Experience Report for the Performance Year 2012 Quality and Resource Use Reports

January 8, 2014

Wilfredo Lim
Lara Converse
Steve Kuncaitis
Kristin Geonnotti
Jeff Ballou
Aimee Valenzuela
Scott McCracken
John Schurrer
Jesse Crosson

MATHEMATICA
Policy Research

Contract Number:
7500107630

Mathematica Reference Number:
40159.020

Submitted to:
Atchut Kanthamani, Program Manager
Northrop Grumman Systems Corporation
2810 Lord Baltimore Drive
Baltimore, MD 21244
Telephone: (443) 546-0047
Atchut.Kanthamani@ngc.com

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Jeff Ballou

Experience Report for the Performance Year 2012 Quality and Resource Use Reports

January 8, 2014

Wilfredo Lim
Lara Converse
Steve Kuncaitis
Kristin Geonnotti
Jeff Ballou
Aimee Valenzuela
Scott McCracken
John Schurrer
Jesse Crosson

MATHEMATICA
Policy Research

CONTENTS

TABLES

This page has been left blank for double-sided copying.

FIGURES

This page has been left blank for double-sided copying.

I. INTRODUCTION

On September 16, 2013, as part of the Physician Feedback Reporting Program, the Centers for Medicare & Medicaid Services (CMS) distributed confidential feedback reports—the 2012 Quality and Resource Use Reports (QRURs)—to medical group practices.[1] Groups of physicians—defined by their Taxpayer Identification Numbers (TINs)—who participated in Medicare fee-for-service (FFS) in 2012 received a QRUR in 2013 if they met the following criteria:

- At least 25 eligible professionals (EPs) billed under the group's TIN in 2012[2]
- The group had at least 20 eligible cases for at least one quality or cost measure

Report recipients include those that participated in the 2012 Group Practice Reporting Option (GPRO) of the Physician Quality Reporting System (PQRS), the 2012 Pioneer Accountable Care Organization (ACO-Pioneer) model, the Medicare Shared Savings Program (ACO-MSSP), or the 2012 Comprehensive Primary Care (CPC) Initiative.[3]

Each QRUR contains detailed information about the quality of care received by Medicare FFS beneficiaries that was attributed to a specific medical group. It also contains the costs associated with that care (resource use). Quality scores include group-reported measures (for groups participating in specific CMS initiatives) and measures calculated directly by CMS from administrative claims.[4] Measures of resource use capture total per capita costs of beneficiaries attributed to the group as well as total per capita costs of beneficiaries with one or more of four common chronic conditions: diabetes, coronary artery disease (CAD), chronic obstructive pulmonary disease (COPD), and heart failure. The reports use a patient attribution methodology that focuses on whether the group of physicians provided the plurality of primary care services to the Medicare FFS beneficiary.

In addition, the 2012 QRUR previews information on how the group would fare under the policies CMS has adopted for the first year of the phase in of the Physician Value-Based Payment Modifier (VBM). The VBM is a new payment adjustment to the Medicare Physician Fee Schedule that will reward higher quality care delivered at lower cost, as required under

[1] More information on the Physician Feedback Program is available at http://www.cms.gov/physicianfeedbackprogram.

[2] Eligible professionals include physicians (doctors of medicine, osteopathy, podiatric medicine, optometry, dental surgery, dental medicine, and chiropractic), practitioners (physician assistants, nurse practitioners, clinical nurse specialists, certified registered nurse anesthetists, certified nurse midwives, clinical social workers, clinical psychologists, registered dieticians, nutrition professionals, and audiologists), and therapists (physical therapists, occupational therapists, and qualified speech-language therapists), as specified in section 1848(k)(3)(B) of the Social Security Act.

[3] CMS did not produce QRURs for the nine groups that were members of multiple ACOs or that were participants in both an ACO and the CPC Initiative in 2012.

[4] Self-reported quality measures are included for groups reporting to the PQRS as a participant in the PQRS GPRO, the Medicare Shared Savings Program, or the Pioneer ACO model.

Section 3007 of the Affordable Care Act. Section 3007 requires CMS to apply the VBM to specific physicians and groups of physicians the Secretary of the Department of Health and Human Services determines appropriate starting January 1, 2015, and to all physicians and groups of physicians paid under the Medicare Physician Fee Schedule by January 1, 2017.

Starting January 1, 2015, Medicare will apply the VBM to groups of physicians with 100 or more EPs based on performance during 2013. In addition, Medicare will use the quality measures that a group reports under the PQRS for the group's VBM. For 2013, the quality reporting mechanisms include (1) a web interface group reporting mechanism consisting of 22 primary and preventive care measures; (2) a qualified registry mechanism, in which the group chooses three quality measures for all EPs in the group; and (3) CMS-calculated administrative claims that consist of 14 primary and preventive care quality indicators. For 2015, these groups may elect to have their VBM calculated based on their performance on cost and quality measures (as will be described in more detail in Sections III and IV), which could result in payments being adjusted up or down. The 2012 QRURs were available prior to the time the group had to decide whether to elect to have their VBM calculated based on 2013 performance so they could see how they could potentially fare under the VBM in 2015. Groups of 100 or more EPs who do not participate in PQRS in 2013 will have their Medicare payments adjusted downward by 1 percent starting in 2015. Medicare will not apply the VBM in 2015 and 2016 to groups of physicians participating in the ACO-MSSP, the ACO-Pioneer Model, or the CPC Initiative.

In this report, we summarize the data contained in the 2012 QRURs so that report recipients, policy makers, researchers, and other stakeholders can see how the VBM is calculated as well as the properties of the cost and quality measures that make up the VBM. This report includes descriptive information about the characteristics of groups for which CMS produced QRURs; the beneficiaries attributed to these groups; and the performance, reliability, and statistical significance for all measures included in the quality and cost composite scores informing the VBM. In addition, we describe some characteristics of groups with high, average, or low quality or cost scores; the relationship between quality and cost scores; and the effects of risk adjustment on VBM measures. Because only larger groups (100 or more EPs) are subject to the VBM in 2015 we report many of our findings for all groups for which CMS produced a QRUR and then separately for groups with 100 or more EPs. We also note that Medicare has refined and enhanced several of its policies for the second year of the VBM phase in (2016 based on 2014 performance), including the addition of a broader set of PQRS measures, elimination of the CMS-calculated administrative claims reporting mechanism, inclusion of an additional cost measure using a new attribution method, and incorporation of a new way to determine the peer group for cost measures comparisons. The refinements can be found at http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Federal-Regulation-Notices-Items/CMS-1600-FC.html.

Key Findings

Of the 6,779 physician groups with 25 or more EPs in 2012, CMS produced QRURs for 3,876 (58 percent of) groups. Many of the groups were large: 1,032 (27 percent) had at least 100 EPs; 1,878 (48 percent) were attributed more than 1,000 beneficiaries. The 2,903 groups that did not receive a QRUR generally had insufficient data to compute meaningful performance measures. Many of these groups consisted of specialists, such as anesthesiologists or radiologists, who do not typically provide primary care services and, therefore, were attributed no or very few patients under the attribution method used for these reports.

More than 80 percent of medical group practices with at least 25 EPs were deemed average performers on both quality and cost under the 2015 VBM methodologies (based on the calendar year [CY] 2013 Medicare Physician Fee Schedule Final Rule). Approximately 8 percent of groups are in quality and cost tiers that would be eligible for an upward adjustment in their Medicare Physician Fee Schedule payments; nearly 11 percent would be eligible to receive a downward payment adjustment. Among those eligible for an upward adjustment, 11 percent would be eligible to receive an additional upward adjustment for treating patients with high average risk.

Using the 2015 VBM methodologies, groups with cost composite scores that were low had higher quality composite scores, on average, than did groups with cost composite scores in the high or average range. In addition to having higher quality scores generally, these groups—which tended to have fewer than 100 EPs, a higher proportion of EPs from the same specialty, a smaller proportion of primary care physicians (PCPs), fewer attributed patients, and patients with fewer risk factors—generally had lower rates of hospitalization for their patients with ambulatory care sensitive conditions (ACSCs) and lower readmission rates.

Among the quality measures, the self-reported PQRS measures had the highest average reliability. Among groups with 25 or more EPs, average reliability was high (above 0.70, a common standard for high reliability) for all PQRS measures—which include those reported by both GPRO participants and ACO participants—as well as for 9 of the 17 rates[5] calculated for the administrative claims–based quality indicators and the acute composite and chronic composite ACSC measures. Reliability was lowest for the hospital all-cause readmissions measure.

Risk adjustment compressed the range of groups' total per capita costs by 83 percent. Before risk adjustment, total per capita costs for physician groups with 25 or more EPs ranged from $477 to $340,516. By contrast, per capita costs after risk adjustment ranged from $756 to $58,945.

Despite the wide range in performance even after risk adjustment, scores for most groups fluctuated in a relatively narrower range. Half of all groups with total per capita cost scores had costs between $8,832 (the 25th percentile score) and $11,243 (the 75th percentile score).

[5] There are 14 CMS-calculated claim-based quality indicators. Because two of them have two component rates 17 rates are calculated.

The total and condition-specific per capita cost measures had high average reliability. The measures all had an average reliability of 0.82 or greater, well above the 0.70 standard for high reliability.

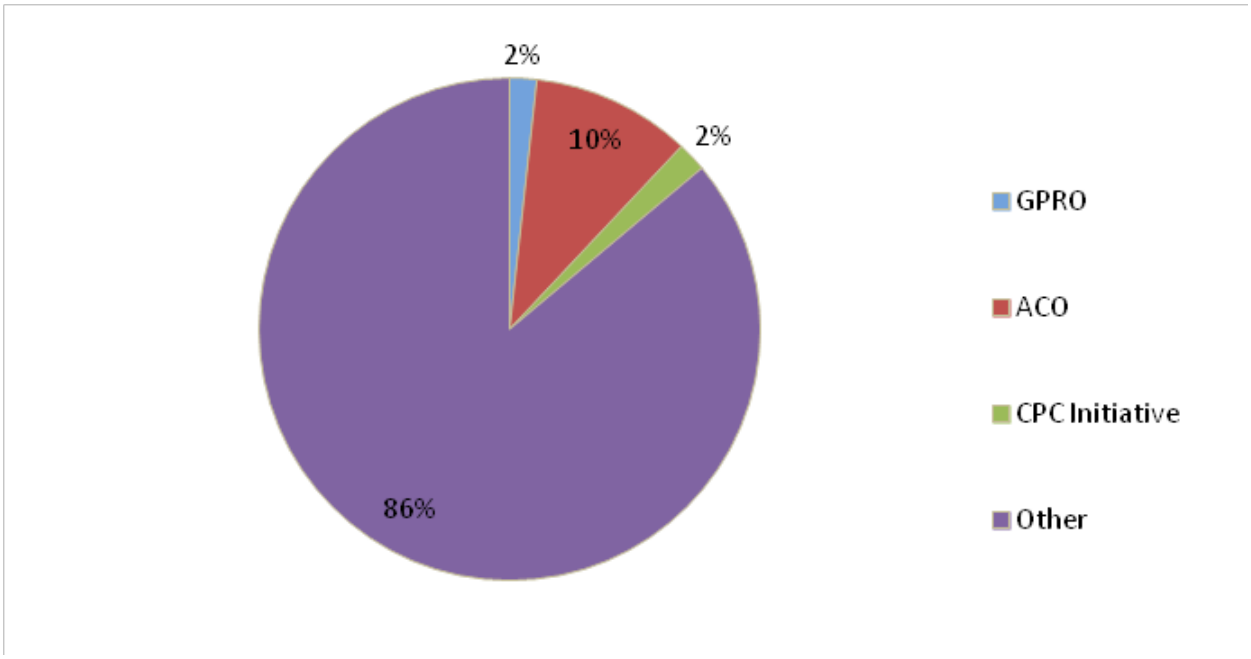
II. CHARACTERISTICS OF GROUPS RECEIVING QRURS AND THEIR ATTRIBUTED PATIENTS

There were 6,779 medical group practices with 25 or more EPs in 2012. The 3,876 groups receiving a 2012 QRUR were distributed across all states, the District of Columbia, Guam, and Puerto Rico. They represent 57 percent of such groups nationwide. These groups had an average of 120 EPs. Among group practices receiving a QRUR, 1,032 (27 percent) had 100 or more EPs, and the average number of EPs associated with these larger groups was 322.

A. Characteristics of Groups That Received a QRUR

The majority of groups that received a QRUR (88 percent) were not participating in one of the payment demonstration projects operated by the Center for Medicare & Medicaid Innovation (see Figure II.1). CMS will not apply the VBM in 2015 to groups participating in the ACO initiatives (ACO-MSSP and ACO-Pioneer) or the CPC Initiative in CY 2013. These groups made up 12 percent of the groups receiving reports in 2012. Groups participating in the PQRS through the GPRO made up 2 percent of groups. The remaining 86 percent of groups were not participating in one of the listed CMS initiatives.

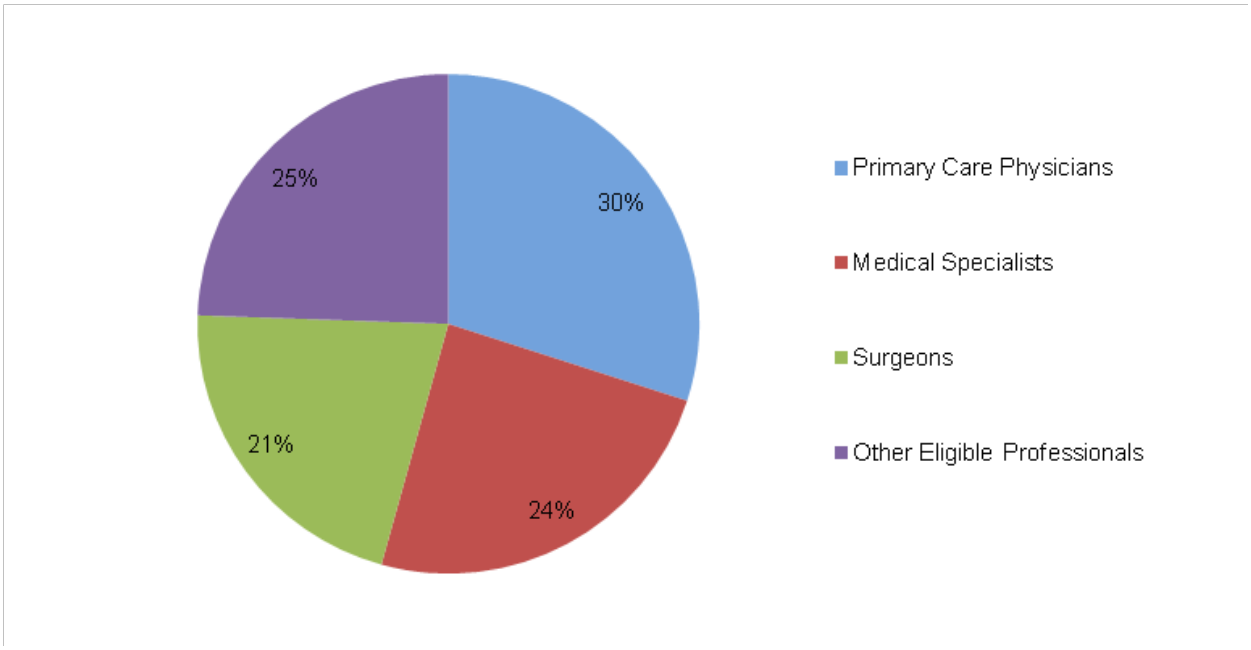
Figure II.1. Type of Groups Receiving 2012 QRURs



Note: Sample of 3,876 groups with at least 25 EPs receiving a 2012 QRUR.

One-quarter of groups (961) had one specialty that represented more than 50 percent of the group's EPs (Figure II.2). The figure shows the percentage of groups with a single specialty constituting a majority of the group's EPs across several broad disciplinary categories.

Figure II.2. Groups with More Than 50 Percent of EPs in the Same Specialty Category



Note: Sample of 961 groups with a specialty that represented more than 50 percent of the group's EPs.

B. Patient Attribution

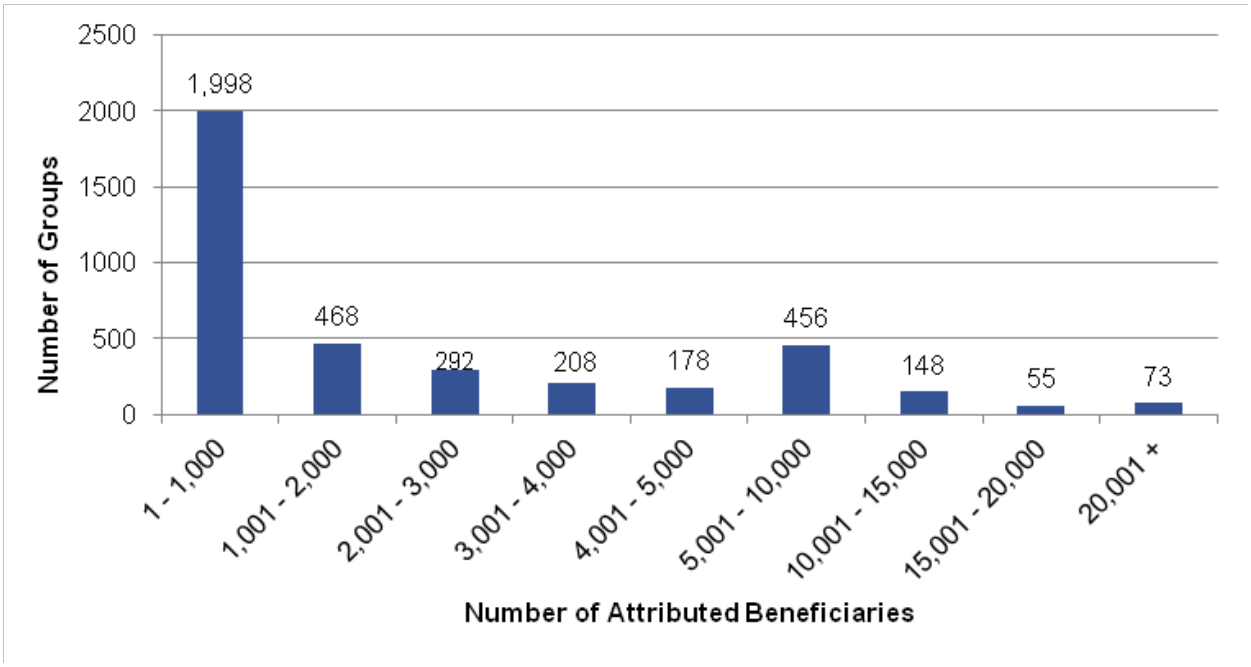
For cost measures and the administrative claims-based quality indicators, the 2012 QRURs use a two-step attribution approach similar to that used to attribute beneficiaries to ACOs in the MSSP. Under this approach, a beneficiary with claims containing procedure codes that represent primary care services provided by one or more PCPs (that is, physicians with specialties of general practice, family practice, internal medicine, or geriatric medicine) is attributed to the group whose PCPs provided the plurality of primary care services (as defined by Healthcare Common Procedure Coding Systems [HCPCS] code) according to the total allowed charges for those services.[6] If a beneficiary does not have any allowable charges for primary care services provided by a PCP, that individual is attributed to the group whose other physicians, clinical nurse specialists, nurse practitioners, and physician assistants provided the plurality of primary care services, as long as at least one physician in the group provided primary care services to the beneficiary. Some patients seen by members of a group might not be attributed to the group because they did not receive primary care services from an eligible professional, or they received primary care services for which there was no allowable charge.

Of the 24,426,141 beneficiaries who were identified on claims submitted by the groups of 25 or more EPs eligible to receive a QRUR, 11,593,241 (47 percent) were attributed to a group. Overall, 86 percent of attributed beneficiaries were attributed in the first step of attribution, meaning that they had primary care services provided by a PCP. There was substantial variation

[6] The following HCPCS Primary Care Service Codes were used: 99201-99205, 99211-99215, 99304-99306, 99307-99310, 99315-99316, 99318, 99324-9928, 99334-99337, 99339-99345, 99347-99350, G0402, G0438-G0439.

in the number of beneficiaries attributed to the groups receiving reports. The average number of beneficiaries attributed to the groups was 2,974 (standard deviation = 5,105). However, slightly more than half (52 percent) were attributed 1,000 or fewer beneficiaries (Figure II.3). Among groups with 100 or more EPs, there was an average of 7,077 attributed beneficiaries (standard deviation = 7,842).

Figure II.3. Number of Beneficiaries Attributed to Groups



Note: Sample of 11,593,241 beneficiaries who were attributed to one of 3,876 groups with at least 25 EPs in 2012.

Attributed beneficiaries received an average of four primary care services in 2012, of which, on average, 63.6 percent were provided by the group to which the beneficiary was attributed. The average allowable charges for the group to which the beneficiary was attributed represented 65 percent of the average allowable charges for all primary care services for a beneficiary.

C. Characteristics of Groups Not Receiving a QRUR

There were 2,903 groups of 25 or more EPs that did not receive a QRUR. Of these, 2,271 groups had no attributed Medicare FFS beneficiaries. An additional 623 groups had fewer than 20 attributed beneficiaries, thus preventing accurate calculation of the quality and resource use measures. Nine groups that were attributed 20 or more beneficiaries were also excluded from receiving a QRUR because they participated in more than one ACO or in both an ACO and the CPC Initiative in 2012. Groups that did not receive a QRUR were dominated by groups with at least 50 percent of EPs in the same specialty category. Many of these groups were made up of physicians unlikely to provide primary care services, such as anesthesiologists or radiologists.

This page has been left blank for double-sided copying.

III. QUALITY OF CARE

The QRURs contain detailed information about the quality of patient care delivered during the 2012 calendar year and evaluate how this quality compares to that of other medical groups. Quality of care is determined by measures submitted through PQRS and by CMS-calculated administrative claims data. This section describes the quality measures reported in the QRURs and analyzes group performance across PQRS measures, CMS-calculated administrative claims-based quality indicators, and CMS-calculated administrative claims-based outcome measures. We also describe the measures and summarize performance and reliability.

A. PQRS Quality Measures

Under PQRS, EPs may report quality measures either as individuals or as a medical practice group under the GPRO. The 2012 QRURs report PQRS measure performance only for EPs participating through the GPRO web interface.[7] Of the 3,876 groups for which CMS produced a QRUR, 66 PQRS GPRO groups and 396 ACOs reported at least one PQRS measure with 20 or more cases.

1. PQRS Quality Measures for GPRO Groups

Groups electing to participate in the PQRS through GPRO are required to report quality measures through a web interface for each of the 29 GPRO quality measures. Only groups with at least 25 EPs were eligible for this reporting option. GPRO measures are patient-centric, and therefore reflect the care a patient receives rather than the care delivered by a particular physician or group. Modules for specific diseases, as well as for more general care, are incorporated into the GPRO. CMS classified subsets of these measures into three domains of care—Clinical Process/Effectiveness, Population/Public Health, and Patient Safety.

2. PQRS Quality Measures for ACO Participants

New to the PQRS in 2012 was a system of reporting for ACOs and MSSP participants. All ACO participant measures are similar or identical to the PQRS GPRO measures reported except for one measure "Preventive Care: Screening for Clinical Depression and Follow-Up Plan (Prev-12), which is unique to the reporting method for ACO participants. Moreover, unlike GPRO participants, the ACO participants' CAD measures were incorporated into a composite measure with "all-or-nothing" scoring; a similar compositing procedure was applied to five of the six diabetes measures.

3. PQRS Measure Reporting and Performance Rates

Benchmarks were computed for each measure from 2011 PQRS quality performance data for all PQRS participants reporting at least 20 eligible cases for the measure. The performance of both those participating as a practice through the GPRO and those submitting a comparable

[7] Because reporting at the individual level is an option in the future, groups will have the opportunity to review their performance on PQRS measures at this level later this year.

PQRS measure as an individual EP were included in the benchmark rate. In Table III.l, we present the reporting and performance rates for each of the PQRS quality measures (along with the benchmark, if it was calculated). For all but one of these measures, rates of reporting at least 20 eligible cases were above 90 percent; the exception was COPD-1: Bronchodilator Therapy (75 percent). Among ACO groups, rates of groups reporting at least 20 cases were consistently in the mid to upper 90s.

Within some measure modules, performance was consistently high across group types and across measures; for others, performance varied by measure. All GPRO groups (including ACOs) tended to perform fairly well on the CAD and heart failure measures. PQRS GPRO groups never had mean performance rates below 70 percent for any individual CAD or heart failure measure; performance rates for the ACOs were never below 60 percent.

More variation was observed among the measures of preventive care. On average, ACO groups had performance rates ranging from less than 20 percent for Depression Screening and Follow-Up Plan (Prev-12) to 79 percent for Tobacco Use: Screening and Cessation Intervention (Prev-10). For PQRS GPRO groups, average performance rates for the preventive care measures ranged from 57 percent for Influenza Immunization (Prev-7) to 88 percent for Prev-10. Performance also varied greatly within each of the domains and did not appear to be systematically different from one domain to another. Among PQRS GPRO measures for which higher performance indicates better quality, performance rates ranged from 56 to 95 percent within the Clinical Process/Effectiveness domain, from 57 to 88 percent for Population/Public Health and from 45 to 79 percent for Patient Safety. Similar results were observed for the ACO groups.

PQRS GPRO groups had more measures than ACO GPRO groups for which both mean performance exceeded the prior year benchmark and, across groups, performance was statistically different from the benchmark on average. Among PQRS GPRO groups, for all five measures for which the average group's performance rate was statistically different from the benchmark, average performance in 2012 was higher than the 2011 benchmark. Of these five measures, four were within the Clinical Process/Effectiveness domain and pertained to preventive care. ACO group performance was statistically different on average only from the prior year benchmark for CAD: Lipid Control (CAD-2) and Medication Reconciliation: Reconciliation after Discharge from an Inpatient Facility (Care-1).

Table III.1. GPRO Quality Measure Performance Rates for Groups with 25 or More Eligible Professionals, by GPRO Measure Type

.	.	PQRS GPRO						ACO					
.	**Domain and Measure Title**	Number of groups reporting ≥20 eligible cases	Percent reporting ≥20 eligible cases	Mean rate (%)	SD	Bench-mark	Average p value	Number of groups reporting ≥20 eligible cases	Percent reporting ≥20 eligible cases	Mean rate (%)	SD	Bench-mark	Average p value
.	**Clinical Process/Effectiveness**	.	.	.	.	.	.	.	.	.	.	.	.
COPD-1	COPD: Bronchodilator Therapy[a]	50	74.6	94.7	9.2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CAD-1	CAD: Antiplatelet Therapy[a]	66	98.5	90.8	6.0	82.8	0.087	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CAD-2	CAD: Lipid Control[b]	66	98.5	77.7	14.0	88.8	0.052	396	99.5	69.3	16.3	88.8	0.042
CAD-7	CAD: ACE Inhibitor or ARB Therapy for Patients with CAD and Diabetes and/or LVSD[b]	66	98.5	83.7	9.6	69.0	0.030	395	99.2	70.7	13.6	69.0	0.099
.	CAD Composite	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	396	99.5	61.8	16.7	n.a.	n.a.
DM-2	DM: Hemoglobin A1c Poor Control in DM (>9.0)[c]	66	98.5	18.7	8.0	21.3	0.085	396	99.5	27.6	14.1	21.3	0.101
DM-3/ DM-13	DM: High Blood Pressure Control in DM[b]	66	98.5	70.8	6.9	69.0	0.208	396	99.5	65.7	11.1	n.a.	n.a.
DM-5/ DM-14	DM: LDL-C Control in DM[b]	66	98.5	56.4	9.6	56.2	0.149	396	99.5	53.1	11.9	n.a.	n.a.
DM-7	DM: Dilated Eye Exam in Diabetic Patient[a]	66	98.5	57.0	18.8	55.6	0.067	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
DM-8	DM: Foot Exam[a]	66	98.5	69.7	19.8	64.3	0.086	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
DM-10/ DM-15	DM: Hemoglobin A1c Control (< 8.0)[b]	66	98.5	71.6	7.8	n.a.	n.a.	396	99.5	63.4	13.0	n.a.	n.a.
DM-11/ DM-16	DM: Daily Aspirin Use for Patients with Diabetes and Ischemic Vascular Disease[b]	65	97.0	83.6	14.8	n.a	n.a.	381	95.7	72.4	18.6	n.a.	n.a.
DM-12/ DM-17	DM: Tobacco Non-Use[b]	66	98.5	80.3	10.9	n.a.	n.a.	396	99.5	68.4	18.3	n.a.	n.a.
.	DM Composite	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	396	99.5	21.9	11.3	n.a.	n.a.
HF-1	HF: Left Ventricular Ejection Fraction (LVEF) Assessment[a]	66	98.5	79.3	22.9	81.6	0.067	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HF-2	HF: Left Ventricular Function (LVF) Testing[a]	63	94.0	85.1	10.6	88.6	0.193	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HF-5	HF: Patient Education[a]	66	98.5	73.2	22.3	54.9	0.063	n.a.	n.a.	n.a.	`	n.a.	n.a.
HF-6	HF: Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	62	92.5	88.3	12.0	76.7	0.085	370	93.0	83.9	12.4	76.7	0.112

Table III.1 *(continued)*

.	.	PQRS GPRO						ACO					
.	**Domain and Measure Title**	Number of groups reporting ≥20 eligible cases	Percent reporting ≥20 eligible cases	Mean rate (%)	SD	Bench-mark	Average p value	Number of groups reporting ≥20 eligible cases	Percent reporting ≥20 eligible cases	Mean rate (%)	SD	Bench-mark	Average p value
HF-7	HF: ACE Inhibitor or ARB Therapy for Left Ventricular Systolic Dysfunction (LVSD)[a]	62	92.5	84.2	12.8	71.2	0.117	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HTN-2	HTN: Controlling High Blood Pressure	66	98.5	69.1	8.6	67.3	0.198	396	99.5	63.5	10.2	67.3	0.227
IVD-1	IVD: Complete Lipid Profile and LDL-C Control	66	98.5	58.0	9.6	n.a.	n.a.	395	99.2	53.1	11.9	n.a.	n.a.
IVD-2	IVD: Use of Aspirin or Another Antithrombotic	66	98.5	83.9	12.8	81.0	0.124	395	99.2	71.3	20.2	81.0	0.093
Prev-5	Prev: Screening Mammography	66	98.5	67.1	11.3	50.8	0.032	396	99.5	61.5	12.6	50.8	0.070
Prev-6	Prev: Colorectal Cancer Screening	66	98.5	62.3	13.4	48.0	0.043	396	99.5	52.9	14.8	48.0	0.105
Prev-8	Prev: Pneumonia Vaccination for Patients ≥ 65	66	98.5	67.1	20.1	52.6	0.049	396	99.5	51.8	21.3	52.6	0.072
.	**Population/ Public Health**	.	.	.	.	.	.	.	.	.	.	.	.
Prev-7	Prev: Influenza Immunization	66	98.5	57.2	18.4	n.a.	n.a.	396	99.5	49.8	15.4	n.a.	.
Prev-9	Prev: BMI Screening and Follow-Up	66	98.5	61.1	18.6	45.2	0.029	396	99.5	52.9	14.0	45.2	0.124
Prev-10	Prev: Tobacco Use: Screening and Cessation Intervention	66	98.5	87.9	15.1	n.a.	n.a.	396	99.5	79.3	16.8	n.a.	n.a.
Prev-11	Prev: Screening for High Blood Pressure	66	98.5	81.4	21.3	n.a.	n.a.	396	99.5	68.5	23.5	n.a.	n.a.
Prev-12	Screening for Clinical Depression and Follow-Up Plan	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	396	99.5	18.9	17.7	n.a.	n.a.
.	**Patient Safety**	.	.	.	.	.	.	.	.	.	.	.	.
Care-1	Medication Reconciliation: Reconciliation After Discharge from an Inpatient Facility	66	98.5	79.2	27.5	95.1	0.064	389	97.7	70.9	29.8	95.1	0.033
Care-2	Falls: Screening for Future Fall Risk	66	98.5	44.7	28.4	n.a.	n.a.	396	99.5	27.8	17.8	n.a,	n.a.

[a] Indicates a 2012 PQRS GPRO measure not included in Quality Composite Score computations because it will not be included in the 2013 web interface set of measures.

[b] Indicates a measure that for ACO GPRO groups is included with one or more other measures for the same condition as part of an “all-or-nothing” composite when computing the Quality Composite Score displayed in this report. For PQRS GPRO groups, although these measures will be treated as an "all-or-nothing" composite in program year 2013 and all years after that, the Quality Composite Score displayed in this report treats these measures as distinct.

Table III.1 *(continued)*

[c] Lower performance rates on these measures indicate better performance. However, the domain score for this domain has been calculated such that positive (+) scores indicate better performance and negative scores (-) indicate worse performance.

SD = standard deviation

4. PQRS Measure Reliability

Reliability scores are represented on a continuum from zero to one. Scores closer to zero indicate lower reliability and scores closer to one indicate higher reliability. Although there is no universally agreed-upon minimum reliability threshold, reliability scores in the 0.40–0.70 range are often considered moderate, and scores greater than 0.70 are considered high. For each of these measures, reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error):

$$\text{Reliability} = \frac{\text{Variation between groups}}{\text{Variation between groups} + \text{ Variation within group}}$$

If a score is deemed highly reliable, we would expect that a group's performance rates would be very similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.

Average reliabilities for all PQRS measures were higher than 0.80 (Table III.2). Among PQRS GPRO groups, COPD: Bronchodilator Therapy (COPD-1) had the lowest reliability (0.85); Medication Reconciliation: Reconciliation After Discharge from an Inpatient Facility (Care-1) had the highest reliability (1.00). For ACO GPRO groups, results were similar, but the range of reliabilities was slightly narrower. The measure with the lowest reliability, Heart Failure: Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) (HF-6), had a score of 0.87; Care-1 still had the highest possible reliability (1.00). Across all group types, the Care Coordination domain exhibited the highest reliability, as both of its measures (Care-1, Care-2) always had reliability of at least 0.99. In contrast, across group types, Clinical Process/Effectiveness was the only domain with reliability for any PQRS measure less than 0.90.

Table III.2. GPRO Quality Measure Reliability and Statistical Significance for Groups with 25 or More Eligible Professionals by GPRO Measure Type

.	.	PQRS GPRO		ACO	
.	Domain and Measure Title	Reliability	Percent of groups different than the benchmark (p<0.05)	Reliability	Percent of groups different than the benchmark (p<0.05)
.	Clinical Process/Effectiveness	.	.	.	.
COPD-1	COPD: Bronchodilator Therapy[a]	0.85	n.a.	n.a.	n.a.
CAD-1	CAD: Antiplatelet Therapy[a]	0.94	78.8	n.a.	n.a.
CAD-2	CAD: Lipid Control[b]	0.98	90.9	0.98	90.9
CAD-7	CAD: ACE Inhibitor or ARB Therapy for Patients with CAD and Diabetes and/or LVSD[b]	0.96	87.9	0.94	64.1
.	CAD Composite	n.a.	n.a.	0.98	n.a.
DM-2	DM: Hemoglobin A1c Poor Control in DM (>9.0)[c]	0.98	86.4	0.98	68.7

Table III.2 *(continued)*

.	.	PQRS GPRO		ACO	
.	Domain and Measure Title	Reliability	Percent of groups different than the benchmark (p<0.05)	Reliability	Percent of groups different than the benchmark (p<0.05)
DM-3/ DM-13	DM: High Blood Pressure Control in DM[b]	0.90	51.5	0.96	n.a.
DM-5/ DM-14	DM: LDL-C Control in DM[b]	0.93	65.2	0.96	n.a.
DM-7	DM: Dilated Eye Exam in Diabetic Patient[a]	0.99	81.8	n.a.	n.a.
DM-8	DM: Foot Exam[a]	0.99	78.8	n.a.	n.a.
DM-10/ DM-15	DM: Hemoglobin A1c Control (< 8.0) [b]	0.92	n.a.	0.97	n.a.
DM-11/ DM-16	DM: Daily Aspirin Use for Patients with Diabetes and Ischemic Vascular Disease[b]	0.93	n.a.	0.96	n.a.
DM-12/ DM-17	DM: Tobacco Non-Use[b]	0.96	n.a.	0.99	n.a.
.	DM Composite	n.a.	n.a.	0.98	n.a.
HF-1	HF: Left Ventricular Ejection Fraction (LVEF) Assessment[a]	0.99	84.8	n.a.	n.a.
HF-2	HF: Left Ventricular Function (LVF) Testing[a]	0.91	57.6	n.a.	n.a.
HF-5	HF: Patient Education[a]	0.99	86.4	n.a.	n.a.
HF-6	HF: Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	0.91	78.8	0.87	67.1
HF-7	HF: ACE Inhibitor or ARB Therapy for Left Ventricular Systolic Dysfunction (LVSD) [a]	0.90	69.7	n.a.	n.a.
HTN-2	HTN: Controlling High Blood Pressure	0.95	51.5	0.94	46.0
IVD-1	IVD: Complete Lipid Profile and LDL-C Control	0.96	n.a.	0.96	n.a.
IVD-2	IVD: Use of Aspirin or Another Antithrombotic	0.99	68.2	0.99	76.3
Prev-5	Prev: Screening Mammography	0.97	89.4	0.97	75.0
Prev-6	Prev: Colorectal Cancer Screening	0.98	90.9	0.98	74.7
Prev-8	Prev: Pneumonia Vaccination for Patients ≥ 65	0.99	86.4	0.99	82.6
.	Population/ Public Health	.	.	.	.
Prev-7	Prev: Influenza Immunization	0.98	n.a.	0.98	n.a.
Prev-9	Prev: BMI Screening and Follow-Up	0.98	87.9	0.98	75.8

Table III.2 *(continued)*

.	.	PQRS GPRO		ACO	
.	Domain and Measure Title	Reliability	Percent of groups different than the benchmark (p<0.05)	Reliability	Percent of groups different than the benchmark (p<0.05)
Prev-10	Prev: Tobacco Use: Screening and Cessation Intervention	0.99	n.a.	0.99	n.a.
Prev-11	Prev: Screening for High Blood Pressure	0.99	n.a.	0.99	n.a.
Prev-12	Screening for Clinical Depression and Follow-Up Plan	n.a.	n.a.	0.99	n.a.
.	Patient Safety	.	.	.	.
Care-1	Medication Reconciliation: Reconciliation After Discharge from an Inpatient Facility	1.00	80.3	1.00	87.0
Care-2	Falls: Screening for Future Fall Risk	0.99	n.a.	0.99	n.a.

[a] Indicates a 2012 PQRS GPRO measure that is not included in Quality Composite Score computations because it will not be included in the 2013 web interface set of measures.

[b] Indicates a measure that, for ACO GPRO groups, is included with one or more other measures for the same condition as part of an “all-or-nothing” composite when computing the Quality Composite Score displayed in this report. For PQRS GPRO groups, although these measures will be treated as an "all-or-nothing" composite in program year 2013 and in the years after that, the Quality Composite Score displayed in this report treats these measures as distinct.

[c] Lower performance rates on these measures indicate better performance. However, the domain score for this domain has been calculated such that positive (+) scores indicate better performance and negative scores (-) indicate worse performance.

n.a. = not applicable

B. Administrative Claims-Based Quality Indicators

For the 3,411 groups of physicians with at least 25 EPs that did not participate in the PQRS GPRO or in an ACO initiative, CMS computed 14 quality indicators from data reported in Medicare administrative claims.[8] We note that many of these groups have EPs who submitted data on quality measures individually through the PQRS. Medicare will use performance on these individually reported PQRS measures in the 2016 VBM (based on 2014 performance); it will no longer use performance on the 14 CMS-calculated administrative claims-based quality indicators for the VBM.

[8] Two indicators (Antidepressant Treatment for Depression and Use of High Risk Medications in the Elderly) have two component rates. A total of 17 rates are calculated for the 14 indicators.

As with GPRO measures, performance computations on the 14 quality indicators are patient-centric; beneficiaries are attributed to groups that submit at least one Medicare Part B claim on their behalf, allowing for one beneficiary to be attributed to more than one group. The indicators focus on care for chronic conditions, preventive screenings, and care coordination. Particular indicators are specific to patients with such conditions as diabetes, CAD, and mental illness, and indicators are grouped into the same three domains of care (clinical process/effectiveness, population/public health, and patient safety) as the PQRS measures discussed in Section A.

1. Administrative Claims-Based Quality Indicator Performance Rates

The variation across indicators with at least 20 eligible cases for a measure is much greater for administrative claims-based quality indicators than for the PQRS measures (Table III.3). For groups with 25 or more EPs and for groups with 100 or more EPs, Osteoporosis Management in Women ≥ 67 Who Had a Fracture had the lowest rate of groups, with at least 20 eligible cases (19 percent and 44 percent, respectively), closely followed by indicators on acute phase and continuation phase of antidepressant treatment for depression (24 percent and 57 percent). In contrast, for both of these group sizes, the two component indicators for Use of High-Risk Medications in the Elderly had the highest rate of having at least 20 eligible cases (89 percent and 95 percent).

Performance also varied greatly across these measures, although trends are fairly apparent by disease. The quality of care received by patients with diabetes, ischemic vascular disease, or CAD, as measured by performance rates on claims-based indicators, is quite high. Among the indicators for which a higher performance rate indicates higher quality, for groups with at least 25 EPs and for those with 100 or more EPs, the three indicators with the highest performance rates focused on care for diabetes; the next two highest performance rates were observed for the ischemic vascular disease and CAD indicators. All five of these indicators are classified into the Clinical Process/Effectiveness domain, as is the measure with the lowest performance rate regardless of group size and excluding inverse measures, Osteoporosis Management in Women ≥ 67 Who Had a Fracture. Less variation is apparent within the Patient Safety domain; quality is high among its three measures with average performance rates never exceeding 33 percent. (For these three measures lower rates indicate higher quality.)

Prior-year benchmarks were also computed for the claims-based quality indicators, and none of the measures differed significantly at the 5 percent level from the prior year benchmark. A weighted average (based on eligible cases) of performance for groups with 25 or more EPs serves as the benchmark for all groups of this size, whereas a comparable weighted average among groups with at least 100 EPs forms the benchmark for larger groups (100 or more EPs). For a number of measures, particularly within the Patient Safety and Clinical Process/Effectiveness domains, more than half of groups had a performance rate that significantly differed from the prior year benchmark at the 5 percent level. For groups with at least 25 EPs, the two measures for which a statistical difference from the mean was the most common were Lack of Monthly INR Monitoring for Beneficiaries on Warfarin and Breast Cancer Screening for Women 40–69. For groups with 100 or more EPs, a statistical difference from the mean was most likely to be observed for Patients Who Receive at Least Two Different Drugs to Be Avoided and Lack of Monthly INR Monitoring for Beneficiaries on Warfarin.

Table III.3. Administrative Claims-Based Quality Indicator Performance, by Group Size

.	25+ EPs							100+ EPs						
Domain and Measure Title	Number of Groups Reporting ≥20 Eligible Cases	Percent Reporting ≥20 Eligible Cases	Mean	SD	Average Reliability	2011 Benchmark	Percent of Groups Different Than the Benchmark (p<0.05)	Number of Groups Reporting ≥20 Eligible Cases	Percent Reporting ≥20 Eligible Cases	Mean	SD	Average Reliability	2011 Benchmark	Percent of Groups Different Than the Benchmark (p<0.05)
Clinical Process/ Effectiveness	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Osteoporosis Management in Women ≥ 67 Who Had a Fracture	653	19.1	14.8	7.0	0.53	18.4	47.3	363	44.3	15.4	6.7	0.43	19.1	33.4
Use of Spirometry Testing to Diagnose COPD	2,064	60.5	29.6	11.3	0.73	31.7	45.6	693	84.5	30.6	9.7	0.83	31.9	47.1
Dilated Eye Exam for Beneficiaries ≤ 75 with Diabetes	2,632	77.2	51.5	12.4	0.79	56.1	51.6	755	92.1	53.7	10.2	0.91	56.6	64.5
Hba1c Testing for Beneficiaries ≤ 75 with Diabetes	2,632	77.2	82.8	12.8	0.82	87.9	51.4	755	92.1	84.7	12.5	0.93	88.1	61.6
Nephropathy Screening Test or Evidence of Existing Nephropathy for Beneficiaries ≤ 75 with Diabetes	2,632	77.2	75.2	10.0	0.76	77.4	48.3	755	92.1	78.2	8.4	0.90	78.5	62.8
Lipid Profile for Beneficiaries ≤ 75 with Diabetes	2,632	77.2	76.7	14.2	0.84	82.3	55.8	755	92.1	78.6	14.2	0.94	82.4	69.2
Lipid Profile for Beneficiaries with Ischemic Vascular Disease	2,559	75.0	72.8	13.3	0.81	77.8	52.1	750	91.5	73.0	13.4	0.92	77.5	62.5
Adherence to Statin Therapy for Beneficiaries with Coronary Artery Disease	1,315	38.6	67.5	8.9	0.45	66.7	30.1	581	70.9	67.2	8.7	0.49	66.8	23.2
Antidepressant Treatment for Depression:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1. Acute phase treatment (at least 12 weeks)	803	23.5	58.3	10.0	0.39	56.6	32.4	467	57.0	57.6	9.9	0.36	57.1	21.2
2. Continuation phase treatment (at least 6 months)	803	23.5	41.0	9.9	0.39	39.1	32.9	467	57.0	40.2	9.8	0.35	39.8	22.4

Table III.3 *(continued)*

.	25+ EPs							100+ EPs						
Domain and Measure Title	Number of Groups Reporting ≥20 Eligible Cases	Percent Reporting ≥20 Eligible Cases	Mean	SD	Average Reliability	2011 Benchmark	Percent of Groups Different Than the Benchmark (p<0.05)	Number of Groups Reporting ≥20 Eligible Cases	Percent Reporting ≥20 Eligible Cases	Mean	SD	Average Reliability	2011 Benchmark	Percent of Groups Different Than the Benchmark (p<0.05)
Lipid Profile for Beneficiaries Who Started Lipid-Lowering Medications	2,500	73.3	37.5	9.3	0.68	40.3	40.9	734	89.5	37.8	8.3	0.85	39.9	52.8
Breast Cancer Screening for Women Ages 40–69	2,884	84.6	58.4	12.8	0.82	65.2	55.9	71	94.0	61.7	10.9	0.92	65.7	66.0
Patient Safety	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Use of High-Risk Medications in the Elderly[a]	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1. Patients who receive at least 1 drug to be avoided[a]	3,028	88.8	22.8	7.4	0.76	19.7	49.2	776	94.6	22.2	6.0	0.89	19.4	62.3
2. Patients who receive at least 2 different drugs to be avoided[a]	3,028	88.8	6.2	3.2	0.69	3.6	54.9	776	94.6	6.0	2.5	0.84	3.5	73.2
Lack of Monthly INR Monitoring for Beneficiaries on Warfarin[a]	2,644	79.0	36.5	14.5	0.89	32.9	63.0	756	92.2	35.0	11.8	0.95	32.5	70.0
Care Coordination	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Follow-Up After Hospitalization for Mental Illness	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1. Percentage of patients receiving follow-up within 30 days	977	28.6	62.1	13.8	0.63	63.7	39.4	521	63.5	63.7	12.7	0.69	64.1	34.2
2. Percentage of patients receiving follow-up within 7 days	977	28.6	34.7	12.7	0.59	35.9	38.2	521	63.5	35.9	12.1	0.64	36.1	31.4

[a] Lower performance rates on these measures indicate better performance. However, the domain score for this domain has been calculated such that positive (+) scores indicate better performance and negative scores (-) indicate worse performance.

SD = standard deviation

2. Administrative Claims-Based Quality Indicator Reliability

The average reliability of the claims-based quality indicators was markedly lower than for the PQRS quality measures, but was still quite high for some indicators, particularly among groups with 100+ EPs. For groups of both sizes, the indicators within the Patient Safety domain had higher reliability with scores of at least 0.68. The Clinical Process/Effectiveness domain was the only domain in which any indicators had reliability less than 0.5, a result that occurred for two indicators among groups with 25 or more EPs and for three indicators among groups with at least 100 EPs.[9]

C. Administrative Claims-Based Outcome Measures

In addition to either the PQRS quality measures or the CMS-calculated administrative claims-based quality indicators, the 2012 QRURs also reported on three administrative claims-based outcome measures: hospitalizations for acute and chronic ambulatory care sensitive conditions (ACSCs) and all-cause readmission. These measures were includes in all QRURs—for groups that reported quality measures through PQRS (discussed in Section A) and for those for which CMS calculated 14 administrative claims-based indicators (discussed in Section B).

1. Hospitalizations for Ambulatory Care Sensitive Conditions

The QRURs contained each group practice's performance on measures of potentially avoidable hospitalizations for ACSCs. These are conditions for which good outpatient care can prevent complications or more serious disease that lead to hospitalization. The Medicare claims-based measures were derived from Prevention Quality Indicators developed by the Agency for Healthcare Research and Quality. CMS reported on potentially avoidable hospitalizations for six individual ACSCs (bacterial pneumonia, urinary tract infection, dehydration, diabetes, COPD or asthma, and heart failure) and two composite measures of hospital admissions for acute and chronic ACSCs. The two composite measures were included in the Care Coordination domain. The acute conditions composite is a combined measure based on admissions for bacterial pneumonia, urinary tract infection, and dehydration. The denominators for these rates consist of Medicare beneficiaries attributed to the group being assessed. The chronic conditions composite is a combined measure based on admissions for diabetes, COPD or asthma, and heart failure. The denominators for these rates consist of attributed beneficiaries who are diagnosed with the corresponding condition.

The ACSC measures were risk adjusted for age and sex by comparing the medical group practice's actual rate of potentially avoidable hospitalizations to an expected rate. The expected rate reflects the average experience of Medicare beneficiaries in the same age category and of the same gender as those attributed to the group. The risk-adjusted rate is calculated as the ratio of the actual rate to the expected rate multiplied by the average actual rate per 1,000 beneficiaries.

On average, all group types improved upon hospitalization rates for chronic conditions relative to their prior year benchmark, but performance relative to the acute hospitalization rate

[9] Component rates for the Antidepressant Treatment for Depression indicator were below 0.50.

benchmark was mixed. The average hospitalization rate for acute conditions across all groups with 25 or more eligible professionals was 8.7; the rate for chronic conditions was 55.8 (Table III.4). Notably, GPRO groups performed better, on average, than their benchmarks for both composites. The range of performance rates was substantial, particularly for the chronic composite, with a standard deviation of 34.8 per 1,000 beneficiaries for groups of 25 or more EPs.

Table III.4. Performance on ACSC Composite Measures, by Group Size and Type

Group Size	Number of Groups	Mean	SD	Average Reliability	Percent Statistically Significant	2011 Benchmark
25+ EPs	.	.	.	.	.	.
Acute ACSC Composite	.	.	.	.	.	.
All groups	3,840	8.7	9.0	0.71	52.3	8.4
GPRO	67	7.1	1.9	0.95	59.7	8.2
ACO	398	7.9	6.6	0.79	58.9	8.3
Other	3,405	8.8	9.3	0.70	51.4	8.4
Chronic ACSC Composite	.	.	.	.	.	.
All groups	3,560	55.8	34.8	0.79	40.1	58.5
GPRO	67	52.8	12.6	0.98	58.2	58.6
ACO	352	51.7	24.5	0.84	44.8	58.5
Other	3,141	56.3	36.0	0.78	39.2	58.5
100+ EPs	.	.	.	.	.	.
Acute ACSC Composite	.	.	.	.	.	.
All groups	1,027	8.5	6.0	0.84	57.3	8.2
GPRO	61	7.2	1.9	0.96	59.0	8.2
ACO	148	8.4	5.3	0.87	58.9	8.2
Other	818	8.6	6.3	0.83	56.8	8.2
Chronic ACSC Composite	.	.	.	.	.	.
All groups	998	57.2	26.3	0.90	44.5	58.6
GPRO	61	53.9	11.8	0.98	54.1	58.6
ACO-MSSP	146	56.4	19.6	0.94	45.3	58.6
Other	791	57.6	28.1	0.89	43.7	58.6

Note: The number of groups is restricted to groups with at least 20 beneficiaries in the composite measure. Not all groups had at least 20 cases. Percent statistically significant is the percentage of groups whose performance was statistically significantly different from the 2011 benchmark (p-value<0.05).

SD = standard deviation

The average reliability for both measures across all groups was higher than 0.70. Larger groups had, on average, higher reliabilities. Among groups with 25 or more EPs, 64 and 74 percent had reliabilities over 0.70 for the acute and chronic ACSC composites, respectively. The percentages for the subset of groups with 100 or more EPs were higher, at 84 and 91 percent, respectively.

Overall, 52 and 40 percent of all groups had performance rates for the acute and chronic ACSC composites, respectively, that were statistically different from the benchmark mean at the 5 percent level. The percentages were larger for groups of 100 or more EPs than for groups with 25–99 EPs, which is consistent with larger groups having more precisely estimated rates.

2. 30-Day All-Cause Readmissions

The QRURs contained medical group practice-specific 30-day all-cause rate of acute care hospital readmissions for beneficiaries discharged from an acute care or critical access hospital. Only unplanned readmissions for any cause within 30 days from the date of discharge of an index admission between January 1, 2012 and December 1, 2012 are included in the measure. This measure was risk adjusted via hierarchical logistic regression models that estimate the number of readmissions predicted for the specific medical group practice, given its case mix.

The range of performance was relatively narrow and the reliability was generally low for the readmission measure (Table III.5). The average readmissions rate was 15.6 across all groups, varying only slightly by group type. Performance among the subset of groups with at least 100 EPs was similar to that of all groups. On average, each group type performed better (that is, lower) than its benchmark. Average reliability across all groups was 0.27. GPRO groups had the highest reliabilities, on average. Average reliability among the subset of larger groups was higher, at 0.48.

Table III.5. Performance on All-Cause Readmissions Measure, by Group Size and Type

Group Size	Number of Groups	Mean	Standard Deviation	Average Reliability	2011 Benchmark
25+ EPs	.	.	.	.	.
All groups	3,111	15.6	0.9	0.27	16.2
GPRO	67	15.6	1.4	0.69	16.1
ACO	315	15.7	1.0	0.34	16.2
Other	2,729	15.6	0.8	0.25	16.2
100+ EPs	.	.	.	.	.
All groups	967	15.6	1.0	0.48	16.1
GPRO	61	15.7	1.3	0.71	16.1
ACO	143	15.7	1.1	0.54	16.1
Other	763	15.6	1.0	0.45	16.1

Note: The number of groups is restricted to groups with at least 20 beneficiaries in measure. Not all groups had at least 20 cases.

IV. COSTS OF CARE

The QRURs include five cost-of-care measures derived from 2012 administrative claims data: total per capita costs and per capita costs for beneficiaries with four common chronic conditions of diabetes, heart failure, COPD, and CAD.[10] The per capita (per beneficiary) cost measure assesses health care services for all Medicare FFS attributed beneficiaries and for those with chronic conditions. The measure includes all Medicare Part A and Part B costs during a calendar year and is payment standardized and risk adjusted to account for any potential differences in costs among providers that result from circumstances beyond the physician's control. Under CMS's attribution rule, beneficiaries are attributed on the basis of the plurality of primary care services, to medical group practices with the greatest potential to influence the quality and cost of care delivered to Medicare FFS beneficiaries.

These unadjusted per capita costs are first calculated as the sum of all Medicare Part A and Part B costs for all attributed beneficiaries, divided by the number of attributed beneficiaries. All unadjusted costs are then payment standardized and risk adjusted to accommodate differences in costs between peers that result from circumstances beyond physicians' control. Risk-adjusted costs are computed as the ratio of a medical group practice's payment-standardized (but not risk-adjusted) per capita costs to its expected per capita costs, as determined by the risk adjustment algorithm. Finally, to express the risk-adjusted cost in dollars and for ease of interpretation, the ratio is multiplied by the mean cost of all beneficiaries attributed to all groups.

All Medicare payments have been payment standardized such that a given service is assigned the same dollar value across all providers within the same facility type or setting, regardless of geographic location or differences in Medicare payment rates among facilities. More information about how CMS standardized payments can be found in the document describing the methodologies used in the 2012 QRURs, which can be accessed at http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/PY2012-Detailed-Methods.pdf. Moreover, payments have also been risk adjusted to account for patient differences in risk.

A. Overall Per Capita Costs Before and After Risk Adjustment

In addition to payment standardization, CMS also risk adjusts costs to account for patient differences that can affect medical costs, regardless of the care provided. Risk adjustment is intended to enable groups to be compared more fairly to their peers; for instance, groups that serve a disproportionate number of high-risk beneficiaries will have risk-adjusted costs that are lower than unadjusted costs. The QRUR risk-adjustment model is prospective—meaning that it uses 2011 risk scores to predict 2012 costs—to ensure that the model measures the influence of health on treatment provided (costs incurred), rather than the reverse.

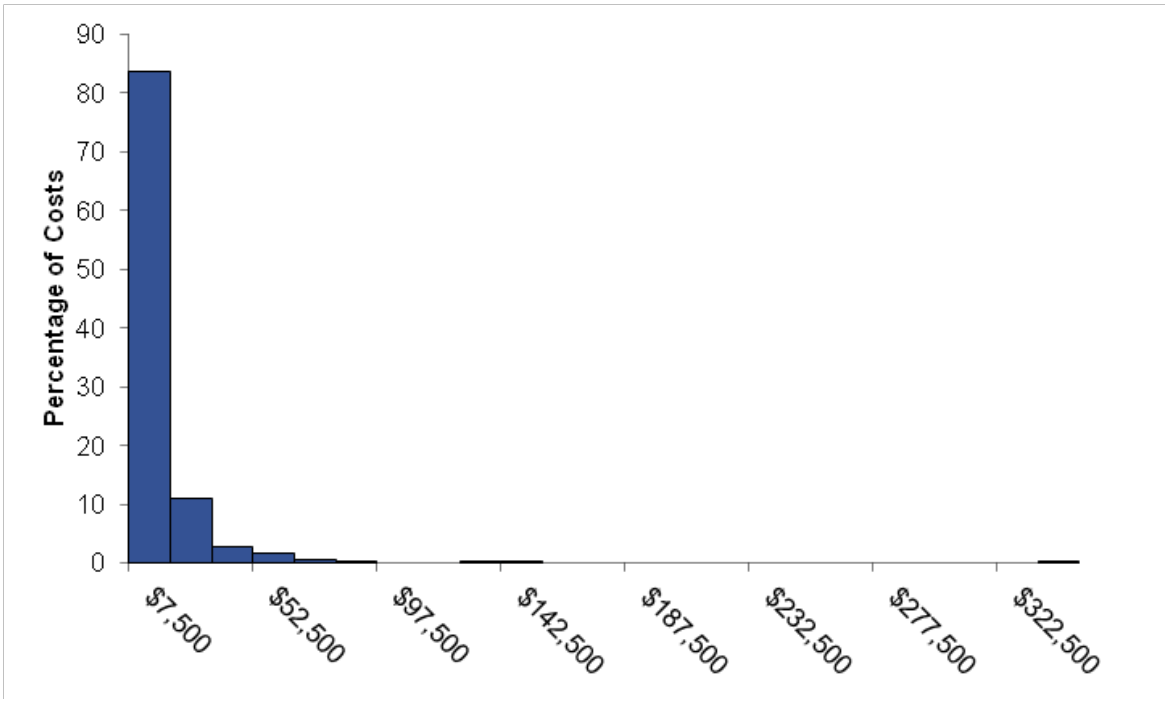
[10] The 2012 administrative claims data include inpatient hospital; outpatient hospital; skilled nursing facility; home health; hospice; durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS); and Medicare carrier (noninstitutional provider) claims. Costs associated with Medicare Part D (outpatient prescription drugs) were not included.

Before calculating risk-adjusted costs, the influence of outliers is limited by resetting the costs of beneficiaries with costs exceeding the 99th percentile to the value of costs in the 99th percentile (a process known as *Winsorization*) and dropping costs below the first percentile (that is, truncation). In the QRURs, 105,788 beneficiaries had their costs reset to the 99th percentile value, which reduced their costs from an average per capita cost of $146,000 to $99,230—a reduction of 32 percent.

After this process, per capita cost measures for the QRUR are risk adjusted using CMS's hierarchical condition categories (HCC) risk scores. The CMS-HCC model produces a risk score that summarizes in a single number each Medicare beneficiary's expected cost of care relative to other beneficiaries, given the beneficiary's demographic profile and medical history. The International Classification of Diseases, Ninth Edition (ICD-9) assigns diagnosis codes to 70 clinical conditions to capture medical condition/cost risk. The model also incorporates patient age, gender, reason for Medicare eligibility (aged or disabled), and Medicaid eligibility. The QRUR risk-adjustment model includes the risk score produced by the CMS-HCC model and an indication of whether a beneficiary was diagnosed in the previous year with end-stage renal disease. More information about how CMS calculates the per capita risk-adjusted costs measure can be found in the document describing the methodologies used in the 2012 QRURs, which can be accessed at http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/PY2012-Detailed-Methods.pdf.

The risk-adjustment process reduced both the overall average per capita costs from $12,815 to $10,788 and the skewness of the cost distribution. In Figure IV.1, we show the distribution of the average unadjusted (payment standardized, but non-Winsorized and non-risk adjusted) total per capita costs across all groups. The costs ranged from $477 to $340,516. In Figure IV.2, we show the distribution of the costs after risk adjustment (and after Winsorization and truncation). Although the distribution after risk adjustment is still skewed, the range of costs was $756 to $58,945.

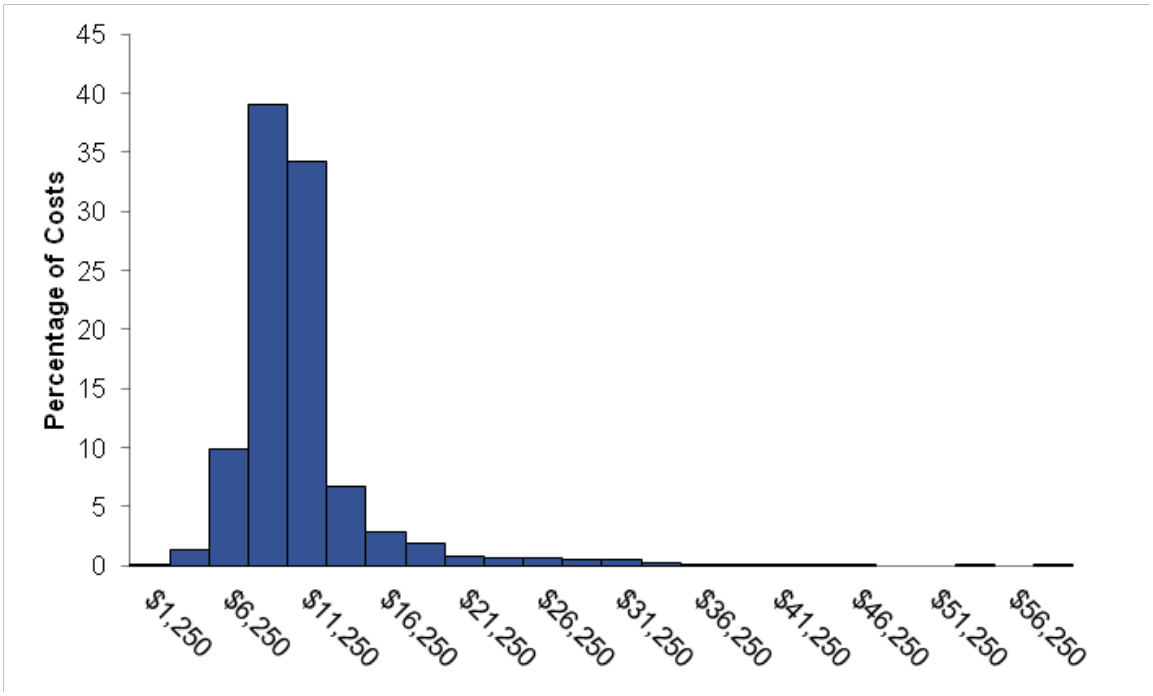
Figure IV.1. Distribution of Payment-Standardized but Non-Risk-Adjusted Costs for Groups with 25 or More Eligible Professionals



n = 3,876 groups.

Note: Non-risk-adjusted and non-Winsorized costs.

Figure IV.2. Distribution of Payment-Standardized and Risk-Adjusted Costs for Groups with 25 or More Eligible Professionals

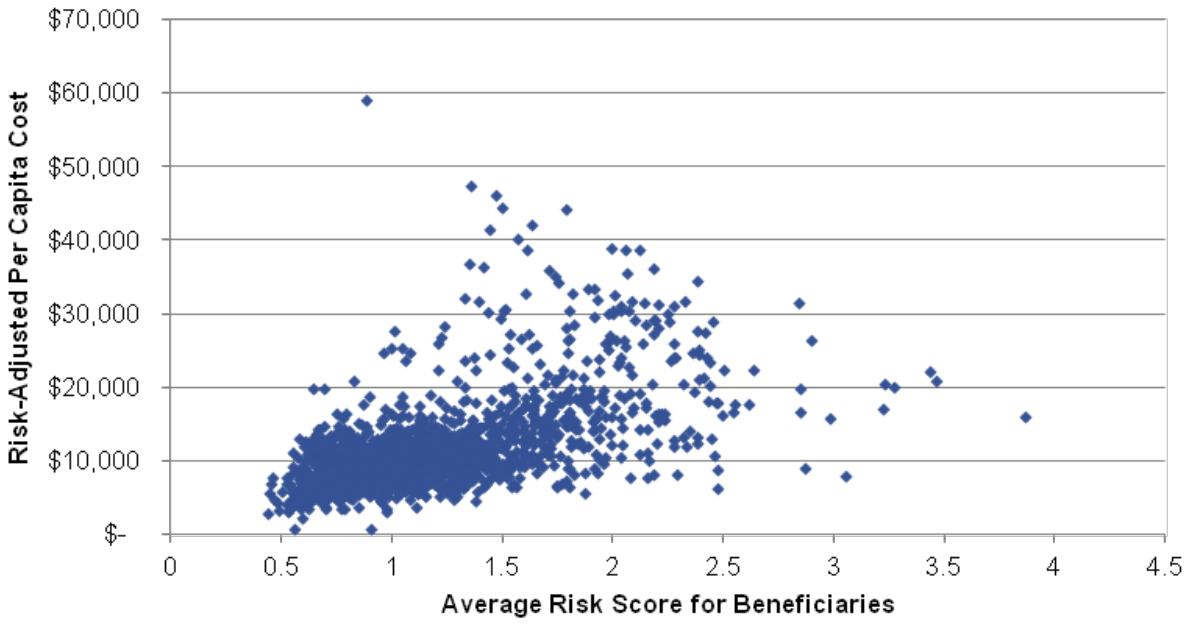


n = 3,876 groups.

B. Relationship Risk-Adjusted Costs and HCC Scores

The correlation between HCC score and risk-adjusted total per capita cost is 0.58 for all groups (Figure IV.3). For the subset of groups with 100 or more EPs, the correlation is 0.61. Risk-adjusted total per capita costs are fairly dispersed at higher levels of risk.

Figure IV.3. Relationship Between Risk Score and Risk-Adjusted Per Capita Costs for Groups with 25 or More Eligible Professionals



n = 3,876 groups.

Note: A score of 1 represents average risk.

C. Condition-Specific Risk-Adjusted Per Capita Costs

The analyses in the remaining parts of this section are restricted to the 3,802 groups that had 20 or more beneficiaries after risk adjustment. Only groups with 20 or more beneficiaries have a cost composite score calculated. In Table IV.1, we present summary statistics for per capita costs by chronic condition—the average cost of providing care for beneficiaries with heart failure being the highest ($26,082), followed by COPD ($24,901), CAD ($17,672), and diabetes ($15,563). As expected, the total costs for treating beneficiaries with these conditions are greater, on average, than the overall per capita costs ($10,734). These disease-specific mean per capita costs did not vary significantly by group size.

Table IV.1. Performance Rates on Cost Measures for Groups with 25 or More Eligible Professionals

.	Number of Groups Reporting	Mean	Standard Deviation	Number of Groups Minus 1 Standard Deviation*	Number of Groups Plus 1 Standard Deviation **	5th Percentile	95th Percentile
All Beneficiaries	3,802	$10,734	$4,282	190	311	$6,452	$17,980
Diabetes	3,251	$15,563	$5,449	164	283	$10,107	$24,800
COPD	2,689	$24,901	$7,471	161	248	$16,906	$38,579
CAD	3,245	$17,672	$6,018	196	289	$11,396	$28,400
Heart Failure	2,846	$26,083	$8,003	211	247	$17,046	$39,912

n = 3,802 groups.

Note: Restricted to the 3,802 groups that had cost composite score.

* Number of groups whose rates are equal to or less than one standard deviation from the mean.

** Number of groups whose rates are equal to or more than one standard deviation from the mean.

D. Statistical Significance

Among groups with 25 or more EPs, slightly more than 57 percent had total per capita costs statistically different at the 5 percent level from the group mean of $10,734. The percentage of groups with 25 or more EPs with available condition-specific per capita costs that were statistically different from the mean was 37 percent for heart failure, 34 percent for COPD, 38 percent for diabetes, and 43 percent for CAD.

Larger groups were more likely to have per capita costs that were statistically different from the national mean. For the subset of groups with at least 100 EPs, 68 percent were statistically different from the mean on the overall cost measure at the 5 percent level. The percentages of groups statistically different from the mean for heart failure, COPD, diabetes, and CAD were 51, 44, 52, and 55 percent, respectively.

E. Reliability of Cost Measures

Reliability describes the overall consistency of a measure; a performance measure is considered reliable if it produces similar results with a different sample of practice beneficiaries. The vast majority of groups had average reliabilities for the total per capita cost measure and the condition-specific measures above 0.7—a commonly used threshold for high reliability.

All group practices with 25 or more EPs achieved an average reliability score of 0.94 for the total per capita cost measure. For all groups, average reliabilities for the condition-specific cost measures ranged from 0.72 to 0.83. The proportion of groups with reliabilities above 0.7 ranged from 62 percent for COPD to 79 percent for CAD.

For larger groups with 100 or more EPs, average reliability was higher for the total per capita cost measure (0.98), as well as for the condition-specific cost measures (0.88 to 0.94). For

these larger groups, 99 percent had a reliability score above 0.7 for the total per capita cost measure. A large proportion of groups had reliabilities above 0.7 for the four chronic condition cost measures: 88 percent for COPD, 87 percent of heart failure, and 94 percent for both diabetes and CAD.

F. Per Capita Costs by Type of Service

The 2012 QRURs report payment-standardized and risk-adjusted per capita costs for all services in total, as well as by detailed type of service. Table IV.2 shows that hospital and outpatient services together accounted for almost half (49 percent) of mean per capita costs. Emergency department visits accounted for the lowest average service-specific per capita cost ($260). On average, evaluation and management visits accounted for $1,090 of a beneficiary's total annual per capita costs. The average per capita costs for post-acute care is also relatively high ($1,634), but the large standard deviation ($1,555) indicates these costs can vary widely from group to group: some have beneficiaries with very high post-acute care costs.[11]

Table IV.2. Per Capita Costs, by Type of Service for Groups with 25 or More Eligible Professionals

.	Mean	Standard Deviation
All Services	$10,520	$3,117
Evaluation & Management	$1,090	$319
Procedures	$768	$259
Hospital Services	$2,566	$1,025
Outpatient Services	$2,595	$1,292
Emergency Department (non-admission)	$260	$93
All Ancillary Services (non-emergency)	$972	$301
Post-Acute Care	$1,634	$1,555
Other Services[1]	$636	$636

n = 3,802 groups

Note: Restricted to the 3,802 groups that had cost composite score.

[1] Other services include ambulance services, chemotherapy and other Part B-covered drugs, and all other services not otherwise classified.

[11] Post-acute care includes service categories such as skilled nursing facilities; home health; and psychiatric, rehabilitation, or other post-acute care.

V. THE VALUE-BASED MODIFIER, QUALITY TIERING, AND PERFORMANCE

Quality tiers, which will be used to determine the VBM for a specific group, are derived from a quality composite score and a cost composite score. The quality composite score summarizes a group's performance across up to six equally weighted quality domains: Clinical Process/Effectiveness, Patient and Family Engagement, Population/Public Health, Patient Safety, Care Coordination, and Efficient Use of Healthcare Resources. The cost composite score summarizes a group's performance across up to two equally weighted cost domains: Per Capita Costs for All Attributed Beneficiaries and Per Capita Costs for Beneficiaries with Specific Conditions: diabetes, CAD, COPD, and heart failure. We have described these measures in detail in sections III and IV.

The quality and cost composite scores in the QRUR summarize a group's performance on multiple individual quality and cost measures. Only measures with at least 20 cases are eligible for inclusion in composite score calculations. The first step in computing a composite score is to standardize the scores for these eligible individual measures by subtracting the peer group benchmark score from the report recipient's score and dividing the result by the standard deviation.[12] Domain scores are then formed for each measure domain by averaging the standardized scores of the measures within that domain. Next, the domain-level scores are combined into an average domain score. (Only domains with at least one measure with at least 20 cases are included in the average domain score.) Finally, the average domain score itself is standardized by subtracting the report recipient's average domain score from the mean average domain score, computed across all physician groups in the report recipient's peer group, and divided by the standard deviation of the average domain score. The result is a score that reflects the report recipient's performance in terms of number of standard deviations above or below the peer group mean.

A. Quality Tiering

Cost and quality composite scores are used to assign groups to low, medium, or high quality (or cost) categories that will be used in quality tiering for the VBM. The following rules are used:

- Groups with standardized quality composite scores of 1 or higher, and whose scores are statistically different from the mean score, are considered high performers.
- Groups with scores of -1 or lower, and whose scores are statistically different from the mean score, are considered low performers.
- All other groups are considered average performers.

[12] For PQRS measures, the peer group is defined as all PQRS participants reporting data on that measure. For all other measures, medical group practices with between 25 and 99 eligible professionals were compared to all medical group practices with a least 25 eligible professionals. Group practices with at least 100 eligible professionals were compared to group practices with at least 100 eligible professionals. The benchmark for all quality measures is the measure's case-weighted mean performance rate of the peer group in 2011. The benchmark for all cost measures is the measure's case-weighted mean performance rate of the peer group in 2012.

- If a group is missing either composite score (because of insufficient cases for all measures in the composite), that group is assigned to average quality and average cost.

The basic structure of the VBM using this approach is displayed below (Table V.1). Because the VBM must be budget neutral, the precise size of the reward for higher-performing groups—those that are at least average on both quality and cost, and better than average on at least one—will depend on the projected billings of these groups relative to lower performing groups (as captured in the table by the variable x), which will vary from year to year with differences in actuarial estimates and in the number and relative performance of medical group practices electing the quality-tiering option. Higher-performing groups treating beneficiaries with an average risk exceeding the risk of the 75th percentile beneficiary in the Medicare population receive an additional 1.0x percent incentive payment on top of the standard upward adjustment, where x is the CMS-calculated payment adjustment factor.

Table V.1. Quality-Tiering Categories and Adjustment

	Low Quality	Average Quality	High Quality
Low Cost	+0.0%	+1.0x%*	+2.0x%*
Average Cost	-0.5%	+0.0%	+1.0x%*
High Cost	-1.0%	-0.5%	+0.0%

Note: x refers to a payment adjustment factor yet to be determined.

*Higher-performing groups serving high-risk beneficiaries (based on average risk scores) are eligible for an additional adjustment of +1.0x percent.

Of the 3,876 groups for which CMS produced a 2012 QRUR and for whom the quality or cost composite could be calculated, the vast majority of groups (80.6 percent) are in the average quality and average cost tiers. Only 8.1 percent of groups are in tiers that would be eligible for an upward adjustment, and 10.9 percent of groups are in tiers that would be eligible for a downward adjustment (Table V.2). Among the groups eligible for an upward adjustment, 11 percent would be eligible for the additional upward adjustment due to treating high-risk beneficiaries.

Table V.2. Distribution of Quality and Cost Tiers for All Groups for Which a Quality or Cost Composite Score Could Be Calculated (3,876 Groups)

	Low Quality	Average Quality	High Quality
Low Cost	0.3%	3.4%	0.8%
Average Cost	3.3%	80.6%	3.9%
High Cost	3.7%	3.9%	0.2%

Among a subset of 1,032 groups with 100 or more EPs, a slightly higher percentage of groups (81.0 percent) are in the average quality and average cost tier (Table V.3). Moreover, the subset of larger groups is less likely to be in a tier that would be eligible for a payment adjustment, either upward or downward. Approximately 8 percent (7.9 percent) of groups are in tiers that would be eligible for an upward adjustment, resulting in a payment incentive of

between +1.0x and +2.0x percent; 10.4 percent of groups are in tiers that would be eligible for a downward payment adjustment of between -0.5x and -1.0x percent of Medicare Physician Fee Schedule payments.

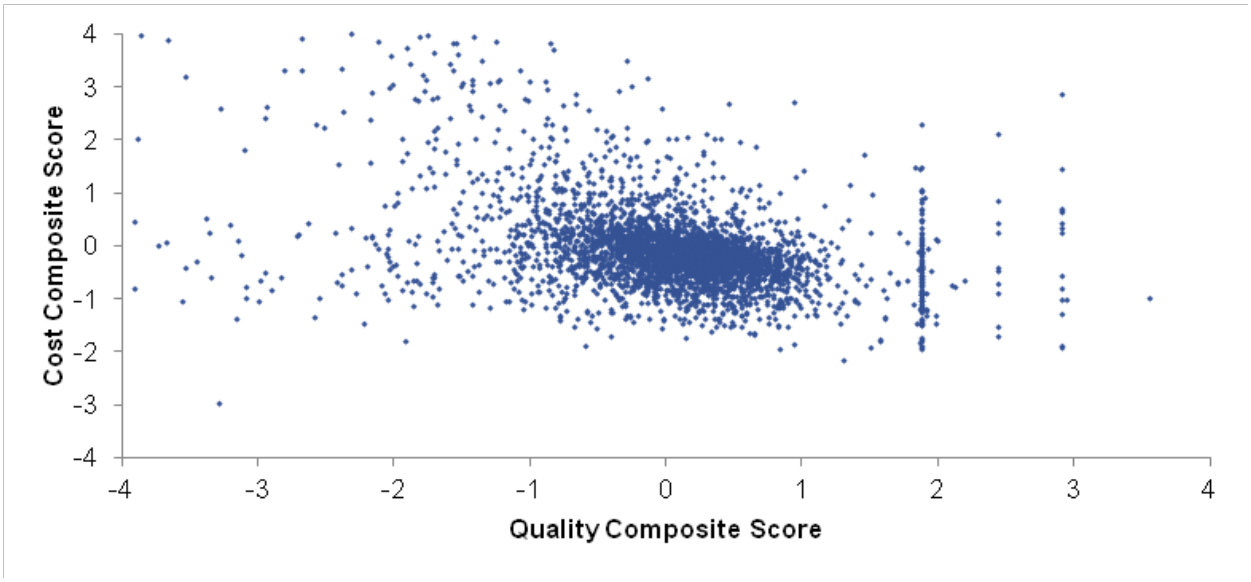
Table V.3. Distribution of Quality and Cost Tiers for Groups with 100 or More Eligible Professionals for Which a Quality or Cost Composite Score Could Be Calculated (1,032 Groups)

	Low Quality	Average Quality	High Quality
Low Cost	0.5%	3.3%	0.7%
Average Cost	4.4%	81.0%	4.0%
High Cost	3.6%	2.4%	0.2%

B. Performance Under the VBM

There is a weak negative relationship between groups' standardized quality and cost composite scores, indicating that, on average, lower quality is weakly associated with higher costs (Figure V.1).

Figure V.1. Relationship Between Standardized Cost Composite Scores Versus Quality Composite Scores, Groups with 25 or More Eligible Professionals



Note: Standardized composite scores between -4 and 4 are shown in the figure. Scores in the figure include those later found not to be statistically different from the mean for the purposes of assigning tiers. Higher cost composite scores indicate lower performance on cost measures.

Groups with high quality scores performed better than groups with average and low quality scores consistently across each of the quality domains (or groupings of quality measures) as well as across the three quality outcomes measures; they also tended to have lower average cost composite scores (Table V.4). Groups with high quality scores had 38 percent fewer admissions for chronic ACSCs and 81 percent fewer admissions for acute ACSCs than groups with average

quality scores. Groups with higher quality scores had a similar readmission rate to groups with average quality scores but a 5 percent lower rate than groups with low quality scores.

Although not a strong relationship, groups with lower cost scores generally performed better on quality measures compared to average- and high-cost groups. The average total per capita cost among low-cost groups was $5,775 compared to $10,013 for average-cost groups and $21,621 for high-cost groups. On average, groups with lower cost scores had lower ACSC hospitalization and readmission rates on average, and had better quality composite scores.

Table V.4. Performance of Groups by Quality and Cost Tier

.	Quality			Cost		
Performance Measure	Low	Average	High	Low	Average	High
Average Quality Composite Score	-2.74	-0.19	1.96	0.19	-0.22	-1.89
Clinical process/effectiveness score	-2.15	-0.43	0.64	-1.06	-0.45	-1.15
Population/public health score	-0.12	0.46	1.81	0.23	0.53	0.56
Patient safety score	-3.28	-0.60	1.13	-0.02	-0.68	-2.03
Care coordination score	-2.33	0.37	2.14	1.25	0.33	-2.15
Acute ACSC composite	23.50	7.55	1.44	3.96	7.57	23.57
Chronic ACSC composite	95.09	51.62	32.07	32.30	52.69	87.90
Hospital readmissions	16.38	15.57	15.50	15.55	15.56	16.37
Average Cost Composite Score	3.67	-0.09	-0.93	-2.51	-0.19	5.67
Per Capita Costs for All Attributed Beneficiaries	$17,888	$10,268	$8,731	$5,775	$10,013	$21,621

Note: Quality and cost composite and individual sub scores are presented as standardized scores, which is a standardized average of equally weighted domain scores. Acute and chronic ACSC composite measures are calculated as rates per 1,000 beneficiaries (for all beneficiaries or for those with the chronic conditions). The hospital readmissions measure is presented as a rate of acute care hospital readmissions per 1,000 attributed beneficiaries.

Groups with fewer than 100 EPs, those with a smaller share of PCPs, and those with patients who had fewer risk factors tended to have, on average, higher quality and lower cost scores (see Table V.5). Larger groups (100 or more EPs) were less likely to have higher quality scores than all groups of 25 or more EPs. Among groups in which fewer than 10 percent of EPs were PCPs, 14 percent were categorized as high quality and 10 percent were low cost. In contrast, 25 percent of groups in which at least 80 percent of providers were PCPs were classified as low quality. Thirty-six percent of these groups dominated by PCPs fell into the high cost category.

Moreover, the patients attributed to high quality and low cost groups had fewer risk factors, on average, than those attributed to groups with average and low quality and cost scores. Among groups in the high quality category, the average risk score of 0.96 compared to 1.07 and 1.59 for average and low quality, respectively. (A score of 1 represents average risk.) Groups categorized as low cost had an average risk score of 0.88. The average risk score for average cost groups was 1.06; for high cost groups it was 1.74.

Among groups composed of at least 80 percent of EPs with the same specialty, 22 percent were classified as high quality compared to 10 percent of groups assigned to low quality. The

percentage of groups classified as low cost and high cost was similar for groups dominated by EPs of the same specialty: 15 percent for low cost groups and 16 percent for high cost groups.

Table V.5. Groups in Each Quality and Cost Tier, by Group Size, Specialty Mix, and Healthiness of Attributed Beneficiaries

.	Quality			Cost		
Group Characteristic	Low	Average	High	Low	Average	High
Distribution by Size	.	.	.	.	.	.
Groups with 25+ EPs (all groups)	7%	84%	6%	5%	83%	8%
Groups with 100+ EPs	9%	84%	5%	5%	88%	7%
Distribution by PCP Share	.	.	.	.	.	.
Groups with 0 to 9% PCPs	5%	76%	14%	10%	75%	8%
Groups with 10% to 49% PCPs	7%	88%	3%	3%	89%	6%
Groups with 50% to 79% PCPs	7%	88%	2%	1%	87%	10%
Groups with 80% to 100% PCPs	25%	68%	2%	1%	60%	36%
Distribution of Groups with at Least 80% of the Same Specialty	10%	63%	22%	15%	57%	16%
Distribution of Groups with HCC risk Score in Top Quartile	23%	67%	5%	2%	62%	31%
Average HCC Risk Score[1]	1.59	1.07	0.96	0.88	1.06	1.74

n = 3,876.

Note: Percentages do not add up to 100 percent because of groups that have a missing quality or cost composite (because of insufficient cases for all measures in the composite).

[1] A score of 1 represents average risk

MATHEMATICA
Policy Research

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC