16.2%	16.7%	18.0%	18.5%
PGP 9	15.0%	15.0%	14.8%	15.9%
PGP 10	23.2%	24.2%	17.9%	18.7%

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.4 Trends in Assigned Beneficiaries versus Comparison Groups

9.4.1 Hospital Discharges

Table 9-4 shows hospital discharge rates per 10,000 beneficiaries for both the assigned beneficiaries and the comparison group beneficiaries for the individual PGP sites and combined across all ten PGPs. The overall discharge rate for the assigned beneficiaries is higher than for the comparison group beneficiaries for the base year and all five performance years (*Figure 9-1*). The overall rates for both the assigned beneficiaries and the comparison group increased from the base year through PY 2. Rates flattened starting in PY3 and fell through PY5.

Hospital discharges per 10,000 beneficiaries are greater for the assigned beneficiaries for six of the PGPs relative to their comparison groups in the base year through PY5 (*Figure 9-2*). PGPs 4 and 8 have lower discharge rates than their comparison groups for all time periods. PGP 5 has lower rates in the base year, but higher rates in each performance year of the Demonstration. PGP 7 has lower rates in the base year, roughly the same rates in PY1, and higher rates each year thereafter. Whether or not these differences in assigned beneficiary versus comparison group changes are statistically significant is the subject of Section 9.5.

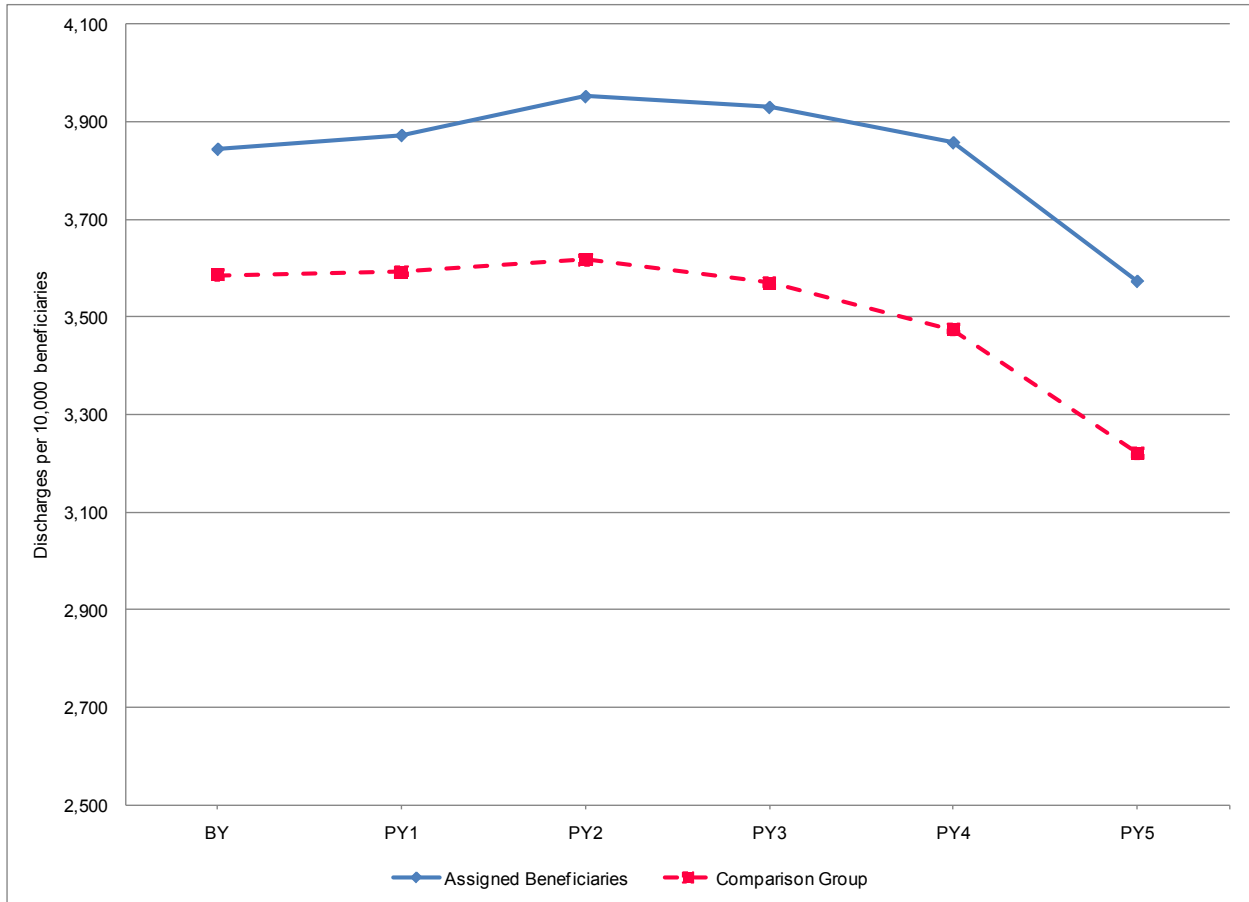
Discharge rates per 10,000 beneficiaries fell for both assigned and comparison beneficiaries after PY3. Possible reasons for this could include a national trend towards lower hospitalizations or a reduction in readmission rates which is the subject of Section 9.4.3.

Table 9-4
Hospital discharge rates per 10,000 beneficiaries by PGP and year

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5
Overall	3,844	3,872	3,951	3,929	3,857	3,573	3,586	3,592	3,617	3,570	3,474	3,221
PGP 1	3,896	3,979	4,001	3,980	3,834	3,577	3,444	3,597	3,577	3,388	3,288	2,946
PGP 2	3,660	3,677	3,790	3,777	3,820	3,418	2,737	2,701	2,732	2,728	2,649	2,354
PGP 3	3,180	3,088	3,131	3,128	3,265	2,883	3,126	3,014	2,938	2,926	2,894	2,674
PGP 4	3,770	3,837	3,934	3,618	3,461	3,315	4,112	4,333	4,243	4,144	3,987	3,608
PGP 5	4,115	4,293	4,272	4,218	4,263	3,885	4,180	4,110	4,123	4,038	3,827	3,575
PGP 6	3,876	3,859	3,977	3,917	3,892	3,422	3,341	3,407	3,448	3,367	3,277	2,950
PGP 7	3,089	3,203	3,523	3,763	3,706	3,448	3,300	3,196	3,393	3,446	3,380	3,323
PGP 8	3,750	3,707	3,629	3,674	3,599	3,273	3,964	3,868	4,019	4,038	3,942	3,701
PGP 9	3,871	3,894	3,752	3,693	3,496	3,376	3,521	3,553	3,451	3,377	3,305	3,111
PGP 10	5,229	5,180	5,503	5,524	5,234	5,135	4,133	4,138	4,243	4,246	4,189	3,967

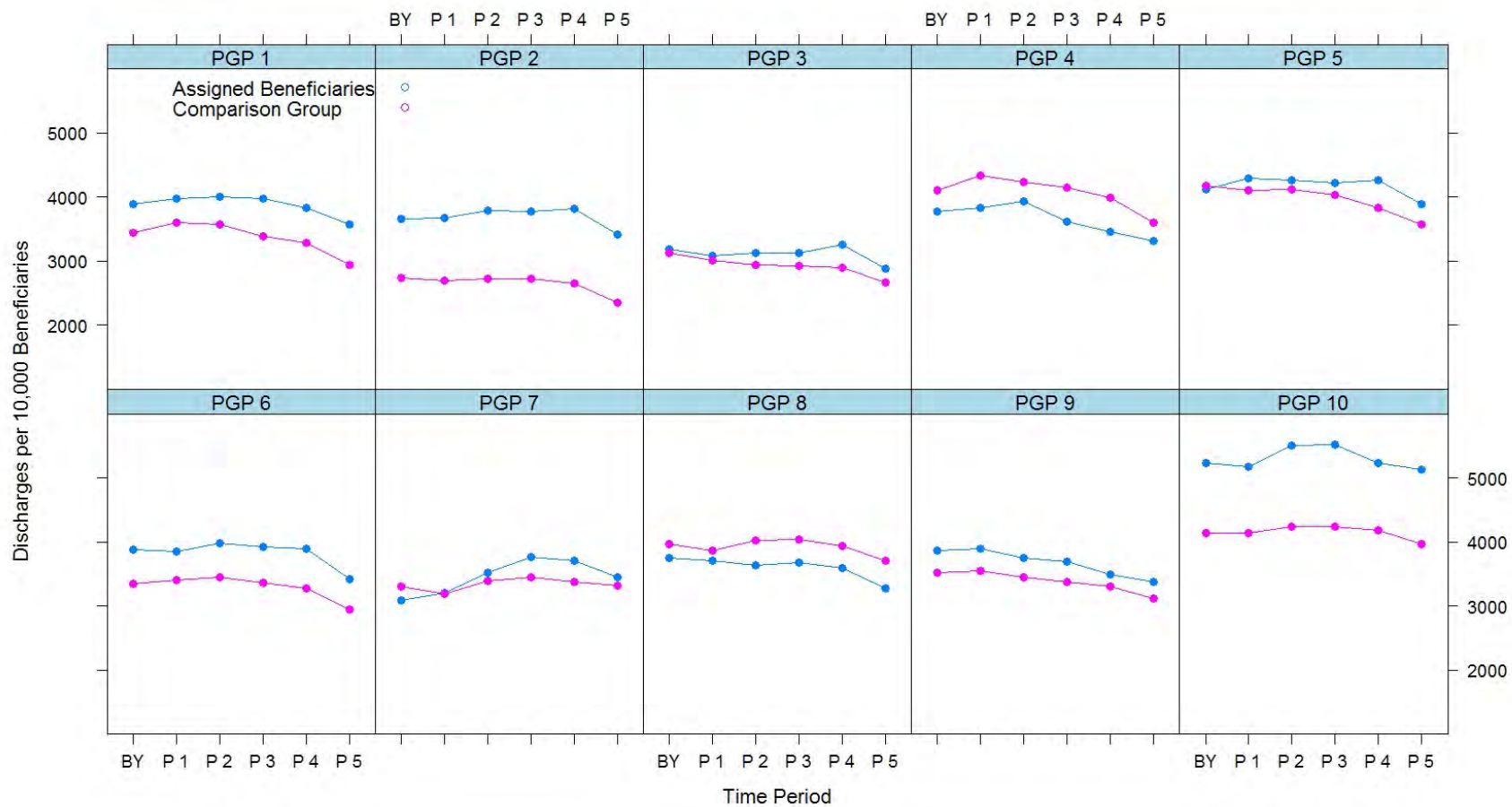
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-1
Hospital discharge rates per 10,000 beneficiaries over all 10 PGPs by year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-2
Hospital discharge rates per 10,000 beneficiaries by PGP and year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.4.2 Emergency Department Visits

Table 9-5 shows emergency department visits per 10,000 beneficiaries for both the assigned beneficiaries and the comparison group over all ten PGPs combined and for each of the PGPs from the base year through PY5. The overall rates for the assigned beneficiaries are higher than for the comparison group for the base year and all five performance years (*Figure 9-3*). The overall rates for both the assigned beneficiaries and the comparison group increased from the base year through PY3 with a flattening of the curve starting in PY4 and falling rates in the last performance year.

Emergency department visits per 10,000 beneficiaries were greater for the assigned beneficiaries for six of the PGPs than their comparison groups in all time periods (*Figure 9-4*). PGPs 4 and 8 had lower rates in all time periods. PGP 3 had lower rates in the base year and PY1, but higher rates in each performance year thereafter. And PGP 9 had lower rates in the base year though PY3, but higher thereafter. Whether or not the difference in changes is statistically significant is the subject of Section 9.5.

For most of the PGPs and their comparison groups, emergency department visits per 10,000 beneficiaries increased between the base year and PY4 and fell between PY4 and PY5. The exception is PGP 7, where visits for both assigned beneficiaries and the comparison group increased between PY4 and PY5.

9.4.3 All Cause Hospital Readmissions

Table 9-6 shows hospital readmission rates for both the assigned beneficiaries and the comparison group beneficiaries for the individual PGP sites and combined across all ten PGPs in the base year and PY5. Between the base year and PY5 the overall readmission rate for assigned beneficiaries increased from 16.5 to 16.9 percent. There was a small dip between the base year and PY1 for the assigned beneficiaries and the readmission rate peaked at 17.3 percent in PY4 (*Figure 9-5*). The overall readmission rate for the comparison group began in the base year at 16.0 percent and ended in PY5 at the same rate, 16.0 percent. There was a slight dip between the base year and PY1 for the comparison group and it peaked at 16.6 percent in PY3.

Readmission rates for five PGPs were always higher than the comparison groups (*Figure 9-6*). Readmission rates for three PGPs were always lower than the comparison groups.

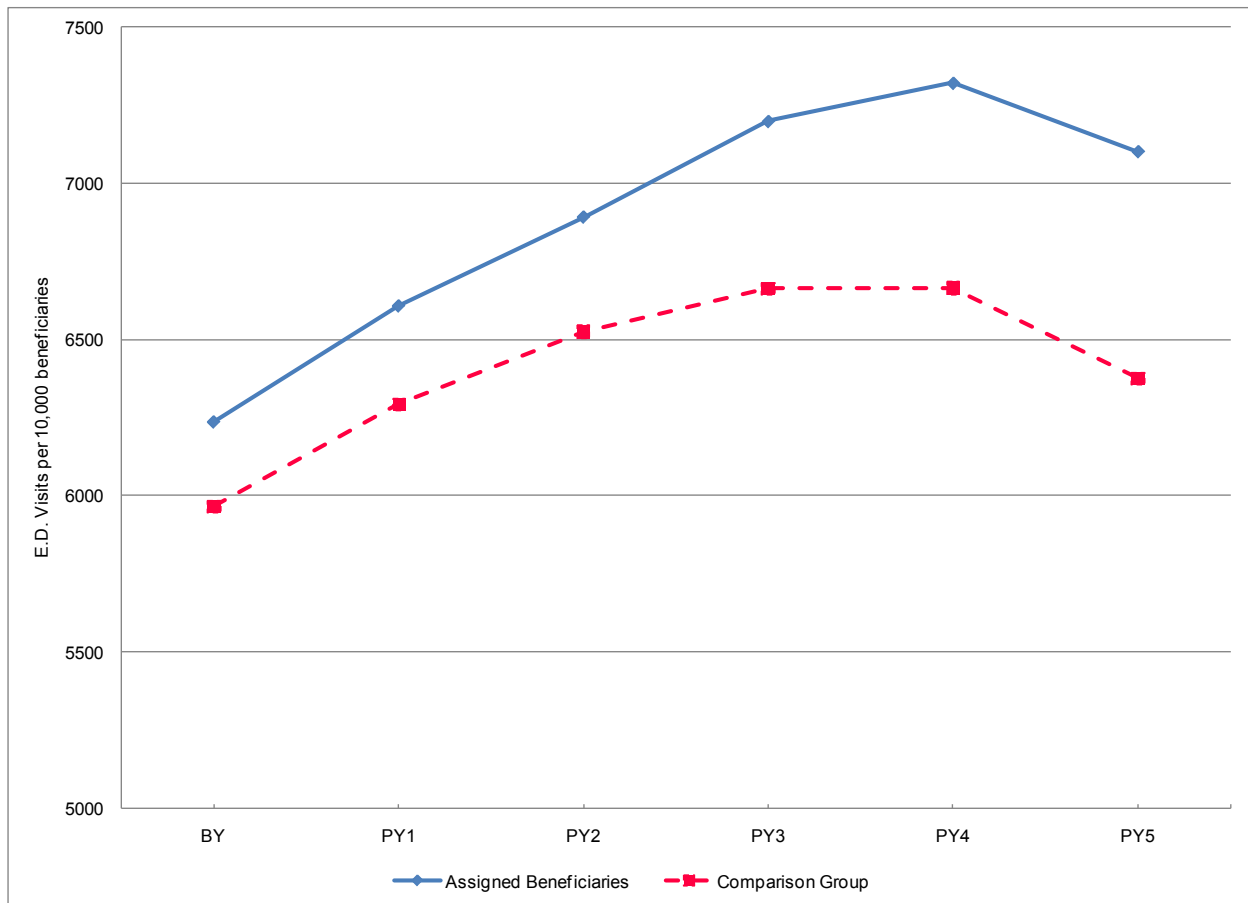
The readmission rates for six PGPs increased from the base year to PY5. PGP 7 had the largest increase in hospital readmission rates for its assigned beneficiaries, peaking in PY4, but still increasing 2 percentage points between the base year and PY5. Readmission rates for the assigned beneficiaries decreased for PGP 1 and PGP 4, and remained constant for PGP 5 and PGP 9. The readmission rates for the comparison groups exhibited somewhat less annual variation than for the assigned beneficiaries.

Table 9-5
Emergency department visits per 10,000 beneficiaries by PGP and year

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5
Overall	6,236	6,608	6,892	7,198	7,320	7,101	5,965	6,292	6,525	6,663	6,664	6,375
PGP 1	5,743	6,232	6,220	6,478	6,442	6,322	4,917	5,471	5,681	5,551	5,531	5,297
PGP 2	7,263	7,805	7,828	8,159	8,480	8,323	5,829	6,367	6,332	6,568	6,505	6,110
PGP 3	5,400	5,641	6,055	6,413	6,665	6,317	6,133	6,010	5,919	6,100	6,123	5,872
PGP 4	6,192	6,622	7,162	7,163	7,165	6,880	6,708	7,337	7,714	7,880	7,896	7,339
PGP 5	6,955	7,229	7,353	7,719	7,997	7,847	6,016	6,042	6,281	6,379	6,356	6,075
PGP 6	5,381	5,827	6,284	6,608	6,834	6,641	5,083	5,657	5,871	6,084	6,142	5,797
PGP 7	6,604	6,936	7,239	7,247	7,464	7,608	5,820	6,043	6,419	6,301	6,431	6,477
PGP 8	4,934	5,259	5,351	5,745	5,831	5,470	5,867	6,054	6,484	6,744	6,883	6,771
PGP 9	6,263	6,455	6,857	7,333	7,473	6,989	6,852	7,218	7,470	7,606	7,449	6,983
PGP 10	7,628	8,071	8,573	9,115	8,851	8,617	6,430	6,720	7,075	7,416	7,323	7,033

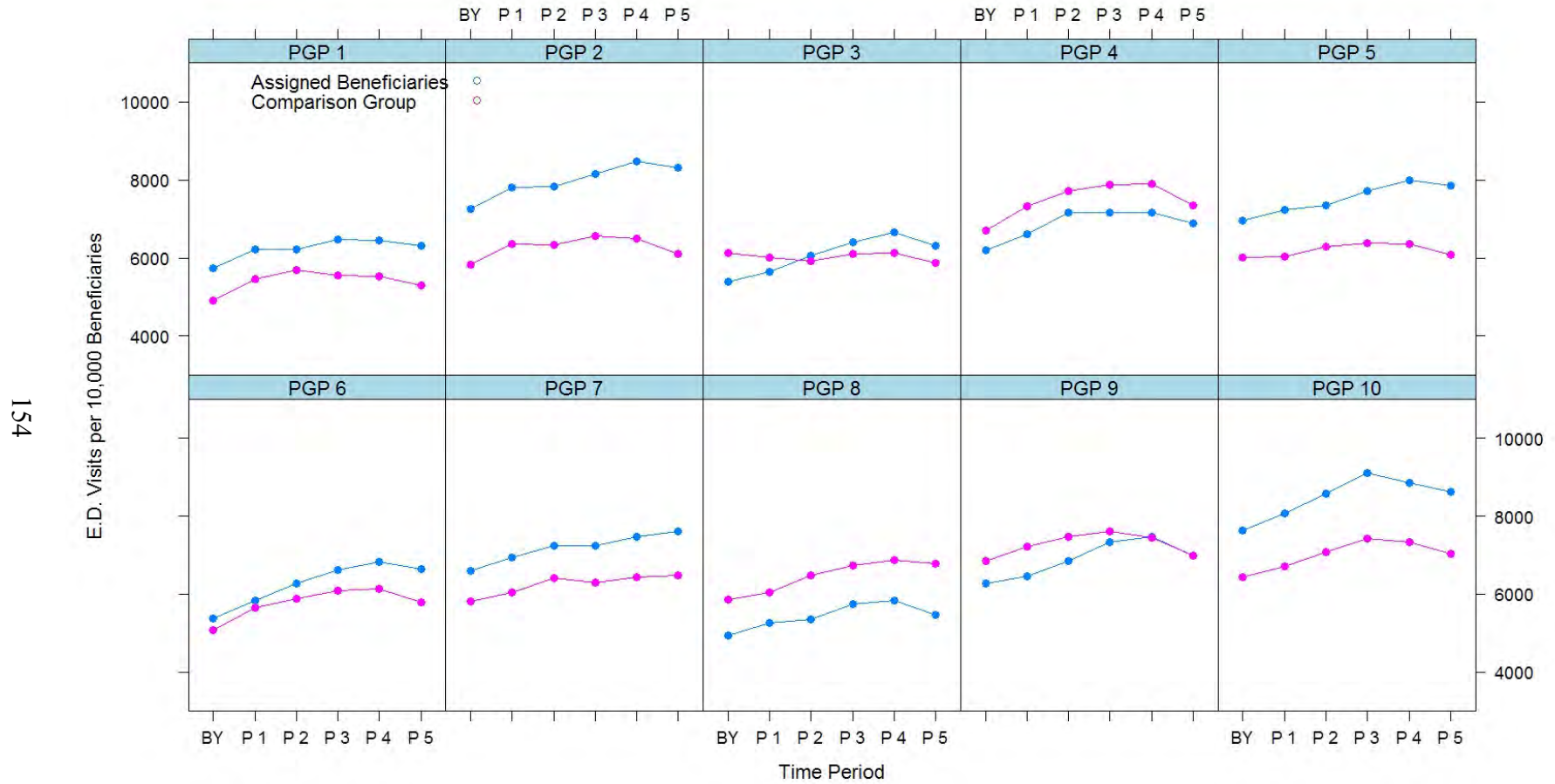
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-3
Emergency department visits per 10,000 beneficiaries over all 10 PGPs by year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-4
Emergency department visits per 10,000 beneficiaries by PGP and year



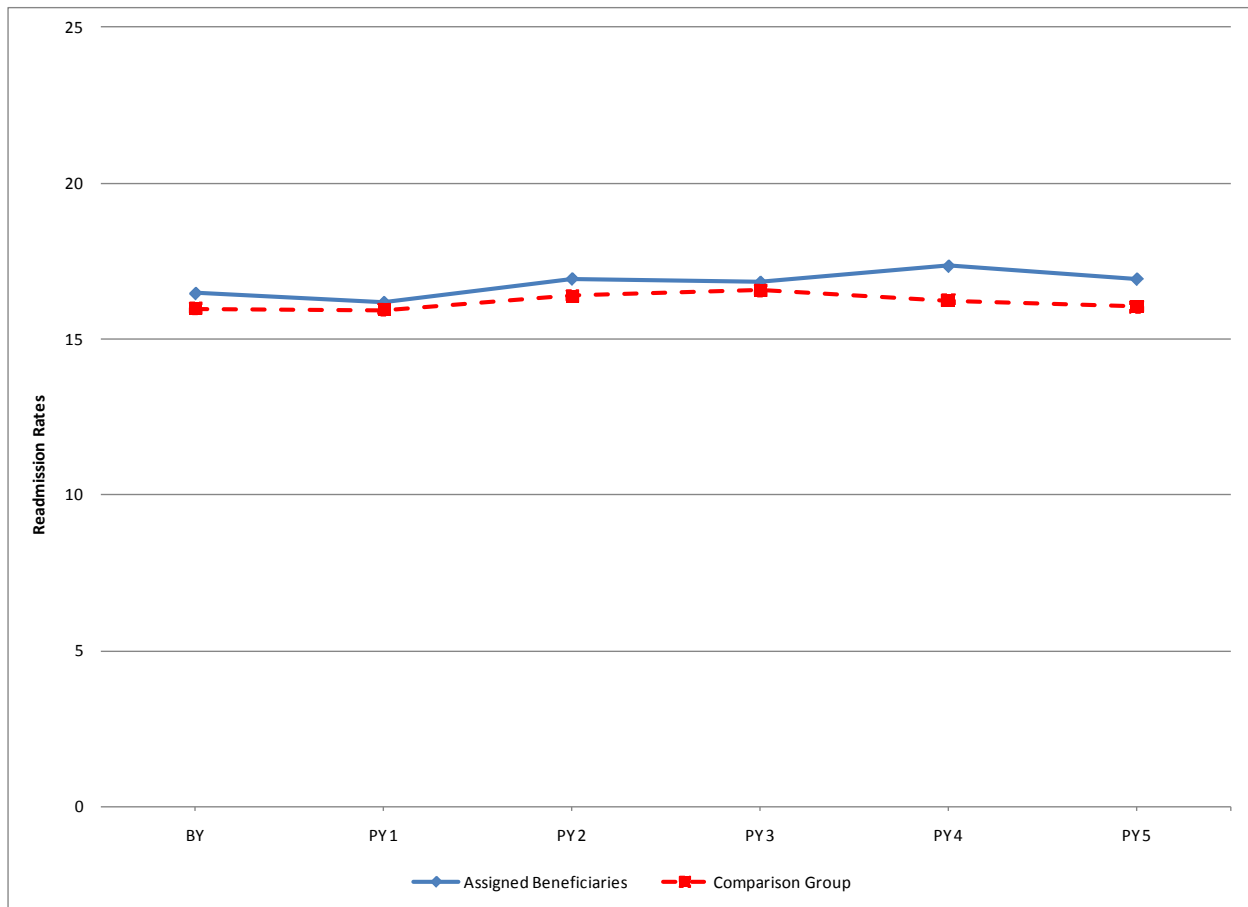
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Table 9-6
Readmission rates by PGP and year

PGP	Assigned Beneficiaries BY	Assigned Beneficiaries PY1	Assigned Beneficiaries PY2	Assigned Beneficiaries PY3	Assigned Beneficiaries PY4	Assigned Beneficiaries PY5	Comparison Group BY	Comparison Group PY1	Comparison Group PY2	Comparison Group PY3	Comparison Group PY4	Comparison Group PY5
Overall	16.5%	16.2%	16.9%	16.8%	17.3%	16.9%	16.0%	15.9%	16.4%	16.6%	16.2%	16.0%
PGP 1	15.5%	15.8%	16.7%	15.7%	16.5%	15.2%	13.0%	12.7%	13.7%	14.5%	13.5%	13.0%
PGP 2	16.3%	17.1%	16.9%	18.2%	17.9%	17.8%	13.4%	13.0%	15.0%	15.1%	13.2%	13.2%
PGP 3	15.2%	13.0%	14.5%	14.7%	15.9%	15.3%	16.0%	15.1%	15.4%	15.4%	15.8%	15.6%
PGP 4	16.0%	15.9%	16.4%	15.0%	15.7%	15.7%	18.2%	18.6%	18.4%	18.9%	19.2%	18.0%
PGP 5	17.0%	17.0%	17.7%	17.0%	17.3%	17.0%	16.3%	15.5%	16.3%	16.3%	15.8%	15.8%
PGP 6	16.3%	15.6%	16.5%	16.7%	16.8%	16.4%	14.9%	15.0%	15.3%	15.3%	14.7%	14.7%
PGP 7	13.8%	14.4%	15.3%	16.4%	17.0%	15.8%	17.1%	17.5%	18.0%	17.9%	17.9%	17.1%
PGP 8	16.2%	16.5%	16.9%	16.2%	18.0%	16.7%	18.0%	18.2%	18.9%	18.7%	19.1%	18.5%
PGP 9	15.0%	14.0%	14.5%	14.6%	15.4%	15.0%	14.8%	15.4%	14.5%	14.9%	14.9%	15.9%
PGP 10	23.2%	22.4%	23.9%	23.6%	22.9%	24.2%	17.9%	18.3%	18.4%	18.6%	18.3%	18.7%

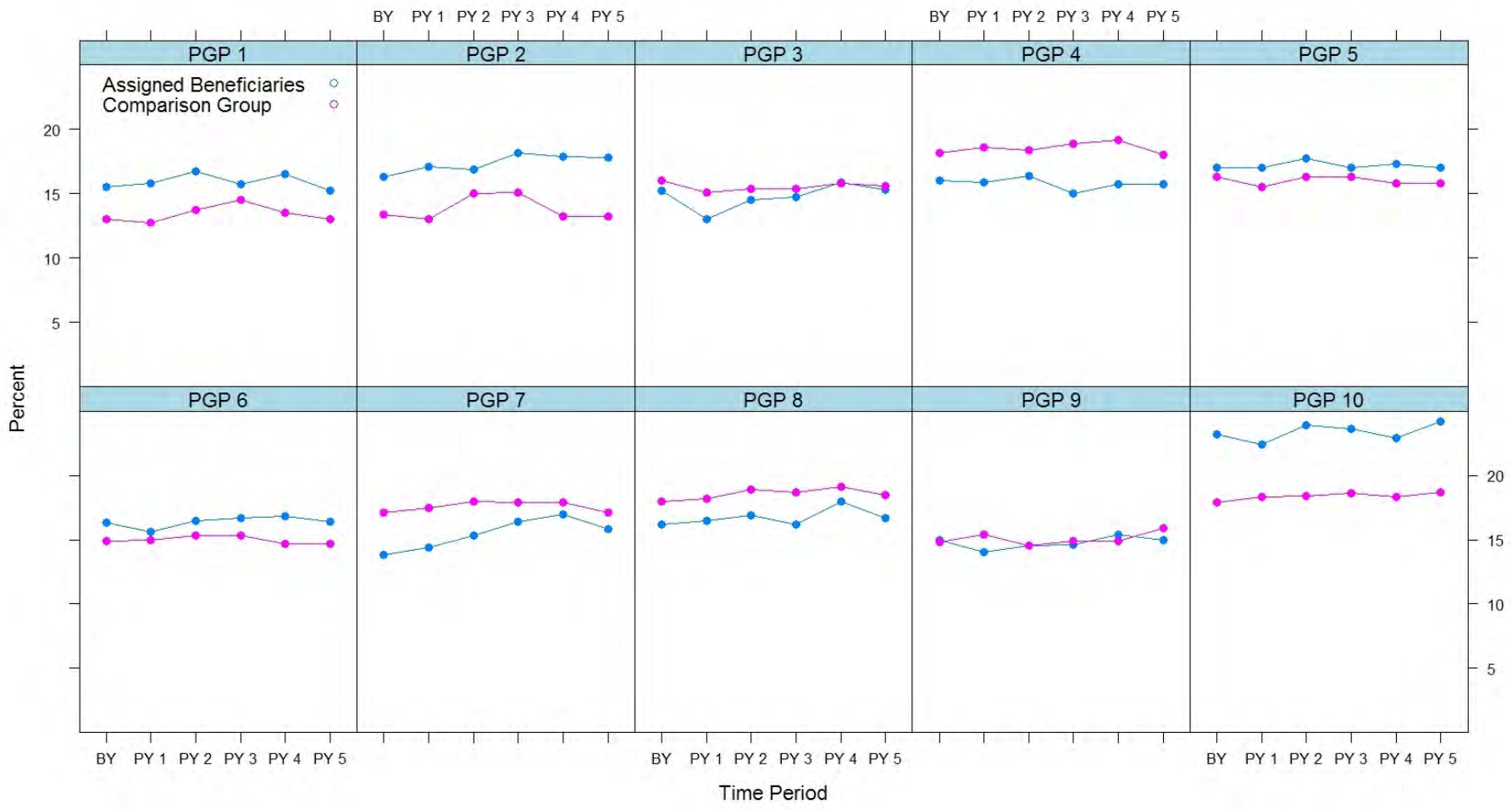
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-5
Readmission rates over all 10 PGPs by year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-6
Readmission rates by PGP and year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.5 Statistical Estimates of Demonstration Impacts

In this section we calculate the difference-in-differences in utilization between the base year and each performance year for each participating PGP and its comparison group. The difference-in-differences were calculated in two steps. The first step was to compute the change between the performance year and the base year for each participating PGP (ΔAB) and its comparison group (CG). The change for the participating PGP (ΔAB) was then subtracted from the change in its comparison group (ΔCG):

$$\text{difference-in-differences} = \Delta\text{CG} - \Delta\text{AB}$$

Values greater than zero indicate the participating PGPs performed better than their comparison groups. Conversely, negative values indicate the participating PGPs performed worse than their comparison groups. Statistical testing was performed on all differences-in-differences.

The basic method for performing statistical tests on difference-in-differences for hospital admission rates, ED visits, and hospital readmission rates was the same as used for expenditures. This is the case even though units of these three measures differ that for expenditures. That is, for instance, hospital admissions (ED visits) are measured as the number of admissions (ED visits) per 10,000 beneficiaries. These values are means just like any other mean. The only significant difference is for readmission rates which are proportions. The formula for calculating standard errors for proportions differs from those for other means.

9.5.1 Hospital Discharges

Table 9-7 shows the difference-in-differences of mean discharge rates per 10,000 beneficiaries between the comparison group beneficiaries and the assigned beneficiaries from the base year to each performance year (computed from *Table 9-4*). Thirty-six of the 55 values, including all five overall values, have negative signs of which 12 were statistically significant at the ten percent level or better. Of the 19 positive values, only three were statistically significant. The overall trend (*Figure 9-7*) shows the difference-in-differences growing more negative from PY1 through PY4, then less negative in PY5. These overall differences were not statistically significant from zero.

The difference-in-differences for the individual PGP sites are shown in *Figure 9-8*. Only PGP 9 had positive differences in differences for all five performance years, but only one was statistically significant (PY4). PGP 8 had positive results from PY 2 through PY5, with statistically significant differences only in PY2 and PY5. PGPs 1, 4, and 10 started the Demonstration in PY1 with positive difference-in-differences, and ended the Demonstration with negative results. Similar to the overall trend, PGPs 2, 3, 5, 6 and 7 fell to their most negative points in PY4 then improved slightly in PY5. The negative values for PGP 5 and PGP 7 are significant in all five years of the Demonstration.

Table 9-7
Hospital discharge rates: Difference in differences between the base year and each performance year

PGP	$\Delta CG - \Delta AB$ PY1	$\Delta CG - \Delta AB$ PY2	$\Delta CG - \Delta AB$ PY3	$\Delta CG - \Delta AB$ PY4	$\Delta CG - \Delta AB$ PY5
Overall	-22	-77	-102	-125	-95
PGP 1	69	28	-140	-93	-179
PGP 2	-53	-136	-126	-247 ^b	-141
PGP 3	-19	-139	-148	-316 ^b	-154
PGP 5	155	-33	184	184	-50
PGP 5	-249 ^b	-214 ^b	-245 ^b	-501 ^a	-376 ^a
PGP 6	83	6	-15	-80	63
PGP 7	-219 ^c	-341 ^a	-528 ^a	-537 ^a	-335 ^a
PGP 8	-53	176 ^c	151	129	214 ^b
PGP 9	10	50	35	160 ^c	85
PGP 10	54	-165	-183	50	-72

NOTES:

^a denotes statistical significance at the 1% level.

^b denotes statistical significance at the 5% level.

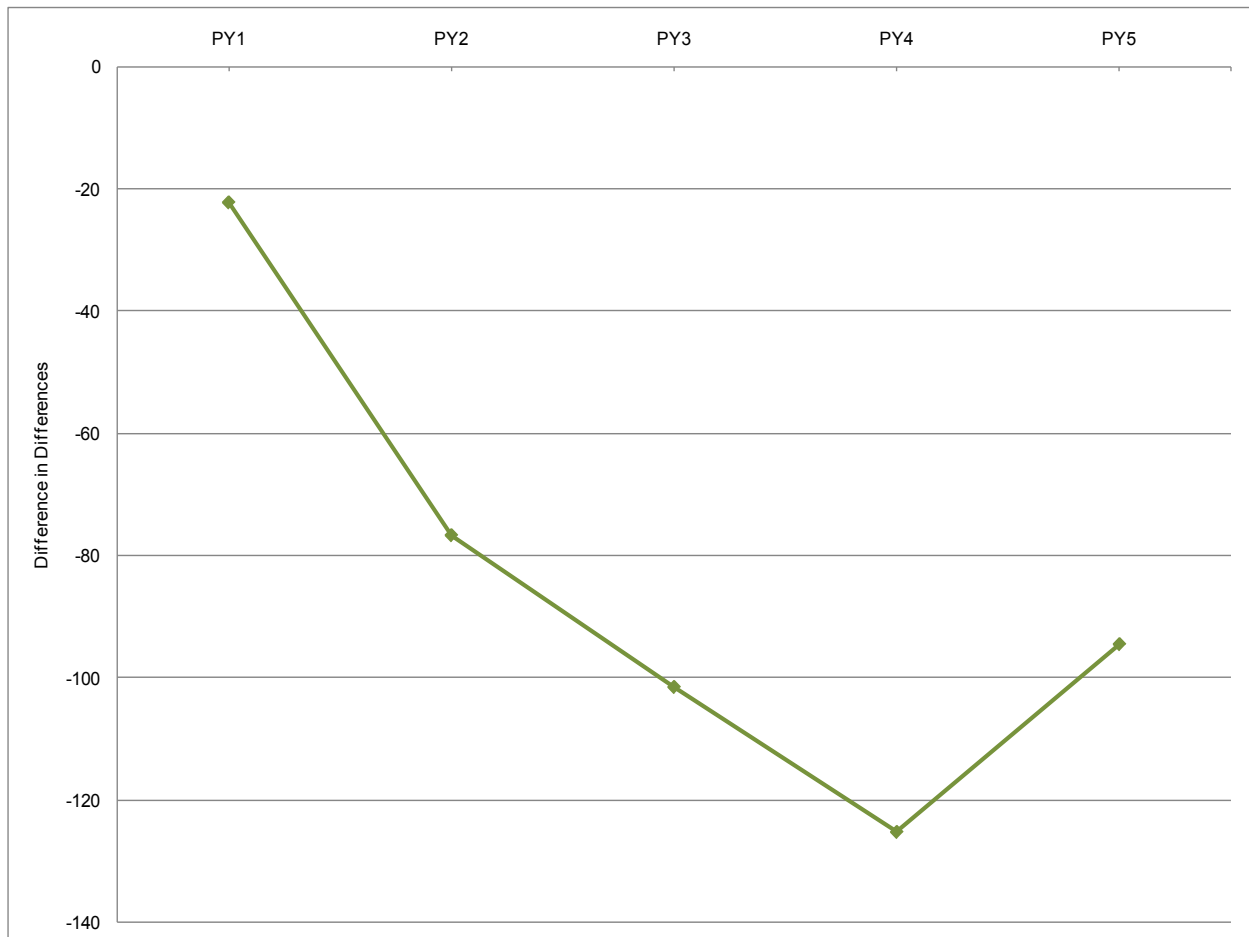
^c denotes statistical significance at the 10% level.

Positive Values = PGPs performed better than their comparison groups.

Negative Values = PGPs performed worse than their comparison groups.

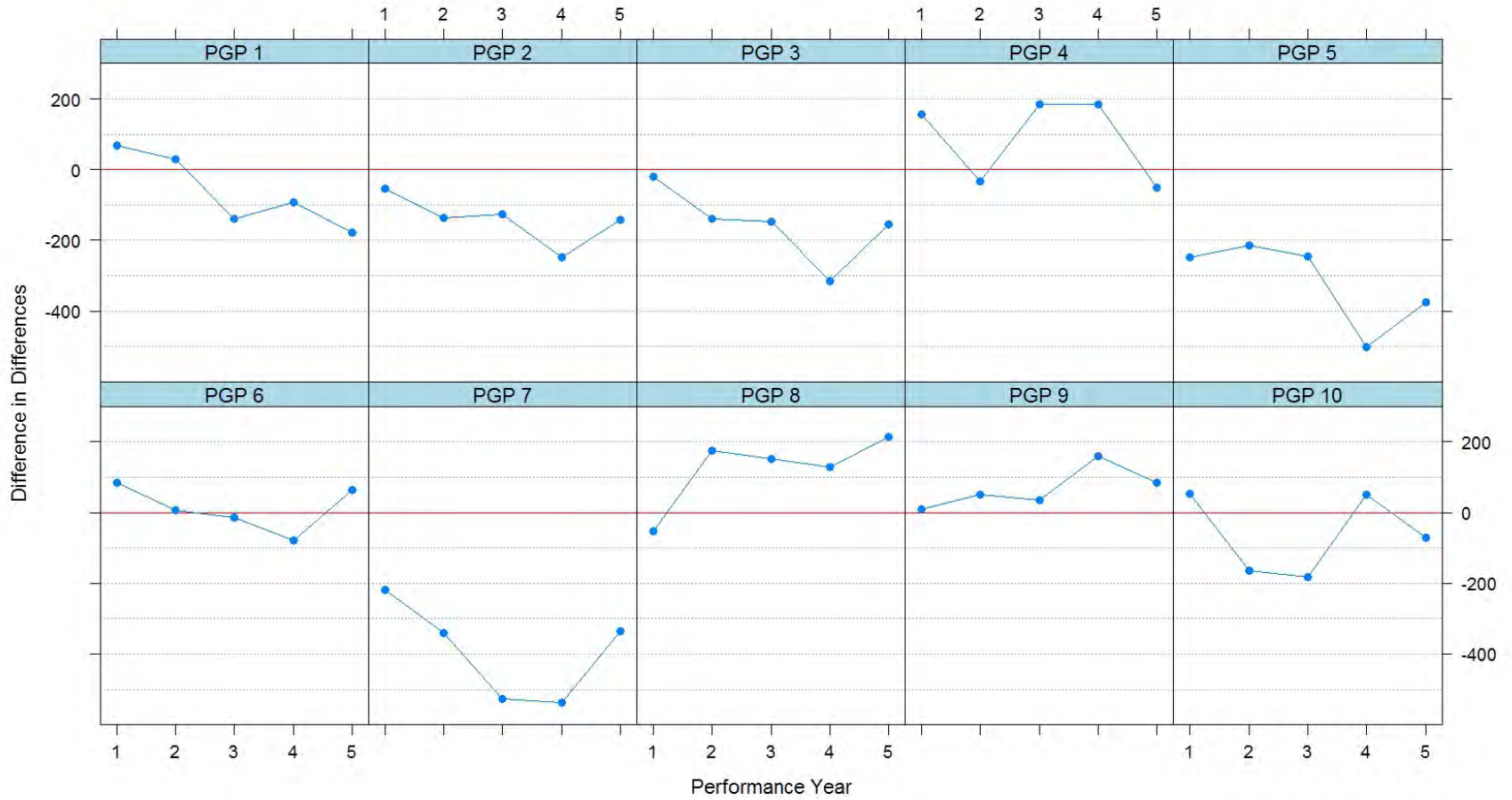
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-7
Hospital discharge rates: Overall difference in differences between the base year and each performance year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-8
Hospital discharge rates: Difference in differences between the base year and each performance year by PGP



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.5.2 Emergency Department Visits

Table 9-8 shows the differences in differences of mean ED visits per 10,000 beneficiaries between the comparison group beneficiaries and the assigned beneficiaries from the base year to each performance year (computed from *Table 9-5*). Forty-two of the 55 values, including all five overall values, have negative signs of which 22 were statistically significant. Of the 13 positive values, only one was statistically significant. The overall values became progressively more negative during the course of the Demonstration (*Figure 9-9*) with the values for PY3 through PY5 being statistically significant. With only two exceptions, PGPs 4 and 8 (*Figure 9-10*), individual PGPs generally had progressively more negative values during the course of the Demonstration. PGP 4 had positive values for all performance years except PY5. PGP 8 had a negative value in PY1 and positive values thereafter, although they did not steadily increase. And PGP 8 was the only participant that had a positive value in PY5.

Table 9-8
Emergency department visits: Difference in differences between the base year and each performance year

PGP	$\Delta CG - \Delta AB$ PY1	$\Delta CG - \Delta AB$ PY2	$\Delta CG - \Delta AB$ PY3	$\Delta CG - \Delta AB$ PY4	$\Delta CG - \Delta AB$ PY5
Overall	-45	-96	-264	-385 ^b	-455 ^b
PGP 1	66	287	-100	-85	-198
PGP 2	-4	-62	-156	-540 ^a	-779 ^a
PGP 3	-363	-868 ^a	-1,045 ^a	-1,275 ^a	-1,178 ^a
PGP 4	200	36	202	216	-56
PGP 5	-249 ^c	-133	-401 ^a	-703 ^a	-834 ^a
PGP 6	128	-115	-226 ^b	-394 ^a	-546 ^a
PGP 7	-108	-35	-162	-249	-346 ^c
PGP 8	-139	200	66	120	368 ^b
PGP 9	174	24	-317 ^b	-614 ^a	-595 ^a
PGP 10	-153	-300	-501 ^a	-330	-386 ^b

NOTES:

^a denotes statistical significance at the 1% level.

^b denotes statistical significance at the 5% level.

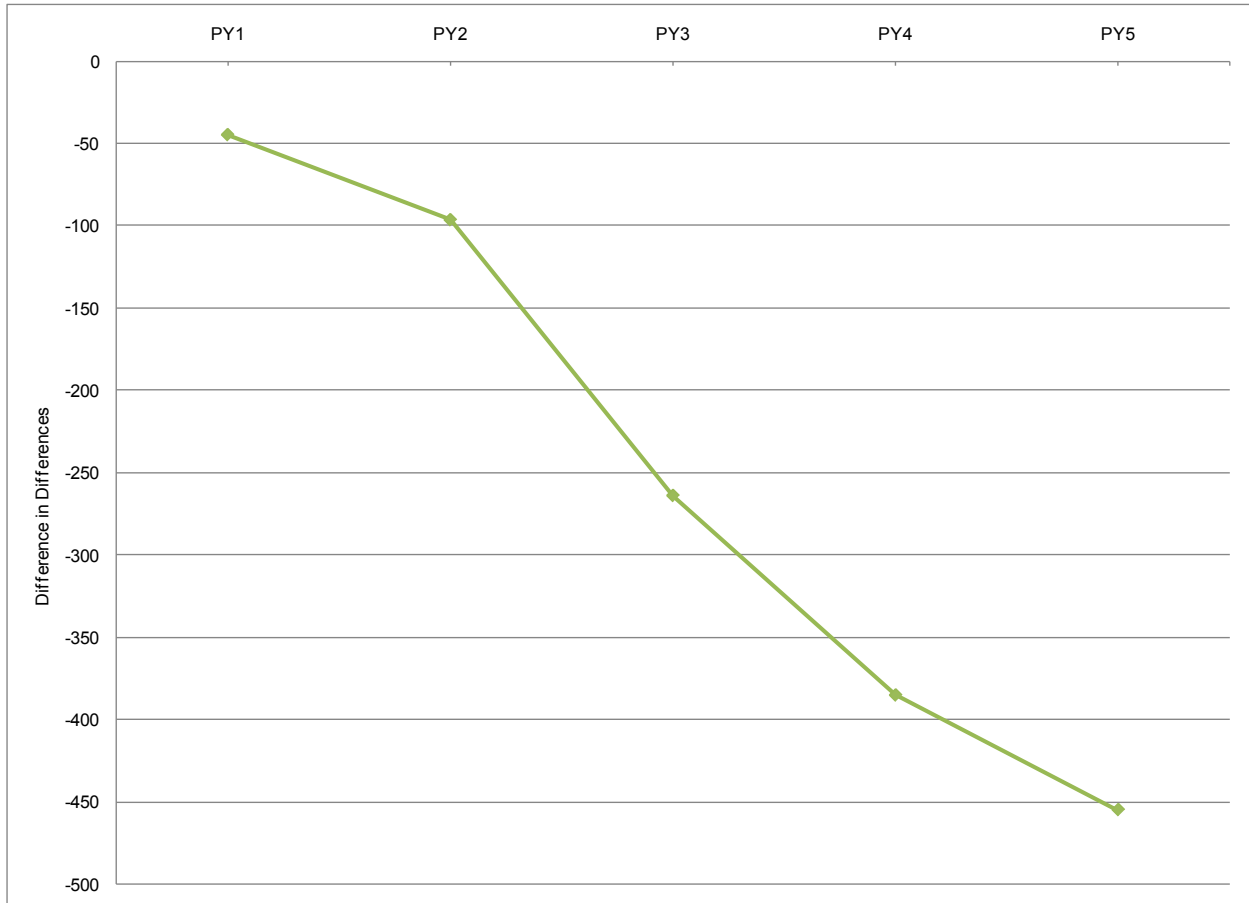
^c denotes statistical significance at the 10% level.

Positive Values = PGPs performed better than their comparison groups.

Negative Values = PGPs performed worse than their comparison groups.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

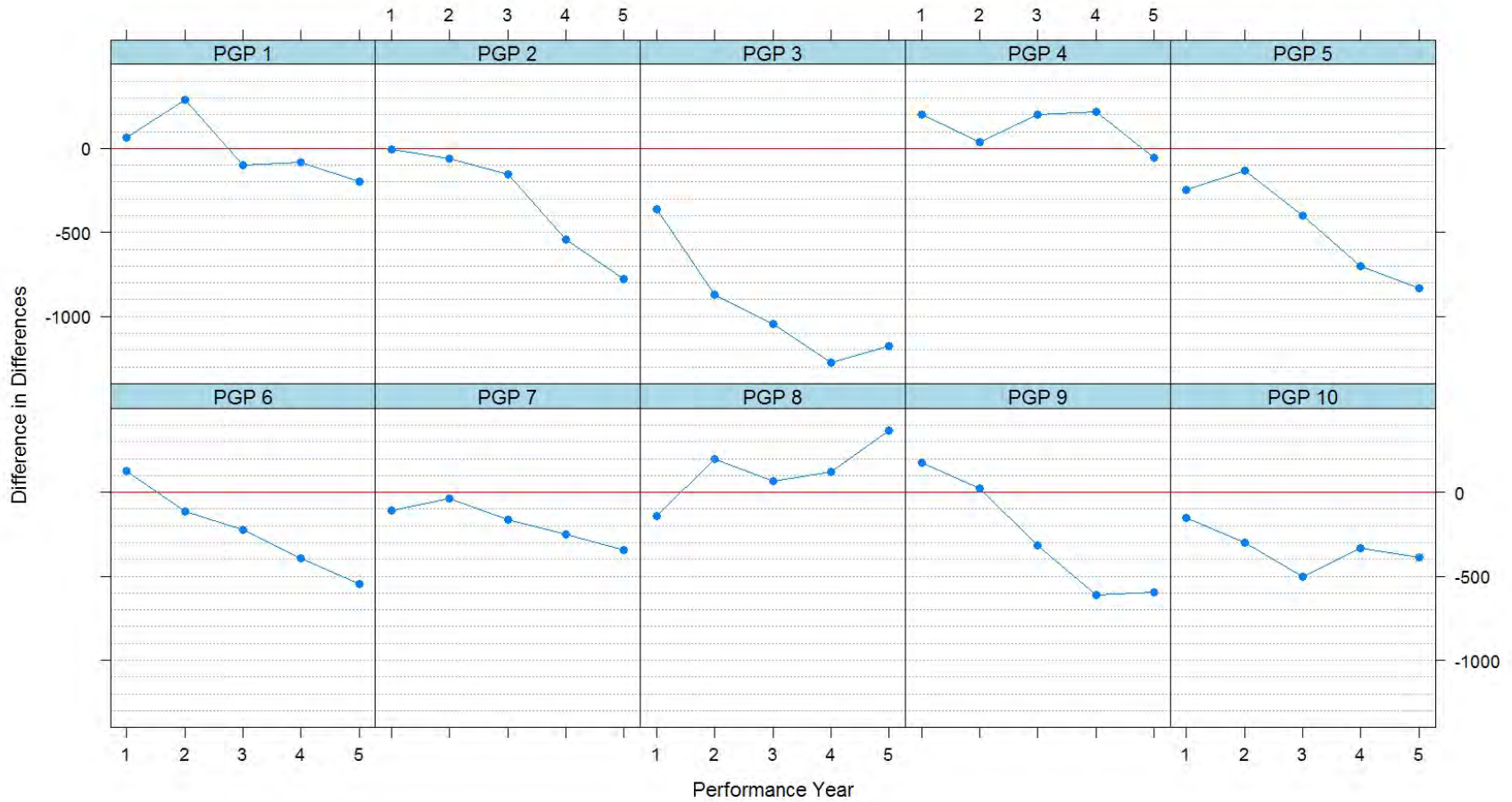
Figure 9-9
Emergency department visits: Difference in differences between the base year and each performance year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-10
Emergency department visits: Difference in differences between the base year and each performance year by PGP

164



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.5.3 All Cause Hospital Readmissions

Table 9-9 shows the difference-in-differences, in percentage points, in readmission rates between the comparison group beneficiaries and the assigned beneficiaries from the base year to each performance year (derived from Table 9-6). Twenty-three of the 55 values have positive signs of which three were statistically significant at the ten percent level or better. Of the 32 negative values, only four were statistically significant.

The overall trend (**Figure 9-11**) shows the percentage point difference-in-differences growing slightly more negative after PY3. None of these overall differences were statistically significant from zero. The percentage point difference-in-differences for the individual PGP sites are shown in **Figure 9-12**. None of the individual PGPs had positive values in all five performance years whereas two (PGP 5 and PGP 7) had negative values in all five years. Seven of the PGPs' difference-in-differences improved between PY4 and PY5, but nonetheless generally remained negative. In relative terms, PGP 4 and PGP 9 performed the best with positive and non-negative values during most performance years.

Table 9-9
All Cause Readmission rates: Difference in differences between the base year and each performance year

PGP	$\Delta CG - \Delta AB$ PY1	$\Delta CG - \Delta AB$ PY2	$\Delta CG - \Delta AB$ PY3	$\Delta CG - \Delta AB$ PY4	$\Delta CG - \Delta AB$ PY5
Overall	0.3%	-0.1%	0.2%	-0.6%	-0.4%
PGP 1	-0.6%	-0.5%	1.3%	-0.5%	0.3%
PGP 2	-1.3%	1.0%	-0.3%	-1.8% ^b	-1.8% ^b
PGP 3	1.3%	0.1%	-0.1%	-0.9%	-0.6%
PGP 4	0.5%	-0.2%	1.6% ^c	1.3%	0.0%
PGP 5	-0.9%	-0.8%	-0.1%	-0.8%	-0.6%
PGP 6	0.8%	0.2%	0.0%	-0.7%	-0.3%
PGP 7	-0.1%	-0.6%	-1.7%	-2.4% ^b	-2.0% ^c
PGP 8	-0.2%	0.2%	0.7%	-0.8%	0.0%
PGP 9	1.6% ^b	0.2%	0.5%	-0.3%	1.0%
PGP 10	1.3% ^c	-0.1%	0.4%	0.8%	-0.1%

NOTES:

^a denotes statistical significance at the 1% level

^b denotes statistical significance at the 5% level

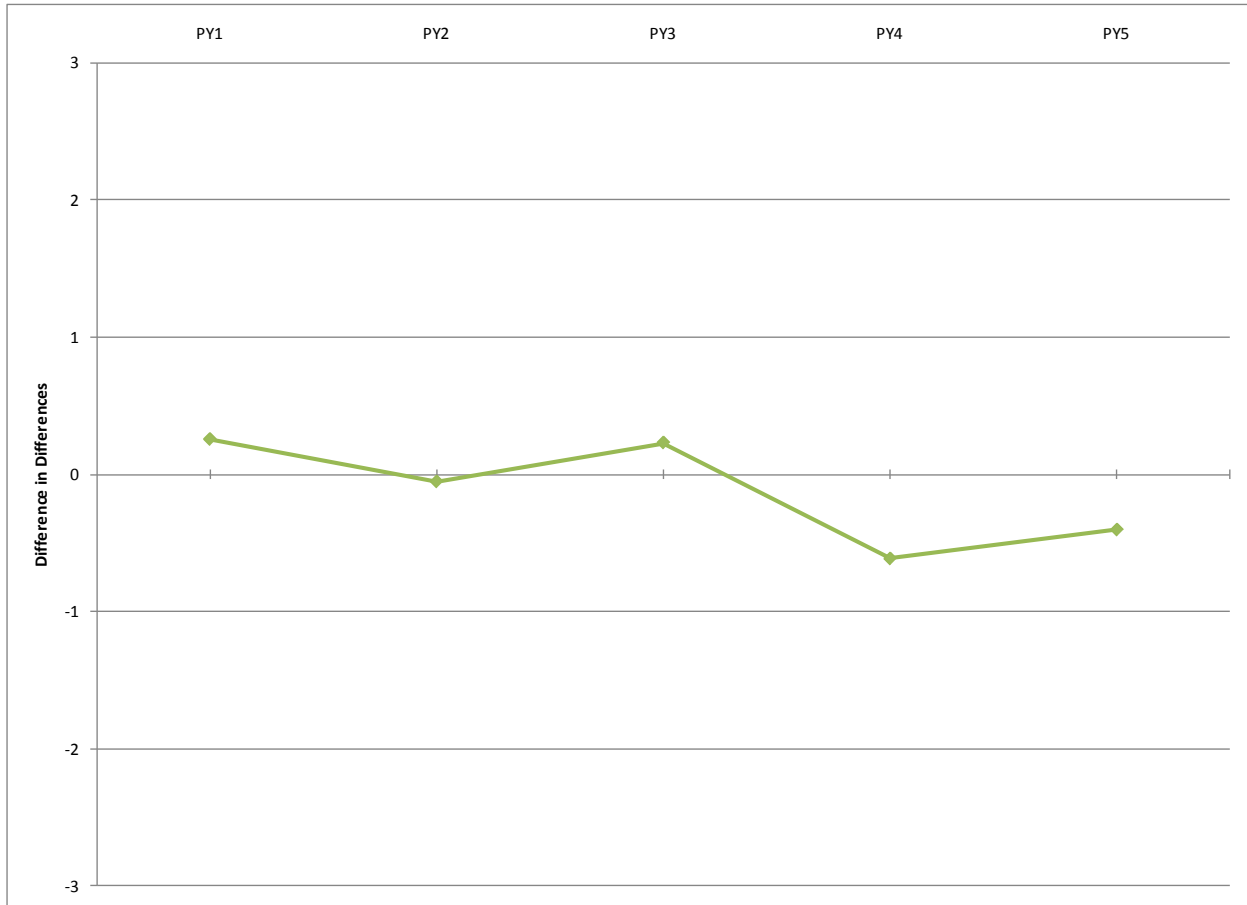
^c denotes statistical significance at the 10% level

Positive Values = PGPs performed better than their comparison groups.

Negative Values = PGPs performed worse than their comparison groups.

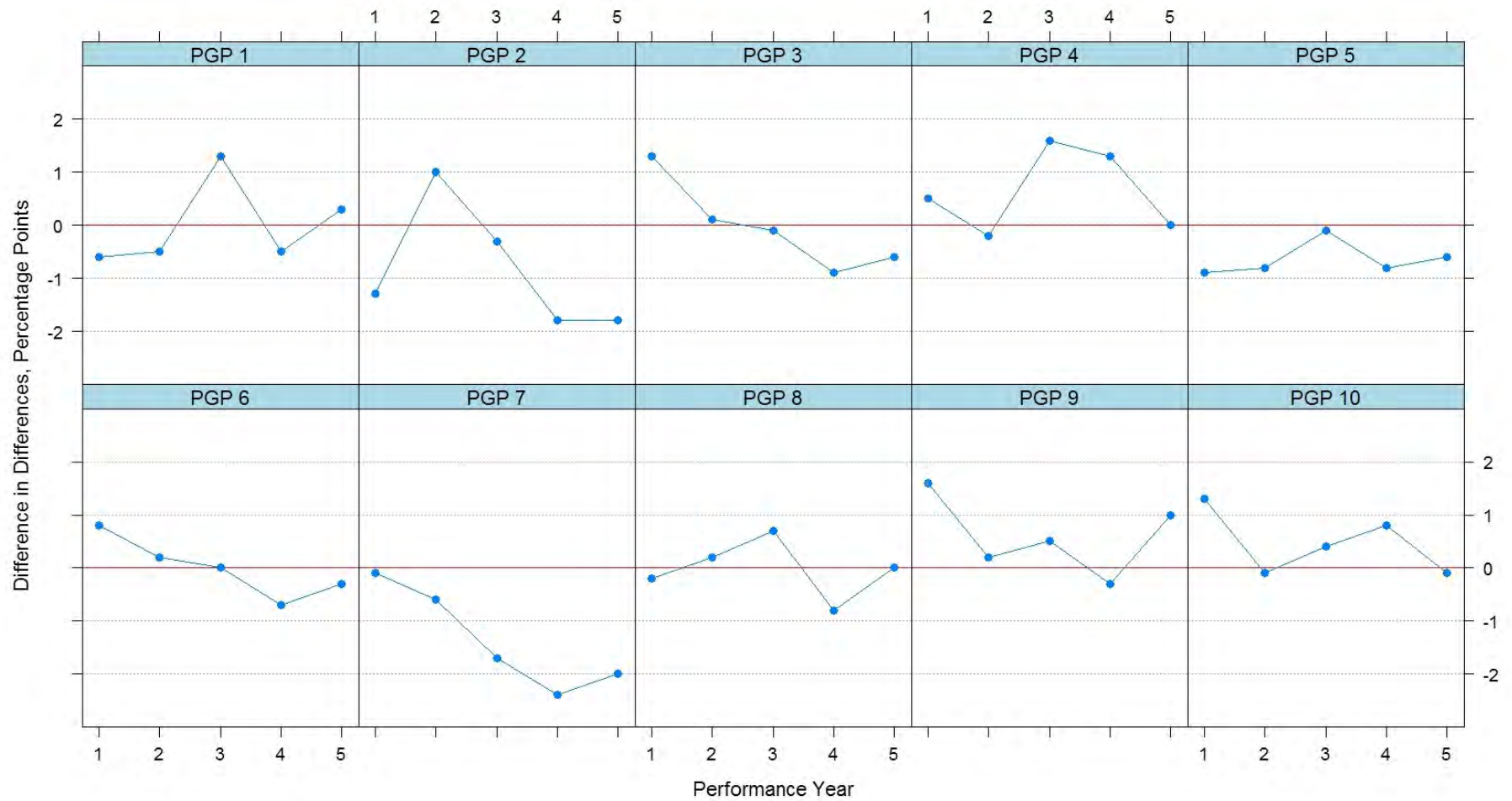
SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-11
All Cause Readmission rates: Difference in differences between the base year and each performance year



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 9-12
Readmission rates: Difference in differences between the base year and each performance year by PGP



SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

9.6 Summary

Overall, the analyses presented in Chapter 9 do not indicate that participating PGPs were able to achieve savings by managing utilization better than non-participating providers. During the course of the PGP Demonstration, hospital discharges for both the assigned beneficiaries and the comparison groups first rose and then started falling in PY3 to rates, ultimately, below the base year rates. Emergency department visits for both the assigned beneficiaries and comparison groups rose between the base year and PY4 before falling in PY5. For both sets of beneficiaries, ED visits in PY5 were above the base year rates.

Results of the difference-in-differences (2D) analyses of hospital discharges indicate most participating PGPs did worse than their comparison groups. Only PGP 8 and PGP 9 consistently did better. Results of the 2D analyses of ED visits discharges indicate most participating PGPs did worse than their comparison groups. Only PGP 4 and PGP 8 consistently did better.

With one possible exception, hospital readmission rates did not exhibit any particular pattern or trend. Seven of the difference-in-differences improved between PY4 and PY5. This, in turn, might have slightly affected hospital discharge rates.

CHAPTER 10

MULTIVARIATE ANALYSES OF DEMONSTRATION IMPACTS

In Chapter 5, an accounting analysis of the Demonstration payment model and period was used to measure the Demonstration impact on Medicare expenditures and identify Medicare program savings. In Chapter 7, the impacts of the PGP Demonstration incentives on annual Medicare expenditures were determined by difference-in-differences simulation analyses contrasting a single follow-up year (PY5) with a baseline year (2004). The same comparison group used for the Demonstration financial reconciliation (see Chapter 2) is employed in the Chapter 5 and 7 analyses.

In this chapter, we present a multivariate methodology for determining the impact of the Demonstration on Medicare expenditures. This methodology involved three significant alterations to the original approach. First, the definition of each comparison group was revised to more closely match the county distribution of the PGPs' assigned beneficiaries, which we refer to as the "county-balancing" method. Second, propensity scores were estimated and propensity score weights were applied to the data. The goal of the propensity score technique is to generate groups that are balanced with respect to key beneficiary characteristics prior to conducting the impact analyses. Third, a multivariate regression model combining nine years of data was used to estimate the impact of the demonstration on annual Medicare expenditures. This multivariate Repeated Cross Sections difference-in-differences regression model (Imbens and Wooldridge, 2009) estimates the effect of the Demonstration on expenditures during the Demonstration period after controlling for beneficiary characteristics and time trends throughout the entire observation period. These three alterations are described in Sections 10.1 to 10.4.

The multivariate methodology is used to estimate Demonstration impacts on expenditures separately for each PGP, as well as an overall effect for all sites combined. The multivariate results are described in Section 10.5. In Section 10.6 we discuss a potential refinement to the county-balancing method and present results of a corresponding sensitivity analysis. Then in Section 10.7 we compare the results in this chapter to those from earlier chapters to evaluate the influence of methodological approach on estimated financial outcomes.

Demonstration expenditure effects are further elaborated in the next two sections. Section 10.8 examines effects separately for each of the six major cost components of total Medicare expenditures. Section 10.9 explores whether Demonstration effects are more pronounced in selected subgroups of beneficiaries.

Finally, the last two sections present the results of using the multivariate methodology to re-analyze Demonstration impacts on quality indicators (Section 10.10) and hospital and emergency department utilization (Section 10.11). Logistic regression models are used to estimate impacts on quality indicators, and hurdle regression models are used to estimate impacts on utilization measures.

10.1 Revised Comparison Group

In the original design, the comparison group consisted of all beneficiaries residing in any county that contained at least 1% of the beneficiaries assigned to the Demonstration PGP. This

approach pulled in all beneficiaries from the entire PGP service area regardless of the geographic distribution of the assigned PGP beneficiaries. In nearly all sites, the groups were being drawn from different geographic areas. In some large counties, nearly all qualifying beneficiaries were assigned to the comparison group. In general, this approach led to considerable imbalances between the groups with regard to the proportion of beneficiaries drawn from individual counties. This in turn produced a) severe skews in the propensity scores, b) propensity models that consisted primarily of large county effects, c) a general lack of group overlap (“common support”) in the propensity distributions, and d) an average constant difference of \$500 in the expenditure levels between the demo and comparison groups even after regression adjustment.

The revised approach was to randomly select comparison beneficiaries from each county to match the number residing in that county in the PGP in that year. Compared to the original approach, the “county-balancing” method had three important consequences. First, it reduced the size of each CG by discarding many beneficiaries from counties that were not well represented among the PGP assigned beneficiaries. Second, it equalized the distribution of beneficiaries drawn from each county in each group (assigned and comparison beneficiaries). Third, it reduced the degree of year-to-year clustering among CG beneficiaries because the same individuals were much less likely to be randomly selected in multiple years. Fourth, it reduced group differences in expenditures to an average of \$151 per year. The effects of county-balancing were also evident in the propensity score models. The alternative approach a) produced far better balance in between the groups in propensity score distributions, b) greatly reduced the effects of individual counties, and c) provided much wider ranges of common support so that fewer beneficiaries were eliminated due to extreme scores.

A refinement to county-balancing would be to create separate propensity models for each county in a PGP's service area that might identify additional comparisons. Section 10.6 provides a discussion and sensitivity analysis for this refinement.

Table 10-1 shows the derivation of the sample of beneficiaries used for the multivariate regression analyses, which primarily reflects reductions in Demonstration comparison beneficiaries due to the revised comparison group. Table 10-1 shows for all 10 PGPs the beneficiary/year observations used in the regressions for the entire 2001 to PY5 analysis period, for assigned and comparison beneficiaries. Focusing on the "Total" column in Table 10-1, 90.6 percent of the Demonstration assigned beneficiary/years were used in the regression analyses (1,776,387 of 1,961,034 person/years). Nine (9.0) percent of assigned beneficiary person/years (175,844) were excluded because these beneficiaries resided outside of their PGP's service area, and 0.4 percent (8,803) were excluded because of an extreme propensity score value.

Only 11.6 percent of Demonstration comparison group beneficiary person/years (1,579,080 of 13,638,460) were used in the regression analyses. This is primarily because 12,053,878 (= 13,634,047 – 1,580,169) beneficiary/years (88.4 percent) were excluded when the number of comparison group beneficiaries was matched to the number of assigned beneficiaries by service area county. Additionally, 4,413 comparison beneficiary/years (0.03 percent) were excluded because they resided outside their PGP's service area, and 1,089 (0.008 percent) were excluded because of an extreme propensity score value (probability < 0.10). In total, the 2001 to

Table 10-1a
Number of demonstration assigned beneficiary years used in the regression analysis,
total, 2001-PY5

Sample	Total	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
Total demonstration beneficiary/years	1,961,034	123,565	262,683	81,225	124,825	233,859	375,405	151,841	164,070	279,358	164,203
Step 1. Dropping out-of-service area beneficiaries											
Number dropped	175,844	18,564	16,195	3,882	5,731	14,899	46,410	6,037	12,975	26,762	24,389
New number observations	1,785,190	105,001	246,488	77,343	119,094	218,960	328,995	145,804	151,095	252,596	139,814
Step 2. Sampling number of comparison group = number of assigned beneficiaries by county											
Number sampled	1,785,190	105,001	246,488	77,343	119,094	218,960	328,995	145,804	151,095	252,596	139,814
Step 3. Dropping extreme Propensity Score beneficiaries											
Number dropped	8,803	0	2,095	0	4	187	5,893	0	2	0	622
Final number observations	1,776,387	105,001	244,393	77,343	119,090	218,773	323,102	145,804	151,093	252,596	139,192

Table 10-1b
Number of demonstration comparison group beneficiary years used in the regression analysis,
total, 2001-PY5

Sample	Total	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
Total demonstration beneficiary/years	13,638,460	154,210	953,967	1,406,261	868,098	1,449,834	415,572	1,950,464	1,133,674	432,695	4,873,685
Step 1. Dropping out-of-service area beneficiaries											
Number dropped	4,413	4,413	0	0	0	0	0	0	0	0	0
New number observations	13,634,047	149,797	953,967	1,406,261	868,098	1,449,834	415,572	1,950,464	1,133,674	432,695	4,873,685
Step 2. Sampling number of comparison group = number of assigned beneficiaries by county											
Number sampled	1,580,169	102,217	203,971	77,343	119,094	212,951	249,506	72,059	151,095	252,119	139,814
Step 3. Dropping extreme Propensity Score beneficiaries											
Number dropped	1,089	0	274	0	81	24	634	5	11	12	48
Final number observations	1,579,080	102,217	203,697	77,343	119,013	212,927	248,872	72,054	151,084	252,107	139,766

172

Table 10-1c
Final sample size (number of beneficiary years) used in the regressions (assigned and comparison)

Total	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
3,355,467	207,218	448,090	154,686	238,103	431,700	571,974	217,858	302,177	504,703	278,958

SOURCE: RTI International Analysis of 2001-PY5 Medicare claims and enrollment information.

PY5 regression sample includes 1,776,387 assigned beneficiary person/years and 1,579,080 comparison group person/years, for a total of 3,355,467 beneficiary person/years. The number of assigned beneficiary person/years in the regression sample is larger than the number of comparison beneficiary person/years because in some service area counties the number of comparison beneficiaries available to be sampled was less than the number of assigned beneficiaries. For example, for PGP 7, all the eligible non-PGP-assigned beneficiaries (4,623) in its home county were included in its regression sample comparison group, but this was less than the number of assigned beneficiaries in the county (13,281).

Some differences by PGP are observed in Table 10-1. For example, PGP 6 accounts for the largest number of assigned beneficiary/years excluded because of out of service area residence or extreme propensity score. The only comparison beneficiaries dropped because of out of service area residence are associated with PGP 1. But the biggest differences are in the percentage of comparison group beneficiaries dropped. In the regression analysis, 66.3 percent of PGP 1's comparison beneficiary/years are included, compared to only 5.5 percent of PGP 3's, 3.7 percent of PGP 7's, and 2.9 percent of PGP 10's that are included. These differences arise because some of the PGPs' service areas include large metropolitan counties where the PGP's assigned beneficiaries (and hence the matched number of comparison beneficiaries) are a small share of the total eligible Medicare beneficiary population.

10.2 Propensity Score Methodology

While the county-balancing method helps to ensure geographic comparability of the two groups, it is possible that the demonstration and comparison groups may differ in some respects that influence both the likelihood of group selection and expenditure levels. Comparison beneficiaries, for example, may be more likely to be drawn from solo practices or have less contact with their primary care physicians than those in the Demonstration group.

To adjust for these potential group differences, we employed a widely-used statistical technique known as propensity score analysis. A propensity score is the predicted probability that a beneficiary is a member of the PGP assigned beneficiaries, conditional on a set of observed covariates. Stronger inferences about intervention effects can be made in situations in which the distribution of propensity scores is the same in the assigned and comparison groups (Imbens and Wooldridge, 2009). To achieve this, we weighted beneficiaries in both groups on the basis of their propensity scores, and then estimated Demonstration effects for the weighted groups. Propensity weights provide a different form of group adjustment than the covariate adjustments in regression models.

The following series of steps were executed for each propensity model.

1. We began by combining PGP assigned and comparison beneficiaries for each site and for each year of the Demonstration (or pre-Demonstration year).
2. For each year, we estimated a logistic regression model to predict propensity scores. The dichotomous outcome for the logistic model was coded 1 for assigned beneficiaries and 0 for comparison beneficiaries. The specification for the model consisted of disease severity, diagnosis groups, demographic characteristics, and

county residence indicators (see Section 10.4 for a list of covariates). Predicted propensity scores were generated from the logistic regression equation and stored for each beneficiary.

3. Inverse propensity weighting (Schafer and Kang, 2008) was used to weight the data. In this method, the weight for the assigned beneficiaries is $1/PS$ (where PS is the propensity score) and the weight is $1/(1-PS)$ for the comparison beneficiaries. The weights were further adjusted by multiplying them by eligibility fractions to account for cases with less than a full year of Medicare eligibility.
4. A well-known problem with inverse probability weighting is that results can be distorted by extreme PS values that produce very large or very small weights. To circumvent this, we removed from further analysis beneficiaries in either group who had low (less than .10) or high (greater than .90) propensity scores.
5. The results for each year were then aggregated to form an analysis file containing data for all 9 years of the study.
6. The distribution of propensity scores was plotted by group to check for balance between the groups. The mean values for individual covariates were also examined to ensure that the propensity score weighting was producing equivalent values for these measures.

10.3 Propensity Model Evaluation

The propensity model estimates for each PGP are shown in *Table 10-2*, using the 2004 base year as an example. The models exhibited varying levels of success in distinguishing Demonstration and comparison beneficiaries. The measures with the greatest impact on propensity scores were always the county indicators. Diagnostic groups, particularly for cancer, were also related to differences in most areas. Demographic factors generally had little influence on the probabilities.

Figure 10-1 overlays the distributions of predicted propensity scores for the PGP assigned and comparison groups for each of the 10 PGPs aggregated over nine years. In these plots, the horizontal axis is the predicted propensity score probability (ranging from zero to one), and the vertical axis is the “density” of the distribution scaled to show the relative frequency of beneficiaries adjusted for differences in the sizes of the groups.

These plots reveal several important features of the groups. First, in all PGPs, the distributions for both groups cover similar ranges of propensity scores. Each group contains a mixture of beneficiaries with low, moderate, and high predicted propensity scores. Second, in nearly all of the sites the propensity score distributions of the two groups track one another quite closely. This indicates that covariate values were well balanced by equalizing the county distributions. Third, spikes appear in the frequencies for many of the plots. These were usually caused by counties that contained large numbers of beneficiaries, increasing the frequencies in both the assigned and comparison groups. The high propensity scores in PGP 6 are caused by the concentration of most of the Demonstration Assigned Beneficiaries in a single county.

Table 10-2
Propensity score models for regression weighting and sample selection, by PGP, 2004
(Table entries are coefficient estimates with p-values for statistical significance beneath)

Variable	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
N	25,180	52,201	19,996	33,540	52,110	71,778	25,281	39,705	61,006	37,710
R ²	0.008	0.061	0.009	0.016	0.019	0.049	0.056	0.012	0.008	0.032
Risk score	0.0143	0.0873	0.0314	-0.0632	0.0392	0.0972	-0.0393	0.0336	-0.0061	0.1409
	0.345	0.000	0.033	0.000	0.000	0.000	0.009	0.002	0.500	0.000
Age group (0-54)	0.0071	0.4189	0.0349	-0.2578	0.3482	0.1601	-0.0666	-0.1848	-0.1500	0.5609
	0.907	0.000	0.565	0.000	0.000	0.000	0.343	0.000	0.000	0.000
Age group (55-64)	-0.0707	0.2381	0.0839	-0.2661	0.1031	0.0777	-0.0863	-0.3304	-0.0305	0.2385
	0.270	0.000	0.194	0.000	0.009	0.048	0.232	0.000	0.363	0.000
Age group (75-84)	0.0135	-0.0326	0.0617	0.0110	-0.0565	0.0088	0.2287	-0.1753	0.0675	-0.0823
	0.653	0.135	0.081	0.683	0.009	0.633	0.000	0.000	0.001	0.001
Age group (>85)	-0.0471	-0.2211	0.1775	-0.0323	-0.2934	-0.0639	0.2601	-0.2496	0.0342	-0.1430
	0.257	0.000	0.000	0.402	0.000	0.010	0.000	0.000	0.236	0.000
Male (0=no; 1=yes)	0.0155	-0.0004	-0.0361	-0.0533	0.0026	-0.0137	-0.0749	-0.0124	-0.0493	0.0279
	0.558	0.983	0.231	0.024	0.889	0.397	0.010	0.557	0.004	0.204
CMS Hierarchical Condition Category (HCC)= cancer	0.3153	0.4499	-0.2034	-0.3131	0.1769	0.3512	0.0747	-0.0469	0.2058	0.3265
	0.000	0.000	0.000	0.000	0.000	0.000	0.079	0.166	0.000	0.000
HCC = diabetes	-0.0279	-0.0692	0.0453	0.1278	-0.0263	0.0080	-0.0550	-0.0158	-0.0429	-0.2954
	0.413	0.003	0.226	0.000	0.214	0.687	0.120	0.567	0.039	0.000
HCC = AMI	0.0639	0.1373	0.3531	0.2791	-0.1011	-0.0618	-0.0815	-0.1888	0.0558	-0.2009
	0.470	0.014	0.000	0.000	0.025	0.213	0.285	0.006	0.202	0.000
HCC= CHF	0.0699	0.2361	-0.1636	0.1260	-0.0916	-0.0851	0.1812	-0.0494	0.1419	-0.0930
	0.134	0.000	0.002	0.001	0.002	0.002	0.000	0.208	0.000	0.008

(continued)

Table 10-2 (continued)
Propensity score models for regression weighting and sample selection, by PGP, 2004
(Table entries are coefficient estimates with p-values for statistical significance beneath)

Variable	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
HCC = Stroke	-0.1200 0.108	-0.0531 0.316	-0.0851 0.256	0.2234 0.000	0.2902 0.000	-0.4933 0.000	0.1676 0.021	-0.2878 0.000	-0.1636 0.000	-0.1974 0.000
HCC = Vascular disease	0.0556 0.243	0.2621 0.000	0.0270 0.585	-0.0430 0.253	-0.0236 0.346	0.0784 0.005	0.0450 0.298	-0.1175 0.002	-0.1092 0.000	-0.1619 0.000
HCC = COPD	0.1762 0.000	-0.2346 0.000	0.0686 0.157	0.0276 0.415	-0.2484 0.000	-0.0408 0.129	0.0466 0.287	-0.0685 0.076	-0.0565 0.020	-0.5376 0.000
Medicaid status (0=no; 1=yes)	0.0403 0.400	-0.1612 0.000	0.0223 0.629	-0.1910 0.000	0.0361 0.202	0.0385 0.160	-0.0087 0.859	-0.3116 0.000	-0.1426 0.000	0.0115 0.767
Originally disabled (0=no; 1=yes)	-0.0548 0.304	0.0761 0.065	0.1373 0.036	-0.1309 0.004	-0.0058 0.873	0.0826 0.015	-0.1735 0.008	-0.4116 0.000	-0.0494 0.119	0.0722 0.116
ESRD status (0=no; 1=yes)	-0.1044 0.550	0.3019 0.035	-0.5998 0.002	-0.8159 0.000	0.3208 0.008	-0.0520 0.627	-0.4543 0.014	-0.2541 0.047	-0.0792 0.474	0.1597 0.111
Race = black	-0.4883 0.114	0.3148 0.070	-0.3684 0.025	-0.4140 0.000	0.1619 0.194	-0.4482 0.091	-0.2777 0.000	-0.7919 0.000	-0.2863 0.005	-0.7109 0.000
Race = Asian	0.1111 0.722	0.7898 0.001	-1.1173 0.000	0.0587 0.823	0.0740 0.744	-0.2237 0.150	-0.5362 0.007	-0.0248 0.811	0.0552 0.779	0.8284 0.000
Race =other race	-0.5080 0.000	0.0616 0.576	-0.2695 0.004	-0.1755 0.218	0.1945 0.106	-0.5216 0.000	-0.6088 0.000	-0.2096 0.017	0.0683 0.495	0.3861 0.000
_lcounty_2	0.0407 0.598	1.2895 0.000	0.6570 0.000	-0.0286 0.593	0.2491 0.000	-0.3834 0.000	1.0407 0.000	-0.2059 0.067	-0.0600 0.236	-0.2459 0.016
_lcounty_3	(omitted) —	-0.2282 0.004	0.0547 0.699	0.0894 0.157	-0.0983 0.115	-0.3285 0.000	-0.0162 0.799	-0.0211 0.828	-0.3456 0.000	-0.0796 0.309

(continued)

Table 10-2 (continued)
Propensity score models for regression weighting and sample selection, by PGP, 2004
(Table entries are coefficient estimates with p-values for statistical significance beneath)

Variable	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
_Icounty_4	-0.1990 0.071	-0.1603 0.021	0.2149 0.002	0.0606 0.228	-0.1232 0.002	-0.3553 0.000	-0.2259 0.001	0.0947 0.275	-0.1706 0.000	-0.2265 0.018
_Icounty_5	-0.4105 0.001	-0.0927 0.177	— —	0.5232 0.000	(omitted) —	1.2647 0.000	— —	0.1411 0.152	-0.0686 0.347	-0.0260 0.713
_Icounty_6	(omitted) —	-0.1854 0.008	— —	0.3394 0.003	0.3314 0.000	(omitted) —	— —	-0.0081 0.933	-0.4259 0.000	-0.3622 0.001
_Icounty_7	-0.4119 0.000	-0.1489 0.110	— —	-0.0124 0.861	0.2065 0.000	-0.5434 0.000	— —	-0.0494 0.685	-0.3851 0.000	-0.2562 0.013
_Icounty_8	0.1462 0.094	-0.3225 0.000	— —	0.0407 0.514	0.2116 0.001	(omitted) —	— —	— —	-0.1734 0.000	0.0261 0.718
_Icounty_9	-0.3571 0.000	0.3751 0.000	— —	-0.0321 0.646	-0.1162 0.023	-0.2535 0.000	— —	— —	-0.4085 0.000	-0.0276 0.670
_Icounty_10	-0.1221 0.043	-0.3485 0.000	— —	(omitted) —	-0.1845 0.002	0.0367 0.188	— —	— —	-0.1238 0.022	0.1220 0.063
_Icounty_11	-0.3085 0.001	-0.0483 0.585	— —	— —	1.3758 0.000	-0.3933 0.000	— —	— —	-0.0291 0.560	— —
_Icounty_12	-0.2338 0.001	(omitted) —	— —	— —	0.1612 0.000	-0.6886 0.000	— —	— —	-0.7147 0.000	— —
_Icounty_13	0.3616 0.000	(omitted) —	— —	— —	-0.1565 0.000	1.1339 0.000	— —	— —	-0.1054 0.107	— —
_Icounty_14	— —	-0.3644 0.000	— —	— —	-0.1631 0.007	1.6548 0.000	— —	— —	-0.5409 0.000	— —

(continued)

Table 10-2 (continued)
Propensity score models for regression weighting and sample selection, by PGP, 2004
(Table entries are coefficient estimates with p-values for statistical significance beneath)

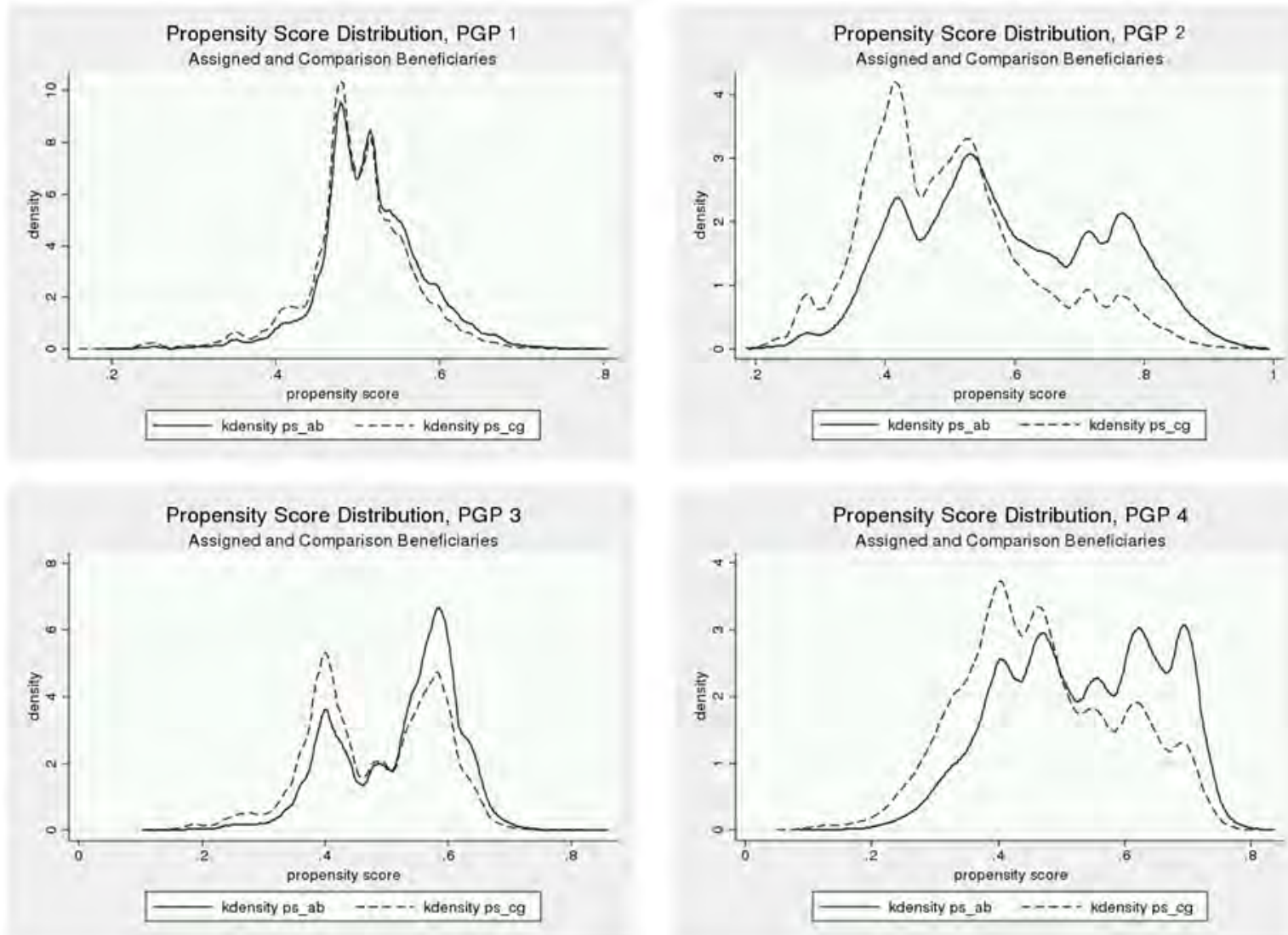
Variable	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
_Icounty_15	—	—	—	—	—	—	—	—	—	—
_Icounty_16	—	—	—	—	—	—	—	—	—	—
_Icounty_17	—	—	—	—	—	—	—	—	—	—
_Icounty_18	—	—	—	—	—	—	—	—	—	—
_Icounty_19	—	—	—	—	—	—	—	—	—	—
_Icounty_20	—	—	—	—	—	—	—	—	—	—
	0.001	-0.2148	-0.6462	-0.1513	-0.2222	0.1766	-0.0152	-0.0741	0.1004	-0.2016
Constant	0.987	0.001	0.000	0.003	0.000	0.000	0.797	0.396	0.011	0.002

NOTES:

1. The sample includes PGP assigned and comparison group beneficiaries, except that out of service area assigned beneficiaries are excluded.
2. The regression is weighted by each beneficiary's eligibility fraction.
3. The model is estimated as a logistic regression.
4. Similar propensity score regressions were estimated for each year 2001 to PY5.

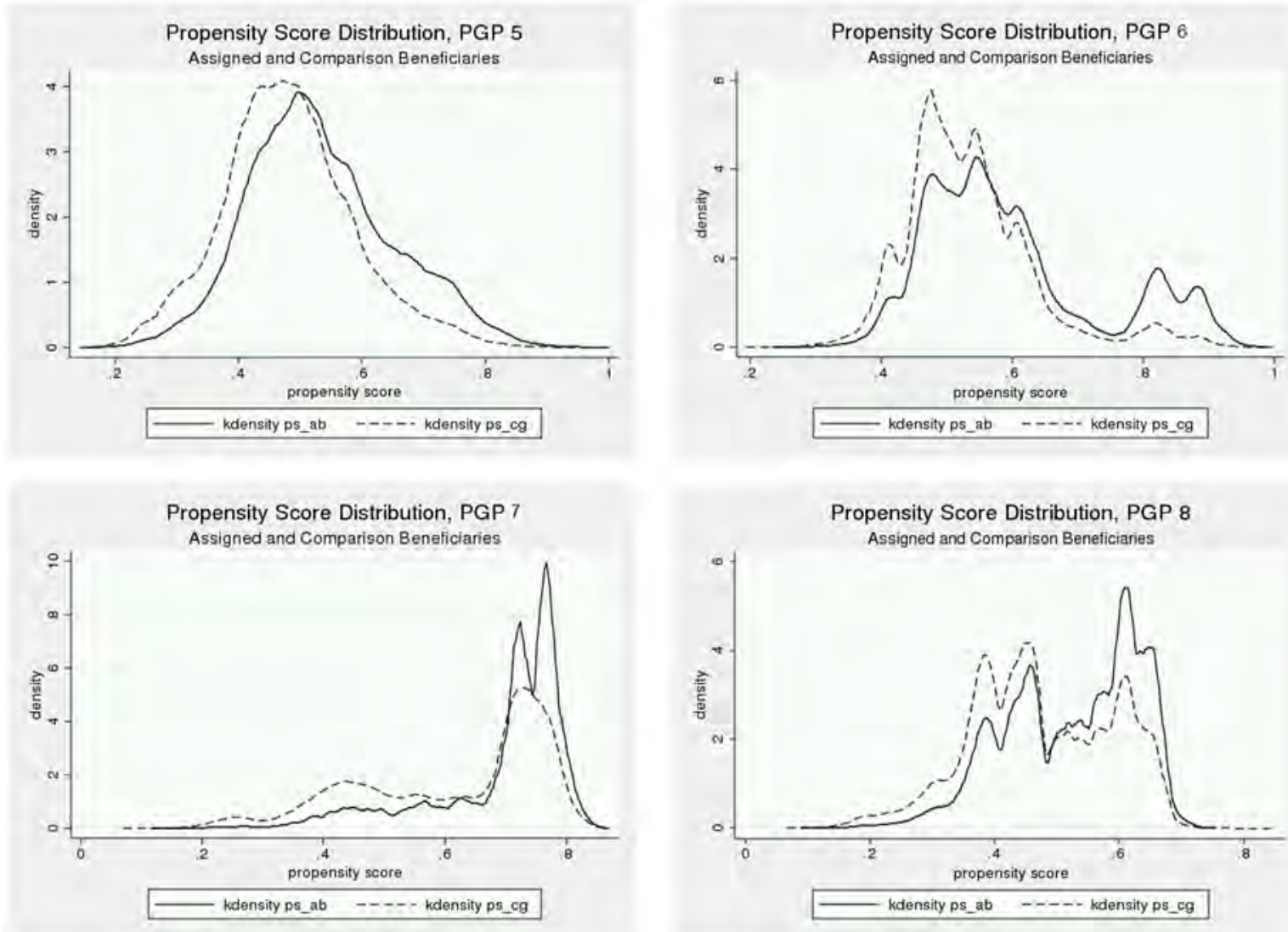
SOURCE: RTI International analysis of Medicare claims 2001–2010

Figure 10-1
Distributions of predicted propensity scores for the assigned and comparison beneficiaries for each of the 10 PGP sites, aggregated over 2001 to PY5



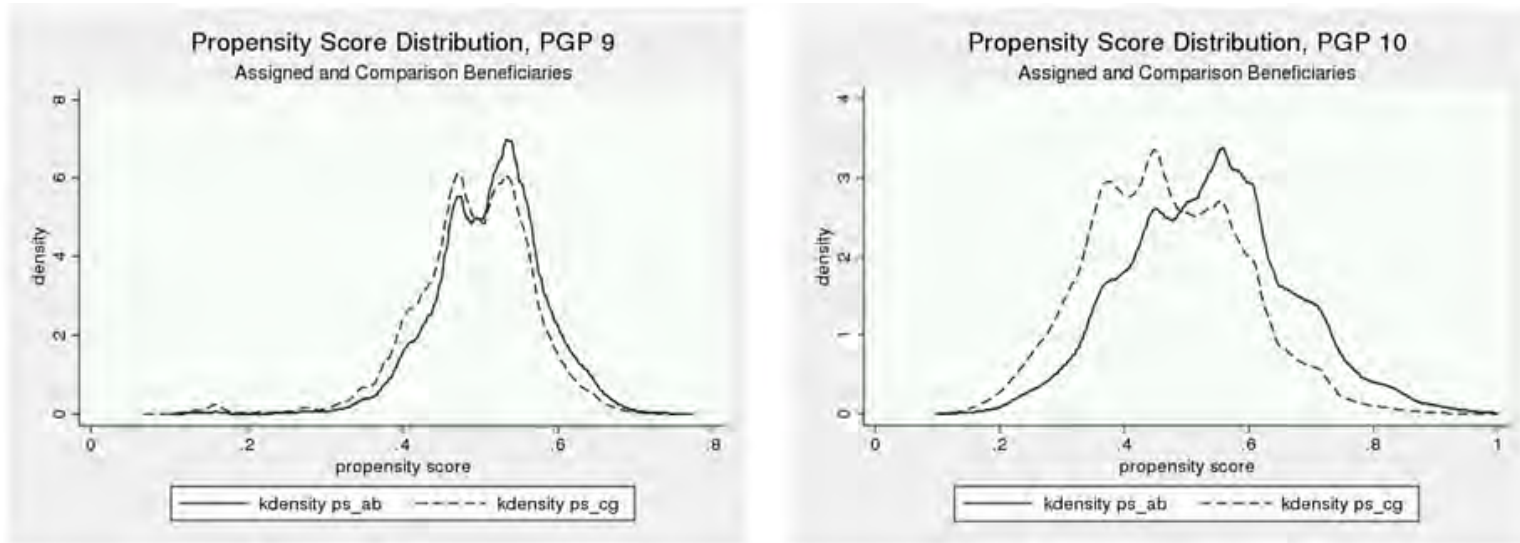
(continued)

Figure 10-1 (continued)
Distributions of predicted propensity scores for the assigned and comparison beneficiaries for each of the 10 PGP sites, aggregated over 2001 to PY5



(continued)

Figure 10-1 (continued)
Distributions of predicted propensity scores for the assigned and comparison beneficiaries for each of the 10 PGP sites, aggregated over 2001 to PY5



SOURCE: RTI analysis of 2001–2010 Medicare administrative data.

Applying the exclusions for extreme propensity scores removed only a small percentage of the assigned beneficiaries. Because of the considerable overlap in propensity score distributions shown in Figure 10-1, the propensity score weighting had only a minor impact on the mean covariate values in each group.

10.4 Repeated Cross Sections Difference-in-Differences Model

To estimate the impact of the PGP Demonstration on outcomes of interest (e.g., financial, quality, utilization), we employ the Repeated Cross Sections difference-in-differences model (Imbens and Wooldridge, 2009), which has three distinct advantages over simpler pre-demonstration and demonstration period contrasts. First, it permits us to incorporate data from all nine years of observation. Second, it adjusts for secular trends occurring during the pre-demonstration and demonstration periods. Third, it permits us to adjust for beneficiary characteristics that influence expenditure levels. These adjustments also help to improve the precision of the estimated Demonstration effects.

The general form of the model as implemented for this analysis was as follows:

$$E_{iy} = a + b_1 * D + b_2 * P + b_3 * D * P + b_j * X_j + b_k * X_k + b_m * X_m + e \quad (\text{Eq.10-1})$$

where: E_{iy} = the annualized Medicare expenditure amount for beneficiary i in year y ,
 a = an intercept term,
 D = an indicator coded 1 for PGP assigned beneficiaries and 0 for comparison beneficiaries,²⁴
 P = a period indicator coded 1 for the Demonstration performance years (PY1-PY5) and 0 for the pre-Demonstration period (2001–2004),
 X_j = a vector of j beneficiary-level covariates,
 X_k = a vector of k year indicators coded for each of the years from 2002 to PY5, with 2001 serving as the reference year,
 X_m = a vector of indicators for individual counties within each geographic area,
 $b_1, b_2, b_3, b_j, b_k,$ and b_m are regression coefficient vectors, and
 e = a residual term.

Three broad sets of covariates were employed. The beneficiary-level covariates included measures of disease severity (concurrent HCC risk score, ESRD, disability as original Medicare eligibility status), diagnostic groups (cancer, diabetes, AMI, CHF, stroke, vascular disease, and COPD classifications), and demographic characteristics (gender, age group, race/ethnicity), and

²⁴ Our preferred model specification includes a single interaction term for the Demonstration effect, which is an average effect over the 5-year Demonstration period. For a given outcome of interest (e.g., Medicare expenditures), our preferred model specification 1) allows us to present a single estimate of the impact of the Demonstration over the entire Demonstration period, and 2) increases statistical reliability by averaging over all years in the Demonstration period. However, an alternative model specification would be to include a full set of Year x PGP Assigned Beneficiaries interactions. The alternative model would imply the same average effect as our preferred model, but would also provide estimates of the Demonstration effect for each year of the Demonstration period. Although in Chapter 10 we generally present results for our preferred model, for Medicare expenditures we also present the results of the alternative model graphically (see Figure 10-2).

Medicaid status. The second covariate set consisted of indicators delineating each observation year, and the third set was the county of residence indicators. The performance year 5 risk score cap is not reflected in the regression results, that is, the uncapped PY5 risk scores are used.

Individual observations in the data are not independent because many beneficiaries appear in multiple years. We used the cluster option in Stata to correct standard errors for this clustering. Beneficiaries with repeated measures were identified on the basis of their HIC numbers. The data were also weighted by the fraction of the year that beneficiaries were alive and eligible for both Medicare Part A and B services. The regression models were estimated by weighted least squares for each PGP site.

In this model, the coefficient of primary interest is b3. This interaction coefficient estimates the annual effect of the Demonstration on annual expenditures during the Demonstration performance years compared to comparison group beneficiaries during that period. Coefficient b1 adjusts for constant annual differences between the groups that persist throughout the study period, and b2 estimates increased expenditures during the performance years that were common to both groups.

10.5 Demonstration Medicare Expenditure Effects Based on Repeated Cross Sections Difference-in-Differences Model

The Repeated Cross Sections difference-in-differences regression model was estimated using the inverse propensity score weights to derive adjusted estimates of the Demonstration effects. These analyses were also adjusted for beneficiary clustering over time and for Medicare eligibility fraction. The regression model estimates are detailed in *Table 10-3*. The overall sample size across all PGP sites and years (assigned beneficiaries and comparison groups) was 3,355,467 and the R-squared value was 0.584.

In the regression models, the annual Demonstration effect is estimated by the coefficient for (assigned beneficiary)*(performance period). This interaction dummy variable identifies PGP assigned beneficiaries during the Demonstration performance period. A negative coefficient indicates savings. This is because if the coefficient is negative, it means that estimated Medicare expenditures are lower for PGP assigned beneficiaries relative to comparison beneficiaries during the Demonstration performance period (after adjusting for covariates in the model). On the other hand, a positive coefficient indicates dis-savings because estimated Medicare expenditures are higher for PGP assigned beneficiaries relative to comparison group beneficiaries in the Demonstration performance period (after adjusting for covariates in the model). The overall impact of the Demonstration across all PGP sites was a savings of \$171 per assigned beneficiary per person year (standard error = \$22, 95% confidence interval = \$127 to \$215) during the Demonstration performance period. This estimate was statistically significantly different from zero ($p < 0.01$). These estimates also show statistically significant expenditure savings ($p < 0.01$) ranging from \$188 to \$818 per assigned beneficiary per year for PGPs 2, 5, and 6. PGPs 3 and 8 had statistically significant savings ($p < 0.05$) of \$229 and \$142, respectively. A statistically significant dis-savings of \$323 per year was found for PGP 1 ($p < 0.01$), while no significant effects were found for the other four PGPs ($p > 0.10$).

Table 10-3
Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures
(standard errors and p-values for statistical significance are shown below coefficient estimates in dollars)
(regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)

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

(continued)

Table 10-3 (continued)
Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures
(standard errors and p-values for statistical significance are shown below coefficient estimates in dollars)
(regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)

Variable	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	PGPs										
	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v
Post-demonstration period	1,479	1,101	1,654	1,136	1,251	1,699	1,615	1,500	1,268	1,358	2,053
	23	81	69	102	79	57	54	112	71	50	96
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PY2	143	96	233	170	282	50	148	179	226	20	150
	22	74	67	97	78	58	52	100	70	48	91
	0.000	0.194	0.000	0.078	0.000	0.387	0.005	0.073	0.001	0.685	0.101
PY3	243	10	356	139	286	185	191	472	325	220	251
	23	79	70	100	80	64	57	104	73	51	97
	0.000	0.901	0.000	0.165	0.000	0.004	0.001	0.000	0.000	0.000	0.010
PY4	485	289	580	468	686	367	385	857	526	331	642
	25	83	72	105	84	68	63	109	77	53	101
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PY5	378	286	486	486	656	439	93	647	189	268	563
	25	85	71	106	85	70	61	110	80	54	98
	0.000	0.001	0.000	0.000	0.000	0.000	0.126	0.000	0.018	0.000	0.000
Male (0=no; 1=yes)	-212	-157	-277	-143	-253	-153	-231	-309	-178	-164	-271
	13	48	37	55	43	34	31	59	39	29	52
	0.000	0.001	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Age group (0-54)	-360	-579	-229	-476	-356	-309	-730	455	-529	-357	-404
	30	133	87	144	89	73	83	164	99	63	116
	0.000	0.000	0.009	0.001	0.000	0.000	0.000	0.005	0.000	0.000	0.000

(continued)

Table 10-3 (continued)
Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures
(standard errors and p-values for statistical significance are shown below coefficient estimates in dollars)
(regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)

Variable	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.
	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v
Age group (55-64)	-466	-473	-315	-532	-695	-432	-498	-445	-646	-386	-561
	32	127	95	128	94	77	90	156	123	60	127
	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000
Age group (75-84)	-102	-133	91	-128	-5	-335	-134	68	-44	-140	-65
	13	48	38	57	43	36	31	59	40	30	53
	0.000	0.006	0.017	0.025	0.914	0.000	0.000	0.250	0.267	0.000	0.215
Age group (> 85)	-206	-576	270	-452	-61	-587	-374	246	38	-269	-169
	20	69	61	88	69	55	47	93	64	47	83
	0.000	0.000	0.000	0.000	0.372	0.000	0.000	0.008	0.552	0.000	0.043
Medicaid status (0=no; 1=yes)	-294	-660	-179	-333	63	-428	-380	-694	-59	-204	125
	24	100	74	108	66	57	61	120	93	48	102
	0.000	0.000	0.016	0.002	0.338	0.000	0.000	0.000	0.526	0.000	0.221
Originally disabled (0=no; 1=yes)	-251	-14	-114	-331	-196	-392	-322	-406	-372	-97	-251
	31	115	98	133	95	76	77	167	134	61	126
	0.000	0.906	0.243	0.013	0.040	0.000	0.000	0.015	0.006	0.111	0.047
ESRD status (0=no; 1=yes)	-5,196	-3,511	-10,537	-1,192	-2,839	-5,804	-5,148	-7,883	-1,942	-2,038	-7,924
	186	740	654	705	594	525	495	802	575	473	495
	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Race = black	242	658	-759	452	-134	104	439	-177	535	48	253
	48	501	288	303	72	245	451	196	130	210	92
	0.000	0.189	0.008	0.137	0.063	0.672	0.331	0.367	0.000	0.819	0.006

(continued)

Table 10-3 (continued)
Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures
(standard errors and p-values for statistical significance are shown below coefficient estimates in dollars)
(regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)

Variable	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.
	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v	p-v
Race = Asian	-771	-750	-210	-559	-804	103	-1,080	-580	-1,209	-554	-981
	89	465	352	163	584	482	263	409	188	335	231
	0.000	0.107	0.551	0.001	0.168	0.831	0.000	0.157	0.000	0.098	0.000
Race = other race	-288	-24	-73	-184	-27	-364	-235	-732	-185	-145	-540
	61	158	201	238	263	206	162	223	158	187	178
	0.000	0.878	0.716	0.438	0.918	0.076	0.146	0.001	0.241	0.438	0.002
CMS Hierarchical	102	654	-315	557	-488	72	387	-200	447	-36	329
Condition	26	96	73	113	86	66	69	111	90	58	93
Category (HCC) =	0.000	0.000	0.000	0.000	0.000	0.275	0.000	0.072	0.000	0.532	0.000
cancer	-157	-84	-263	-490	30	-152	-205	-181	-207	14	-88
HCC = diabetes	17	64	50	73	52	41	42	80	59	38	67
	0.000	0.194	0.000	0.000	0.571	0.000	0.000	0.024	0.000	0.721	0.192
HCC = AMI	743	1,579	621	1,160	615	-1,055	2,358	144	912	1,163	614
	60	255	191	273	206	139	174	279	227	121	190
	0.000	0.000	0.001	0.000	0.003	0.000	0.000	0.606	0.000	0.000	0.001
HCC = CHF	436	524	374	361	541	347	404	627	418	26	1,488
	31	108	95	155	100	76	76	142	113	64	116
	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.683	0.000
HCC = stroke	440	-85	584	259	-144	-41	1,883	479	1,768	131	187
	50	177	168	205	148	111	160	228	202	103	167
	0.000	0.633	0.001	0.207	0.329	0.715	0.000	0.035	0.000	0.204	0.261

(continued)

Table 10-3 (continued)
Overall and individual PGP multivariate financial outcomes regression models for per capita expenditures
(standard errors and p-values for statistical significance are shown below coefficient estimates in dollars)
(regression is estimated for assigned and comparison beneficiaries on 2001-PY5 data)

Variable	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
	PGPs										
	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v	s.e. p-v
HCC = vascular disease	226 27 0.000	150 99 0.129	306 85 0.000	-155 116 0.182	695 97 0.000	-26 56 0.641	408 72 0.000	20 111 0.861	35 102 0.732	109 58 0.060	861 100 0.000
HCC = COPD	58 25 0.020	192 86 0.025	-89 74 0.231	260 115 0.024	30 77 0.701	40 61 0.512	100 69 0.147	5 112 0.967	146 101 0.148	51 51 0.312	334 99 0.001

NOTES:

1. Dependent variable is Medicare annualized expenditures.
2. The regression is estimated on 2001 to PY5 data (2001 to 2010 data) for PGP assigned and comparison group beneficiaries (simulated assigned and comparison group beneficiaries before PY1), selected as described in the text.
3. Regression is weighted by Medicare eligibility fraction and by beneficiary propensity scores as described in the text.
4. Demonstration impact is estimated by the coefficient of (assigned beneficiary)*(performance year). Negative coefficients indicate savings, and positive coefficients indicate dis-savings, on a per person per year basis.
5. P-values for statistical significance of regression coefficient estimates presented below coefficient estimates. A p-value of '0' indicates that the coefficient is significantly different from zero at better than the 0.1% level of significance. A p-value of for example 0.006 indicates a 0.6% level of significance, a p-value of 0.015 indicates a 1.5% level of significance, a p-value of 0.077 indicates a 7.7% level of significance, a p-value of 0.325 indicates a 32.5% level of significance, etc.
6. Regression models also include dummy variables for county of residence of beneficiaries and a constant term (not shown in table). The 2001 and PY1 year dummy variables are omitted to avoid collinearity.
7. Statistical significance levels (p-values) and coefficient standard errors are adjusted for beneficiary-level clustering.
8. Results do not reflect the Demonstration PY5 risk score cap.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

The effects of the covariates on expenditures are consistent with what one would expect in these analyses. Expenditures were heavily influenced by HCC risk scores and diagnostic classifications, particularly CHF and stroke.²⁵ The effects of individual years of observation follow the time trends observed in the earlier analyses. In addition, there were many large county effects in each site (not shown).

In addition to the average annual Demonstration effect during the Demonstration performance years as analyzed in Table 10-3, the time pattern of the annual Demonstration effect is of interest. To analyze this, we estimated the same regression model as in Table 10-3, except that we replaced the single interaction of assigned beneficiary status with a Demonstration period indicator (Post*AB), with 9 interactions of assigned beneficiary status with each of the 9 annual periods, 2001 to PY5 (AB*2001, AB*2002, AB*2003, etc.). (One of the assigned beneficiary/year interactions were omitted from the regression to avoid perfect collinearity.) The 9 assigned beneficiary/year interactions allow year-by-year analysis of the relationship of assigned beneficiary to comparison group expenditures.

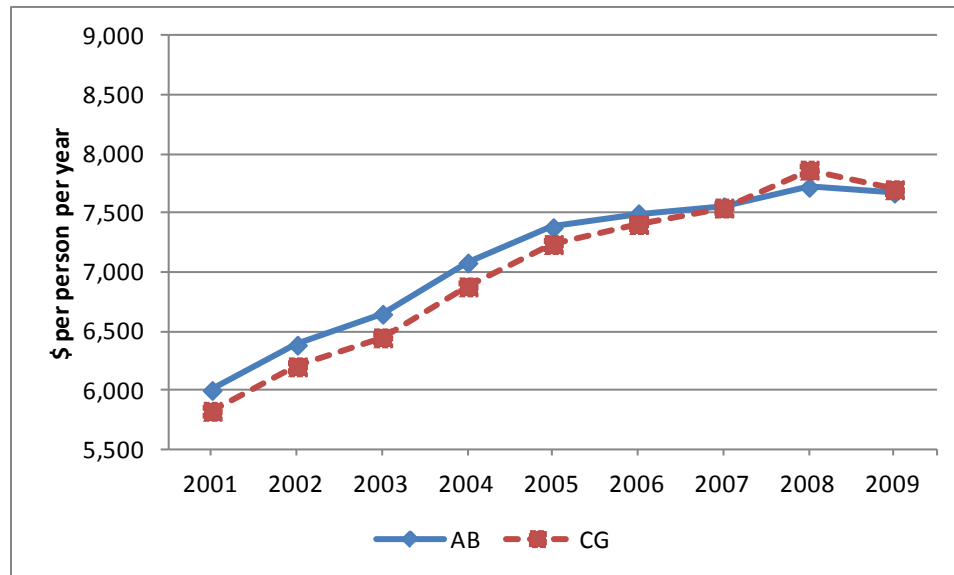
We do not show the alternative regression results in a table, but we used it to produce the graph shown in **Figure 10-2**. Figure 10-2 simulates assigned and comparison beneficiary expenditures from 2001 through 2009 (PY5). We used the alternative regression coefficients (not shown) to predict expenditures for a beneficiary with 9-year average entire sample (combined intervention/comparison group) characteristics assigned alternatively to the intervention group or to the comparison group.²⁶ Figure 10-2 shows that the assigned beneficiary expenditures are higher than comparison beneficiary expenditures in 2001, and that the gap is roughly constant to the Demonstration base year of 2004. Beginning in 2005, the first Demonstration performance year, the gap between assigned and comparison beneficiary expenditures narrows. By 2007, assigned and comparison beneficiary expenditures are virtually equal. From 2007 to 2008, comparison beneficiary expenditures rise considerably more than assigned beneficiary expenditures, reversing the gap that existed in 2001. However, from 2008 to 2009, the final Demonstration performance year, comparison beneficiary expenditures decline more than assigned beneficiary expenditures. The Demonstration ends in 2009 (technically March 2010), then, with comparison beneficiary expenditures slightly higher than assigned beneficiary expenditures, which is a sharp contrast from 2001 in which the assigned beneficiary expenditures were markedly higher than the comparison beneficiary expenditures. In short, Figure 10-2 shows that the Demonstration effect gradually grew from 2005 (the first

²⁵ The PGP Demonstration financial reconciliation accounting model (see chapters 2 and 5) used concurrent risk scores to adjust for a beneficiary's casemix in a given year. We use the same concurrent risk scores in our multivariate analyses to adjust for a beneficiary's casemix in a given year (although we do not apply a risk score cap in PY5 as did the accounting model). Note that the concurrent risk adjustment model used to calculate the concurrent risk scores only includes the HCC disease groups from a standard prospective risk adjustment model (Olmsted, Pope, and Kautter, 2006), so in some sense the concurrent risk scores are "prospectivized" because they incorporate diagnostic information mainly from chronic conditions.

²⁶ The 9-year sample means were calculated for the intervention and comparison groups combined for all variables except for flag_ab, flag_post_ab, year dummies, and AB_year interactions. The latter variables were allowed to take on their group- and/or year-specific values, and were used to trace out predicted intervention and comparison group expenditures by year.

Demonstration performance year) through 2007, spiked upwards in 2008, then fell somewhat in 2009, the last Demonstration performance year.

Figure 10-2
Predicted Expenditures by Year, Assigned and Comparison Beneficiaries, 10 PGPs
Combined, Assigned Beneficiary Effects by Year



NOTE: Regression covariates set at their sample means.

SOURCE: RTI analysis of 2001-2010 Medicare administrative data.

Numerically, in 2001 assigned beneficiary expenditures as simulated in Figure 10-2 are 3.0 percent higher than simulated comparison beneficiary expenditures. In 2004, in the base year of the Demonstration, simulated assigned beneficiary expenditures are still 3.0 percent greater than simulated comparison group expenditures. However, by 2009 at the end of the Demonstration, simulated assigned beneficiary expenditures are 0.4 percent lower than simulated comparison beneficiary expenditures. The alternative regression estimates that assigned beneficiary expenditures are \$176 greater than comparison beneficiary expenditures in 2001 ($p < 0.01$). The change in this difference from 2002 to 2009 is: 2002: \$7 ($p = 0.84$); 2003: \$23 ($p = 0.51$); 2004: \$26 ($P = 0.482$); 2005: -\$33 ($P = 0.405$); 2006: -\$85 ($P = 0.044$); 2007: -\$165 ($P < 0.01$); 2008: -\$322 ($P < 0.01$); and 2009: -\$204 ($p < 0.01$). These trends show that the excess of assigned over comparison beneficiary expenditures is not statistically different from 2001 to 2004, begins to narrow with Demonstration implementation in 2005 (although the change is not statistically significant), narrows by a statistically significant amount in 2006, narrows more in 2007, reverses in 2008 (i.e., $-\$322 + \$176 < 0$), and is still reversed although less so at the end of the Demonstration in 2009 (i.e., $-\$204 + \$176 < 0$, meaning that simulated assigned beneficiary expenditures are \$28 less than simulated comparison beneficiary expenditures at the end of the Demonstration, which is statistically significant).

10.6 Possible Refinement to the County-Balancing Method

One element of the county-balancing approach is that the number of comparison beneficiaries is limited by the number of PGP assigned beneficiaries in a county. There may be additional suitable comparison beneficiaries in a county who could be added to the analysis as long as their propensity scores fell within an acceptable range. While the county balancing approach is unbiased and simple, adding comparison beneficiaries might increase the precision of the regression estimates. An alternative approach to county-balancing would be to create separate propensity models for each county in a PGP's service area that might identify additional comparisons. This alternative, however, is much more complex and involves considerably more analyses. Given the analytic time period covers 9 years and there are 117 counties across the 10 PGPs, it follows that 1,053 (9 x 117) county-level propensity models would need to be evaluated.

To explore the ramifications of the alternative approach, we applied it to estimate the overall savings across the 10 PGPs. For each county in each PGP service area in each year, beneficiaries with extreme estimated propensities (smaller than 0.05 or larger than 0.95) were removed. The alternative method resulted in a 47% increase in the PGP-wide analytic sample, from 3.6 million to 4.9 million. However, this increase only had a marginal impact on the magnitude and precision of the estimated PGP-wide effect on expenditures—the per beneficiary per year savings estimates for the alternative approach and the county-balancing approach were \$166 and \$171, respectively, and the corresponding standard errors were \$21.09 and \$22.03, respectively. The less parsimonious alternative method, which is more complex and involves considerably more analyses, delivered similar results as the county-balancing approach and therefore does not alter our findings and conclusions about the impact of the PGP demonstration.

10.7 Comparison to Previous Demonstration Effect Estimates

The overall Demonstration impact estimates reported in Table 10-3 for the multivariate methodology can be compared to the estimates of Demonstration impacts made in other chapters using accounting or descriptive analytical methods. These approaches exhibited imbalances between the intervention and comparison groups. They also did not adjust for beneficiary characteristics when measuring the impact of the Demonstration on outcomes. Thus a multivariate methodology was developed to address these limitations. Our best estimate of the impact of the Demonstration on expenditures is \$171 per assigned beneficiary per year. This is the estimate based on the multivariate methodology (Table 10-3). However, it is still useful to compare results across the various methodologies.

The accounting estimate of the Demonstration impact (Table 11-7, scenario 1: target minus actual expenditures counting amounts within the 2 percent corridor as savings or losses) is \$252 per assigned beneficiary person year in gross savings offset by \$46 in PGP-generated losses, for a savings impact of \$207.²⁷ This savings estimate is higher than the regression savings estimate of \$171 shown in Table 10-3. We also estimate the Demonstration effect (target minus actual expenditures) in Chapter 7, Table 7-3, utilizing a descriptive difference-in-

²⁷ This savings estimate reflects the PY5 risk score cap, which is not incorporated in the regression analysis presented in this chapter. Without the PY5 risk score cap, the Chapter 11 demonstration savings estimate would be higher.

difference analysis removing the effect of the pre-Demonstration expenditure trend. The 5-year average of the overall Demonstration savings in the third row of Table 7-3 is \$51 per beneficiary per year.²⁸ This savings estimate is smaller than the regression-based estimate.

We conclude that estimates of overall Demonstration savings using multiple methodologies are broadly consistent, and are in the range of \$50 to \$210 per assigned beneficiary per performance year when all ten sites are combined. This represents a small, but not negligible reduction in the level of medical expenditures. These savings estimates measure the annual reduction in Medicare fee-for-service payments to medical providers for the care of Demonstration assigned beneficiaries. They do not subtract Medicare’s performance payments to participating PGPs under the Demonstration but include savings or losses within the 2 percent corridor. For an analysis of performance payments and their effect in reducing estimated Medicare program savings, see Chapters 5 and 11.

10.8 Demonstration Effects on Medicare Expenditure Components

We performed analyses to determine if the impact of the Demonstration was greater on certain expenditure components. The six expenditure components we analyzed were inpatient hospital, skilled nursing facility (SNF), institutional (hospital) outpatient, Part B physician/supplier, home health, and durable medical equipment (DME). We also analyzed total inpatient, consisting of inpatient hospital plus SNF, and total outpatient, consisting of all other expenditure categories. The PGP Demonstration did not include hospice payments; hence the six payment components analyzed sum to total eligible Demonstration expenditures. We show overall total expenditure results from Table 10-3 (inpatient plus outpatient excluding hospice) for reference.

The method of our analysis was to rerun the repeated cross sections model described in Section 10.5, Table 10-3, replacing total eligible Medicare payments with each of the payment components in turn. Regressions were run for all PGPs combined, and for each of the 10 PGPs individually. Results are shown in **Table 10-4**. Table 10-4 shows only the Demonstration impact coefficient from each regression, the interaction of the Demonstration assigned beneficiaries and the Demonstration performance period (“post” period). For example, the per capita Demonstration impact for all PGPs, total expenditures, in the upper left hand corner of Table 10-4, is 171 dollars, the same value as shown in Table 10-3 and discussed in Section 10.5 as the overall Demonstration effect on expenditures.

²⁸ Replacing the \$162 estimated PY5 savings in the first row of Table 7-3 with the \$225 estimated savings without the risk score cap, to make the analysis in Chapter 7 comparable to the analysis in this chapter.

Table 10-4
Demonstration impacts on components of per capita expenditures (standard errors and p-values are below coefficients)

PGP	Total	Inpatient Total	Inpatient Inpatient	Inpatient SNF	Outpatient Total	Outpatient Institutional (Hospital)	Outpatient Part B Physician/Supplier	Outpatient Home Health	Outpatient Durable Medical Equipment
All	-171	-228	-176	-68	25	85	39	22	0
	22	18	16	8	12	7	7	3	3
	0.000	0.000	0.000	0.000	0.043	0.000	0.000	0.000	0.935
1	323	-24	-49	25	352	448	-78	-25	10
	79	58	51	22	51	34	34	11	12
	0.000	0.675	0.341	0.267	0.000	0.000	0.022	0.027	0.384
2	-188	-328	-227	-119	126	224	-106	1	9
	64	53	46	24	34	24	16	11	10
	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.935	0.347
3	-229	-216	-170	-73	-101	-144	56	12	-20
	94	78	69	31	53	28	35	16	16
	0.015	0.006	0.014	0.021	0.054	0.000	0.108	0.442	0.199
4	87	57	88	-37	31	39	-14	-1	4
	74	62	57	23	38	19	22	14	12
	0.244	0.353	0.122	0.100	0.414	0.044	0.513	0.951	0.710
5	-310	-400	-324	-87	42	237	-160	-38	2
	59	49	44	18	31	19	19	9	8
	0.000	0.000	0.000	0.000	0.184	0.000	0.000	0.000	0.748
6	-818	-56	-431	-170	-332	-440	157	-30	-22
	53	44	37	23	30	18	19	5	9
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012
7	-26	-84	-44	-54	35	-43	68	-3	12
	102	84	71	39	51	26	34	18	11
	0.798	0.319	0.536	0.173	0.500	0.098	0.044	0.889	0.244

(continued)

Table 10-4 (continued)
Demonstration impacts on components of per capita expenditures (standard errors and p-values are below coefficients)

PGP	Total	Inpatient Total	Inpatient Inpatient	Inpatient SNF	Outpatient Total	Outpatient Institutional (Hospital)	Outpatient Part B Physician/Supplier	Outpatient Home Health	Outpatient Durable Medical Equipment
8	-142	-188	-217	12	7	37	8	-11	-32
	69	57	51	22	38	22	25	9	12
	0.041	0.001	0.000	0.603	0.856	0.083	0.759	0.263	0.007
9	21	4	0	-1	5	100	-40	-24	-32
	49	38	34	15	28	17	16	8	8
	0.675	0.911	0.994	0.957	0.873	0.000	0.013	0.003	0.000
10	120	-187	-139	-48	289	494	-231	-85	106
	91	76	70	28	53	34	28	16	19
	0.191	0.013	0.048	0.084	0.000	0.000	0.000	0.000	0.000

NOTES:

1. Expenditure component regressions are estimated for assigned and comparison beneficiaries on 2001 to PY5 data using the same sample, methods, and specification as the regression for total expenditures (Table 10-3).

Coefficient estimates are in dollars.

Standard errors and p-values are shown below coefficient estimates.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

Across all 10 PGPs, Demonstration savings were achieved totally from the inpatient setting (savings = \$228). In fact, the estimated Demonstration impact on total outpatient expenditures indicates slight dis-savings (dis-savings = \$25), possibly indicating some degree of substitution of outpatient for inpatient services among the Demonstration PGPs. The finding of inpatient savings is consistent with the PGPs' cost-saving interventions, which focused on reducing avoidable inpatient admissions (see Chapter 4). Among the individual PGPs, 6 of 10 achieved significant inpatient savings, and none showed dis-savings. Of the 5 that achieved significant overall savings, 3 achieved all of their total savings through inpatient reductions, and 2 achieved total savings through both inpatient and outpatient reductions. Two PGPs—PGPs 3 and 6—realized statistically significant savings in total outpatient expenditures. In fact, 3 PGPs incurred statistically significant total outpatient dis-savings.

The PGPs can be described as follows in terms of the statistical significance of their cost control performance relative to their comparison groups:

- Two PGPs achieved lower total expenditures because of reduced inpatient expenditures, with flat outpatient expenditures;
- Three PGPs showed no Demonstration effect on inpatient, outpatient, or total expenditures;
- Two PGPs achieved lower total, inpatient, and outpatient expenditures;
- One PGP achieved a decrease in overall expenditures because of lower inpatient expenditures, despite an increase in outpatient expenditures;
- One PGP showed a negative impact on inpatient expenditures offset by a positive effect on outpatient expenditures, resulting in no net effect on overall expenditures; and
- One PGP incurred an increase in overall expenditures, with flat inpatient expenditures and higher outpatient expenditures.

In short, the PGPs had much greater success in controlling inpatient than outpatient expenditures.

Within the inpatient setting, inpatient hospital contributed about three-quarters of the total savings, SNF the other one-quarter. Among individual PGPs, inpatient and SNF savings were highly correlated: for the most part, PGPs that controlled inpatient costs also controlled SNF costs. Inpatient/SNF savings proportions were fairly similar to the overall savings proportions for most individual PGPs.

Outpatient cost components showed varying effects. Across all 10 PGPs, hospital outpatient costs rose more rapidly for Demonstration beneficiaries than their comparison group, by \$85 per capita. This could indicate a substitution of outpatient hospital services for inpatient services. Conversely, Part B physician/supplier and home health services were constrained for

assigned beneficiaries relative to comparison beneficiaries, by 39 and 22 dollars per capita, respectively. DME expenditures showed no overall Demonstration effect.

Several individual PGPs—PGPs 1, 2, 5, and 10—showed quite large and statistically significant increases in hospital outpatient expenditures relative to their comparison groups. Other PGPs—PGPs 6 and 3—showed large and statistically significant relative declines in hospital outpatient expenditures. The change in hospital outpatient and Part B physician expenditures is strongly negatively correlated across PGPs: when hospital outpatient expenditures rise, Part B physician expenditures fall, and vice versa. A contributing factor here may be that some participating PGPs substituted hospital outpatient for physician office services through the establishment of hospital-based “provider-based clinics,” which can be reimbursed by Medicare at a higher rate than office-based practices. Moreover, there may be substitution between ambulatory surgery centers, which are billed through Part B physician/supplier claims (except for hospital-based ambulatory surgery centers), and the hospital outpatient department (including hospital-based ambulatory surgery centers), which is billed through hospital outpatient claims.

Five of the 10 Demonstration PGPs achieved statistically significant savings on home health costs, and none show statistically significant dis-savings. Three PGPs achieved statistically significant reductions in DME costs, but one PGP showed a large, statistically significant increase.

10.9 Subgroup Analyses of Demonstration Effects

We also performed a series of analyses designed to determine whether the impact of the Demonstration was greater in selected beneficiary subgroups. A subgroup refers to beneficiaries who share a particular characteristic. Subgroups of interest included 7 diagnosis groups (cancer, CHF, diabetes, CAD²⁹, COPD, stroke, vascular disease), beneficiaries with one or more of the 7 diagnosis groups, ESRD, beneficiaries with any inpatient expenditures, top 10% of HCC risk scores, top 25% of HCC risk scores, and Medicaid enrollee. The percentages that each of these subgroups comprise of total assigned beneficiaries are shown in **Table 10-5** for all 10 PGPs combined and for each PGP. (The percentages are for all person years 2001 to PY5 and are weighted by the regression weight of the eligibility fraction multiplied by the propensity score weight). For all 10 PGPs, the size of the subgroups ranges from 0.8 percent of assigned beneficiaries (ESRD) to 51.1 percent (any of the 7 major conditions). Most of the subgroups comprise from 10 to 25 percent of assigned beneficiaries. The subgroup proportions by PGP are generally fairly similar, although PGP 10 stands out as having a sicker patient population.

²⁹ CAD is defined as acute myocardial infarction and stable and unstable angina. It does not include chronic ischemic heart disease.

Table 10-5
Subgroup percentages of assigned beneficiaries, total 2001-PY5, regression sample

Variable	All PGPs	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
Number of beneficiary/years in regression sample	1,776,387	109,643	237,601	81,631	125,670	228,417	302,897	115,739	160,423	266,895	147,471
Subgroup											
Cancer, %	12.8	12.9	13.3	11.9	10.9	13.4	12.3	14.2	11.6	12.6	15.1
CHF, %	12.6	11.5	11.4	11.0	12.8	14.3	12.6	13.5	10.1	13.2	14.7
Diabetes, %	22.1	19.1	21.1	21.1	25.5	26.5	21.2	20.4	18.2	22.1	24.2
COPD, %	13.2	13.1	13.1	11.5	14.9	15.7	10.9	13.7	8.5	15.9	13.9
Acute ischemic heart disease, %	3.6	2.2	3.5	3.3	3.5	4.8	2.9	3.5	2.6	4.1	5.5
Stroke, %	3.9	3.4	3.5	4.4	4.9	5.5	2.3	4.2	2.9	4.3	5.2
Vascular disease, %	13.2	10.7	11.7	12.3	11.8	20.0	11.0	15.6	9.8	12.8	15.3
Any of 7 above conditions, %	51.1	48.5	49.6	48.7	52.4	58.7	47.9	52.3	42.6	52.8	55.6
Medicaid, %	14.4	10.6	12.0	17.5	21.1	17.2	13.4	11.9	12.5	15.7	13.3
Originally disabled, %	6.3	6.5	5.8	5.8	7.1	7.1	6.1	4.8	3.6	7.9	6.5
ESRD, %	0.8	0.7	0.5	0.9	1.0	0.7	0.7	0.6	0.8	0.7	1.7
Disabled, %	15.0	11.2	14.8	16.1	17.2	16.4	12.1	10.7	15.3	17.6	18.2
Upper 10% risk score	9.9	8.6	8.9	10.6	9.9	11.1	8.7	10.3	9.2	9.7	13.5
Upper 25% risk score	25.6	23.3	24.4	25.4	24.4	28.9	23.9	26.3	22.7	26.0	30.8
Inpatient \$ > 0, %	21.2	22.3	18.6	19.1	21.4	21.4	21.0	19.9	21.6	22.2	24.1

NOTE: The risk score percentiles are defined for the assigned and comparison group beneficiaries combined, so the proportions of assigned beneficiaries are not exactly 10% and 25%. These percentages are weighted by the regression weight, which is the product of the eligibility fraction and the propensity score weight. CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ESRD = end stage renal disease; PGP = physician group practice.

SOURCE: RTI analysis of 2001-2010 Medicare administrative data.

If a subgroup has an important impact on expenditures, then the demonstration effect should be larger for members of the subgroup than for non-members. To estimate subgroup effects, four additional terms were added to the model specified in equation 10-1. These consisted of:

- a main effect for the subgroup,
- two-way interactions of the subgroup by assigned beneficiary status, and the subgroup by Demonstration period, and
- the three-way interaction of subgroup by assigned beneficiary status by Demonstration period.

If expenditure effects differ by subgroup, then the coefficient for the 3-way interaction should be significantly different from zero. For each subgroup, separate models were estimated for each PGP group and for all 10 PGPs combined, i.e., the subgroups were studied one by one, not simultaneously. These models were estimated on the full sample of assigned and comparison group beneficiaries ($n = 3,355,467$ for the combined 10 PGPs, the same sample as used for Table 10-3), not on the subsamples consisting of subgroup members only.

The results of the subgroup regression analyses are summarized in **Table 10-6**. Since most of the terms in the regression model are the same as those in the earlier model, the table displays only the coefficients for the overall demonstration effects and the subgroup-specific effects. The primary coefficients of interest are the “demo effect” (the coefficient of “post*AB”—the product of post-demonstration period and assigned beneficiary) and “demo subgroup effect” (the coefficient of “assigned beneficiary*post*subpopulation”). The former shows the demonstration effect on assigned beneficiaries who are not members of the subgroup under study in the particular regression. The latter shows the effect of demonstration on subgroup beneficiaries relative to the demonstration effect on non-subgroup beneficiaries.³⁰

³⁰ The total effect of the demonstration on the subgroup is the sum of the “demo effect” and “demo subgroup effect” coefficients.

Table 10-6

Demonstration impacts on assigned beneficiary subpopulations (standard errors and p-values are below coefficients)

Subgroups	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
1. Cancer subgroup	-170	125	-189	-254	105	-324	-831	78	-58	46	157
Post*AB (Demo effect)	21	75	63	92	74	58	51	100	68	48	91
	0.000	0.097	0.002	0.006	0.159	0.000	0.000	0.434	0.388	0.341	0.084
assigned beneficiary*post* subpopulation (demo subgroup effect)	-11	1,536	-36	216	-145	95	125	-731	-695	-184	-353
	92	347	262	415	319	241	241	394	310	206	332
	0.905	0.000	0.891	0.602	0.650	0.694	0.604	0.064	0.025	0.371	0.288
2. CHF subgroup	-97	241	18	-217	82	-205	-611	47	-86	10	86
Post*AB (Demo effect)	20	75	57	86	67	54	48	90	64	45	82
	0.000	0.001	0.750	0.012	0.221	0.000	0.000	0.602	0.183	0.823	0.294
assigned beneficiary*post* subpopulation (demo subgroup effect)	-590	678	-1,838	-167	39	-740	-1,602	-506	-546	82	289
	106	372	341	489	365	263	256	488	379	223	403
	0.000	0.069	0.000	0.733	0.915	0.005	0.000	0.300	0.149	0.714	0.474
3. Diabetes subgroup	-86	433	-53	-258	-10	-138	-587	-5	-168	69	179
Post*AB (Demo effect)	23	83	66	99	78	62	54	106	71	51	95
	0.000	0.000	0.421	0.009	0.894	0.027	0.000	0.964	0.017	0.178	0.061
assigned beneficiary*post* subpopulation (demo subgroup effect)	-371	-608	-648	225	369	-599	-1,010	-89	201	-205	-358
	62	234	189	270	198	154	155	295	220	138	248
	0.000	0.009	0.001	0.404	0.063	0.000	0.000	0.762	0.362	0.139	0.148
4. COPD subgroup	-119	317	-56	-151	105	-250	-704	-46	-120	72	124
Post*AB (Demo effect)	21	78	63	91	73	58	51	99	68	48	89
	0.000	0.000	0.375	0.098	0.148	0.000	0.000	0.643	0.077	0.133	0.163
assigned beneficiary*post* subpopulation (demo subgroup effect)	-403	39	-1,032	-678	-83	-384	-977	126	-280	-305	-119
	89	305	260	424	280	221	247	411	367	180	353
	0.000	0.897	0.000	0.109	0.767	0.082	0.000	0.759	0.446	0.090	0.743
5. Acute Ischemic Heart Disease¹ subgroup	-146	313	-176	-199	99	-272	-771	-7	-122	58	164
Post*AB (Demo effect)	21	78	61	91	72	57	51	99	67	48	87
	0.000	0.000	0.004	0.028	0.167	0.000	0.000	0.940	0.071	0.224	0.060
assigned beneficiary*post* subpopulation (demo subgroup effect)	-456	886	41	-1,108	-168	-818	-1,573	-844	-780	-606	478
	222	966	707	1,019	768	490	630	1,023	849	439	695
	0.040	0.359	0.954	0.277	0.827	0.095	0.012	0.409	0.358	0.168	0.492

(continued)

Table 10-6 (continued)
Demonstration impacts on assigned beneficiary subpopulations (standard errors and p-values are below coefficients)

Subgroups	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
6. Stroke subgroup	-146	291	-110	-173	69	-219	-804	-45	-151	15	172
Post*AB (Demo effect)	21	78	62	91	73	58	52	99	68	48	90
	0.000	0.000	0.075	0.058	0.341	0.000	0.000	0.650	0.026	0.749	0.055
assigned beneficiary*post* subpopulation (demo subgroup effect)	-629	985	-2,222	-1,316	362	-1,676	-456	271	529	124	-851
	191	674	637	784	566	422	618	873	774	389	647
	0.001	0.144	0.000	0.093	0.522	0.000	0.461	0.757	0.494	0.750	0.189
7. VASCD subgroup	-120	256	-91	-256	50	-232	-624	-18	-123	39	134
Post*AB(Demo effect)	21	76	59	89	69	56	50	94	66	47	86
	0.000	0.001	0.122	0.004	0.468	0.000	0.000	0.847	0.062	0.399	0.118
assigned beneficiary*post* subpopulation (demo subgroup effect)	-416	588	-873	353	326	-387	-1,657	18	-159	-154	-441
	97	374	318	433	365	199	260	408	357	213	363
	0.000	0.116	0.006	0.415	0.372	0.052	0.000	0.965	0.657	0.468	0.225
8. Any 7 subgroup	-14	105	94	-126	-9	-44	-325	40	-87	134	139
Post*AB(Demo effect)	18	68	51	77	61	49	40	80	58	41	76
	0.424	0.123	0.066	0.102	0.886	0.367	0.000	0.620	0.131	0.001	0.067
assigned beneficiary*post* subpopulation (demo subgroup effect)	-323	424	-620	-166	197	-458	-1,008	-102	-110	-226	-163
	43	160	129	190	143	106	109	197	151	94	171
	0.000	0.008	0.000	0.384	0.169	0.000	0.000	0.607	0.468	0.016	0.338
9. Medicaid subgroup	-186	294	-204	-227	7	-301	-786	-130	-104	-32	152
Post*AB(Demo effect)	23	81	67	98	78	64	56	104	71	52	95
	0.000	0.000	0.002	0.021	0.926	0.000	0.000	0.213	0.147	0.544	0.112
assigned beneficiary*post* subpopulation (demo subgroup effect)	96	279	34	-29	382	-40	-248	827	-323	313	-351
	71	306	221	288	212	172	172	376	252	152	310
	0.176	0.361	0.878	0.921	0.071	0.814	0.151	0.028	0.201	0.039	0.257
10. Originally disabled subgroup	-158	355	-176	-193	73	-275	-798	-76	-120	30	152
Post*AB(Demo effect)	22	81	65	95	76	61	54	103	70	50	93
	0.000	0.000	0.007	0.043	0.336	0.000	0.000	0.460	0.086	0.545	0.103
assigned beneficiary*post* subpopulation (demo subgroup effect)	-203	-494	-227	-639	195	-495	-319	1,047	-728	-122	-549
	110	397	341	482	342	274	275	565	509	214	449
	0.065	0.214	0.507	0.185	0.569	0.071	0.247	0.064	0.152	0.567	0.221

(continued)

Table 10-6 (continued)
Demonstration impacts on assigned beneficiary subpopulations (standard errors and p-values are below coefficients)

Subgroups	All	PGP 1	PGP 2	PGP 3	PGP 4	PGP 5	PGP 6	PGP 7	PGP 8	PGP 9	PGP 10
11. ESRD subgroup	-177	329	-185	-223	38	-304	-825	-41	-182	3	176
Post*AB(Demo effect)	22	78	63	92	72	58	52	101	68	48	88
	0.000	0.000	0.003	0.016	0.597	0.000	0.000	0.683	0.008	0.956	0.046
assigned beneficiary*post* subpopulation (demo subgroup effect)	674	-1,406	-690	-1,522	4,486	-107	-192	2,505	4,870	2,187	-2,950
	0.310	0.602	0.777	0.556	0.028	0.956	0.916	0.384	0.009	0.177	0.100
12. Disabled subgroup	-220	287	-235	-181	-48	-384	-810	-96	-150	-22	24
Post*AB(Demo effect)	23	82	68	99	79	63	55	105	72	52	95
	0.000	0.000	0.001	0.068	0.545	0.000	0.000	0.361	0.037	0.677	0.797
assigned beneficiary*post* subpopulation (demo subgroup effect)	286	270	310	-380	730	483	-128	526	85	196	560
	0.000	0.380	0.126	0.190	0.001	0.006	0.505	0.169	0.719	0.170	0.050
13. Inpatient>0 subgroup	-60	211	35	-215	-12	-153	-349	39	-56	-57	233
Post*AB(Demo effect)	12	48	33	51	38	31	28	49	37	27	49
	0.000	0.000	0.289	0.000	0.751	0.000	0.000	0.422	0.123	0.035	0.000
assigned beneficiary*post* subpopulation (demo subgroup effect)	-342	270	-554	66	437	-364	-1,384	-150	-80	-26	-863
	0.000	0.311	0.038	0.865	0.119	0.090	0.000	0.708	0.755	0.884	0.005
14. Upper 10% subgroup	13	230	68	-54	23	103	-400	4	-12	52	296
Post*AB(Demo effect)	18	67	52	76	60	46	41	85	57	41	72
	0.479	0.001	0.188	0.475	0.706	0.025	0.000	0.963	0.835	0.202	0.000
assigned beneficiary*post* subpopulation (demo subgroup effect)	-1,934	207	-2,293	-1,708	1,153	-3,580	-4,011	-1,321	-1,805	-404	-2,330
	0.000	0.769	0.000	0.035	0.077	0.000	0.000	0.117	0.005	0.334	0.000
15. Upper 25% subgroup	63	124	131	-11	17	171	-205	64	24	77	264
Post*AB(Demo effect)	14	54	40	61	47	37	33	65	45	33	58
	0.000	0.021	0.001	0.858	0.718	0.000	0.000	0.321	0.589	0.020	0.000
assigned beneficiary*post* subpopulation (demo subgroup effect)	-1,316	814	-1,635	-1,235	542	-2,650	-3,067	-464	-708	-397	-1,179
	0.000	0.024	0.000	0.005	0.111	0.000	0.000	0.323	0.042	0.064	0.001

NOTE:

¹Acute ischemic heart disease is defined as acute myocardial infarction and unstable angina and other acute ischemic heart disease. It does not include stable angina or other chronic ischemic heart disease.

SOURCE: RTI analysis of 2001–2010 Medicare administrative data.

The analyses show that overall per person year savings are greatest primarily for those diagnosed with cardiovascular conditions (CHF, stroke, and vascular disease) as well as with diabetes and COPD. The overall impact for these conditions was driven largely by the findings in only three sites—PGPs 2, 5, and 6. Similar savings effects were observed for those who had any of the 7 chronic conditions compared to those who had none. Once the Demonstration effect on several other subgroups was accounted for, the Demonstration effect on the remaining (i.e., non-subgroup) beneficiaries was greatly attenuated. For example, the demonstration effect for CHF beneficiaries was $-\$97 + (-\$590) = -\$687$, and for non-CHF assigned beneficiaries was only $-\$97$. This compares to an all-sample Demonstration effect of $-\$171$. In other words, the Demonstration impact on non-CHF assigned beneficiaries was only $-\$97$ rather than the overall $-\$171$ effect. Other subgroups which accounted for a large portion of the overall Demonstration effect were:

- inpatient $\$ > 0$ (effect = $-\$60 + (-\$342) = -\$402$), inpatient $\$ = 0$ (effect = $-\$60$)
- diabetes (effect = $-\$86 + (-\$371) = -\$457$), non-diabetes (effect = $-\$86$)
- COPD (effect = $-\$119 + (-\$403) = -\$522$), non-COPD (effect = $-\$119$)
- vascular disease (effect = $-\$120 + (-\$416) = -\$536$), non-VD (effect = $-\$120$)
- Upper 10% risk score (effect = $\$13 + (-\$1,934) = -\$1,921$), non-upper 10% (effect = $\$13$)
- Upper 25% risk score (effect = $\$63 + (-\$1,316) = -\$1,253$), non-upper 25% (effect = $\$63$).

Especially striking here is that beneficiaries with risk scores in the upper 10 percent or upper 25 percent. These findings indicate that all Demonstration savings were generated among chronically ill beneficiaries and beneficiaries with high expected expenditures, and that the majority of overall savings (about two-thirds) were generated among beneficiaries who were hospitalized at least once during a year. Not surprisingly given the PGP's interventions, Demonstration savings were achieved among sick, high-cost beneficiaries.

There were no statistically significant effects for cancer, ESRD (which has a relatively small sample size), or Medicaid patients. Expenditures were somewhat higher under the Demonstration for assigned beneficiaries who were disabled. All of the Demonstration savings were achieved among the elderly. Many of the disease management programs implemented by the PGPs focused on conditions of the elderly, and this emphasis may account for the more favorable results for cardiovascular conditions and for the higher costs for the generally younger disabled subgroup. The originally disabled (disabled when non-elderly but who are now elderly), however, did show a Demonstration savings impact.

10.10 Quality Outcomes

Similar to the financial outcomes modeling, we conducted multivariate analysis on a number of quality indicators to ascertain any Demonstration effect, as well as to identify patterns and elements that influenced the quality of care patients received. We ran logistic regressions on each of the seven claims-based measures for which we had data from both ‘treatment group’ (i.e., PGP assigned beneficiaries) and ‘control group’ (i.e., comparison group beneficiaries). As in the financial models, the control group beneficiaries were matched by propensity scoring to the PGP assigned beneficiaries, and outliers (e.g., clear mismatches) from the propensity modeling were dropped from analysis.

For each measure, we included all available observations in the topic’s denominator population from the 4-year pre-demonstration period (2001 to 2004) through the 5-year demonstration period (PY1-PY5). The dependent variables were the 0/1 flags (0=no; 1=yes) indicating whether the beneficiary received the given quality indicator of interest (e.g., DM-7 foot exam). Independent variables included a 0/1 flag for whether the beneficiary was assigned to a PGP, a flag for whether the information came from the Demonstration performance period, the interaction effect of the two variables representing the effect from the Demonstration (i.e., “post*AB), annual flags covering each year in the pre-Demonstration and Demonstration periods (reference group = 2001), the beneficiary’s risk score (in each year if the beneficiary participated in multiple years), gender (reference group = female), age categories (reference group = 65-74 years old), Medicaid (dual eligible) status, race (reference group = white), selected Hierarchical Condition Category diagnostic groups, and whether the beneficiary was entitled by disability.

Results from our logistic models, pooling across all PGPs, are presented in *Table 10-7*. Models for the seven claims-based quality indicators are presented in the columns, and the independent variables, or predictors, are presented in the rows. The table contains odds ratios (OR), standard errors and indications of statistical significance, represented by p-values.

Table 10-7
Demonstration impact on the seven claims-based quality indicators (Logistic regression results)
(standard errors and p-values are below the odds ratio coefficients)

Variable	DM-1 (obs=) 370,779	DM-4 (obs=) 370,739	DM-6 (obs=) 370,861	DM-7 (obs=) 245,904	CAD-5 (obs=) 514,340	HF-2 (obs=) 32,837	PC-5 (obs=) 254,786
Assigned beneficiary (0=no; 1=yes)	1.58	1.16	1.29	1.35	1.00	1.07	1.26
	0.02	0.01	0.01	0.02	0.01	0.05	0.02
	0.000	0.000	0.000	0.000	0.723	0.164	0.000
Post-demonstration period (0=no; 1=yes)	1.35	2.00	1.31	1.13	1.49	1.70	0.89
	0.03	0.02	0.02	0.02	0.02	0.08	0.02
	0.000	0.000	0.00	0.000	0.000	0.000	0.000
post*AB (Demo effect)	1.11	1.24	1.40	1.05	1.19	1.05	1.16
	0.03	0.02	0.02	0.02	0.02	0.07	0.02
	0.000	0.000	0.000	0.039	0.000	0.530	0.000
Year=2002 (0=no; 1=yes)	1.10	1.29	1.13		1.18	1.30	
	0.02	0.02	0.02	—	0.01	0.07	—
	0.000	0.000	0.000		0.000	0.000	
Year=2003 (0=no; 1=yes)	1.23	1.53	1.20	1.09	1.31	1.43	0.98
	0.03	0.02	0.02	0.02	0.01	0.07	0.02
	0.000	0.000	0.000	0.000	0.000	0.000	0.253
Year=2004 (0=no; 1=yes)	1.29	1.82	1.25	1.07	1.44	1.59	0.90
	0.02	0.02	0.02	0.02	0.01	0.07	0.02
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year=2006 (0=no; 1=yes)	1.05	1.12	1.21	1.03	1.13	1.18	1.04
	0.02	0.02	0.02	0.02	0.01	0.08	0.02
	0.042	0.000	0.000	0.048	0.000	0.038	0.016
Year=2007 (0=no; 1=yes)	1.15	1.21	1.26	1.00	1.23	1.28	1.10
	0.02	0.02	0.02	0.02	0.01	0.08	0.02
	0.000	0.000	0.000	0.896	0.000	0.003	0.000
Year=2008 (0=no; 1=yes)	1.06	1.25	1.33	1.04	1.24	1.16	1.13
	0.02	0.02	0.02	0.02	0.02	0.09	0.02
	0.013	0.000	0.000	0.032	0.000	0.085	0.000

(continued)

Table 10-7 (continued)
Demonstration impact on the seven claims-based quality indicators (Logistic regression results)
(standard errors and p-values are below the odds ratio coefficients)

Variable	DM-1 (obs=) 370,779	DM-4 (obs=) 370,739	DM-6 (obs=) 370,861	DM-7 (obs=) 245,904	CAD-5 (obs=) 514,340	HF-2 (obs=) 32,837	PC-5 (obs=) 254,786
Year=2009 (0=no; 1=yes)	0.99	1.25	1.43	1.08	1.28	1.36	1.19
	0.02	0.02	0.02	0.02	0.02	0.09	0.02
risk score	0.815	0.000	0.000	0.000	0.000	0.000	0.000
	0.90	0.82	1.38	1.00	0.84	1.13	0.90
	0.00	0.00	0.01	0.00	0.00	0.01	0.01
male (0=no; 1=yes)	0.000	0.000	0.000	0.908	0.000	0.000	0.000
	0.87	0.93	0.95	0.76	1.04	1.00	—
	0.01	0.01	0.01	0.01	0.01	0.04	—
age group (0-54)	0.000	0.000	0.000	0.000	0.000	0.997	—
	0.42	0.50	0.85	0.48	0.56	0.59	0.50
	0.02	0.02	0.02	0.02	0.03	0.13	0.02
age group (55-64)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.56	0.63	0.84	0.57	0.63	0.71	0.53
	0.02	0.02	0.01	0.02	0.02	0.10	0.02
age group (75-84)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.91	0.83	0.97	1.12	0.66	0.71	—
	0.02	0.02	0.01	0.02	0.01	0.06	—
age group (>85)	0.000	0.000	0.041	0.000	0.000	0.000	—
	—	—	—	—	0.29	0.41	—
	—	—	—	—	0.01	0.06	—
	—	—	—	—	0.000	0.000	—
Medicaid status (0=no; 1=yes)	1.24	0.95	1.02	0.93	0.83	0.85	0.74
	0.02	0.01	0.01	0.02	0.01	0.05	0.02
	0.000	0.001	0.085	0.000	0.000	0.002	0.000
Originally disabled (0=no; 1=yes)	0.84	0.75	0.93	0.83	0.79	0.75	0.60
	0.02	0.02	0.02	0.02	0.01	0.06	0.02
	0.000	0.000	0.000	0.000	0.000	0.000	0.000

(continued)

Table 10-7 (continued)
Demonstration impact on the seven claims-based quality indicators (Logistic regression results)
(standard errors and p-values are below the odds ratio coefficients)

Variable	DM-1 (obs=) 370,779	DM-4 (obs=) 370,739	DM-6 (obs=) 370,861	DM-7 (obs=) 245,904	CAD-5 (obs=) 514,340	HF-2 (obs=) 32,837	PC-5 (obs=) 254,786
ESRD status (0=no; 1=yes)	—	—	—	—	0.76 0.03 0.000	0.44 0.12 0.000	1.24 0.08 0.006
race = black	0.87 0.03 0.000	0.70 0.02 0.000	1.12 0.02 0.000	0.93 0.03 0.023	0.77 0.03 0.000	0.99 0.12 0.927	1.09 0.04 0.015
race = Asian	0.96 0.10 0.648	0.83 0.07 0.006	1.03 0.07 0.677	0.73 0.08 0.000	0.97 0.08 0.652	0.54 0.44 0.162	0.54 0.10 0.000
race = other races	0.77 0.05 0.000	0.74 0.04 0.000	1.04 0.04 0.301	0.83 0.04 0.000	0.75 0.04 0.000	1.23 0.23 0.372	0.75 0.05 0.000
CMS Hierarchical Condition Category (HCC) = cancer	0.85 0.02 0.000	0.93 0.02 0.000	1.11 0.02 0.000	1.01 0.02 0.724	0.97 0.01 0.011	1.02 0.06 0.762	1.70 0.02 0.000
HCC = diabetes	—	—	—	—	1.60 0.01 0.000	0.86 0.04 0.000	0.86 0.02 0.000
HCC = AMI	0.91 0.03 0.000	1.69 0.02 0.000	0.74 0.02 0.000	0.98 0.02 0.346	1.22 0.01 0.000	2.15 0.06 0.000	1.05 0.04 0.152
HCC= CHF	0.97 0.02 0.121	0.98 0.01 0.236	0.91 0.01 0.000	1.01 0.02 0.496	0.82 0.01 0.000	— — —	0.79 0.02 0.000
HCC = Stroke	0.89 0.03 0.000	1.01 0.02 0.591	0.79 0.02 0.000	0.99 0.03 0.691	0.96 0.01 0.008	1.04 0.07 0.580	0.82 0.04 0.000

(continued)

Table 10-7 (continued)
Demonstration impact on the seven claims-based quality indicators (Logistic regression results)
(standard errors and p-values are below the odds ratio coefficients)

Variable	DM-1 (obs=) 370,779	DM-4 (obs=) 370,739	DM-6 (obs=) 370,861	DM-7 (obs=) 245,904	CAD-5 (obs=) 514,340	HF-2 (obs=) 32,837	PC-5 (obs=) 254,786
HCC = Vascular disease	1.01 0.02 0.482	1.14 0.01 0.000	0.99 0.01 0.283	1.11 0.02 0.000	1.08 0.01 0.000	1.08 0.04 0.060	0.99 0.02 0.651
HCC = COPD	0.76 0.02 0.000	0.90 0.01 0.000	0.83 0.01 0.000	0.86 0.02 0.000	0.82 0.01 0.000	0.90 0.04 0.005	0.71 0.02 0.000
constant	7.44 0.07 0.000	2.83 0.06 0.000	0.80 0.05 0.000	2.02 0.07 0.000	3.11 0.05 0.000	1.93 0.19 0.000	2.46 0.06 0.000
r-squared	0.052	0.063	0.057	0.048	0.082	0.066	0.063

NOTES:

1. Dependent variable for logistic regression is binary indicator for achieving quality measure.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

As shown by the first two independent variables, the PGP assigned beneficiaries had significantly higher quality of care (i.e., received the recommended care) compared to their comparison group in five of the seven quality indicators, and all PGP and CG beneficiaries tended to have a higher level of care after the 2004 base year, except for mammography screening. However, our Demonstration effect indicator (i.e., the interaction term for assigned beneficiary and Demonstration performance period) clearly shows that the Demonstration, and not trends alone, had a positive impact on the quality of care patients received in six of the seven indicators, and in all four of the diabetes indicators in particular. After the PGP Demonstration began, compared to their CG counterparts, beneficiaries with diabetes in the PGPs had 11 percent higher odds of receiving an HbA1c test (DM-1), 24 percent higher odds of receiving an LDL-C test (DM-4), 40 percent higher odds of receiving medical attention for nephropathy (DM-6), and 5 percent higher odds of being checked for eye problems (DM-7). PGP beneficiaries with coronary artery disease had 19 percent higher odds of receiving a complete lipid panel compared to the comparison group. For breast cancer screening among women, PGP beneficiaries had 16 percent higher odds of receiving a mammogram after the Demonstration took effect. No statistically significant Demonstration effect was found for left ventricular function test among hospitalized heart failure patients (HF-2).

We also ran logistic regressions for each measure by PGP. We found that five of the PGPs had a significant positive demonstration effect in DM-1: HbA1c testing (i.e., that they were more likely to provide the recommended care to patients than their comparison group). Compared to their CGs, seven PGPs provided better care in DM-4: LDL-C testing; nine PGPs performed better in CAD-5: Lipid profile; and five PGPs performed better in PC-5: breast cancer screening. For DM-6: urine protein testing, DM-7: eye exam, and HF-2: LVF testing, the demonstration effect was mixed across the PGPs, but more PGPs showed a positive demonstration effect than those with a negative demonstration effect. Measures with the most improvement due to the demonstration were DM-4: LDL-C testing and CAD-5: lipid profile, where seven and nine PGPs, respectively, were more likely to perform these recommended processes of care than their comparison groups during the Demonstration performance period.

10.11 Demonstration Impacts on Utilization of Medical Services

In this section we conduct a multivariate analysis of the impact of the PGP Demonstration on utilization outcomes. The multivariate analyses allow us to disentangle the effect of the Demonstration from other factors that affect utilization, in particular, beneficiary health status. Beneficiary health status is not accounted for in the descriptive analysis of utilization in Chapter 9, but is controlled for by the risk score in the multivariate analysis. The descriptive analyses were performed on the entire populations of assigned beneficiaries and comparison beneficiaries. The multivariate utilization analyses in this section were performed on the same assigned beneficiary and comparison populations as used in other multivariate analyses in this chapter. The same explanatory (right-hand side) variables were used as in the multivariate expenditures analyses.

Models. Utilization variables are typically counts of visits, discharges, and services. Counts are inherently integer values and are characterized by non-negative values with high frequencies concentrated on a few low discrete values (e.g., 0, 1, 2). When performing ordinary least squares (OLS) regressions on count data, the parameter estimates can be “inefficient, inconsistent, and biased” (Long and Freese, 2006) with heteroskedastic errors. Consequently,

nonlinear regression models (e.g., Poisson and the negative binomial) are typically employed in multivariate analyses of count data.

Although single-equation Poisson and negative binomial models can account for a significant share of observations with a zero value of the dependent variable, these models may not be adequate when a *large* share of a population (e.g., Medicare beneficiaries) does not utilize a service. As was seen in Chapter 9, such is the case of hospitalizations (Table 9-1) and emergency department (ED) visits (Table 9-2) for both the assigned beneficiaries and the comparison groups. That is, a large share of the populations were non-users.

When a large share of population does not utilize a service in a given time period, it is possible that a multi-step process may be involved in determining utilization. The first step divides a population into non-users and users while the second only affects users. The first step can be due to either initial decisions or to initial events. In the case of elective surgery, the decision to have surgery is made by the patient together with the patient's physician (surgeon). The patient and physician need not have equal input into the initial decision. However, relative to the patient it is likely the physician has an increasingly important role in the decisions for subsequent hospitalizations. This could be due to the patient's deteriorating health or decreasing ability to provide input into the decision-making process.

In such cases, a modified count model such as the "hurdle" model is an appropriate approach for multivariate analyses of utilization. A hurdle model, also known as a two-part model, consists of two equations. The first equation is typically a logit or probit regression on all observations (both non-users and users) and serves to estimate the impact of a predictor on which members of the population were non-users and which were users. The second equation is typically a zero-truncated Poisson or negative binomial regression and is estimated on only the population of users. Hurdle models are often used in multivariate analyses of utilization in health services research. The hurdle model is also attractive because it allows a decomposition of utilization effects in the use/non-use versus amount of use conditional on use dimensions.

In the PGP Demonstration, participating PGPs have an incentive to reduce hospitalizations and ED visits. For some conditions, they can do this through medical management to avoid hospitalization. ED visits can be reduced by better access to office visits for assigned beneficiaries.³¹ The extent participating PGPs are able to reduce hospitalizations and ED visits might depend on whether they are able to avoid them altogether or, once a patient has utilization, to reduce the subsequent number. For example, some PGPs identified high-cost beneficiaries for medical management in part through the occurrence of hospitalization. The avoidance of any utilization can be addressed through the first hurdle equation while reduction of utilization subsequent to an initial event can be addressed through the second hurdle equation.

³¹ We do not test these possibilities directly in the regressions.

Results. *Table 10-8* shows both the logit and zero-truncated negative binomial (NBZT) parts of the estimated hurdle models for hospital admissions and ED visits. The estimates in shown in Table 10-8 are based on the pooled set of PGP Demonstration sites.

The negative signs for the demonstration effect variable (post*AB) in the logit regressions indicate that the assigned beneficiaries had, relative to the comparison population, lower likelihoods of any admission or ED visit during the Demonstration performance period. The negative sign on the demonstration effect in the NBZT regression on admissions indicates that assigned beneficiaries with at least one admission had, relative to the comparison population, fewer admissions. These effects are both statistically significant, that is, the Demonstration PGPs as a group reduced both the probability of an admission, and the number of admissions (conditional on at least one), by a statistically significant amount.

Not all of the ten PGPs had reduced admissions (not shown in table). Four of the PGPs (2, 5, 6, and 8) had negative and statistically significant demonstration effects on the probability of an admission (logit coefficients). But only PGPs 2 and 6 had negative and statistically significant demonstration effects on the number of admissions conditional on at least one admission (NBZT coefficients). PGP 1 had a positive and statically significant demonstration effect on the probability of admission (logit coefficient) and PGP 8 had a positive and statistically significant demonstration effect on the number of admissions conditional on at least one (NBZT coefficient).

In the ED visit results (not shown in table), PGPs 2, 5, 6, 7, 9, and 10 had negative and statistically significant demonstration effect logit coefficients. But only PGPs 6 and 7 had negative and statistically significant demonstration effect NBZT coefficients. PGPs 3 and 4 had a positive and statistically significant demonstration effect logit coefficient.

Unlike the expenditure regressions, estimates of the magnitude of the demonstration effects cannot be taken directly from the coefficients for the demonstration interaction terms because the logit and NBZT are nonlinear regression techniques. The standard method to derive numerical estimates involves simulations. In these simulations, four dependent variables are estimated for each observation in the sample. Aside from the demonstration status (EC) and the pre/post (T) variables, actual values for all of the other explanatory variables are used. Since there are 2 values each for EC and T, four separate estimates of the dependent variable are calculated as follows:

1. For each observation i , a simulated dependent variable (\hat{Y}) is calculated as if the observation is for an assigned beneficiary in the pre period by setting EC to one and T to zero—see Cell 1 in *Figure 10-3*.
2. For each observation i , \hat{Y} is calculated as if the observation is for an assigned beneficiary in the post period by setting EC to one and T to one—see Cell 2 in *Figure 10-3*.

Table 10-8
Demonstration impact on hospital admissions and emergency department visits, All PGPs
(Hurdle regression results)

All PGPs	Admissions	Admissions	Emergency	Emergency
	Logit	Negative binomial	department visits	department visits
			Logit	Negative binomial
Assigned beneficiary	0.0210	0.0199	0.0191	0.0469
	0.0055	0.0059	0.0046	0.0080
	0.000	0.001	0.000	0.000
Post*AB(Demo effect)	-0.0473	-0.0138	-0.0362	-0.0159
	0.0074	0.0080	0.0058	0.0099
	0.000	0.084	0.000	0.108
2002	0.0105	0.0183	0.0516	0.0351
	0.0072	0.0078	0.0058	0.0114
	0.147	0.019	0.000	0.002
2003	-0.0098	0.0062	0.0497	0.0317
	0.0073	0.0079	0.0058	0.0115
	0.176	0.431	0.000	0.006
2004	-0.0283	0.0036	0.0584	0.0290
	0.0073	0.0078	0.0057	0.0121
	0.000	0.647	0.000	0.016
Post-demonstration period	-0.0297	-0.0417	0.1151	0.0796
	0.0083	0.0090	0.0065	0.0121
	0.000	0.000	0.000	0.000
PY 2	-0.0433	-0.0180	0.0032	0.0062
	0.0073	0.0077	0.0055	0.0081
	0.000	0.019	0.567	0.442
PY 3	-0.1278	-0.0556	0.0004	0.0135
	0.0075	0.0079	0.0056	0.0081
	0.000	0.000	0.937	0.095
PY 4	-0.2055	-0.0878	-0.0167	0.0116
	0.0078	0.0082	0.0057	0.0082
	0.000	0.000	0.004	0.158
PY 5	-0.3384	-0.1490	-0.0880	-0.0298
	0.0080	0.0086	0.0058	0.0083
	0.000	0.000	0.000	0.000
Risk score	1.3160	0.2549	0.7992	0.2528
	0.0032	0.0012	0.0027	0.0017
	0.000	0.000	0.000	0.000
Male (0=no; 1=yes)	-0.0866	-0.0220	-0.1156	-0.0430
	0.0041	0.0044	0.0035	0.0057
	0.000	0.000	0.000	0.000

(continued)

Table 10-8 (continued)
Results from Hurdle regression models of hospital admissions and emergency department visits, All PGPs

All PGPs	Admissions Logit	Admissions Negative binomial	Emergency department visits Logit	Emergency department visits Negative binomial
Age group (0-54)	0.3297 0.0085 0.000	0.6271 0.0107 0.000	0.8182 0.0068 0.000	1.2204 0.0116 0.000
Age group (55-64)	0.0766 0.0089 0.000	0.2377 0.0092 0.000	0.3405 0.0070 0.000	0.5297 0.0145 0.000
Age group (75-84)	0.1457 0.0047 0.000	0.0468 0.0051 0.000	0.2607 0.0038 0.000	0.1381 0.0065 0.000
Age group (>85)	0.3982 0.0061 0.000	0.0931 0.0061 0.000	0.6821 0.0051 0.000	0.3163 0.0067 0.000
Medicaid status (0=no; 1=yes)	0.0144 0.0064 0.025	0.0716 0.0061 0.000	0.3794 0.0053 0.000	0.3608 0.0076 0.000
Originally disabled (0=no; 1=yes)	-0.0002 0.0084 0.983	0.0794 0.0070 0.000	0.2274 0.0071 0.000	0.2313 0.0096 0.000
Race=Black	0.0437 0.0134 0.001	0.0439 0.0137 0.001	0.2809 0.0110 0.000	0.1732 0.0143 0.000
Race=Asian	-0.1298 0.0322 0.000	-0.0819 0.0374 0.028	-0.2846 0.0260 0.000	-0.2069 0.0383 0.000
Race=Other races	0.0133 0.0187 0.478	0.0020 0.0216 0.926	0.0648 0.0157 0.000	0.1282 0.0230 0.000
ESRD status (0=no; 1=yes)	-4.3856 0.0345 0.000	-0.7757 0.0133 0.000	-2.5768 0.0265 0.000	-0.7933 0.0179 0.000
CMS Hierarchical Condition Category (HCC) = cancer	-0.2144 0.0064 0.000	0.0397 0.0050 0.000	-0.2852 0.0053 0.000	-0.0379 0.0066 0.000
HCC =AMI	1.2524 0.0114 0.000	0.0698 0.0052 0.000	0.7727 0.0100 0.000	0.0553 0.0068 0.000

(continued)

Table 10-8 (continued)
Results from Hurdle regression models of hospital admissions and emergency department visits, All PGPs

All PGPs	Admissions Logit	Admissions Negative binomial	Emergency department visits Logit	Emergency department visits Negative binomial
HCC = Diabetes	-0.0150 0.0048 0.002	0.0707 0.0046 0.000	0.0358 0.0040 0.000	0.0645 0.0064 0.000
HCC= CHF	0.3855 0.0057 0.000	0.3035 0.0045 0.000	0.3247 0.0052 0.000	0.2341 0.0056 0.000
HCC = Stroke	0.5648 0.0093 0.000	0.1527 0.0054 0.000	0.6845 0.0083 0.000	0.1457 0.0065 0.000
HCC = Vascular disease	0.0986 0.0057 0.000	0.1185 0.0043 0.000	0.0545 0.0049 0.000	0.0776 0.0054 0.000
HCC = COPD	0.4072 0.0055 0.000	0.2104 0.0045 0.000	0.3904 0.0048 0.000	0.2794 0.0056 0.000

NOTES:

Standard errors are shown below regression coefficient estimates. The hurdle model used in Stata does not produce pseudo R squared or other goodness of fit measures for the individual logit and zero-truncated negative binomial regressions.

This regression includes county variables not shown in this table.

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

Figure 10-3
Components for Difference-in-Differences Calculations

Group	Time Period		Changes for each group (post minus pre)
	T = 0 (pre)	T = 1 (post)	
Demo Participant (EC = 1)	1 $\hat{Y}_{i,EC=1,T=0}$	2 $\hat{Y}_{i,EC=1,T=1}$	$\Delta PP_{i,EC=1} =$ Cell 2 minus Cell 1
Comparison (EC = 0)	3 $\hat{Y}_{i,EC=0,T=0}$	4 $\hat{Y}_{i,EC=0,T=1}$	$\Delta PP_{i,EC=0} =$ Cell 4 minus Cell 3

NOTES:

- \hat{Y} represents either \hat{P} from the logit regressions or \hat{U} from the zero-truncated negative binomial regressions.
- EC denotes the dummy variable used to distinguish between demonstration participants and the comparison population while T denotes the dummy variable used to distinguish between the pre and post periods.
- The numbers in the shaded boxes are cell numbers.
 - For each observation i , \hat{Y} is calculated as if the observation is for a comparison beneficiary in the pre period by setting EC to zero and T to zero—see Cell 3 in Figure 10-3.
 - For each observation i , \hat{Y} is calculated as if the observation is for a comparison beneficiary in the post period by setting EC to zero and T to one—see Cell 4 in Figure 10-3.

For each observation, pre/post changes (ΔPP) are calculated as if the observation were for an assigned beneficiary ($\Delta PP_{i,EC=1}$) and as if the observation were for a comparison beneficiary ($\Delta PP_{i,EC=0}$). The demonstration difference-in-differences effect for each beneficiary is calculated by subtracting the comparison pre/post change from assigned beneficiary pre/post change:

$$2D_i = \Delta PP_{i,EC=1} - \Delta PP_{i,EC=0} \quad 10-2$$

The average demonstration effect is then estimated by calculating the mean of the individual observation demonstration effects.

The above simulations were performed separately for the logit and NBZT regressions. To obtain the combined hurdle effect, \hat{P} and \hat{U} were simulated as above for each observation and then multiplied together—**Figure 10-4** illustrates this. In particular, all observations in the sample are used in estimating the \hat{P} and \hat{U} . The same formula as above was used for calculating the pre/post for each observation and then the mean was found for the average demonstration effect.

Figure 10-4
Components for Difference-in-Differences for the Combined Hurdle Calculations

Group	Time Period		Changes for each group (post minus pre)
	T = 0 (pre)	T = 1 (post)	
Demo Participant (EC = 1)	1 $\hat{P}_{i,EC=1,T=0} \cdot \hat{U}_{i,EC=1,T=0}$	2 $\hat{P}_{i,EC=1,T=1} \cdot \hat{U}_{i,EC=1,T=1}$	$\Delta PP_{1,EC=1} =$ Cell 2 minus Cell 1
Comparison (EC = 0)	3 $\hat{P}_{i,EC=0,T=0} \cdot \hat{U}_{i,EC=0,T=0}$	4 $\hat{P}_{i,EC=0,T=1} \cdot \hat{U}_{i,EC=0,T=1}$	$\Delta PP_{1,EC=0} =$ Cell 4 minus Cell 3

NOTES:

1. EC denotes the dummy variable used to distinguish between demonstration participants and the comparison population while T denotes the dummy variable used to distinguish between the pre and post periods.
2. The numbers in the shaded boxes are cell numbers.

The results from these simulations are shown for hospital admissions (*Table 10-9*) and ED visits (*Table 10-10*) for the pooled and individual PGP samples. Within each table, the left set of three columns shows the 2D results and their statistical significance. A negative sign for the logits indicates reduced probability of an event due to the demonstration while a negative sign for the NBZTs indicates reduced number of events for beneficiaries who have at least one event. The combined hurdle effect ($\hat{P} \cdot \hat{U}$) shows the change in the expected value of utilization.

The mean difference in differences shown in the left-most columns of Tables 10-9 and 10-10 were converted into percent difference in differences by dividing the estimate on left side of the tables by mean of pre-period values for assigned beneficiaries ($\bar{Y}_{EC=1,T=0}$). The estimates for the percent difference in differences are shown in the three right-most columns of Tables 10-9 and 10-10. For example, the decrease in the probability for a hospital admission for PGP 2 of 0.0097 translates into a 4.87 percent reduction in admission.

For both hospital admissions and ED visits, the largest impacts of the demonstration on utilization were usually in reducing the probability of utilization. That is, the absolute magnitude of the percent difference in differences were largest for the logit results. For PGP 6, for example, the reduction in the probability of a hospital admission was 6.63 percent while the reduction in hospital admissions, for beneficiaries with at least one hospitalization, was 2.41 percent.

The probability of a hospital admission fell 2.15 percent for all PGP sites combined (Table 10-9). Of the five sites with statistically significant coefficients, four experienced reductions in the probability of hospitalization ranging from 2.47 (PGP 3) percent to 6.63 percent (PGP 6) while PGP 1 experienced an increase of 2.23 percent. For beneficiaries with at least one hospitalization, the reduction was 0.40 percent for all PGP sites combined. Only PGPs 2 and 6 had statistically significant reductions, 1.40 percent and 2.41 percent respectively. PGP 8 had a statistically significant increase of 1.38 percent. The combined hurdle effect for all PGP sites combined was a 2.29 percent reduction in expected hospitalizations. Four PGPs (2, 3, 5, and 6) experienced combined hurdle effects that were statistically significant and they all had reductions

Table 10-9
PGP demonstration impacts on hospital admissions: Results from Hurdle regression models

PGP	Difference-in-Differences Logit	Difference-in-Differences Negative Binomial	Difference-in-Differences Combined Hurdle	Percent Difference-in-Differences Logit	Percent Difference-in-Differences Negative Binomial	Percent Difference-in-Differences Combined Hurdle
All	-0.0048***	-0.0056*	-0.0089***	-2.15%	-0.40%	-2.29%
1	0.0051*	0.0044	0.0086	2.23%	0.33%	2.32%
2	-0.0097***	-0.0193**	-0.0208***	-4.87%	-1.40%	-5.94%
3	-0.0051	-0.0112	-0.0110*	-2.47%	-0.86%	-3.30%
4	0.0023	-0.0120	-0.0022	1.03%	-0.85%	-0.53%
5	-0.0063***	0.0015	-0.0085*	-2.87%	0.10%	-2.10%
6	-0.0148***	-0.0330***	-0.0335***	-6.63%	-2.41%	-8.88%
7	-0.0031	0.0158	0.0026	-1.48%	1.20%	0.76%
8	-0.0081***	0.0190*	-0.0023	-3.64%	1.38%	-0.59%
9	0.0016	-0.0035	0.0001	0.69%	-0.26%	0.03%
10	0.0005	0.0123	0.0072	0.19%	0.83%	1.47%

NOTES:

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

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

Table 10-10
PGP demonstration impacts on emergency department visits: Results from Hurdle regression models

PGP	Difference-in-Differences Logit	Difference-in-Differences Negative Binomial	Difference-in-Differences Combined Hurdle	Percent Difference-in-Differences Logit	Percent Difference-in-Differences Negative Binomial	Percent Difference-in-Differences Combined Hurdle
All	-0.0060***	-0.0086	-0.0137***	-1.91%	-0.51%	-2.13%
1	-0.0034	0.0195	0.0047	-1.14%	1.22%	0.83%
2	-0.0048*	0.0125	-0.0011	-1.03%	1.13%	-0.16%
3	0.0093**	0.0269	0.0275	3.44%	1.63%	4.86%
4	0.0068*	-0.0054	0.0085	2.10%	-0.31%	1.20%
5	-0.0093***	-0.0001	-0.0138	-2.84%	-0.01%	-2.04%
6	-0.0193***	-0.0472***	-0.0523***	-6.76%	-2.99%	-9.71%
7	-0.0109***	-0.0377*	-0.0355***	-3.14%	-2.19%	-4.92%
8	-0.0024	0.0204	0.0055	-0.87%	1.29%	0.98%
9	-0.0049*	-0.0080	-0.0122	-1.48%	-0.47%	-1.77%
10	-0.0049*	-0.0107	-0.0105	-1.44%	-0.62%	-1.41%

NOTES:

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

SOURCE: RTI analysis of Medicare claims and enrollment data for 2001 to PY5.

in expected hospitalizations ranging from 2.10 percent (PGP 5) to -8.88 percent (PGP 6). PGP 8's combined hurdle effect was not statistically significant as its reduction in the probability of a hospitalization was offset by the increase in the number of hospitalizations for those beneficiaries with at least one hospitalization.

The probability of a ED visit fell 1.91 percent for all PGP sites combined (Table 10-10). Of the eight sites with statistically significant coefficients, six experienced reductions in the probability of ED visit ranging from 1.03 (PGP 2) percent to 6.76 percent (PGP 6) while PGPs 3 and 4 experienced increases of 3.44 percent and 2.10 percent, respectively. For beneficiaries with at least one ED visit, the 2D effect of -0.51 percent for all PGP sites combined was not statistically significant. Only PGPs 6 and 7 had statistically significant reductions, 2.99 percent and 2.19 percent respectively. None had a statistically significant increase. The combined hurdle effect for all PGP sites combined was a 2.13 percent reduction in expected ED visits. Two PGPs (6 and 7) experienced combined hurdle effects that were statistically significant: they had reductions in expected ED visits of 9.71 percent (PGP 6) and 4.92 percent (PGP 7).

CHAPTER 11

SENSITIVITY ANALYSES FOR REFINEMENTS IN DEMONSTRATION DESIGN

This chapter addresses several Demonstration design issues, including patient attribution, performance benchmarks, savings calculations, target expenditures, diagnostic coding, Medicare payment rates, and quality performance measurement. For each issue, we present sensitivity analyses for refinements in the Demonstration design.

11.1 Patient Attribution

The key to determination of Demonstration savings and PGP performance payments in the PGP Demonstration was a comparison of PGP assigned beneficiaries with its comparison population (see Chapter 2 for details). PGP Participants could earn performance payments in the Demonstration if they were able to keep the expenditure growth rate for their assigned beneficiary population below their comparison population. Thus, patient attribution was an important element in the PGP Demonstration.

11.1.1 Type of Provider

PGPs were consulted during the pre-implementation phase to finalize the patient assignment algorithm that resulted in only using outpatient E&M services provided in physician offices to assign patients to the physician groups. Participating PGPs were interviewed regarding their views on the appropriateness of the PGP beneficiary assignment methodology during a series of site visits conducted to all 10 PGPs during the Fall/Winter of 2005–2006. In general, PGPs found the assignment methodology to be a reasonable approach that resulted in a set of assigned beneficiaries for whom they could be held accountable for cost and quality performance.

The two PGPs that are academic medical centers had some reservations, however. They found that office or other outpatient 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. They recommended that CMS consider revising the assignment algorithm for future shared savings demonstrations and programs, to focus on E&M services provided by primary care physicians only.

Table 11-1 compares the assignment algorithm under the PGP Demonstration which is based on E&M services provided by any specialty, with an assignment algorithm based only on primary care E&M services. Compared to the PGP Demonstration definition, the alternative definition tends to have fewer assigned beneficiaries and lower concurrent risk scores. For example, in PY4 there were 216,977 assigned beneficiaries under the demonstration with a mean score of 1.113, whereas under the alternative definition it would have been 176,717 assigned beneficiaries and a mean score of 1.034.

Table 11-1
Statistics on Alternative Beneficiary Assignment Algorithms, Performance Year 4

PGP	PGP Demo Definition (E&M All Specialties): Number of Assigned Benes	PGP Demo Definition (E&M All Specialties): Mean Concurrent Risk Score	Alternative Definition (E&M Primary Care Specialties): Number of Assigned Benes	Alternative Definition (E&M Primary Care Specialties): Mean Concurrent Risk Score	Difference: Number of Assigned Benes	Difference: Mean Concurrent Risk Score
All PGPs	216,977	1.113	176,717	1.034	40,260	0.079
PGP 1	13,825	0.998	9,911	0.902	3,914	0.096
PGP 2	32,127	1.118	20,709	1.025	11,418	0.093
PGP 3	10,184	1.071	9,835	1.118	349	-0.048
PGP 4	15,285	0.938	18,499	0.955	-3,214	-0.017
PGP 5	25,908	1.265	23,272	1.168	2,636	0.097
PGP 6	35,148	1.141	27,753	1.085	7,395	0.055
PGP 7	16,489	1.053	14,370	1.010	2,119	0.043
PGP 8	18,562	0.940	16,644	0.939	1,918	0.001
PGP 9	31,408	1.027	23,696	0.958	7,712	0.070
PGP 10	18,041	1.483	12,028	1.149	6,013	0.334

NOTES: Primary care specialties include: 1 (general practice), 8 (family practice), 11 (internal medicine), 38 (geriatric medicine).

SOURCE: RTI International analysis of 2008-2009 Medicare Administrative Data.

11.1.2 Quantity of Services

The Demonstration patient assignment algorithm was based on the quantity and type of office and other outpatient E&M services provided by the PGP. In our PGP Demonstration design development (Kautter et al., 2007b), we analyzed historical data to determine the optimal quantity of E&M services to base the patient assignment algorithm on. We concluded that a patient should be assigned to a PGP if that PGP provided a plurality of E&M services to the patient. We now update this analysis on a richer and longer time series covering the 10 PGP participants over the time period 2001 to PY2.

Table 11-2a through 11-2c compares the assignment algorithm under the PGP Demonstration, which is based on a plurality of office or other outpatient E&M services, with two alternative assignment algorithms, one based on “one or more” E&M visits, and the other based on a majority of E&M visits. Note that beneficiaries assigned under the plurality of E&M (PGP Demonstration) method are a subset of beneficiaries assigned under the “one or more” E&M method, and similarly, beneficiaries assigned under the majority of E&M method are a subset of beneficiaries assigned under the plurality method. As shown in the tables, for the assignment algorithm under the Demonstration, between 2001 and PY2: (1) approximately 65 percent of beneficiaries that were provided at least office or other outpatient E&M service at a PGP were assigned to the PGP; (2) PGPs provided around 80 percent of E&M services for their assigned beneficiaries; (3) PGPs generally retained approximately 70 percent of their assigned beneficiaries from one year to the next; and (4) PGPs generally retained approximately 40 percent of their assigned beneficiaries after five years.

While alternative assignment methodologies performed better on our sample size criterion (e.g., assignment based on one or more E&M service) or on our provider responsibility criterion (e.g., assignment based on the majority of E&M services), none of the alternative assignment methodologies performed better on both criteria. Interestingly, these results show that the plurality and majority assignment algorithms are quite similar, and are basically the same. Of course, a key disadvantage with the “one or more” assignment algorithm is that beneficiaries can be assigned to more than one PGP.

Table 11-2a
Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs
One or More E&M Visits

Year	Potential Assigned	# Assigned	% Assigned	% E&M Provided by PGP	Per Capita Dollars	Mean Risk Score	Risk Adjusted Dollars	Prior Year Assigned Beneficiaries—% Assigned	2001 Assigned Beneficiaries—% Assigned
2001	318,857	306,444	96.1	54.4	6,593	1.018	6,476	—	100.0
2002	322,498	309,701	96.0	54.6	7,005	1.010	6,935	72.6	72.6
2003	329,159	317,017	96.3	54.6	7,492	1.031	7,264	73.8	63.0
2004	340,118	327,510	96.3	54.4	8,096	1.090	7,426	74.5	56.2
PY1	351,249	329,110	93.7	53.7	8,844	1.094	8,083	70.6	48.0
PY2	349,330	323,081	92.5	53.5	9,315	1.068	8,723	71.3	41.3

Table 11-2b
Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs
Plurality (PGP model)

Year	Potential Assigned	# Assigned	% Assigned	% E&M Provided by PGP	Per Capita Dollars	Mean Risk Score	Risk Adjusted Dollars	Prior Year Assigned Beneficiaries—% Assigned	2001 Assigned Beneficiaries—% Assigned
2001	318,857	208,875	65.5	81.3	5,813	0.921	6,313	—	100.0
2002	322,498	211,243	65.5	81.7	6,157	0.914	6,737	71.8	71.8
2003	329,159	216,314	65.7	81.9	6,638	0.937	7,082	72.9	62.4
2004	340,118	223,203	65.6	81.3	7,130	1.007	7,077	73.5	55.5
PY1	351,249	223,893	63.7	80.8	7,843	0.998	7,857	69.3	47.1
PY2	349,330	219,577	62.9	80.9	8,251	1.023	8,067	70.5	40.6

Table 11-2c
Alternative PGP assignment algorithms based on quantity of office and other outpatient E&M—All PGPs
Majority

Year	Potential Assigned	# Assigned	% Assigned	% E&M Provided by PGP	Per Capita Dollars	Mean Risk Score	Risk Adjusted Dollars	Prior Year Assigned Beneficiaries—% Assigned	2001 Assigned Beneficiaries—% Assigned
2001	318,857	194,491	61.0	86.4	5,617	0.898	6,254	—	100.0
2002	322,498	196,914	61.1	86.6	5,961	0.893	6,678	71.5	71.5
2003	329,159	202,144	61.4	86.7	6,428	0.915	7,023	72.7	62.1
2004	340,118	207,221	60.9	86.5	6,894	0.984	7,008	72.9	55.1
PY1	351,249	207,774	59.2	86.0	7,586	0.974	7,787	68.8	46.6
PY2	349,330	203,978	58.4	86.0	7,997	1.011	7,912	70.1	40.2

NOTES:

1. E&M = Evaluation and Management.
2. Non-E&M PGP assignment criteria includes (a) A and B for every Medicare eligible month, (b) no HMO, (c) no working aged, (d) no non-US residence, (e) no other Medicare FFS demo, (f) enrollment record.
3. Beneficiaries assigned under the 1 or more E&M assignment criteria if they have at least 1 office or other outpatient E&M visit at the PGP, and satisfy all non-E&M assignment criteria.
4. Beneficiaries assigned under the plurality E&M assignment criteria if they have a plurality of their office or other outpatient E&M visit at the PGP, and satisfy all non-E&M assignment criteria.
5. Beneficiaries assigned under the plural E&M assignment criteria if they have a plurality of their office or other outpatient E&M visit at the PGP, and satisfy all non-E&M assignment criteria.
6. All statistics are enrollment weighted.

SOURCE: RTI Analysis of 2001–2007 100% Medicare Claims Files and Enrollment Data.

11.2 Alternative Performance Benchmarks

The measured savings achieved by each participating PGP depends crucially on the benchmark used for measuring savings. The choice of benchmark is a trade-off between two competing goals. First, the PGP Demonstration was “self-evaluating” to the extent that the assumption, underlying the benchmark, of what expenditures for the groups’ assigned beneficiaries would be in the absence of the Demonstration incentives was valid. The less comparable is the population or time period used to compute the benchmark, the less valid is the benchmark. On the other hand, the more rigorously-defined the comparison group is, the greater are the information needs and the amount of processing time and effort to compute the benchmark, resulting in longer lags between PGP performance and incentive payments for that performance.

This section considers the effect on measured Demonstration total and shared savings from using alternative benchmarks to measure savings. The alternative benchmarks and definitions are presented in *Table 11-3*. In all cases the structure of the Demonstration savings computation algorithm is retained: actual annualized per capita Medicare program expenditures for each group’s assigned beneficiaries are compared to a target level. Only the definition of the target expenditure level differs across alternatives.

For comparison purposes, the first “alternative” is the benchmark that was actually used in the PGP Demonstration: assigned beneficiary expenditures in a performance year are compared to assigned beneficiary base year expenditures, adjusted by the growth rate of expenditures for a set of comparison beneficiaries in the market, determined retrospectively, and by the relative risk score growth. A second benchmark substitutes Medicare program expenditure growth statewide instead of that for a retrospectively-determined group of beneficiaries in the groups’ market areas, and a third uses the growth rate of program expenditures nationally. Although these alternatives also must be computed retrospectively, they are simpler to compute and therefore shorten the lag between services provided and expenditures incurred, and when performance payments can be made.

Two alternative benchmarks could be computed prospectively. The first (Pre-Existing Trend) is computed using the year prior to the Base Year (for the PGP Demonstration, this is 2003) as the basis for computing target expenditures as well as the two-year growth rate in expenditures and risk scores from 2001 to 2003. These quantities are presumably known as of the end of the Base Year and are therefore prospectively set. The second prospective alternative (National Pre-Existing Trend) is computed analogously to the Pre-Existing Trend alternative but instead uses national average program expenditures and growth rates.

Table 11-3
Alternative performance benchmarks/expenditure target

Benchmark/Expenditure Target	Description
PGP Demonstration	Target expenditures are computed exactly as in the PGP Demonstration. For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the risk-adjusted comparison group growth rate.
Statewide Average as Comparison Growth Rate	For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the statewide average Medicare expenditure growth rate.
National Average as Comparison Growth Rate	For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the national average Medicare expenditure growth rate.
Pre-Existing Trend	For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the risk-adjusted assigned group growth rate in the year prior to the Base Year.
National Pre-Existing Trend	For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the national average Medicare expenditure growth rate in the year prior to the Base Year.
PGP Demo Plus Pre-Existing Trend	For each group, target expenditures are computed as Base Year risk-adjusted assigned beneficiary expenditures, multiplied by the gross risk-adjusted comparison group growth rate net of the risk-adjusted Base Year growth rate difference

SOURCE: RTI International.

The final alternative combines the actual PGP Demonstration algorithm with the Pre-Existing Trend alternative. In this alternative, growth rates of comparison group Medicare program expenditures, and in assigned and comparison group risk scores, are first “de-trended” using growth rates in these quantities from before the Base Year (2001 through 2003). These de-trended growth rates are then used in the PGP Demonstration algorithm.

Because these alternative benchmarks were never in force, the alternative measured savings amounts are hypothetical and require assumptions about Medicare program expenditures in these hypothetical situations. The savings measured using the alternative benchmarks in this section assume that all assigned and comparison group expenditures and risk scores are the same as actually incurred.

Table 11-4 presents several metrics for comparing the alternative benchmarks against the actual PGP Demonstration benchmark. The first is the (unweighted) average difference of actual minus target expenditures, expressed as a percentage of target expenditures. The minimum and maximum savings percentages across the 10 PGPs are also presented. Three additional metrics present information about the variation in measured savings percentages across the alternatives, compared to the PGP Demonstration algorithm: the number of groups with measured savings below the PGP Demonstration minimum, the number of groups with measured savings above the PGP Demonstration maximum, and the ratio of the variance in measured savings across the 10 groups under each alternative to that for the PGP Demonstration. The final metric gives the correlation coefficient between savings under each alternative and savings in the PGP Demonstration.

Table 11-4
Actual performance year two per-capita expenditures and percent difference of target over actual expenditures

Growth Rates*	PGP Demo	Statewide Average Growth Rate	National Average Growth Rate	Pre-Existing Trend	National Pre-Existing Trend	PGP Demo Plus Pre-Existing Trend
Average	1.20	3.62	3.84	-3.88	3.34	0.06
Minimum	-4.04	-2.34	-1.00	-19.14	-14.61	-8.89
Maximum	4.91	7.51	6.68	6.14	15.51	5.52
Number Below Demo Minimum	0	0	0	5	1	2
Number Above Demo Maximum	0	3	3	2	4	2
Relative Std. Dev. (PGP Demo = 1.00)	1.00	1.21	0.71	9.16	10.74	3.09
Correlation with PGP Demo	1.00	0.80	0.73	0.15	0.00	0.18

NOTE: $*100\% \times [\text{Actual Expenditures} - \text{Target Expenditures}] \div [\text{Target Expenditures}]$

SOURCE: RTI International.

Changing the benchmark for target expenditures can have dramatic impacts on measured savings. Using the concurrent statewide or national average growth rates would have resulted in higher measured savings, by at least two percentage points on average. The minimum and maximum savings rates would have been higher under either of these alternatives, and three groups would have had measured savings greater than the maximum under the actual PGP Demonstration algorithm. Interestingly, there is more variation in measured savings using statewide expenditures rather than national quantities, suggesting that there is at least as much variation in expenditures within regions as across regions, at least for states with participating PGPs. However, measured savings under these alternatives are highly correlated with savings under the PGP Demonstration, indicating that relative performance among groups would be similar across the alternatives.

As shown in the fourth and fifth columns of Table 11-4, using pre-Demonstration, rather than concurrent, trends in the target benchmarks would produce very different, and much more variable, results. Within-market pre-existing trend benchmarks would result in very negative

average savings, whereas national average pre-existing trends would produce average savings above that measured in the PGP Demonstration. Measured savings under the pre-existing trend alternatives are also much more variable, producing very high and very low maximum and minimum savings rates among the 10 groups—presumably expenditure growth rates are highly variable over time, resulting in a great deal of “noise” in the pre-existing trend benchmarks. Also, the correlation coefficients between the pre-existing trend and PGP Demonstration savings rates are near zero, emphasizing the fact that using pre-Demonstration, rather than concurrent, benchmarks, would produce results very different from that realized in the PGP Demonstration.

The final alternative, the de-trended PGP Demonstration benchmark, unsurprisingly yields results between those of the PGP Demonstration and those of the pre-existing trend benchmarks.

Table 11-5 presents shared savings under each alternative as a percentage of target expenditures. Most notable in this table is that average shared savings are lowest for the PGP Demonstration savings computation algorithm. This is due to two factors. First, average shared savings will be higher when average savings are higher—note that the statewide and nationwide concurrent benchmarks and the nationwide pre-existing trend alternatives would have yielded higher average shared savings. Second, the greater the variance in measured savings, the greater the average shared savings since shared savings are bounded below by zero. The greater the variation in savings, the larger are positive savings. However, negative savings produce zero shared savings, regardless of the magnitude of the negative savings.

Table 11-5
Actual performance year two shared savings as percent of target expenditures

Growth Rates*	PGP Demo	Statewide Average Growth Rate	National Average Growth Rate	Pre-Existing Trend	National Pre-Existing Trend	PGP Demo Plus Pre-Existing Trend
Average	0.50	1.74	1.74	0.68	3.17	0.83
Minimum	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	2.32	4.41	3.74	3.32	10.81	2.81
Number Receiving Bonus	4	8	8	3	7	4
Number Receiving Bonus Under Demo & Alternative	4	4	4	2	2	0
Number Receiving Bonus Under Demo, Not Alternative	0	0	0	2	2	4
Number Receiving Bonus Under Alternative, Not Demo	0	4	4	1	5	4
Number Not Receiving Bonus Under Demo & Alternative	6	2	2	5	1	2

NOTE: $*100\% \times [\text{Shared Savings}] \div [\text{Target Expenditures}]$

SOURCE: RTI International.

11.3 Alternative Calculation of Savings and Sharing Rates

We summarize Demonstration financial results under the alternative assumption that savings or losses within the 2 percent corridor are Medicare program expenditures, or lack of expenditures, and contribute to savings or losses. We present four scenarios with different design assumptions that simulate performance payments to the PGPs and Medicare program savings. In contrast to the PGP Demonstration methodology, the four scenarios take into consideration amounts of target minus actual expenditures within the 2 percent corridor.

For the purposes of this section, we assume that the amounts within the corridor represent real dollars and contribute to savings or losses. Under the actual Demonstration methodology, the corridor serves to guard against shared savings payments for normal variation and measurement imprecision, and amounts within the corridor are not included in financial reconciliation calculations. As discussed in Section 2.3, the 2 percent corridor 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. Actual observed expenditure growth rates are a combination of changes due to a PGP's efforts and those due to events specific to individual patients and entirely outside of a PGP's control. For the purposes of this section, we acknowledge the purpose of the 2 percent corridor in the Demonstration but present savings calculations that consider amounts within the corridor.

In all four scenarios, gross savings, losses and net savings are held constant. In all four scenarios, \$87.2 million is added to net savings as the amount of net savings within the 2 percent corridor. Simulated performance payments to the PGPs are estimated based on varying design assumptions which are discussed below. Medicare program savings remain after subtracting simulated performance payments from the sum of Demonstration net savings and net savings within the 2 percent corridor.

11.3.1 Design Assumptions of the Alternative Scenarios

In this section we describe the design assumptions of the four alternative scenarios. In each scenario, the following assumptions may vary: the sharing rate, the minimum savings requirement, the ability to share in first dollar of savings, the minimum loss rate, and the symmetry of risk.

The first scenario is not an alternative calculation of shared savings to the PGPs but instead is an alternative evaluation of the savings to Medicare. The first scenario uses the actual performance payments to the PGPs under the Demonstration to evaluate the savings to Medicare assuming that amounts within the corridor are counted as savings or losses to Medicare. Under the other three scenarios we present alternative shared savings calculations assuming that the PGPs share in the first dollar of savings under varying assumptions about the minimum savings rate (MSR), the sharing rate, the capping of earned performance payments, and the symmetry of risk. These three scenarios present alternative shared savings calculations—two of which we adapted from the Medicare Shared Savings Program for Accountable Care Organizations shared savings and losses methodology.

The four alternative scenarios we evaluate address the following questions.

Scenario One: What are the savings or losses to Medicare if amounts within the 2 percent corridor contribute to savings and losses, but actual PGP Demonstration earned performance payments and losses are held constant?

Scenario Two: What are the savings or losses to Medicare if PGPs that meet the PGP Demonstration MSR of 2 percent are eligible to earn up to 80% of gross savings starting at the first dollar of savings?

Scenario Three: What are the savings or losses to Medicare if PGPs that meet a sliding scale MSR based on the number of assigned beneficiaries are eligible to earn up to 50% of gross savings starting at the first dollar of savings?

Scenario Four: What are the savings or losses to Medicare if PGPs that exceed a MSR of 2 percent are eligible to earn up to 60% of gross savings starting at the first dollar of savings and PGPs that exceed a minimum loss rate (MLR) of -2 percent are responsible for paying CMS for the first dollar of shared losses up to a final loss rate?

Table 11-6 summarizes the treatment of the 2 percent corridor, the MSR, the sharing rate, and the capping of the earned performance payments under the Demonstration methodology and across the four alternative scenarios. The baseline, risk adjustment methodology, and target expenditures under the four alternatives do not differ from the Demonstration methodology and thus are not described here. Scenarios Three and Four use the same parameters reported in Table 11-6 as the Medicare Shared Savings Program (MSSP) one-sided and two-sided models, respectively, according to the Final Rule for that program promulgated in 2011.³² Note that these presentations of savings do not use the MSSP savings methodology, but rather show the effect of some design assumptions; therefore, this presentation cannot be used to project or compare the PGP experience to the MSSP.

³² Scenarios Three and Four are not comprehensive simulations of the Medicare Shared Savings Program models because they use different baselines, target-setting methodologies, and risk adjustment.

**Table 11-6
Treatment of 2% corridor, MSR, sharing rate, and earned performance payment caps under the PGP demonstration methodology and the four alternative scenarios**

Design Assumptions	Actual PGP Demonstration Methodology	Scenario One: Actual PGP earned performance payments and addition of 2% corridor as savings or losses to Medicare	Scenario Two: 80% quality performance sharing rate at first dollar	Scenario Three: One-sided 50% quality performance sharing rate	Scenario Four: Two-sided 60% quality performance sharing rate
Treatment of amounts within ±2% corridor	Amounts of “Total Target-Actual Expenditures” within the ±2% corridor are not counted as savings or losses.	Positive amounts of “Total Target-Actual Expenditures” within the ±2% corridor contribute to Demonstration gross savings while negative amounts contribute to Demonstration losses.	Same as Scenario One	Same as Scenario One	Same as Scenario One
Minimum savings rate (MSR) and minimum loss rate (MLR)	MSR of 2%.	MSR of 2%.	MSR of 2%. PGPs share in first dollar of savings.	MSR is on a sliding scale based on the number of assigned beneficiaries. PGPs share in the first dollar of savings.	MSR of 2% and MLR of -2%. PGP’s earn first dollar of shared savings if total savings exceeds MSR and are responsible for paying CMS for first dollar of shared losses if total loss is less than MLR.
Sharing rate	PGPs are eligible to share in 80% of Total Target Minus Actual Expenditures outside the 2% corridor. (A portion of this shared savings pool is dependent on quality: 30% in PY1, 40% in PY2, 50% in PY3-PY5. This is also true in Scenarios One and Two).	Same as actual PGP Demonstration methodology.	PGPs are eligible to share in 80% of Total Target Minus Actual Expenditures, beginning at the first dollar of savings. A portion of this shared savings pool is dependent on quality: 30% in PY1, 40% in PY2, 50% in PY3-PY5.	The quality performance sharing rate is set at 50%. The final sharing rate is the product of the quality performance sharing rate and the quality score.	The quality performance sharing rate is set at 60%. The final sharing rate is the product of the quality performance sharing rate and the quality score. The final loss rate equals one minus the final sharing rate, not to exceed 60%.

(continued)

Table 11-6 (continued)
Treatment of 2% corridor, MSR, sharing rate, and earned performance payment caps under the PGP demonstration methodology and the four alternative scenarios

Design Assumptions	Actual PGP Demonstration Methodology	Scenario One: Actual PGP earned performance payments and addition of 2% corridor as savings or losses to Medicare	Scenario Two: 80% quality performance sharing rate at first dollar	Scenario Three: One-sided 50% quality performance sharing rate	Scenario Four: Two-sided 60% quality performance sharing rate
Capping earned performance payments	Capped at 5% of target expenditures	Same as actual PGP Demonstration methodology.	Same as actual PGP Demonstration methodology.	Capped at 10% of target expenditures. ¹	Earned performance payments are capped at 15% of target expenditures. Payments due to CMS are capped at 5% of target expenditures in PY1, 7.5% of target expenditures in PY2, and 10% of target expenditures in PY3-PY5. ¹

NOTES:

¹ Earned performance payments do not exceed the cap under Scenario Three or Scenario Four for any PGP in any performance year.

Establishment of target expenditures, the baseline period, and the risk adjustment methodology under the four alternative scenarios are the same as under the PGP Demonstration methodology. Target expenditures are PGP specific and are based on each PGP's base year expenditure level. Target Expenditures = PGP Base Year Expenditures x (1 + Comparison Group Growth Rate). The baseline period is a one year base year, calendar year 2004. The PGP Demonstration risk adjustment model is concurrent—using diagnoses from the same year as expenditures—rather than prospective. The PGP Demonstration risk model was calibrated only once, using base year 2004 data, and so did not reflect any changes in medical treatment, expenditure, or diagnostic coding patterns over the course of the Demonstration.

SOURCE: RTI International

11.3.2 Net Savings, Simulated Performance Payments, and Medicare Program Savings

Table 11-7 summarizes the Demonstration net savings, simulated performance payments, and Medicare program savings under the actual PGP Demonstration methodology and the four alternative scenarios for PY1 through PY5 combined. This table also summarizes the design assumptions used in each scenario. Calculations for each performance year by PGP under each scenario are presented in an appendix to this report.

Under the actual Demonstration methodology, over the combined five years of the Demonstration, gross savings totaled \$152.9 million. Of this amount, \$15.1 was absorbed by PGP losses. Of the remaining \$137.7 million in net savings, \$107.6 million was distributed to the participating PGPs as performance payments resulting in Medicare Program savings of \$30.2 million. Scenarios One through Four hold net savings from the Demonstration constant. We add the amount of net savings within the 2 percent corridor (\$87.2 million) to Demonstration net savings in all four scenarios.

Simulated performance payments are calculated by including net savings within the 2% corridor and changes for design assumptions. In Scenario One, PGPs do not earn any additional performance payments. This is because Scenario One assumes that performance payments to the PGPs are held constant at the actual Demonstration performance payments. In Scenario Two, PGPs earn an additional \$77.6 million in performance payments compared to the actual Demonstration payments. This increase is explained by the assumption that PGPs share in the first dollar of savings including net savings within the corridor. In Scenario Three, PGPs earn an additional \$12.7 million compared to the actual Demonstration payments.³³ In Scenario Four, PGPs earn an additional \$25.9 million compared to the actual Demonstration payments.³⁴

Medicare program savings in the four scenarios are calculated as the residual savings after subtracting simulated performance payments from the sum of net savings and the \$87.2 million amount of net savings within the corridor. Under Scenario One, Medicare program savings increase by \$82.7 million compared to the Medicare savings reported using the Demonstration methodology, for total savings of \$112.9 million. Under Scenario Two, an additional \$5.1 million is retained by Medicare for a total of \$35.3 million program savings. Under Scenario Three, an additional \$70.0 million is retained by Medicare for total savings of \$100.2 million. Finally, under Scenario Four, an additional \$56.8 million is retained by Medicare for total savings of \$87.0 million.

Adding net savings within the 2 percent corridor to Demonstration net savings increases both total simulated performance payments and Medicare program savings under all four

³³The additional \$12.7 million in performance payments is computed from: 1) \$77.6 million increase due to the distribution of amounts within 2% corridor, 2) \$51.3 million decrease as a result of the design assumptions, and 3) \$13.7 million decrease due to the interaction of the design assumptions and the corridor amounts.

³⁴The additional \$25.9 million in performance payments is computed from: 1) \$77.6 million increase due to distribution of amounts within 2% corridor, 2) \$36.3 million decrease as a result of the design assumptions, and 3) \$15.4 million decrease due to the interaction of the design assumptions and the corridor amounts.

Table 11-7
Net savings, net amounts within the 2% corridor, total simulated performance payments, and Medicare program savings
under PGP Demonstration methodology and Scenarios One-Four, PY1-PY5 combined

Design Assumptions	PGP Demonstration Methodology	Scenario One: Amounts within 2% corridor considered savings to Medicare	Scenario Two: 80% quality performance sharing rate at first dollar	Scenario Three: One-sided 50% quality performance sharing rate	Scenario Four: Two-sided 60% quality performance sharing rate
Sharing rate	80%	80%	80%	50%	60%
Minimum savings rate	2%	2%	2%	Varies	2%
PGPs share in first \$ coverage	No	No	Yes	Yes	Yes
Minimum loss rate	No	No	No	No	Yes
Asymmetrical risk	No	No	No	Yes	No
Impact on Savings					
[A] Net savings from Table 5-12	\$137,762,108	\$137,762,108	\$137,762,108	\$137,762,108	\$137,762,108
[B] Net amount within the 2% corridor	—	82,729,599	82,729,599	82,729,599	82,729,599
[C] Less performance payments to PGP sites from Table 5-12	-107,556,954	-107,556,954	-107,556,954	-107,556,954	-107,556,954
[D] Less simulated additional performance payments	0	0	-77,643,402	-12,690,845	-25,916,266
[E] Total simulated Medicare savings	30,205,153	112,934,752	35,291,350	100,243,908	87,018,486

NOTES:

[A] Net savings from PGP Demonstration Methodology from Table 5-12.

[B] Net amount within the 2% corridor is constant across all scenarios.

[C] Actual performance payments to PGP sites from Table 5-12. Actual performance payments are constant across all scenarios.

[D] Additional performance payments to PGP sites under the scenarios resulting from changes in design assumptions. Under Scenario One, no additional performance payments are estimated. Under Scenario Two, PGPs earn an additional \$77.6 million in performance payments as a result of the inclusion of the 2% corridor. Under Scenario Three, PGPs earn an additional \$12.7 million in performance payments. The inclusion of the 2% corridor increases performance payments by \$77.6 million; the change in design assumptions decreases performance payments by \$51.3 million, and the interaction of the design assumptions and the corridor decrease performance payments by \$13.7 million. Under Scenario Four, PGPs earn an additional \$25.9 million in performance payments. The inclusion of the 2% corridor increases performance payments by \$77.6 million; the change in design assumptions decreases performance payments by \$36.3 million, and the interaction of the design assumptions and the corridor decrease performance payments by \$15.4 million.

[E] The residual amount remaining after subtracting actual performance payments and additional simulated performance payments represents total simulated Medicare savings. [A] + [B] + [C] + [D]

SOURCE: RTI calculations with 2004-2010 Medicare claims and enrollment data

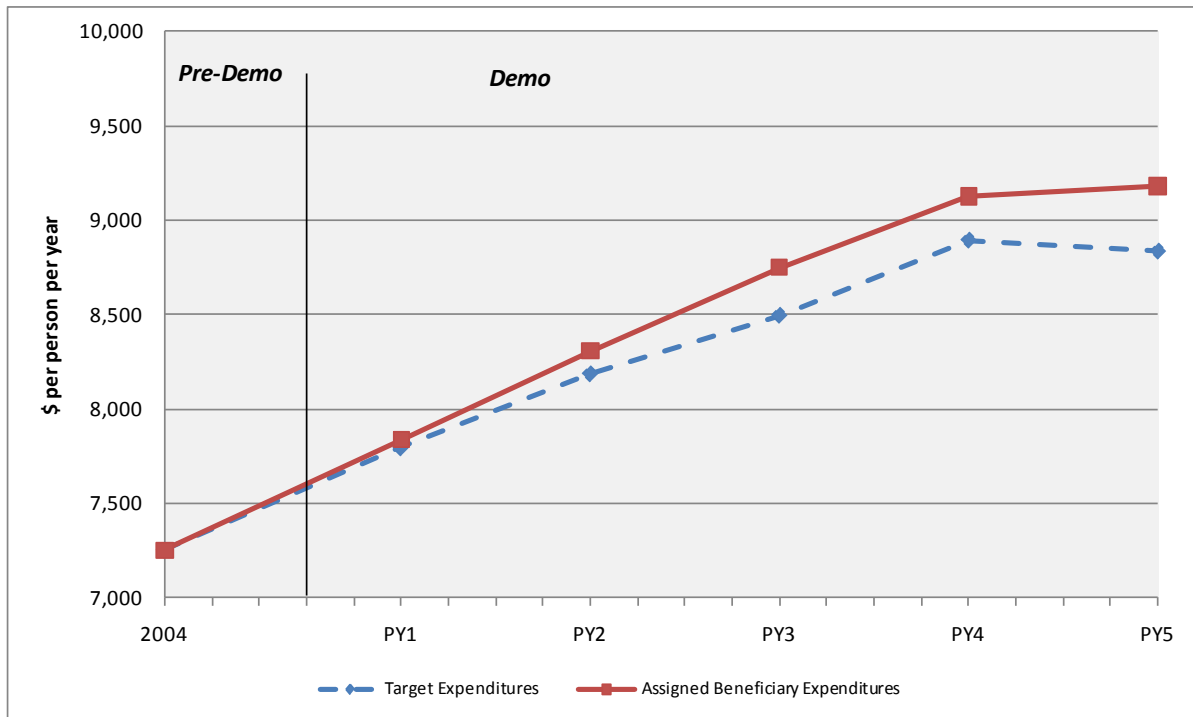
scenarios. The distribution of the \$87.2 million in net savings within the corridor between performance payments and Medicare program savings varies based on the assumptions about the MSR, the sharing rate, the capping of performance payments, and the symmetry of risk. Higher net earned performance payments result in lower Medicare program savings, with the greatest earned performance payments simulated in Scenario Two and the lowest earned performance payments simulated in Scenario One. Performance payments to PGPs are maximized under Scenario Two because PGPs are eligible to earn up to 80% of net savings starting at the first dollar of savings. Savings retained by Medicare as program savings are maximized under Scenario One where the total amount of net savings in the 2 percent corridor is retained as Medicare program savings.

11.4 Target Expenditures Without Risk Adjustment

The Demonstration methodology adjusts for risk using a version of the CMS-HCC model implemented for Medicare Advantage risk adjustment as described in Chapter 2. Each beneficiary is assigned a risk score based on diagnosis information. The final average risk scores for the assigned and comparison group populations at each PGP are applied to per capita expenditure growth rates. In this section, we briefly look at the impact of risk adjustment on Demonstration results by simulating a non-risk adjusted target.

Figure 11-1 illustrates the trend in assigned beneficiary expenditures versus a simulated non-risk adjusted target expenditures. In contrast to the results presented in Chapter 5, here assigned beneficiary expenditures exceed target expenditures. This shows the important effect of risk adjustment on Demonstration results. Overall, the PGPs would have a negative value of total target minus actual expenditures, indicating that the actual expenditures exceeded target expenditures. *Figure 11-2* graphs the difference between PGP expenditures and non-risk adjusted target expenditures by individual PGP site. Six out of the ten PGPs have actual expenditures in excess of non-risk adjusted target expenditures from the BY through PY5. PGP 6 is the only site with actual expenditures less than non-risk adjusted target expenditures for all five performance years of the Demonstration. Without risk adjustment, the PGPs overall would have actual expenditures in excess of target expenditures.

Figure 11-1
Assigned beneficiary versus non-risk-adjusted target expenditures, 2004 base year to PY5,
all 10 PGPs

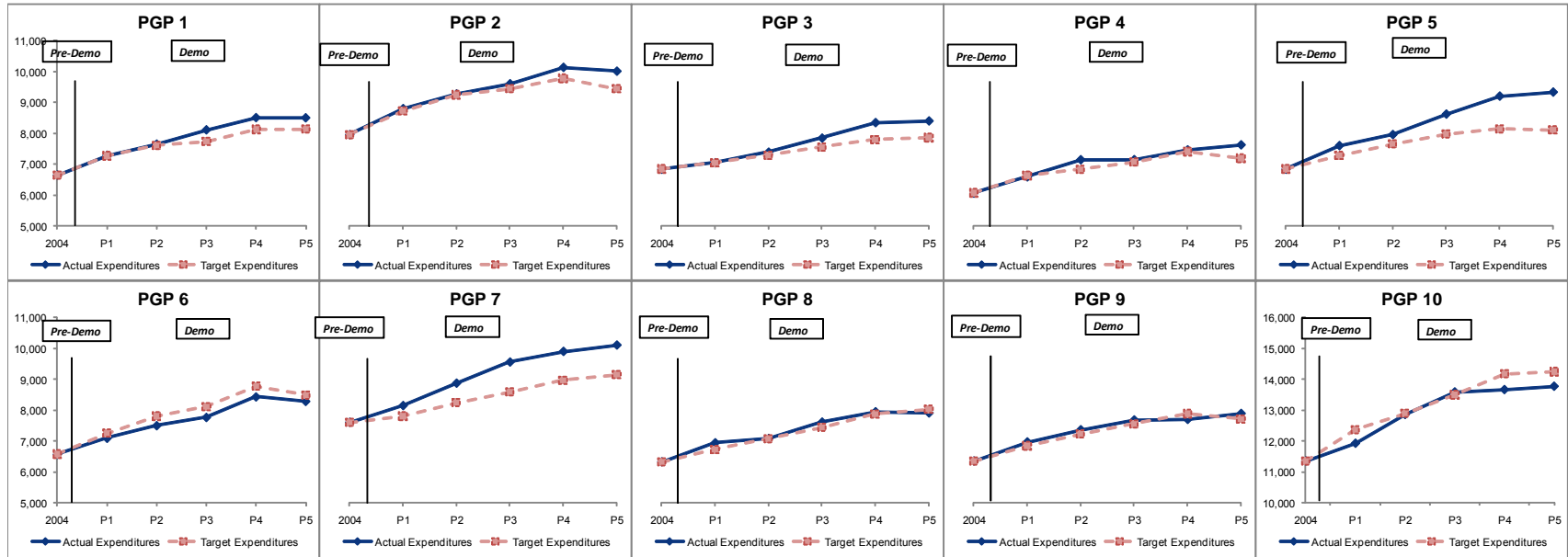


NOTES:

Unweighted average across the 10 PGPs.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 11-2
Assigned beneficiary versus non-risk-adjusted target expenditures, by PGP, 2004 base year to PY5



NOTES:

The vertical scale is the same for all PGPs except PGP 10, for which it is higher.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

11.5 Diagnostic Coding

11.5.1 Importance of Coding in the Demonstration

Changes in expenditures for the Demonstration sites and their comparison groups were risk adjusted during the Demonstration. A modification of the CMS-HCC risk adjustment model was developed for the Demonstration. It was a concurrent risk adjustment model, which uses this year's diagnoses and demographics to predict this year's expenditures.³⁵ The results of the Demonstration thus depended on a) changes in expenditures for the Demonstration sites and comparison groups and b) changes in risk of the populations in these groups. The changes of expenditures are tightly linked to services provided and are derived from the claims of all types of providers. The risk measures derive from reporting of diagnoses on three types of claims, hospital inpatient, hospital outpatient and physician/clinician. There is an implicit assumption in the PGP model of computing risk adjusted savings that the coding patterns for equivalent patients were the same in the Demonstration sites and the comparison groups.

There has been a pattern of increased intensity and specificity of coding among FFS providers, particularly physicians on average over the last ten years or so. CMS has posted on its website a list of acceptable ICD-9-CM codes for several years now.³⁶ Guidelines have always indicated that the coding should be at the highest level of specificity available for each condition. The degree to which three- and four-digit codes are reported when five-digit codes are available has diminished. Specificity can result in increased risk scores because a more specific code will often indicate that a patient is in a higher cost category than a vague code. In addition there has been a movement to detect chronic diseases at earlier stages. Additional screening for diabetes, for example, results in recording the disease at a higher frequency. This trend toward more and more specific coding has been noted because the prospective form of the CMS-HCC risk adjustment model is run over multiple years of data for both Medicare Advantage and FFS Medicare beneficiaries.

If the growth pattern were the same in the Demonstration and comparison groups there would be no net effect on risk adjusted savings. If there are differences there is reason to determine the causes for the differences. True changes in case mix should be allowed to stand in the calculation, but other changes should be adjusted out if possible.

11.5.2 Reasons for Coding Differences

One possible reason for coding in the PGPs and comparison groups to differ was the incentive for more scrupulous and thorough coding in the Demonstration sites. The computation that determined whether a PGP had met a target depended on expenditures after adjustment for risk. A higher coding rate increased the apparent risk and reduced the risk adjusted expenditure growth. A target is easier to meet if apparent risk is growing more quickly than spending. Presumably only the Demonstration sites and not their comparison groups had an incentive to pay particular attention to coding intensity.

³⁵ A prospective risk adjustment model, like the CMS-HCC model used for risk adjustment of Medicare Advantage capitation rates, uses last year's diagnoses and demographics to predict this year's expenditures.

³⁶ ICD-10 diagnostic coding will replace ICD-9 diagnostic coding in the US beginning in fiscal year 2014.

The phenomenon of incentives resulting in extra coding efforts has been observed in the Medicare Advantage program. Capitated payments to these health plans are risk adjusted using the same diagnosis grouper as was used in the PGP Demonstration. It is known that data mining techniques are used to develop reminders to physicians that certain patients were not reported with a chronic disease that was reported the year before; and to seek out laboratory results and drug claims that indicate the presence of a diagnosis. There are firms offering consulting services and software to “optimize” coding. This program has produced much higher growth in coding indicating prevalence of diabetes and of higher severity diabetes, than has occurred in the FFS program. In fact, because of the higher growth in coding in Medicare Advantage relative to FFS, CMS has recently implemented a coding adjustment for the Medicare Advantage program.

PGP sites had indicated that they were aware of the importance of coding. Most had electronic medical record systems of some form that assisted in the collection of codes. Similar coding changes have been observed in the Medicare inpatient hospital payment system that also depends on diagnoses. The phenomenon is generally known as “coding creep”.

An alternative hypothesis is that the sites were making efforts to attract patients with particular diseases and enroll them in programs of disease management. If the sites always attracted sicker patients because of perceived quality, that risk level would not create an issue for the calculation. Only a growth in attractiveness over time would have had an effect because the Demonstration compared rates of growth. True increased risk profiles of patients should be retained in the growth. Distinguishing average coding creep, incentive-related creep and true risk change is a challenge.

11.5.3 Observed Coding Changes

There are a number of dimensions of the changes in coding for the PGPs relative to their comparison groups that can be tabulated.

1. The proportion of the assigned beneficiaries who had none of the 71 Hierarchical Condition Categories (HCCs) used to calculate risks scores in the Demonstration. This is termed the rate of “NOHCC.”
2. The rate at which particular diseases were reported in the model.
3. The net affect the coding changes would have on the risk score of a typical group.

NOHCC. The percentage of beneficiaries coded with NOHCC decreased from the base year to PY2 in both the Demonstration and comparison groups. The changes were non-uniform. Typically about 40 percent of Medicare beneficiaries do not have at least one of the 71 HCCs included in the model. There has been shrinkage in the NOHCC group. In *Table 11-8* the two-year change in the rate of NOHCC is displayed. It is clear that in all but one site the change for the PGPs was much larger than for the comparisons; the average was twice as large. The largest differentials were for PGPs 2, 3, and 4.

The base NOHCC rates were generally lower than the FFS average of 40 percent because all the beneficiaries included in the Demonstration groups had to have at least one E&M visit to

be assigned to a group and be included in the study population. In most, but not all cases the PGPs had lower rates of NOHCC in the base period and the rates were further reduced over the two-year Demonstration period. The increased coding in the Demonstration sites and comparison groups was apparent, but the cause of the higher rate of increase in the PGPs is not yet clear.

Table 11-8
Percent change in the percentage of beneficiaries with NOHCC,
base year to performance year two

PGP	PGP Assigned Base Rate	PGP Assigned Change	Comparison Group Base Rate	Comparison Group Change
PGP 1	35%	-4.4%	41%	-5.9%
PGP 2	32	-9.8	40	-3.0
PGP 3	34	-5.7	36	-0.6
PGP 4	35	-8.3	32	-2.7
PGP 5	28	-9.4	31	-3.5
PGP 6	35	-9.8	42	-6.5
PGP 7	33	-9.4	34	-4.5
PGP 8	41	-7.8	39	-4.2
PGP 9	32	-9.0	34	-5.0
PGP 10	29	-9.1	32	-6.1
Mean	34	-8.2	36	-4.2

NOTE: No HCC = Did not have at least one of the 71 hierarchical condition categories used for Demonstration risk adjustment.

SOURCE: RTI analysis of Demonstration data.

Coding of Particular Disease Groups. The prior analysis looked only at the change in how many people were coded with a diagnosis in the 71 HCCs used for risk adjustment in the PGP Demonstration. We also looked at specific HCCs to determine which were increasing in prevalence at a noteworthy rate, presented in *Tables 11-9a and 11-9b*. The most consistently growing HCC was that for Renal Failure. The pattern for this code was influenced, however, by the change in definition of the ICD-9 codes that occurred in FY 2006 (Oct 2005). There has been an upward trend in the prevalence of these codes, however, even without the change. The change in October 2005 altered the name of ICD-9 code 585 from Chronic Renal Failure to Chronic Kidney Disease, and the higher detail levels of the codes are stages of kidney disease prior to and including renal failure. The change affected all providers, but the rates of change in the PGPs were greater than in the comparison groups in 7 out of 10 sites and the magnitudes of

the changes were large. The smallest growth rate was 37 percent and the largest, 121 percent. The rates of change for HCC131 and other HCCs are in Tables 11-9a and 11-9b.

Table 11-9a
Percentage change in rate of coding of four HCC groups,
base year to performance year two, PGP 1 – PGP 5

HCC	PGP 1 PGP	PGP 1 Comp	PGP 2 PGP	PGP 2 Comp	PGP 3 PGP	PGP 3 Comp	PGP 4 PGP	PGP 4 Comp	PGP 5 PGP	PGP 5 Comp
HCC131	78.9	88.1	70.9	45.0	54.0	44.6	77.8	45.7	37.2	55.0
HCC81	-4.8	-3.1	11.4	8.3	-20.9	-15.6	12.3	-4.3	-6.7	-15.7
HCC15	-8.7	35.2	46.8	8.9	29.8	21.9	24.2	13.8	5.6	10.0
HCC80	5.4	4.8	-4.9	-0.4	-1.4	-3.7	1.3	-1.3	-1.5	-0.8

Table 11-9b
Percentage change in rate of coding of four HCC groups,
base year to performance year two, PGP 7 – PGP 10

HCC	PGP 6 PGP	PGP 6 Comp	PGP 7 PGP	PGP 7 Comp	PGP 8 PGP	PGP 8 Comp	PGP 9 PGP	PGP 9 Comp	PGP 10 PGP	PGP 10 Comp
HCC131	88.7	96.0	121.8	42.8	62.7	60.2	110.3	79.1	62.4	46.5
HCC81	6.1	-9.9	-9.9	-3.6	-5.4	-11.7	-15.9	-2.3	11.1	-10.1
HCC15	22.6	41.6	4.1	-1.5	-0.1	-3.8	32.9	32.6	31.5	20.8
HCC80	-1.0	1.7	13.9	3.2	10.5	-6.1	0.8	1.8	-3.5	1.3

NOTES:

- HCC = Hierarchical Conditions Category
- HCC131 = Renal Failure / Chronic Kidney Disease
- HCC81 = Acute Myocardial Infarction
- HCC15 = Diabetes with Renal or Peripheral Circulatory Manifestation
- HCC80 = Congestive Heart Failure

SOURCE: RTI analysis of Demonstration data.

The next HCC in the table, HCC81, Acute Myocardial Infarction, is less subject to influence of discretionary coding. It might have been a factor in selection into a PGP if patients with known ischemic heart disease gravitated to the PGPs or did so after the heart attack. The pattern of coding growth is mostly negative and does not show systematic greater growth in the PGPs. This HCC is not a chronic condition and so does not persist from year to year. Its relative frequency is low compared to chronic diseases.

Diabetes with complications is a group subject to attention to detail in specifying codes, has tests that can be done to establish a complication even if not clinically reported by the patient, and is often the focus of a disease management program. Several of the PGPs had programs focusing on diabetes (see Chapter 4). Table 11-9 also presents the growth in coding for HCC15, Diabetes with Renal or Peripheral Circulatory Manifestation. This is the diabetes

group with the largest associated disease score. As with the Kidney codes 7 out of 10 PGPs outgrew the comparison groups. The growth rates were generally smaller than growth for the kidney codes and a few were negative.

Heart Failure is a chronic disease which has little coding flexibility in ICD-9 and the CMS-HCC risk adjustment model. Unlike diabetes, which has 5 levels in cost severity, Heart Failure has only one. It is a disease that lends itself to a management program and the PGPs might have been expected to wish to attract this population. Many PGPs implemented a Heart Failure program as part of their Demonstration interventions (see Chapter 4). In growth of coding frequency, however, only half the PGPs showed greater growth than the comparisons. In this case PGP 1, with lower growth in the other illustrative cases, had higher growth; PGP 10, stronger in the other 3 cases, showed a reduction. With little flexibility the coding intensity growth was modest in most cases. PGPs 7 and 8 showed atypically high growth both absolutely and relatively.

Looking only at these four disease classes we see that PGPs 1, 5, and 10 have shown lower growth than the comparisons in three of the four; PGPs 2, 3, 4, 7, 8, and 10 have shown higher growth in at least three.

Overall Effect of Code Changes. Another way of viewing the effect of coding is to sum up the changes related to coding rates over all the groups. In this case the frequency changes are weighted by the relative factors for each HCC in the risk adjustment model. Diseases with higher cost implications have a greater impact as the coding frequency changes. Noting that 1.000 is the relative risk factor for the average beneficiary, the marginal contribution to the relative risk factor for the Renal Failure group is 0.618, for Diabetes in HCC15, 0.302, for HCC81, AMI, 1.893, and for HCC80, Heart Failure, 0.433. The coding changes for Renal Failure will have a large impact related to the magnitude of the changes and the high weight. Heart attacks are relatively infrequent but have a large weight related to the very high likelihood of hospitalization and a form of surgery.

In *Tables 11-10a and 11-10b* the net effect of coding rate changes for each group, over the two-year interval, is viewed as the percentage change in the component of the risk score related to disease coding. In nine out of ten cases the apparent growth in the average risk of the PGPs exceeded that of the comparison groups. The average two-year growth over the 10 PGPs was 8.6 percent, whereas the comparison groups' apparent average growth was 5.1 percent. If the one exception site is omitted the averages are 9.1 percent and 4.9 percent, respectively. These are substantial differences in averages. The variation in the growth rate differences across sites is large, with one site at 13.3 percent for the PGP and 3.9 percent for the comparison.

In the current computation of whether or not a PGP had achieved a target, the full amount of these differences is included. To the extent they capture true changes in populations this is proper. To the extent they capture changes in coding practices adopted by PGPs they should be adjusted. The analysis of changes by disease gives mixed evidence. The interviews with PGPs, however, lend credence to the view that changes in coding practice played a part in the average coding growth disparities observed.

Table 11-10a
Percentage change in rate of coding of four HCC groups,
base year to performance year two

PGP 1 PGP	PGP 1 Comp	PGP 2 PGP	PGP 2 Comp	PGP 3 PGP	PGP 3 Comp	PGP 4 PGP	PGP 4 Comp	PGP 5 PGP	PGP 5 Comp
3.6%	6.8%	10.2%	3.6%	3.2%	0.3%	10.2%	5.7%	8.5%	4.5%

Table 11-10b
Percentage change in rate of coding of four HCC groups,
base year to performance year two

PGP 6 PGP	PGP 6 Comp	PGP 7 PGP	PGP 7 Comp	PGP 8 PGP	PGP 8 Comp	PGP 9 PGP	PGP 9 Comp	PGP 10 PGP	PGP 10 Comp
9.9%	8.7%	13.3%	3.9%	7.0%	4.7%	10.7%	7.4%	9.0%	5.4%

NOTE:

HCC = Hierarchical Condition Category

SOURCE: RTI analysis of Demonstration data.

11.5.4 Changes in Risk Scores in Pre-Demonstration and Demonstration Periods

Another approach to examining changes in coding and the risk scores is to compare score changes for the years prior to the Demonstration and during the Demonstration period. In this analysis we look for a change in the rate at which scores were growing in the period before and during the Demonstration. One might expect year-to-year increases would be larger when there was an incentive for coding or a PGP initiated a new program to attract a sicker population.

Tables 11-11a and 11-11b show the PGP and comparison group mean concurrent risk scores for each PGP site and year of the Demonstration. The growth rate in risk score between BY and PY5 ranged from 11.7 to 25.5 percent for the PGPs, and ranged from 4.2 to 16.5 percent for their comparison groups. Thus it is clear that the PGP risk scores grew faster during the Demonstration than the comparison group risk scores. The difference in risk score growth rates ranged from 1.4 to 17.3 percent. This can be seen graphically in *Figure 11-3*, which compares the risk ratio for the assigned and comparison beneficiaries. The risk ratio between BY and a performance year is the risk score in the performance year divided by the risk score in the base year. It is equivalent to one plus the growth rate. As shown in the figure, the gap between the risk ratios for the PGPs versus the comparison groups widens as the Demonstration progresses.

It is noteworthy that the differential growth rate in risk scores between the PGPs and their comparison groups existed prior to the start of the Demonstration. This can be seen in *Figure 11-4*, which shows the risk ratios between 2001 and PY 5. As seen in the figure, the risk ratio for the PGP and comparison group widens during the pre-Demonstration period, as the degree of separation becomes more pronounced during the Demonstration period.

Table 11-11a
Risk scores base year to performance year five—Assigned beneficiaries

PGP Name	BY	PY1	PY2	PY3	PY4	PY5	Growth BY to PY5
PGP 1	0.871	0.876	0.902	0.966	0.998	0.985	13.1%
PGP 2	0.991	1.042	1.086	1.095	1.118	1.109	11.9
PGP 3	0.972	0.972	1.011	1.039	1.071	1.086	11.7
PGP 4	0.850	0.900	0.937	0.930	0.938	0.958	12.8
PGP 5	1.026	1.081	1.103	1.205	1.265	1.258	22.6
PGP 6	0.916	0.960	1.004	1.055	1.141	1.149	25.5
PGP 7	0.904	0.964	1.023	1.049	1.053	1.111	22.9
PGP 8	0.821	0.869	0.872	0.923	0.940	1.011	23.1
PGP 9	0.897	0.930	0.988	1.025	1.027	1.044	16.4
PGP 10	1.308	1.328	1.415	1.500	1.483	1.491	14.0

Table 11-11b
Risk scores base year to performance year five—Comparison group

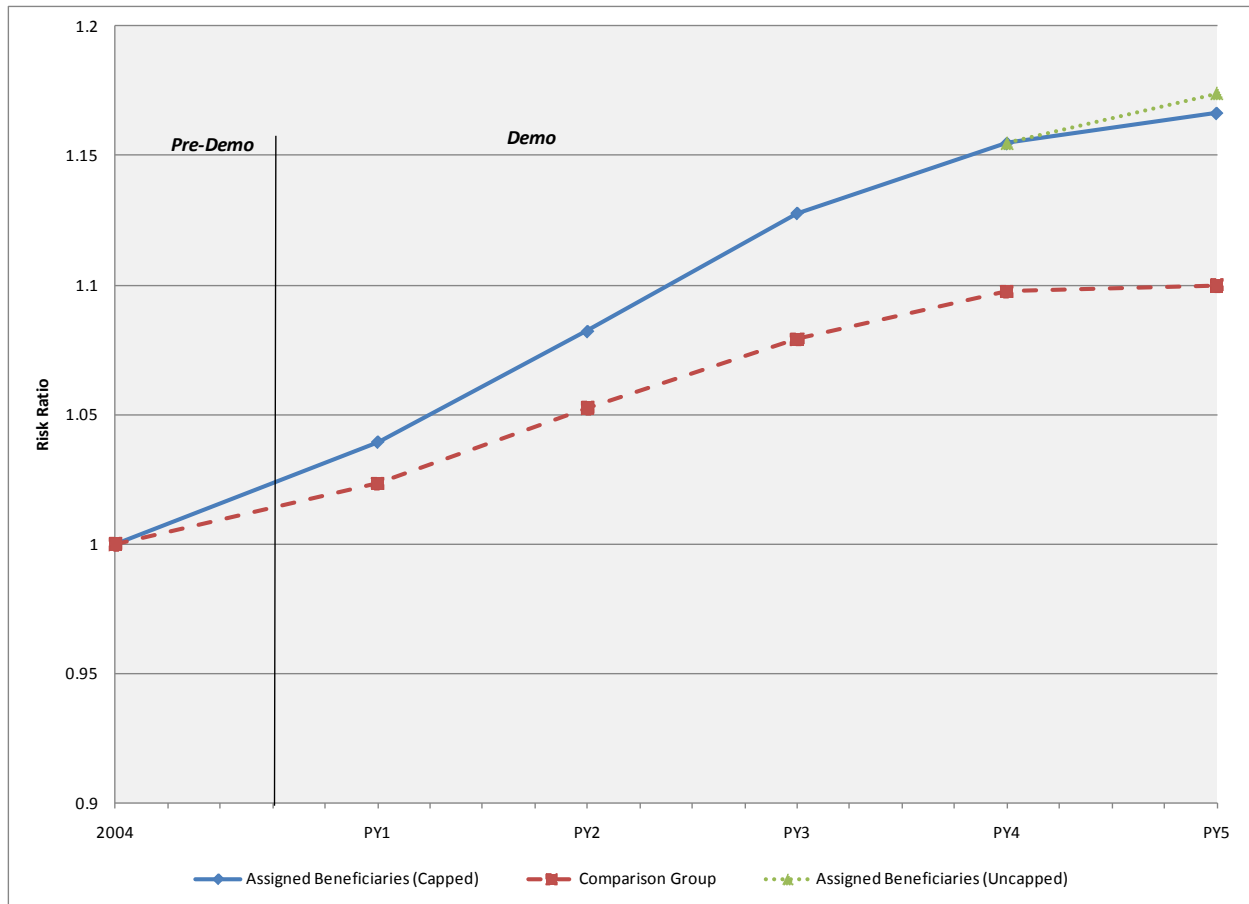
PGP Name	BY	PY1	PY2	PY3	PY4	PY5	Growth BY to PY5	CG Minus AB Growth
PGP 1	0.772	0.801	0.827	0.849	0.864	0.862	11.7%	1.4%
PGP 2	0.735	0.758	0.764	0.770	0.783	0.781	6.3	5.6
PGP 3	0.936	0.930	0.938	0.975	0.981	0.976	4.2	7.5
PGP 4	0.990	1.047	1.051	1.076	1.086	1.067	7.7	5.1
PGP 5	0.958	0.961	0.996	1.008	1.015	1.008	5.2	17.3
PGP 6	0.757	0.780	0.825	0.841	0.874	0.867	14.4	11.1
PGP 7	0.940	0.941	0.975	1.012	1.004	1.043	10.9	12.0
PGP 8	0.888	0.906	0.932	0.971	1.011	1.035	16.5	6.6
PGP 9	0.898	0.934	0.968	0.972	0.990	0.989	10.1	6.3
PGP 10	1.018	1.036	1.075	1.120	1.141	1.147	12.7	1.3

NOTE:

*Risk scores do not reflect the cap put in place during PY5.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 11-3
Risk ratio¹ of assigned beneficiaries vs. risk ratio of comparison group,
2004 base year to PY5²



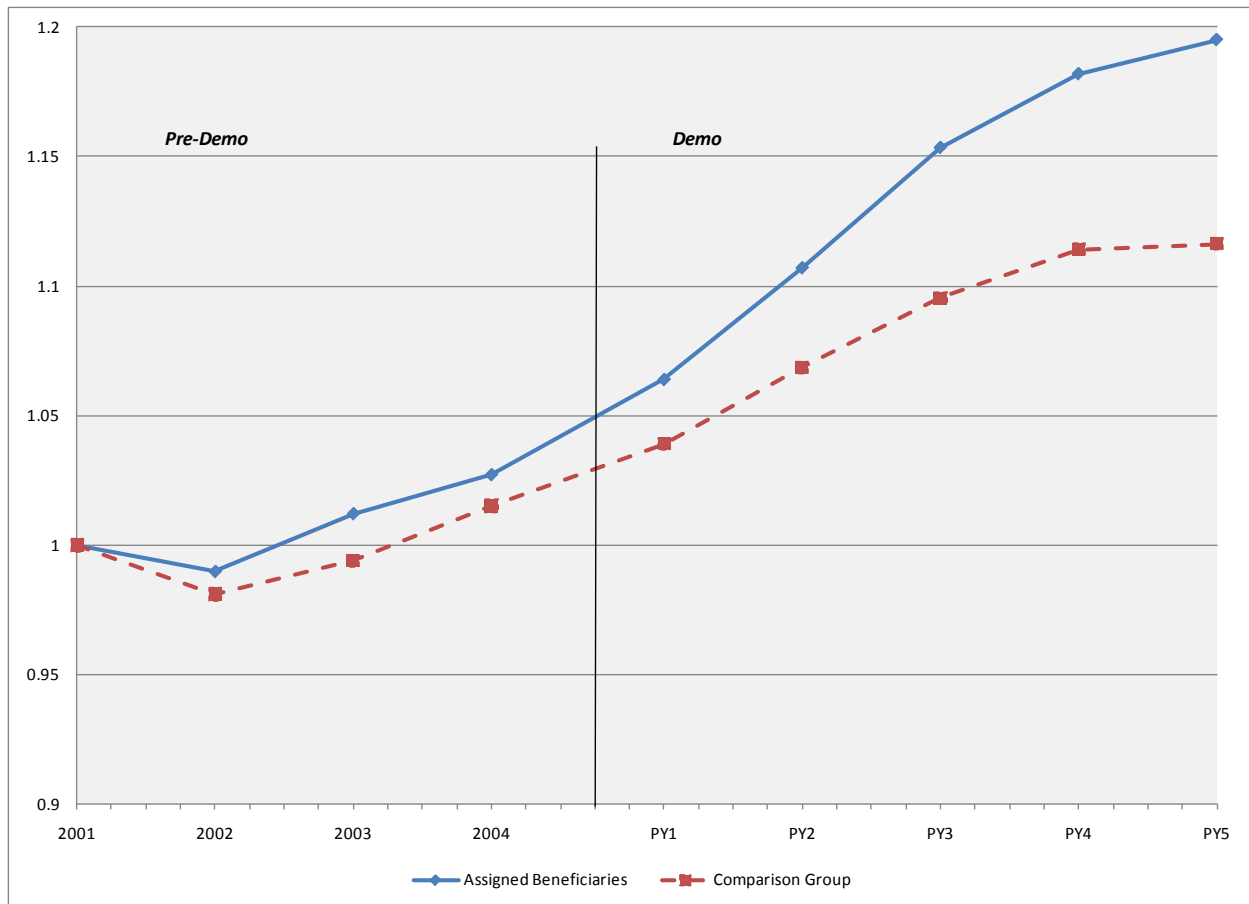
NOTE:

¹ Risk ratio is the ratio of current year risk score to the 2004 risk score. The risk ratio for assigned beneficiaries reflects the risk score cap put in place in PY5 only. The projected uncapped ratio is also shown.

² Unweighted average of the 10 PGPs.

SOURCE: RTI calculations with 2004–2010 Medicare claims and enrollment data.

Figure 11-4
Risk ratio¹ of assigned beneficiaries vs. risk ratio of comparison group,
2001 base year to PY5²



NOTE:

¹ Risk ratio is the ratio of current year risk score to 2001 risk score. The risk ratio for assigned beneficiaries does not reflect the PY5 risk score cap.

² Unweighted average of the 10 PGPs.

SOURCE: RTI calculations with 2001–2010 Medicare claims and enrollment data.

Table 11-12 summarizes the above information across all PGP sites, as well as for PGPs that earned performance bonuses in PY5 versus PGPs that did not earn performance bonuses in PY5. The first row shows that the PGP risk scores were growing at 0.9 percent per year in the three years prior to the start of the Demonstration. The comparison group scores were growing at 0.5 percent per year. The difference in the rates is 0.4 percentage points. In the Demonstration period the growth rates increased substantially to 3.3 percent and 1.9 percent per year, respectively, a difference of 1.4 percentage points. This shows the PGP score growth was not only higher but increased more in going from the pre-Demonstration period to the Demonstration period.

Table 11-12 also includes this information for PGPs that shared in savings in PY5 and PGPs that did not share in savings in PY5. On average, for PGPs sharing in savings in PY5, the risk scores were growing at 0.9 percent per year in the three years prior to the start of the Demonstration. The comparison group scores were growing at 0.6 percent per year. The difference in the rates is 0.3 percentage points. In the Demonstration period the risk score growth rates increased substantially to 3.7 percent and 2.6 percent per year, respectively, a difference of 1.1 percentage points per year growth during the demonstration, compared to 0.3 percentage point difference before.

For PGPs not sharing in savings in PY5, prior to the start of the Demonstration risk scores grew at a rate of 0.9 percent per year, while the comparison group scores were growing at 0.5 percent per year, a difference of 0.4 percentage points. However, in the Demonstration period, the risk score growth rates for the PGPs not sharing in savings in PY5 were 3.0 percent per year, and 1.5 percent growth per year for the comparison groups, a difference of 1.5 percentage points.

It thus appears that although the PGPs had higher risk score growth rates than their comparison groups during the Demonstration period, and had higher increases going from the pre-Demonstration period to the Demonstration period, these trends were broadly similar for PGPs that shared in savings in PY5 and those that did not.

11.5.5 Concurrent versus Prospective Risk Scores

Performance payments in the PGP Demonstration were based on comparing PGP target expenditures against PGP actual expenditures for their assigned patients. In the absence of risk adjustment, health status changes over time can cause the comparison of target expenditures and actual expenditures to be invalid. Risk adjustment can partially remove the effect of health status changes over time, and thus creates a more “apples to apples” comparison between target expenditures and actual expenditures

Table 11-12
Average Annual Percent Change in PGP risk score, pre-demonstration and demonstration periods

PGP	Assigned Beneficiaries Pre-Demo (2001-BY)	Assigned Beneficiaries Demo (BY-PY5)	Assigned Beneficiaries Difference (Post-Demo minus Pre-Demo)	Comparison Group Beneficiaries Pre-Demo (2001-BY)	Comparison Group Beneficiaries Demo (BY-PY5)	Comparison Group Beneficiaries Difference (Demo minus Pre-Demo)	Difference in Difference (AB minus CG)
Average	0.90%	3.28%	2.38%	0.51%	1.93%	1.41%	0.97%
<i>Sharing in Savings in PY5</i>							
Average	0.93	3.69	2.76	0.58	2.56	1.99	0.77
PGP 6	3.04	4.68	1.65	1.55	2.76	1.21	0.44
PGP 8	2.52	4.29	1.77	0.19	3.11	2.92	-1.15
PGP 9	1.12	3.10	1.98	1.48	1.95	0.47	1.51
PGP 10	-2.94	2.70	5.64	-0.92	2.43	3.34	2.30
<i>Not Sharing in Savings in PY5</i>							
Average	0.88	3.00	2.12	0.47	1.50	1.03	1.09
PGP 1	0.96	2.54	1.58	-0.46	2.25	2.71	-1.13
PGP 2	0.81	2.31	1.49	-0.30	1.24	1.54	-0.05
PGP 3	1.81	2.25	0.44	1.53	0.84	-0.69	1.12
PGP 4	1.56	2.47	0.91	1.86	1.53	-0.33	1.24
PGP 5	1.31	4.21	2.89	-0.53	1.04	1.57	1.32
PGP 7	-1.18	4.23	5.41	0.75	2.11	1.36	4.05

NOTE:

Risk scores are uncapped for all years.

SOURCE: RTI calculations with 2001–2010 Medicare claims and enrollment data.

A limitation of diagnosis-based risk adjustment is that changes in risk scores over time can not only reflect changes in health status, but also changes in coding patterns. The ideal risk adjuster would only remove the effect of health status changes, and would not be impacted by coding pattern changes.

To adjust costs for health status changes, the Demonstration used Medicare Advantage (MA) Hierarchical Condition Categories (CMS-HCC) risk adjustment model. This model uses demographic information and diagnoses found on administrative claims to assign HCC indicators to each beneficiary. These indicators along with the 2004 per capita cost of each beneficiary were used in a regression model to compute coefficients or weights for each HCC category. Using these weights and the HCC indicators for each beneficiary, costs can be predicted at the beneficiary level during the demonstration period. Since beneficiary costs are predicted for a performance year using the HCC information from that same year, the resulting model is termed a “concurrent” mode. The demonstration uses a concurrent risk model. If the HCC indicators from the prior year were used to predict immediate future year costs, the model would be termed a “prospective” model.

One refinement of the PGP model is to adjust for health status changes using the same prospective risk adjustment models as used in the Medicare Advantage (MA) program. A prospective model may initially dampen the rate of risk score growth due to coding incentives because the claims information used to calculate risk score would be measured in the prior year. Not only is the MA model prospective, but the coefficients or weights are computed for different time periods and use the entire Medicare population. The demonstration risk model was fairly costly to run and maintain; therefore, using an established model may enhance the use of the PGP concept.

Table 11-13 compares concurrent risk scores of the model used in the demonstration to with a prospective risk scores determined with an MA model. The relative growth in PGP concurrent risk scores (relative to the local comparison group) was high for some PGPs (third column on top of Table 11-13 labeled “assigned bene – comparison bene risk score growth”). Between PGP Demonstration performance years 2 and 4, the relative growth rate in concurrent risk scores for three PGPs was over 5 percentage points higher than the risk score growth in the local market area. However, the relative growth rate in PGP prospective risk scores (relative to a national FFS comparison group) was not excessively high for any of the PGPs during this period (third column on bottom of table labeled “assigned bene FFS normalized prospective risk score³⁷”), with none of the PGPs over 5 percentage points.

³⁷ Prospective “FFS normalized” risk scores are prospective risk scores adjusted for coding patterns and population changes in the Medicare FFS population. Therefore, comparing the difference in concurrent risk score growth rates between the PGP and its comparison group with the growth rate in the prospective FFS normalized risk score growth rates for the PGP is appropriate.

Table 11-13a
Concurrent (Demo model) risk scores

PGP	PY2 to PY3*	PY3 to PY4*	PY2 to PY4*	PY2 to PY3†	PY3 to PY4†	PY2 to PY4†	PY2 to PY3§	PY3 to PY4§	PY2 to PY4§
PGP 1	7.17	3.25	10.65	2.65	1.77	4.46	4.52	1.48	6.19
PGP 2	0.80	2.12	2.94	0.80	1.72	2.53	0.01	0.40	0.41
PGP 3	2.74	3.04	5.86	3.85	0.71	4.59	-1.11	2.33	1.27
PGP 4	-0.75	0.90	0.15	2.45	0.86	3.33	-3.20	0.04	-3.18
PGP 5	9.20	5.02	14.69	1.20	0.68	1.89	8.00	4.34	12.80
PGP 6	5.08	8.17	13.66	1.88	4.03	5.99	3.20	4.14	7.67
PGP 7	2.55	0.40	2.95	3.75	-0.73	2.99	-1.21	1.13	-0.04
PGP 8	5.86	1.82	7.78	4.18	4.12	8.47	1.67	-2.30	-0.69
PGP 9	3.78	0.22	4.02	0.47	1.81	2.29	3.31	-1.59	1.73
PGP 10	6.05	-1.14	4.84	4.14	1.95	6.17	1.91	-3.09	-1.33

* Assigned beneficiary risk score growth.

† Comparison beneficiary risk score growth.

§ Assigned beneficiary – comparison beneficiary risk score growth.

Table 11-13b
Prospective (MA model) risk scores

PGP	PY2 to PY3*	PY3 to PY4*	PY2 to PY4*	—	—	—	PY2 to PY3†	PY3 to PY4†	PY2 to PY4†
PGP 1	-2.87	2.93	-0.01	—	—	—	-0.89	1.83	0.92
PGP 2	-1.66	2.31	0.61	—	—	—	0.33	1.21	1.54
PGP 3	-2.02	1.88	-0.17	—	—	—	-0.03	0.78	0.75
PGP 4	-2.50	-0.41	-2.90	—	—	—	-0.51	-1.48	-1.99
PGP 5	-1.46	3.37	1.85	—	—	—	0.53	2.26	2.80
PGP 6	-2.67	2.83	0.08	—	—	—	-0.70	1.72	1.01
PGP 7	-2.36	1.88	-0.53	—	—	—	-0.38	0.79	0.41
PGP 8	-2.21	3.47	1.18	—	—	—	-0.22	2.35	2.12
PGP 9	-1.64	3.31	1.62	—	—	—	0.36	2.20	2.57
PGP 10	1.44	2.28	3.75	—	—	—	3.48	1.15	4.67

NOTES: The assigned beneficiary concurrent risk score growth rates are analogous to the assigned beneficiary unnormalized prospective risk score growth rates. Also, the difference in assigned and comparison concurrent risk score growth rates is analogous to assigned beneficiary FFS normalized prospective risk score growth rates. MA, Medicare Advantage.

* Assigned beneficiary unnormalized risk score growth.

† Assigned beneficiary FFS normalized risk score growth.

SOURCE: RTI analysis of Medicare administrative data.

11.5.6 Stayer Analysis

We tested for diagnostic coding pattern changes during the PGP Demonstration using a “stayer” analysis. The strategy was borrowed from the Medicare Advantage (MA) program, which empirically has shown that MA coding intensity has been substantial, and has incorporated an adjustment for MA coding intensity into the MA risk adjustment methodology (CMS, 2009). The MA program uses a “prospective” risk adjustment system, which means that risk adjustment in Year T is based on diagnoses measured in Year T-1. Stayers in Year T are defined as beneficiaries enrolled in the MA plan in both Years T and T-1. Under prospective risk adjustment, the MA plan controls the diagnostic coding used to measure the Year T stayer’s risk score, because by definition the stayer was enrolled in the MA plan in Year T-1 when the diagnoses were measured for the prospective risk score.

The proposition made by CMS is that if a beneficiary is enrolled in the MA plan for Years T-1, T, and T+1, then we know these beneficiaries are stayers in Years T and T+1, and thus the MA plan will control the diagnostic coding used to calculate the prospective risk scores for Years T and T+1 stayers. Thus if there is a large increase in the prospective risk scores from Years T to T+1 for MA plan stayers relative to FFS stayers,³⁸ then this must be due to MA coding intensity, so goes the proposition made by the MA program (CMS, 2009). Importantly, note that an implicit assumption in this proposition is that health status remains constant.

We conducted a stayer analysis for the PGP Demonstration based on the strategy used in the MA program (CMS, 2009). We identified beneficiaries assigned to the PGP in each year 2006, 2007, and 2008, and then calculated the prospective risk score growth from 2007 to 2008. We repeated this for the Medicare FFS population, which acts as the comparison population. As shown in **Table 11-14**, what we found was the differential in risk score growth rate between PGP and FFS in one year was -0.2 percentage points, implying that coding intensity by the PGPs was less than FFS overall. To test the robustness of these results, we repeated the stayer analysis by identifying beneficiaries assigned to the PGP in each year 2007, 2008, and 2009, and then calculated the prospective risk score growth from 2008 to 2009. We repeated this for the FFS population. What we found was the differential risk score growth rate between PGP and FFS was 0.9 percentage points, implying that coding intensity by the PGPs was more than FFS overall.

³⁸ A FFS stayer in Year T is enrolled in FFS in both Years T and T-1.

Table 11-14
PGP vs. FFS differential in prospective risk score growth for stayer cohorts

PGP	PGP vs. FFS Differential (percentage points) 2007–2008	PGP vs. FFS Differential (percentage points) 2008–2009	PGP vs. FFS Differential (percentage points) Average
PGP 1	-0.2	1.6	0.7
PGP 2	-0.7	0.0	-0.4
PGP 3	-0.8	-0.2	-0.5
PGP 4	-1.1	-1.7	-1.4
PGP 5	0.4	3.1	1.8
PGP 6	1.3	9.4	5.4
PGP 7	-0.6	0.3	-0.2
PGP 8	-0.1	-0.5	-0.3
PGP 9	0.5	-0.7	-0.1
PGP 10	-0.8	-2.6	-1.7
Average	-0.2	0.9	0.3

NOTES:

1. Stayer cohort for 2007–2008 statistics are beneficiaries assigned to the PGP in 2006–2008
2. Stayer cohort for 2008–2009 statistics are beneficiaries assigned to the PGP in 2007–2009
3. Risk scores are FFS normalized prospective risk scores
4. PGP vs. FFS differential defined as PGP minus FFS risk score growth rate
5. Differential is percentage points.

SOURCE: RTI analysis of 2006–2009 Medicare administrative data.

With PGPs having lower risk score growth than the FFS average in 2007-2008, the results from the first cohort imply that the PGPs did not have additional coding intensity compared to FFS, and possibly had less. The results from 2008-2009 are somewhat contradictory to the first cohort, since they imply that the PGPs had non-trivial additional coding intensity, 0.9 percentage points on average, compared to FFS.

In summary, participating in shared savings depends on changes in risk scores as well as expenditures. Section 11.5 focused on the risk scores, which are the result of coding. It was shown that the PGPs generally had higher coding growth during the Demonstration than the comparison groups, and that the overall difference in the risk score growth rates grew from the pre-Demonstration to Demonstration periods. However, these patterns seemed to be similar for PGPs sharing in savings versus those not sharing in savings, and the stayer analysis was inconclusive.

11.6 Medicare Payment Rates

The reduction in Medicare expenditures through reduction of utilization of Medicare-covered health services is an important goal of the PGP Demonstration. Sufficiently high savings per Medicare beneficiary result in performance payments to participating PGPs. As noted elsewhere in this report, changes in per beneficiary expenditures for beneficiaries assigned to participating PGPs, however, can be the result of changes other than in utilization. The subject of this section is the potential effects in Medicare payment policy on the attainment of Medicare savings.

Changes in Medicare payment rates and payment factors can result in savings that are not due to lower utilization. Of particular concern are payment factor changes made since the 2004 base year, which might disproportionately affect the PGPs and the local comparison groups. The demonstration savings methodology does not adjust for such changes. Adjustments for patient casemix (i.e., risk adjustment) were made because of the possibility that PGP and beneficiary behavior, as well as chance, might account for changes in relative casemix between the PGP and comparison populations.

These changes in payment rates and rules usually affect all providers belonging to a specific subclass such as all teaching hospitals or all hospitals in a given metropolitan area. In the PGP Demonstration, Medicare expenditures from nearly all types of providers were used in determining Medicare savings. Each type of provider (e.g., hospitals, physicians, skilled nursing facilities, etc.) is subject to its own payment system. Each of these payment systems has many components: the treatment (e.g., procedure or hospitalization) given patients, provider characteristics, and providers' geographic location. While many of the annual changes in payment rates and rules can offset each other, it is also possible that the changes can reinforce each other so that the cumulative changes might have a large impact on Medicare savings attained by the participating PGPs in the Demonstration. Unless there have been no year-to-year changes whatsoever, the direction of the combined impact of changes to various payment systems is usually unknown *a priori* and has to be ascertained empirically.

Routine changes in Medicare payment rates, such as the Geographic Practice Cost Indices (GPCIs) used to adjust physician payments, and rules that favored the PGP participants could have resulted in larger performance payments than would be justified by changes in utilization alone. Conversely, payment rate changes could have adversely affected the participants by lowering Medicare savings to the point that performance payments were not achieved. Savings can be affected by more than changes in payment rates and rules. For instance, the type of setting in which patients are treated might change between the base year and performance year. For example, a participating PGP might decide to admit a greater share of their patients to community hospitals instead of teaching hospitals. Since community hospitals likely cost less than teaching hospitals for the same DRG, this substitution would likely result in savings to a PGP that is not related to lower utilization. Another example would be a change in the place of service for routine examinations (e.g., evaluation and management visits) from rural health clinics (RHCs) to physician offices (or vice versa). In this case, the practice expense in physician offices would likely be lower than in a health clinic; consequently, savings could accrue to a PGP purely as a result of a change in the site of service.

This section, then, considers two issues: (1) the effects of annual changes in Medicare “prices” and (2) analytic challenges arising from different sites of care for, ostensibly, the same services (e.g., routine examinations).

11.6.1 Inpatient Prospective Payment System

To examine the impact of Medicare payment rates on the Demonstration, we focus on the prices paid under Medicare’s inpatient prospective payment system (IPPS) for short-term, acute-care hospitalizations. IPPS hospital prices were selected for analysis because IPPS hospital expenditures account for the largest share of Medicare expenditures at 40 percent. Excluding outlier payments, the Medicare price per discharge consists of the sum of the operating (OP) and capital (CP) payments. Each part of the OP and CP components of the Medicare price per discharge consists of subcomponents. These include standardized payments that are adjusted by wage and non-labor input price indices, indirect medical education (IME) adjustments, disproportionate share of poor patients (DSH) adjustment, and, for capital payments, a large urban area adjustment. Additionally, there is a common subcomponent of OP and CP, the DRG weight for a discharge. Unlike the other OP and CP subcomponents, the DRG weight does not vary by hospital and/or geographic location.

The calculation of relative prices, using MedPAR inpatient claims, was as follows. First, for each IPPS discharge, outlier payments were subtracted from total Medicare payments (which excludes the so-called “pass-throughs”). The resulting amount was divided by the DRG weight. The result, for each hospital, is its basic price. Second, for each time period (BY, PY 1 and PY 2), the mean price for assigned beneficiaries and the mean price for comparison beneficiaries were calculated. Third, the relative price for each time period was calculated by dividing the mean price for assigned beneficiaries by the mean price for the comparison beneficiaries.

Changes in relative prices reflect changes in all components of the IPPS payment system other than the DRG weights. Changes in the relative prices also might result of other changes such as the possible substitution of admissions from teaching hospitals to community hospitals—this type of change is discussed below.

As shown in *Table 11-15*, the annual changes in relative IPPS prices for the PGPs and comparison groups were usually less than one percent in either direction. Only one participating PGP (PGP 2) experienced an annual change in relative IPPS prices greater than two percent. On average, the cumulative changes in relative IPPS prices from the base year to PY2 were close to zero. Four of the ten participating PGPs experienced cumulative changes in relative IPPS prices of less than one percent in either direction. Of the five participating PGPs that experienced cumulative changes in relative IPPS prices between one and two percent, two had positive cumulative changes while the other three had negative cumulative changes. At 4.78 percent, only PGP 2 experienced a cumulative change in relative IPPS prices greater than two percent.

Since payments for IPPS hospitalizations constitute only about 40 percent of the Medicare expenditures on health care service received by Medicare beneficiaries, the changes in relative Medicare (trust-fund) payments per beneficiary were smaller than the changes in relative IPPS prices. All but one of the annual changes in relative trust-fund payments was less than one

Table 11-15
Changes in relative IPPS prices and relative Medicare payments per beneficiary

PGP	BY*	PY1*	PY2*	BY†	PY1†	PY2†	BY§	PY1§	PY2§	BY to PY1‡	PY1 to PY2‡	BY to PY2‡	BY to PY1**	PY1 to PY2**	BY to PY2**
PGP 1	4,192	4,395	4,569	4,136	4,295	4,530	1.014	1.023	1.009	0.96	-1.43	-0.48	0.39	-0.57	-0.19
PGP 2	5,426	5,870	6,332	4,816	5,059	5,363	1.127	1.160	1.181	2.98	1.74	4.78	1.19	0.70	1.91
PGP 3	5,285	5,536	5,637	5,263	5,524	5,681	1.004	1.002	0.992	-0.18	-1.00	-1.18	-0.07	-0.40	-0.47
PGP 4	4,557	4,831	4,933	4,973	5,244	5,317	0.916	0.921	0.928	0.52	0.72	1.25	0.21	0.29	0.50
PGP 5	4,606	4,828	5,104	4,229	4,508	4,735	1.089	1.071	1.078	-1.67	0.64	-1.04	-0.67	0.26	-0.42
PGP 6	4,459	4,673	4,773	4,319	4,550	4,641	1.033	1.027	1.028	-0.54	0.14	-0.39	-0.22	0.06	-0.16
PGP 7	5,450	5,535	5,794	6,032	6,163	6,364	0.904	0.898	0.910	-0.60	1.37	0.76	-0.24	0.55	0.30
PGP 8	4,786	5,008	5,118	5,285	5,540	5,626	0.906	0.904	0.910	-0.17	0.63	0.46	-0.07	0.25	0.18
PGP 9	4,166	4,419	4,589	4,390	4,657	4,806	0.949	0.949	0.955	0.00	0.62	0.62	0.00	0.25	0.25
PGP 10	6,409	6,544	6,684	5,544	5,772	5,898	1.156	1.134	1.133	-1.92	-0.05	-1.97	-0.77	-0.02	-0.79

NOTES:

*Mean IPPS “price” per discharge, assigned beneficiaries, with discharge price equal to Medicare IPPS payment on inpatient claims minus outliers and then divided by the DRG weight. Also excludes the inpatient deductible.

†Mean IPPS “price” per discharge, comparison population, with discharge price equal to Medicare IPPS payment on inpatient claims minus outliers and then divided by the DRG weight. Also excludes the inpatient deductible. Service area county values are weighted by county share of assigned beneficiaries.

§Relative prices (the relative price is the “PGP” price divided by the “comparison” price).

‡Percent change in relative IPPS prices.

**Percent change in relative Medicare payments per beneficiary, equal to the percent change in relative IPPS prices multiplied by 0.40, the hospital share of Medicare payments per beneficiary.

SOURCE: RTI analysis of 2004–2007 Medicare claims.

percent in either direction, with PGP 2 being the exception. On average, the cumulative effects of changes in relative IPPS prices on relative trust-fund payments from the base year to PY2 were close to zero. At 1.91 percent, only PGP 2 experienced a cumulative change greater than one percent.

Only PGP 2 experienced a large change, annual or cumulative, in relative IPPS prices. And even for PGP 2, the effect of the change in relative IPPS prices on relative Medicare payments was muted. These findings indicate that changes in relative IPPS prices had little effect on the ability of participating PGPs to achieve Medicare savings *in totum*. And, in the case of PGP 2, which was most at a disadvantage due to relative price changes, PGP 2 was still able to generate sufficient Medicare savings to earn Demonstration performance payments. The significance or impact of these relative price changes can be judged against the two percent payment thresholds. That is, changes less than two percent, in absolute value, can be considered as “statistical noise” just as differences between target and actual expenditures are.

Changes in mean prices and relative prices might not just reflect changes in actual standard payments. In particular, changes in the distribution of hospitals to which patients were admitted can affect mean prices and, hence, relative prices. This can happen because the PGP Demonstration provides incentives for participants to substitute admissions from higher cost teaching hospitals (normally higher payment resulting from higher IME and DSH) to lower cost community hospitals. For example, suppose a participating PGP, after the base year, decides to admit a greater share of their patients to community hospitals instead of teaching hospitals. Absent changes in the components of IPPS payments, a shift in the share of admissions from teaching to community hospitals would lower the relative price during the course of the demonstration, resulting in savings to the PGP. Another way the relative price could fall is through a greater share of comparison patients being admitted to teaching hospitals while the distribution of assigned beneficiaries among teaching and non-teaching hospitals does not change. Unlike the first example, where the participating PGP was proactive in admitting its patients to less-expensive hospitals, in this second example the participating PGP is either passive or is resisting a trend to admit more patients to teaching hospitals.

The foregoing analysis, then, has several limitations. One limitation is that it examined changes for just two time periods. Another limitation is that changes in relative prices might also reflect changes in the distribution of hospitals to which patients were admitted. That is, the changes in relative prices presented here include the effects of changes in the distribution of hospitals to which patients were admitted as well as changes in Medicare payment rates. The results presented here, then, do not purely represent changes in Medicare payment rates. A more comprehensive analysis would need to be conducted to disentangle the effects of changes in relative prices from the effects arising from changes in the distribution of hospitals to which patients were admitted.

11.6.2 Other Medicare Payment Policies: Site of Service

The preceding section was concerned with changes in the relative prices of IPPS discharges. IPPS prices include IME, DSH, and geographic payment adjustments. Payments for physician, hospital outpatient, and other services also have geographic payment adjustments. Some of these other services can be performed in multiple settings such as physician offices,

hospital outpatient departments, RHCs, Federally Qualified Health Centers (FQHCs), and ambulatory surgical centers (ASCs). Medicare typically does not pay providers the same amount across settings for these otherwise identical services because of site of service payment differences. Analyses of the impact of payment policies for services that can be provided in multiple settings can be challenging, as it is difficult to reconcile and standardize service definitions and payment amounts across multiple settings between the base year and performance years. The complexity and challenges of analyzing these systems is the subject of this section.

The site of service payment differential can be explained by the following example of an office or outpatient visit for a new patient (CPT code 99202). Ignoring geographic price adjustments, payment for this service if provided in a physician office would have been \$67.08, of which \$31.83 (47 percent) is for the practice expense (PE) component (see box at right). The same service provided in a hospital outpatient department would have been paid \$107.09, almost 60 percent greater than if the service had been provided in a physician office. The difference between the two payments, \$40.01, reflects the OPPS facility payment of \$60.48, less the difference between the physician fee schedule practice expense, ($\$31.83 - \$11.37 = \$20.46$).

Medicare payments for physician services and facility use, 2007

Categories	RVUs (99202) Office	RVUs (99202) Facility	Payments Office	Payments Facility
Physician Fee Schedule components				
Work	0.88	0.88	\$33.35	\$33.35
Practice expense	0.84	0.30	31.83	11.37
Malpractice	0.05	0.05	1.89	1.89
Total RVUs	1.77	1.23	—	—
Total physician payments	—	—	\$67.08	\$46.61
OPPS facility fee, 2007	—	—	—	60.48
Physician plus OPPS payments	—	—	67.08	107.09
Site of service differential	—	—	—	40.01
PE+facility share of payments	—	—	47%	67%

When measuring the changes in relative prices, adjusting the payment for site of service payment differences (could restate in terms of only one, the physician fee schedule; as illustrated in the above example) is the preferred theoretical approach. That is, it is not sufficient to only rely on changes in the fee schedules. In some cases, this restatement is relatively easy. When all payments are based on a common metric, such as a CPT code, the restatement can be readily done. However, there are many challenges in restating payments, which makes the theoretically preferred approach impossible to implement, as shown below.

If a common code or payment schedule does not underlie services provided in multiple settings, then there are additional analytic challenges. In particular, determining or standardizing the unit of payment is fraught with analytic problems. The foregoing example of CPT 99202 can be used to illustrate an aspect of this problem. RHCs report services using Uniform Billing (UB)

revenue center codes. The UB revenue center code 0521 (clinic visit at a RHC/FQHC) encompasses CPT codes 99201 through 99215. Because resource use and payments for CPT codes 99201 through 99215 differ greatly, it is difficult to perform any types of comparisons between individual CPT procedures and RHC procedures.

Another complication is that Medicare payments for RHC (and FQHC) services are cost-based rather than fee-based. This makes it difficult to control for changes in geographic payment adjustments.

Another problem is how to adjust for physician practices that are converted into Provider-Based Clinics (PBCs). A physician practice purchased by a hospital can receive PBC status if the practice site is within 35 miles of the hospital's main campus as well as satisfying financial and administrative integration and other criteria. One of the participating PGPs is a PBC and would have received the \$107.09 combined payment (not including geographic price adjustments) for CPT 99202 services discussed earlier. The existence of PBCs in other participating PGPs and at the providers for their comparison groups is not known. And, if some of them have PBCs, it is not known when the conversion from an independent PGP to PBC occurred (affects comparisons between the base year and performance years).

Because of these complexities, it was not possible to determine the impact of the site of service payment differentials.

11.7 Alternative Quality Performance Methodology

We also tested different quality performance benchmarking methodologies using PY4 quality measurement results (as PY5 quality results were not yet available at the time we conducted the analysis for this chapter).

Under the current PGP Demonstration methodology, performance targets for each topic were set equal to the lowest of the following three values:

1. The higher of 75 percent compliance OR the Medicare HEDIS mean for the measure
2. The 70th percentile Medicare HEDIS level for the measure
3. The quality improvement target, which is defined as a 10 percent reduction in the gap between base year performance and 100 percent compliance.

In this section, we explored three additional quality performance methodologies: (1) using prior year results as quality improvement targets, (2) setting the benchmark to be the highest of all possible targets, and (3) using “all-or-nothing” or “graduated” composite scores to assess quality achievements. We compared the performance using these alternative methodologies with current results, and studied how changing the methodology would affect the PGPs' quality performance.

11.7.1 Setting Yearly Quality Improvement Targets

The quality improvement target used in the PGP Demonstration was set during the base year and remained constant for the duration of the Demonstration. This had a couple of implications. First, a PGP could have intentionally underperformed during the base year so that performance targets would be easier to achieve in subsequent performance years; this would be especially easy to do in the chart-based quality measures, since we could not verify by claims analysis or by our auditing methodology. For example, a group with a baseline score of 30 percent in a measure would only have to attain 37 percent or more in each subsequent year to be credited for this measure. Second, if the PGP had a low quality result for a measure during the base year, the group would not have to perform at a high level in any subsequent years to reach the performance threshold (i.e., 37 percent from our example). One way of encouraging progressively higher performance would be to update the quality improvement target based on the previous year's performance. That is, during Performance Year 1, the quality improvement target would still be equal to a 10 percent reduction in the gap between base year and 100 percent compliance. For Performance Year 2, however, the target would be equal to a 10 percent reduction in the gap between Performance Year 1 and 100 percent compliance. Using this alternative methodology, a PGP's quality improvement targets would increase over the course of the Demonstration. For those measures with quality improvement targets lower than (A) or (B) above-- hence the QI target selected as the threshold to surpass -- this means that in order to "pass" the topic, the PGPs would have to improve their performance over the course of the Demonstration.

Results of our sensitivity analysis using yearly rebasing, by updating the PY4 quality targets to be based on PY3 results instead of the base year, are presented in *Figures 11-5* and *11-6*. Across all measures, PGPs achieved on average 11 percent fewer targets than under the original methodology. Only PGP 7's overall performance was not impacted by the rebasing, but both PGP 4 and PGP 9 would have experienced sharp declines in their performance using the rebasing method (attaining 6 and 7 fewer quality targets, respectively). For these two practices, for example, the percentage of quality points earned, then, would be 68 percent instead of 96 percent in PGP 4 and 74 percent instead of 98 percent in PGP 9. This was notable as PQRI incentives and, if applicable, PGP shared savings bonuses, would have been largely and negatively impacted. In fact, under this rebasing QI method, the percent of points earned (hence, incentives earned) across the ten PGPs would be 12 percent lower than the original PGP method.

We also found some differing results among the topics using the quality improvement rebasing method. For example, the PGPs' perfect performance on HF topics was not affected by the change in methodology, but PGPs met an average of 27 percent fewer targets in the DM topic with the rebasing QI method (data not shown). For the CAD topic, PGPs met an average of 12 percent fewer targets using the alternative method. PGPs also performed worse in the HTN and PC topics when we rebased the quality improvement target from PY3.

Figure 11-5
Number of quality targets met in PY4, original PGP method vs. rebasing QI targets

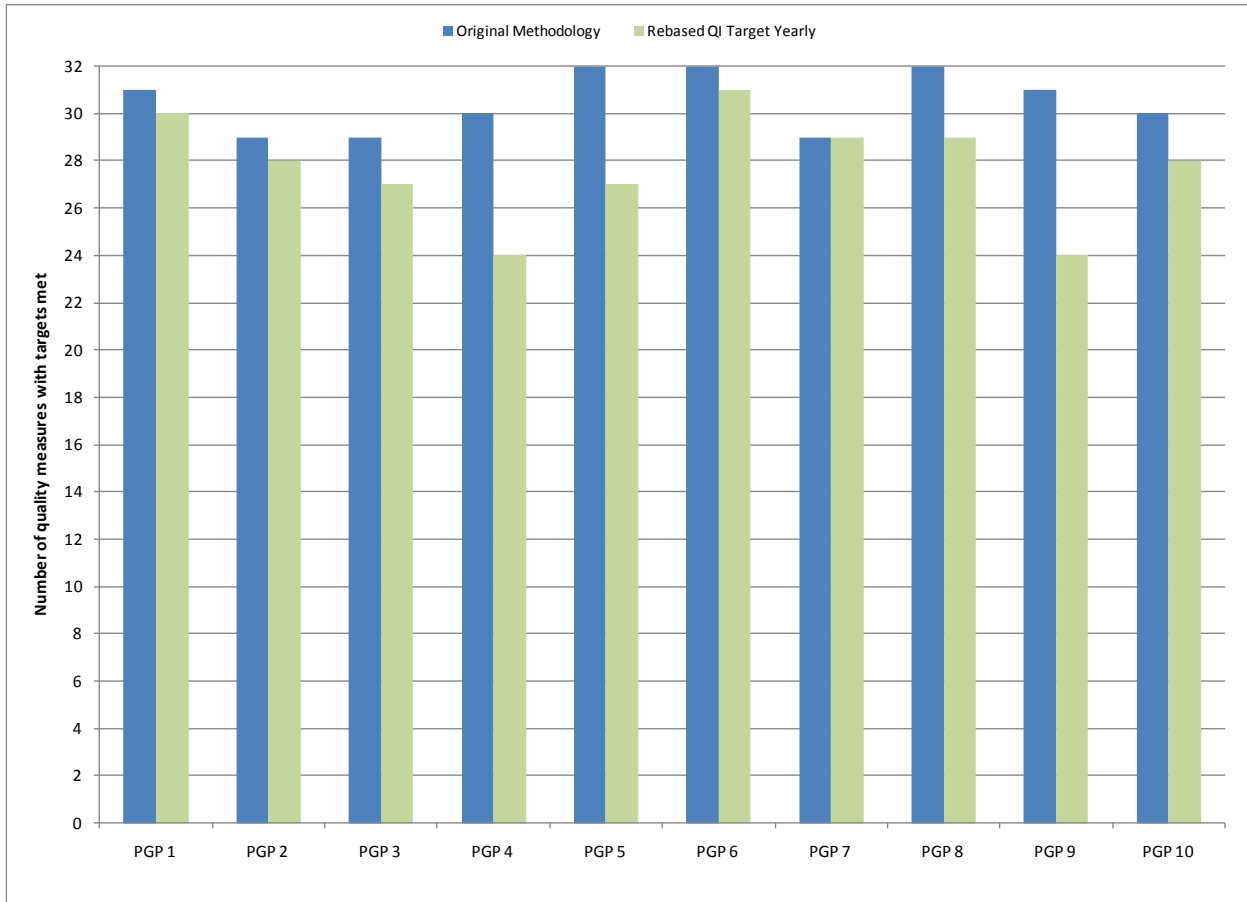
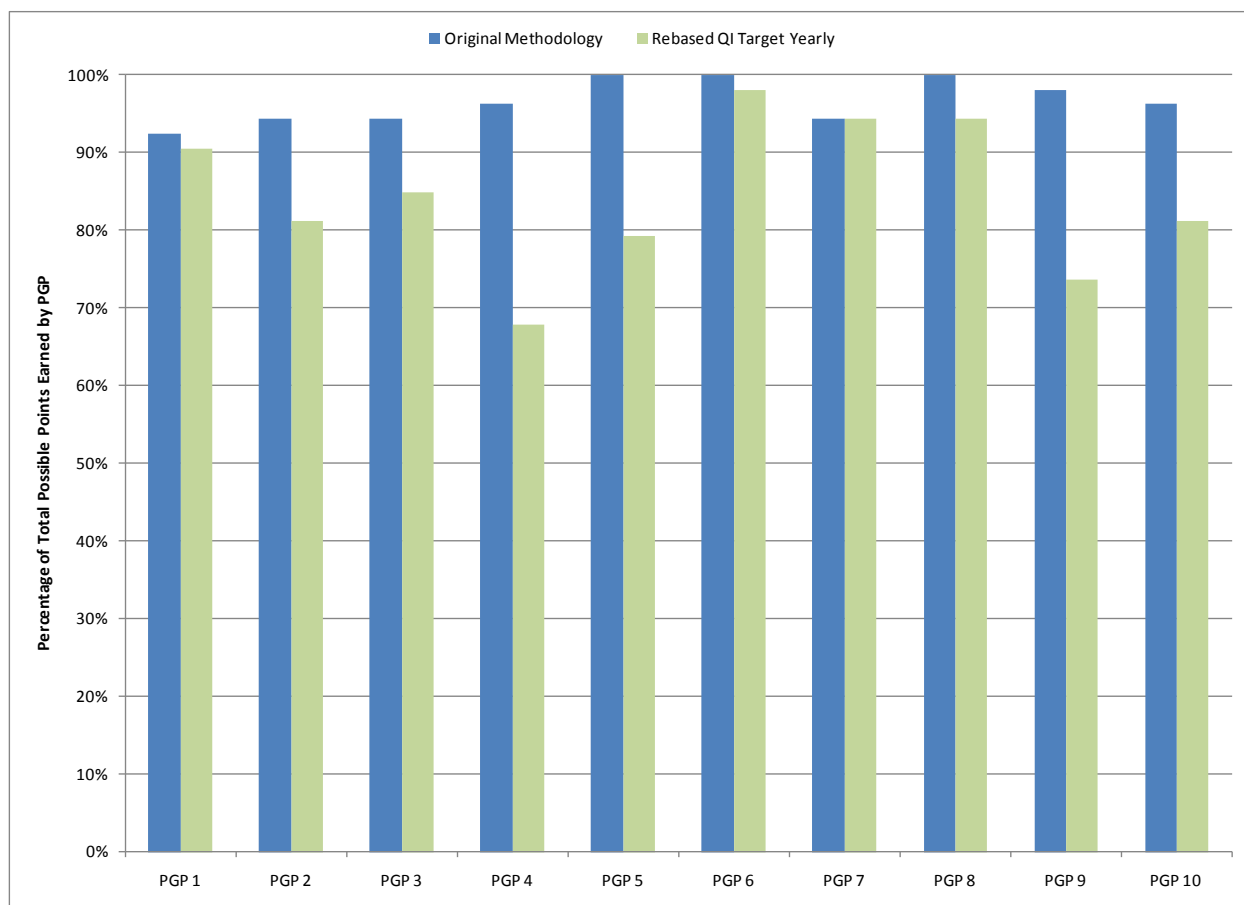


Figure 11-6
Percent of total points earned in PY4, original PGP method vs. rebasing QI targets



11.7.2 Setting Benchmark to be Highest of All Possible Targets

In contrast to the current methodology, which set the performance target as the lowest of the three possible targets (A), (B), and (C), this second alternative methodology would set the performance target as the highest of the three possible targets – thus ‘raising the bar’ to a higher level. Currently, PGPs who performed poorly during the base year, for instance at 50 percent, would only need to perform at 55 percent in all subsequent years to meet their performance target, because the lowest target – the quality improvement target – would remain at 55 percent. Using the revised methodology, PGPs would need to achieve the highest of 75 percent compliance, the Medicare HEDIS mean, the 70th percentile Medicare HEDIS level, or the quality improvement target. Thus, the low 55 percent quality improvement target will no longer satisfy. This methodology also would likely take away the incentive to underperform during the base year.

Results of our sensitivity analysis using the highest pre-specified thresholds are presented in *Figures 11-7* and *11-8*. All PGPs would have performed worse under the alternative method, attaining seven to twelve fewer quality targets. Across all measures, PGPs achieved 36.1 percent fewer targets than under the original methodology, meeting an average of 19.5 out of 32 quality

measures. PGP 2, PGP3, and PGP4 were particularly affected by the change in methodology, with over 40 percent reduction in quality attainment. Financially, this decrease in quality performance translated to the PGPs earning an average of 56 percent of their PQRI and/or shared savings incentives, compared to an average of 97 percent under the PGP Demonstration methodology. PGPs 1, 3 and 4 all would receive less than 50 percent in incentive payments; these sites, along with PGP 9 would experienced a difference of greater than 40 percent in financial incentives compared to what they actually received under the PGP Demonstration.

Figure 11-7
Number of quality targets met in PY4, original PGP method vs. setting to highest target

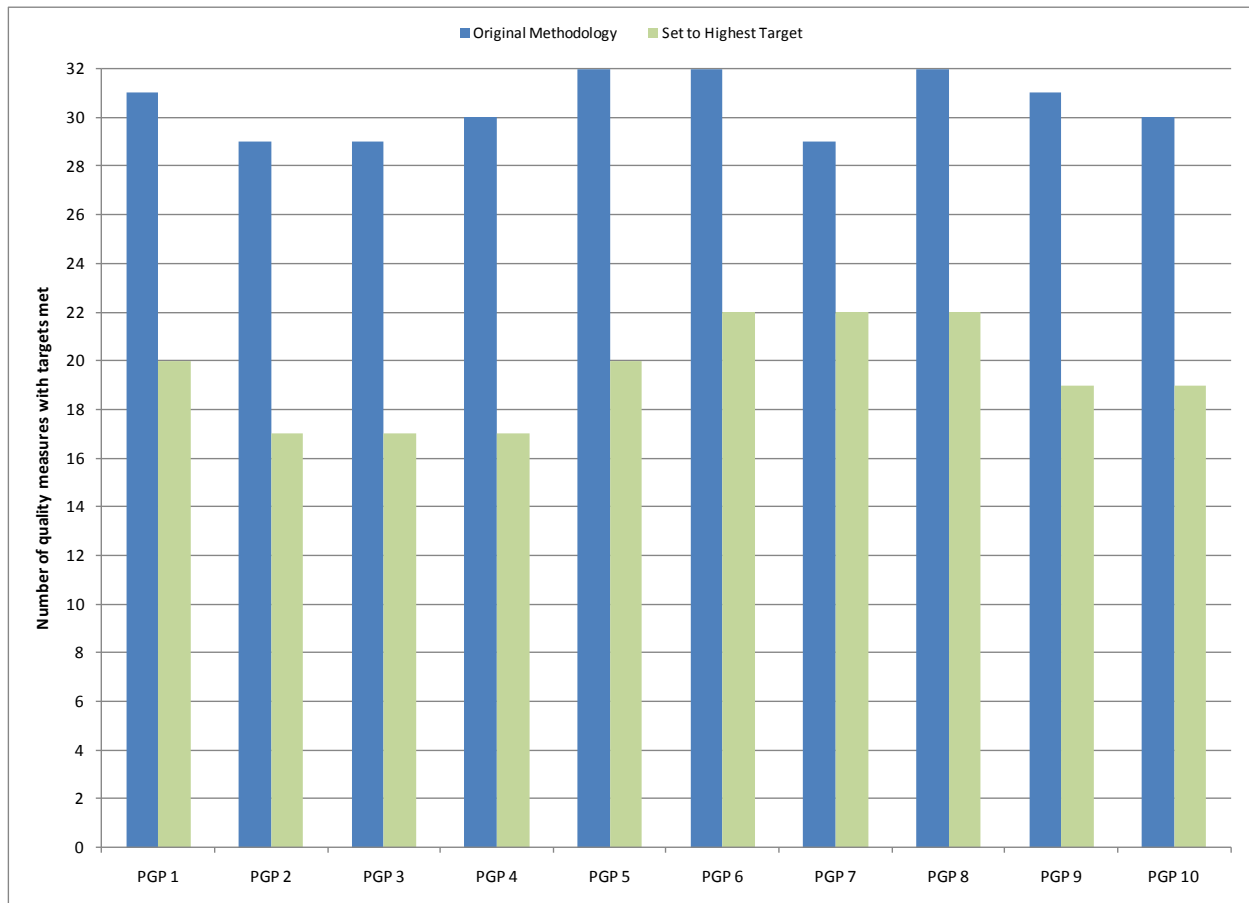
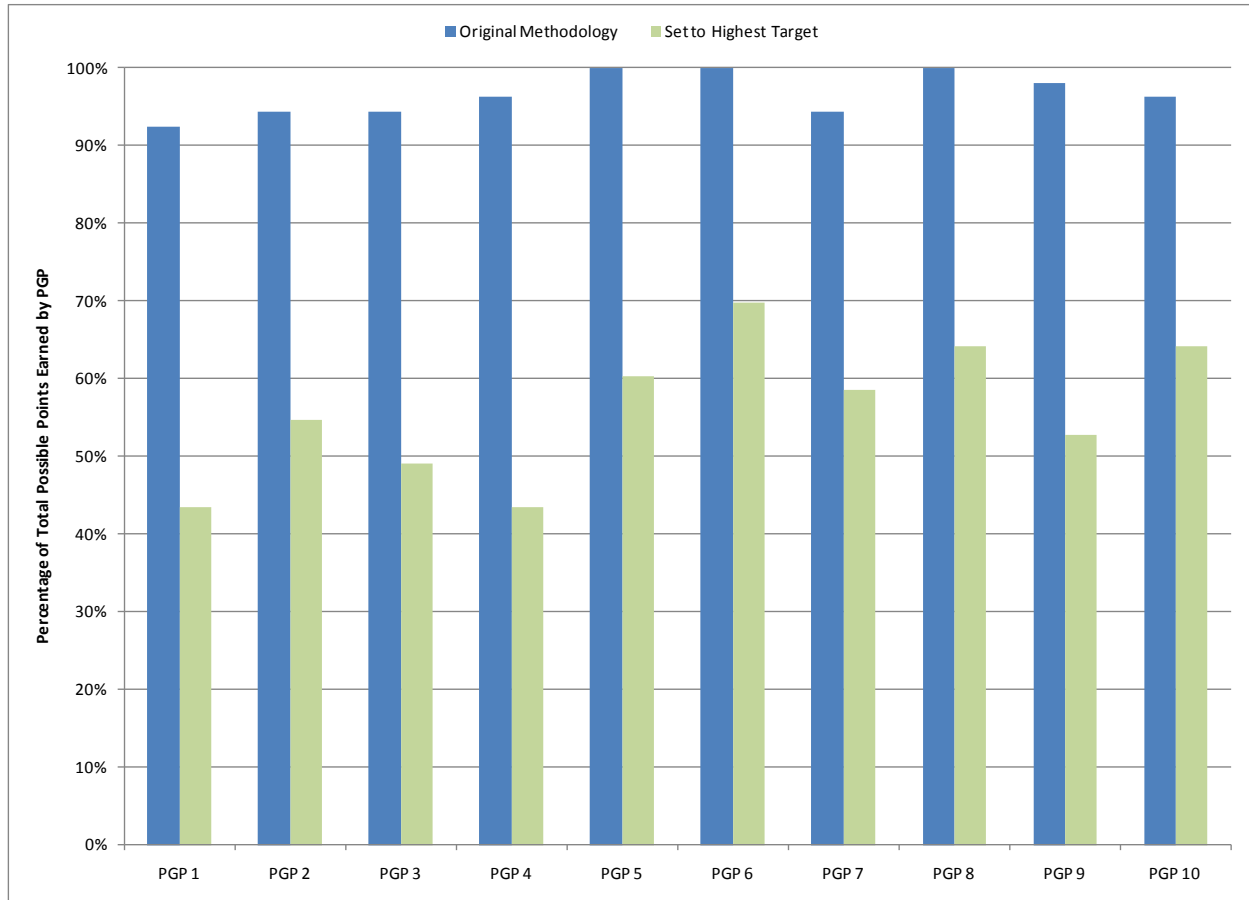


Figure 11-8
Percent of total points earned in PY4, original PGP method vs. setting to highest target



Findings by topic follow the same pattern as in Section 11.5.1. The PGPs’ performance on the HF topic was not affected as much as in the DM or CAD topics. In DM, no PGPs met all 10 quality measure targets using the revised methodology compared to seven out of ten having done so under the original method. PGPs met an average of 52 percent fewer DM topics under the new methodology. In CAD, the PGPs met an average of 40 percent fewer CAD measures under the new methodology

11.7.3 Composite Scoring

Composite scores are created when two or more quality measures are combined into one global score. Composite scores allow for a comprehensive evaluation of the quality of care patients receive. In the case of the PGP Demonstration, composites scores can be created for each condition topic (i.e., each pool of eligible patients). For example, all ten of the diabetes mellitus quality indicators can be combined into one score in order to evaluate whether a group’s diabetes patients received all of the recommended diabetes care. Similarly, a composite score may be calculated from the seven quality indicators in the CAD topic to generate one score for measuring comprehensive CAD patient care. Since the quality indicators selected for use in a topic were based on evidence-based guidelines, PGP consensus, and expert opinions, the

measures included in the composite score would indicate appropriate and comprehensive care for patients with the disease.

For our analyses, we evaluated quality results using the all-or-none composite scoring methodology. The all-or-none composite scores differ substantially from the current PGP Demonstration quality scores because they are patient-level, rather than organization-level, assessments. In this method, each PGP's eligible patient in the topic was assessed for all quality measures in the topic. If the patient met the numerator requirements for all the measures, the patient was considered to have met the composite measure. If the patient had one or more numerator misses in the topic, then they did not meet the composite measure. For this sensitivity analysis for the DM and CAD topics, patients' composite scores were calculated over all ten of the DM measures, and all seven of the CAD measures. For the HF measures, we excluded the every visit measure (HF3 and HF4) when calculating the HF composite score.

Table 11-16 shows the comparison in PGP performance across the DM, HF and CAD topics, based on the original PGP Demonstration method and based on the 'all-or-none' composite scoring method. The original PGP scores for each topic was simply the average of the organization-level scores across the measures in the topic (i.e., the percent of eligible cases in the organization that met the numerator criteria), and the all-or-none composite score is the percent of patients meeting all the numerator requirements in the topic. Here, a target threshold was not necessary. The difference in performance was the most notable in the diabetes topic: the 10 PGPs, on average, achieved 82 percent of the diabetes-related quality indicators under the original method, but really only 20 percent of the PGPs' diabetes patients received all of their recommended care. In fact, the composite results in DM were strikingly low among the PGPs, ranging from 15.1 percent in PGP 4 to only 27.7 percent in PGP 6. Composite results for the HF and CAD topics were higher, with average scores of 67.7 percent (HF) and 64.2 percent (CAD). But the PGPs still fared worse by over 20 percentage points using the patient-level all-or-none composite method compared to the original organizational-level measurement. This finding sheds important light on the quality of care provided to chronically ill patients by the PGPs. The low composite scores in DM, HF and CAD indicated that a large portion of patients with these chronic conditions were not receiving all of their recommended care.

Table 11-16
Comparison of quality performance, using ‘All-or-None’ composite scoring

PGP	Diabetes Mellitus (DM) Topic: Original PGP Demo method (A)	Diabetes Mellitus (DM) Topic: All or none Composite Score (B)	Diabetes Mellitus (DM) Topic: Difference (A) - (B)	Heart Failure (HF) Topic: Original PGP Demo method (C)	Heart Failure (HF) Topic: All or none Composite Score (D)	Heart Failure (HF) Topic: Difference (C) - (D)	Coronary Artery Disease (CAD) Topic: Original PGP Demo method (E)	Coronary Artery Disease (CAD) Topic: All or none Composite Score (F)	Coronary Artery Disease (CAD) Topic: Difference (E) - (F)
PGP 1	82.5%	22.4%	60.2%	94.1%	79.6%	14.5%	93.1%	68.9%	24.3%
PGP 2	82.5%	22.5%	60.1%	90.2%	60.3%	30.0%	88.6%	52.7%	36.0%
PGP 3	81.9%	17.5%	64.4%	92.7%	65.2%	27.5%	91.0%	59.9%	31.1%
PGP 4	77.9%	15.1%	62.8%	87.7%	46.6%	41.1%	91.6%	66.7%	24.9%
PGP 5	85.5%	18.4%	67.0%	89.8%	63.3%	26.4%	89.0%	56.3%	32.7%
PGP 6	85.9%	27.7%	58.2%	92.5%	72.0%	20.5%	93.9%	74.5%	19.5%
PGP 7	83.0%	18.9%	64.1%	93.4%	69.8%	23.6%	94.2%	72.9%	21.4%
PGP 8	83.6%	16.3%	67.4%	94.3%	74.0%	20.3%	92.3%	66.3%	25.9%
PGP 9	79.0%	18.5%	60.5%	90.7%	65.3%	25.4%	86.8%	53.8%	33.0%
PGP 10	81.6%	23.4%	58.2%	95.1%	81.1%	14.0%	91.7%	70.6%	21.0%
Average	82.3%	20.1%	62.3%	92.1%	67.7%	24.3%	91.2%	64.2%	27.0%
Minimum	77.9%	15.1%	58.2%	87.7%	46.6%	14.0%	86.8%	52.7%	19.5%
Maximum	85.9%	27.7%	67.4%	95.1%	81.1%	41.1%	94.2%	74.5%	36.0%

CHAPTER 12

DISCUSSION OF DEMONSTRATION MODEL AND GENERALIZABILITY

In this concluding chapter, we first discuss the strengths and weaknesses of the PGP Demonstration model, and then consider issues in generalizing it to other large physician groups and to smaller physician groups.

12.1 Discussion of Demonstration Model: Strengths and Limitations

With continuing strong interest in improving its quality and controlling its costs, Medicare is exploring alternative approaches to reform. In the 1990s, managed care was a favored approach, but has suffered a backlash in recent years (Robinson, 2001). More recently, pay for performance has been heavily promoted (CMS, 2005), including the Accountable Care Organization (ACO) Medicare Shared Savings Program (MSSP) (CMS, 2011), which was implemented by CMS in 2012, and its predecessor the PGP Demonstration.

The PGP Demonstration was Medicare's first pay for performance initiative for physicians. Unlike some other pay for performance initiatives, the PGP Demonstration explicitly established incentives for efficiency as well as quality. It is a provider-based model that relies on the physician group as the organizational means to improve the quality and efficiency of care.

The Demonstration model changes provider payment, not the insurance arrangements of Medicare beneficiaries, who remain enrolled in the traditional FFS program with complete freedom of provider choice. Disruptions to providers are minimized by the maintenance of standard FFS Medicare payments to them. The innovation of the PGP Demonstration model is that participating provider groups have the opportunity to earn an additional performance payment for providing high quality and efficient care. They share savings they create in the care of beneficiaries assigned to them with the Medicare program, and retain more of the savings the higher their measured quality of care. The financial risk to providers is mitigated by the continuance of FFS payment, the use of provider-specific base costs as a starting point for measuring savings, and the lack of penalties for underperformance. Providers do face the business risk of investments to improve quality and efficiency without any upfront payments from Medicare, and the risk of foregone FFS revenues.

The PGP Demonstration payment model starts from traditional unmanaged FFS payment and moves it part of the way towards a capitated, enrolled managed care model. It retains important FFS advantages: for beneficiaries, freedom of provider choice, and for providers, reduced financial risk and lesser incentives--as compared to capitation--for stinting on services and avoiding the sickest patients. The PGP Demonstration model can be seen as a means to transition from FFS to managed care capitation as provider groups take increasing responsibility for their assigned patients and "capture" and manage more of the services provided to them. Theoretical analyses by health economists indicate that "hybrid" or "mixed" capitation/FFS models, such as the PGP Demonstration model, may be preferable to either pure FFS or pure capitated payment (Ellis and McGuire, 1986). Hence, the PGP Demonstration model may be an end in itself, not just a transitional model.

The PGP Demonstration model has several other noteworthy characteristics. In addition to establishing incentives to control volume of services, it takes advantage of the Medicare FFS program's market (monopsony) power to hold down the FFS rates paid to providers. In a model relying on beneficiary enrollment in competing private plans, Medicare's market power is dissipated among multiple private insurers, who may not be able to obtain as favorable rates from providers as a single government insurer. Rather than relying on absolute cost control targets, the PGP Demonstration model establishes "yardstick competition" (Schliefer, 1985) among providers by employing a local comparison group to judge cost-control performance. Expenditure growth performance targets are by definition reasonable and feasible, since they have been achieved by other providers in the local market. In variants of the model, the comparison group could be other similar provider groups, local or not, so that provider groups are competing against similar groups in cost control and, potentially, quality.

The PGP Demonstration model faces several challenges. First, the model presumes the existence of large physician group practices or integrated delivery systems as participants. But most physicians are not affiliated with such organizations, instead engaging in solo or small group practice. Therefore, the wide generalizability of the model depends on the formation of "network model" organizations that aggregate the experience of many small physician practices. Such physician organizations may also need to be integrated with hospitals and other institutional providers to optimize coordination of care. We have more to say about the generalizability of the PGP Demonstration model in the second part of this chapter.

An alternative—or complement—to a provider-based model is to rely on an independent third-party such as a private insurer or a disease management company to coordinate, integrate, and manage care. Such third parties have advantages of economies of scale in risk pooling, and in developing and implementing care management systems, although they lack a PGP's relationship with and influence over providers and patients.

A second issue is how much control a Demonstration-participating organization can exert over its assigned beneficiaries when they retain freedom of provider choice and have no incentives beyond those in traditional FFS Medicare to choose high-quality or efficient providers or to restrain their use of services. As a provider payment model, the PGP Demonstration model could be combined with revised beneficiary cost sharing or other requirements to encourage beneficiaries to use services efficiently.

A third issue is that turnover in their assigned patients may limit participating PGPs' incentives to engage in quality-improving and cost-saving activities that will only pay off in the long run. This issue also faces private insurance plans whose enrollees may disenroll periodically. To obtain provision of activities with longer-range payoffs, the Medicare program may need to explicitly pay for these results, for example, diabetic blood sugar control.

Fourth, the PGP Demonstration model focuses on expenditure growth from baseline, not the starting level of costs or efficiency. The most inefficient providers may have the greatest potential for the largest performance payments under the model. A provider-specific base encourages voluntary participation, reduces risk to providers, and focuses incentives where the greatest potential for improving cost control exists. But it may be seen as unfair for historically inefficient providers to earn the highest performance payments. Similarly, the quality indicators

performance improvement targets--while giving poor performers a realistic hope of earning a quality bonus--may be seen as rewarding substandard, albeit improving, performance.

Fifth, measuring cost and quality performance remains technically challenging, data intensive, and administratively burdensome, a challenge faced by many pay for performance approaches. Lacking randomized patient assignment to participating PGPs and control groups, the comparability of assigned beneficiary and comparison group 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). Expenditure growth comparisons may also be biased by changes in expenditures unrelated to patient care, such as those arising from changes in Medicare payment policies, statistical and accounting fluctuations in measured per capita expenditures, and inadequate risk adjustment. Measuring performance requires processing massive amounts of administrative claims data, and abstracting medical records for many of the Demonstration quality indicators. These activities pose burdens on both the Medicare program and participating provider groups. The process quality indicators used in the Demonstration, while state of the art, are not a comprehensive measurement of all aspects of the quality of care provided nor of outcomes achieved.

12.2 Generalizability of Demonstration Model

This section discusses the potential for generalizing the PGP Demonstration model beyond the Demonstration providers groups. Generalizing to small physician groups poses some additional issues, which are discussed in a separate subsection.

12.2.1 Large Physician Groups

The Demonstration model could be offered to other large physician organizations participating in Medicare. Criteria similar to those required of Demonstration participants could be required for new organizations. One particularly important criterion would be large size to ensure the statistical reliability of financial reconciliation results—a minimum of 150 to 200 affiliated physicians, expected to result in at least 15,000 assigned beneficiaries per year.³⁹ 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 PGP Demonstration model were offered on a voluntary basis, self-selected participation by physician groups that expected to do well financially under the Demonstration model would be anticipated. This is not a negative, however, because performance payments under the Demonstration model are contingent on efficiency and quality improvements.

If the Demonstration model were extended, several refinements could be implemented, some of which were discussed in Chapter 11. Implementation on a wide scale would require simplified target setting procedures, possible refinement in patient attribution algorithms, and strategies or adjustments to deal with differential increases in intervention provider group diagnostic coding intensity on Medicare claims, which inappropriately raises intervention provider risk scores. Schedules for rebasing the Demonstration's financial reconciliation

³⁹ MSSP only requires 5,000 assigned beneficiaries for a participating ACO (CMS, 2011).

algorithm would need to be established. The Demonstration quality indicators can be refined by adding, deleting, and grouping process measures as appropriate, adding outcome measures if possible, and changing the weight in scoring algorithms between rewarding excellence versus improvement.

If participation in the Demonstration model became widespread, non-participating comparison groups might become hard to identify. In this case, the Demonstration targets could be changed to be based on other participating providers, rather than the current “usual Medicare fee-for-service care” comparison group. That is, participating providers could be put in competition with each other in a “yardstick competition” or “tournament” approach. Performance payments would be made only to participating organizations that performed better than their peers. Another option, which will be used in MSSP, is to use a national comparison group (CMS, 2011).

12.2.2 Additional Issues for Small Physician Groups

For the PGP Demonstration model to be extended to small or solo physician organizations, they would need to be aggregated into larger units for the purposes of applying the Demonstration model. Demonstration financial, and to a lesser extent quality, performance cannot be judged reliably for physician groups with less than roughly 150 to 200 affiliated physicians, or about 15,000 assigned beneficiaries. One of the PGP Demonstration participants, Middlesex Health System, is a network model combining numerous smaller physician practices. It did not earn any performance payments under the Demonstration.

In addition to network models such as Middlesex Health System, other organizational forms could participate in the Demonstration model as representatives of multiple smaller physician practices. These could include Independent Practice Associations (IPAs), physician-hospital organizations (PHOs), Management Services Organizations (MSOs), Accountable Healthcare Organizations (AHOs), medical homes, and medical foundations. Another possibility would be to group physicians by their hospital medical staff affiliations or by county medical societies or other geographic units. Physicians might band together in “virtual” groups formed expressly for the purpose of participating in the Demonstration model.

Participating in the Demonstration financial model is transparent to participating providers and imposes no special burden on small practices. The financial model is based on the standard Medicare fee-for-service claims processing system, and requires no additional data submission on the part of participating practices, other than the tax IDs that identify the participating provider group on Part B Medicare physician claims. However, many of the Demonstration quality indicators require medical chart abstraction, which may impose high costs and burdens on small practices. Medicare’s “Care Management Performance” Demonstration tested the feasibility of obtaining reporting of PGP-Demonstration-like quality indicators from smaller practices. Applying the lessons of this Demonstration to an extension of the PGP Demonstration quality measurement model to smaller physician practices would be important.

REFERENCES

- Boden W et al. 2007. Optimal Medical Therapy with or without PCI for Stable Coronary Disease. *N Engl J Med*; 356:1503-1516.
- CMS, January 31, 2005. Medicare Pay For Performance (P4P) Initiatives. CMS Press Release.
- Centers for Medicare & Medicaid Services, 2009. Advanced Notice of Methodological Changes for Calendar Year (CY) 2010 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies. URL: <http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/Advance2010.pdf>
- Centers for Medicare & Medicaid Services, 2011. Medicare Shared Savings Program: Accountable Care Organizations, Final Rules. Federal Register 76(212): 67,802–67,990.
- Ellis, R.P. and McGuire, T.G.: Provider Behavior Under Prospective Payment. Cost Sharing and Supply. *Journal of Health Economics* 5(2):129-151, June 1986.
- 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.
- Imbens, G.W. and Wooldridge, J.M., 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1): 5-86
- Jencks, S.F., Williams, M.V., and Coleman, E.A., 2009. Rehospitalizations Among Patients in the Medicare Fee-for-Service Program. *The New England Journal of Medicine*, 360: 1418-1428.
- Kautter, J., G.C. Pope, M. Trisolini, et al.: *Physician Group Practice Demonstration Bonus Methodology Specifications*. Final Report to the Centers for Medicare & Medicaid Services under Contract No. 500-00-0024/Task Order # 13. Waltham, MA: RTI International, December 2004.
- Kautter, J., E. Drozd, M. Trisolini, G.C. Pope, and J. Aggarwal. *Physician Group Practice Demonstration Evaluation Design*. Final Report to the Centers for Medicare & Medicaid Services under Contract No. 500-95-0048/Task Order #4. Waltham, MA: RTI International, June 2007a.
- Kautter, J., G.C. Pope, M. Trisolini, and S. Grund. “Medicare Physician Group Practice Demonstration: Quality and Efficiency Pay-for-Performance.” *Health Care Financing Review*, Volume 29, Number 1, Pages 15-29, Fall 2007b.
- Kautter, J., Pope, G.C., Trisolini, M., et al. *Physician Group Practice Second Evaluation Report*. Prepared for the Centers for Medicare & Medicaid under Contract No. 500-00-0024 TO #13. Waltham, MA: RTI International, 2008.

Long, J.S. and J. Freese, 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Second Edition. Stata Press.

MedPAC, March 2011. Medicare Payment Policy, Report to Congress.

Merriman, K. and Caldwell, D. June 2008. *How to Identify Emergency Room Services in Medicare Claims*, ResDAC TN-003.

Olmsted, E., G. Pope, and J. Kautter: *PGP Demonstration Report on Risk Adjustment. Final Report*. Prepared for CMS under Contract No. 500-00-0024, Task Order 13. Waltham, MA: RTI International, 2006.

Pope, G.C., Kautter, J., Ellis R.P., et al.: *Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model*. Health Care Financing Review 25(4):119-141, Summer 2004.

Pope, G.C., M. Trisolini, J. Kautter, W. Adamache: *Physician Group Practice (PGP) Demonstration Design Report*. Final Report to the Centers for Medicare & Medicaid Services under Contract No. 500-95-0048/Task Order # 4. Waltham, MA: RTI International, October 2002.

Pope, G.C., and J. Chromy: *Issues in Demonstrating Group-Specific Volume Performance Standards*. Prepared for the U.S. Health Care Financing Administration. Waltham, Massachusetts: Health Economics Research, Inc., 1997.

Robinson, J.C.: The End of Managed Care. *Journal of the American Medical Association* 285(20):2622-2688, May 2001.

Schafer, J.L., and Kang, J. Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example. *Psychological Methods* 13(4): 279-313, 2008.

Schliefer, T: A Theory of Yardstick Competition. *Rand Journal of Economics* 16(3):319-327, Autumn 1985.

Trisolini, M., J. Kautter, G.C. Pope, et al.: *Physician Group Practice Demonstration Quality Measurement Specifications, Version 2*. Report to the Centers for Medicare & Medicaid Services under Contract No. 500-00-0024/Task Order # 13. Waltham, MA: RTI International, July 2005.