Comprehensive Primary Care Initiative

eCQM Benchmarking Methodology Report

VERSION 1.0

October 2016

Table of Contents

List of Tables

List of Figures

Introduction

Comprehensive Primary Care Initiative

Introduction and Purpose

The Comprehensive Primary Care (CPC) initiative is a 4-year, multipayer initiative designed to strengthen primary care. Since the launch of CPC in October 2012, CMS has collaborated with commercial and state health insurance plans in seven U.S. regions. Through this collaboration, the Centers for Medicare & Medicaid Services (CMS) and other payers each offer population-based care management fees and shared savings opportunities to participating primary care practices in those regions. These opportunities support the provision of a core set of five comprehensive primary care functions: (1) Risk-Stratified Care Management, (2) Access and Continuity, (3) Planned Care for Chronic Conditions and Preventive Care, (4) Patient and Caregiver Engagement, and (5) Coordination of Care Across the Medical Neighborhood. The initiative is testing whether provision of these primary care functions at each practice site—supported by multipayer payment reform, continuous use of data to guide quality improvement, and Meaningful Use of health information technology—can achieve improved care, better health for populations, and lower costs.[1]

Calendar year 2013 was the first program year for the CPC initiative, and CPC practices began to submit annual results from electronic clinical quality measures (eCQMs) at the practice site level. CMS began offering shared savings in 2014.

To earn shared savings, CPC sites must meet certain standards for quality of care. Three types of quality measures are used: (1) eCQMs, (2) claims-derived measures of readmissions and ambulatory care sensitive admissions, and (3) survey-derived measures of patient experience. The claims-derived and patient experience measures are not the subject of this report and so are not described further; information on them is available in a presentation on the CPC shared savings methodology.[2]

CMS did not score eCQM performance in the 2014 performance year, instead requiring only successful reporting as part of the shared savings quality requirement. Beginning with the 2015 performance year, CMS began scoring eCQMs as a part of determining shared savings payments

---

[1] For more information about CPC, see the following: Centers for Medicare & Medicaid Services. Comprehensive Primary Care Initiative [CMS website]. Updated February 25, 2016. http://innovation.cms.gov/initiatives/comprehensive-primary-care-initiative/.

[2] Centers for Medicare & Medicaid Services. Comprehensive Primary Care Initiative Shared Savings Methodology. Version 2.0. September 2015. https://innovation.cms.gov/Files/x/Comprehensive-Primary-Care-Initiative-Shared-Savings-Methodology-PDF.pdf.

by comparing practice sites' performance against benchmarks.[3] CMS expected that 2015 eCQM performance data reported electronically to the Physician Quality Reporting System (PQRS) and eCQM attestation data from 2014 and 2015 reported directly to the CMS Electronic Health Record (EHR) Incentive Program (Meaningful Use) would be sufficient to establish reliable benchmarks.

This report describes the approach that CMS used to create external benchmarks, select appropriate data sources, and calculate the benchmarks for shared savings in the CPC program. The CPC initiative ends on December 31, 2016; eCQMs from performance year 2016 will be submitted in early 2017.

The report is organized as follows. The remainder of this section provides an overview of CPC quality reporting requirements. Section II provides a conceptual framework for the ideal benchmarks and considerations for benchmarking CPC eCQMs. In Section III, we describe the range of potential data sources for benchmarking. Section IV presents a plan for selecting the best data source. Results of the assessments appear in Section V, followed by a discussion and conclusions in Section VI.

Overview of CPC Quality Reporting Requirements

In the CPC initiative, quality measures function to provide information on whether changes made in practice structure and related processes result in improvement in quality of care and patient experience. eCQMs also are used to inform the distribution of shared savings. Measuring and tracking eCQM performance helps practitioners focus on using their EHRs in ways that support quality measurement and increase EHR usefulness at the point of care. They help guide care-delivery improvement activities by activating decision support and data display that rely on underlying data that are structured.

CMS selected the CPC eCQMs after a review of current measures used in CMS programs for quality reporting. The goal was to identify measures that focus primarily on ambulatory care quality and had high clinical impact and known performance gaps. In an effort to align with other federal programs, particularly those focused on primary care, CMS selected measures used in Stages 1 and 2 of the CMS Medicare EHR Incentive Program (often called *Meaningful Use*), to ensure consistency with CMS electronic measure specification requirements and Office of the National Coordinator for Health Information Technology (ONC) certification. By choosing high-impact measures that align with those in other CMS programs, CPC sought to accelerate their adoption and leverage the unique opportunity created by a test bed of technologically enabled practices. On behalf of the Agency, this test provided experience in the actual collection and reporting of evolving e-measures that already had been selected for broader Agency use.

[3] For more information about the CPC shared savings methodology and eCQM scoring, see the following: https://innovation.cms.gov/files/x/comprehensive-primary-care-initiative-shared-savings-methodology-pdf.pdf.

The underlying premise of CPC is that transformation occurs within the four walls of the practice site as clinicians begin to develop systems and workflows that encourage them to work as a team to deliver care. An emphasis on transformation at the *practice site* drove the CPC decision to collect quality measures at the level of the practice site rather than the individual eligible professional (EP). In this respect, CPC quality measurement is unlike the reporting of eCQMs for PQRS or Meaningful Use, which occur at the level of individual EPs or groups defined by TINs. The population on which CPC quality measures are based includes all patients (not just Medicare patients) who (1) had at least one visit at the CPC practice site location during the measurement year and (2) meet the denominator inclusion criteria for the measure. To capture this population electronically, the ONC-certified EHR technology used by each CPC practice must be configured to calculate and report aggregate eCQM results for all patients who have had at least one visit at the CPC practice site location. An ideal benchmark source likewise would feature aggregate eCQM results.

No other CMS program to date has used data from eCQM performance in value based payment. CPC is setting an Agency precedent, and the results may have important implications for future CMS programs.

Table 1 lists the 13 eCQMs used to measure performance in CPC in the 2015 performance year. They represent the 2014 specifications. The first column shows the CMS ID number and version number for each measure, and the second column contains the corresponding National Quality Forum (NQF) measure number. Three of the measures are no longer endorsed by NQF but remain in the CPC measure set for reasons noted below. The eCQMs provide performance information on process (prevention and cardiovascular care), outcomes (diabetes), and structure (medication documentation) that are important to primary care and geriatrics. They span a wide range of common clinical conditions treated in primary care.

Table 1: Selected Characteristics of eCQM Measures in the CPC Program

CMS Measure Number & Version	NQF Measure Number	Measure Name	Available in 2014	Available in 2015	NQS Domain
165v3	0018	Controlling High Blood Pressure	Yes	Yes	Clinical Process/Effectiveness
138v3	0028	Preventive Care and Screening: Tobacco Use; Screening, and Cessation Intervention	Yes	Yes	Population/Public Health
125v3	0031a	Breast Cancer Screening	Yes	Yes	Clinical Process/Effectiveness
130v3	0034	Colorectal Cancer Screening	Yes	Yes	Clinical Process/Effectiveness

CMS Measure Number & Version	NQF Measure Number	Measure Name	Available in 2014	Available in 2015	NQS Domain
147v4	0041	Preventive Care and Screening: Influenza Immunization	Yes	Yes	Population/Public Health
127v3	0043[b]	Pneumonia Vaccination Status for Older Adults	No	Yes	Clinical Process/Effectiveness
122v3	0059[b]	Diabetes: Hemoglobin A1c Poor Control	Yes	Yes	Clinical Process/Effectiveness
163v3	0064[a]	Diabetes: Low Density Lipoprotein (LDL) Management	Yes	Yes	Clinical Process/Effectiveness
182v4	0075[a,b]	Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control	Yes	Yes	Clinical Process/Effectiveness
144v3	0083	Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	Yes	Yes	Clinical Process/Effectiveness
139v3	0101	Falls: Screening for Future Fall Risk	Yes	Yes	Patient Safety
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	Yes	Yes	Population/Public Health
68v4	0419	Documentation of Current Medications in the Medical Record	No	Yes	Patient Safety

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care; eCQM, electronic clinical quality measure; NQF, National Quality Forum; NQS, National Quality Strategy; v, version

[a] NQF no longer endorses these measures.

[b] CMS will not use 2015 benchmarks for NQF measures 0059, 0064, and the second rate of 0075.

The measures fall into three domains of the National Quality Strategy: Clinical Process/ Effectiveness, Population/Public Health, and Patient Safety. For the 2015 performance year, CPC sites that selected at least nine measures across three domains and met additional requirements, could simultaneously meet the reporting requirements of CPC, Meaningful Use, and the PQRS, as described in Section III.

There are two reporting methods for eCQMs. Sites may submit data electronically to the PQRS data collection system, or they may enter data into the CPC Web Application Attestation Module, a process known as *attestation*. Both require aggregated data generated by certified EHR technology.

Electronic submission occurs through a reporting architecture and data transmission method known as Quality Reporting Document Architecture – Category III (QRDA III). Each QRDA III report contains summary information defined by quality measures, patients, a healthcare provider, and a time period. For example, a report could contain information on two eCQMs for 150 patients for a single physician, measured over the 2015 calendar year. The information on each measure in the report is summarized over the patients rather than being reported separately for each patient.

Table 2 summarizes the number of eCQMs that were required for reporting each year.[4] To earn credit in 2014, sites had to report data on at least 9 measures. The level of performance was not scored in 2014; sites needed time to adjust to program requirements, reporting methods, and (in some cases) new EHR systems, and to establish strategies, processes, and workflows.

Starting in 2015, eCQM performance was phased into the shared savings methodology and sites were required to report on at least 9 of 13 measures. CMS waited until 2015 for several reasons. In 2013, practices could have reported from an EHR according to either 2011 or 2013 certified EHR technology (CEHRT) standards.[5] The electronic specifications of the measures were from Meaningful Use Stage 1, which CMS determined to be inappropriate for purposes of setting CPC benchmarks for several reasons: (1) sites were allowed to report on a measurement period of only 90 days, whereas nearly all measures were designed for a 12-month period; (2) some measure specifications contained errors in 2013; (3) electronic specifications changed in 2014 under Meaningful Use Stage 2; and (4) prior to 2014 many practices were learning how to structure their EHRs to collect data for both decision support and quality improvement.[6]

Increasing the CPC measure set from 11 to 13 eCQMs in 2015 provided CPC practices with a larger selection of measures from which to choose. The additional flexibility afforded by the increase was important for two reasons. First, some EHRs do not support reporting for the full set of eCQM measures. Second, some measures were no longer concordant with clinical guidelines. By maintaining the 2014 measure set in later years, practices that were able to report 9 of 11 eCQMs in 2014 were well positioned to report these measures again in 2015 and 2016.

---

[4] For additional reporting requirements, see Centers for Medicare & Medicaid Services. Comprehensive Primary Care Initiative eCQM User Manual. Version 4.1. October 2015. https://innovation.cms.gov/Files/x/PY-2015-CPC-EHR-CQM-Manual-v4.pdf.

[5] EHR certification criteria are published on the ONC web site: https://www.healthit.gov/policy-researchers-implementers/standards-and-certification-regulations.

[6] The CMS web site provides resources that explain the stages of Meaningful Use. Program requirements for 2015 appear at https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/2015ProgramRequirements.html.

Table 2: CPC Quality Reporting Requirements by Performance Year

Year	Standard for Shared Savings	Reporting Options[a]	No. of eCQMs Available	Minimum No. of eCQMs Required to Report	No. of eCQMs in Shared Savings Calculation[b]
2014	Reporting eCQMs	QRDA III, attestation	11	9	Any 9 successfully reported
2015	Performance on eCQMs	QRDA III, attestation	13	9	Up to 9
2016	Performance on eCQMs	QRDA III, attestation	13	9	Up to 9

Abbreviations: CPC, Comprehensive Primary Care; eCQM, electronic clinical quality measure; QRDA III: Quality Reporting Document Architecture, category 3

[a] QRDA III represents electronic reporting to the Physician Quality Reporting System data collection system. Sites attest eCQM aggregate results to the CPC Web Application Attestation Module.

[b] Sites could report more than nine measures each year. In 2014, sites simply had to successfully report nine measures.

Section 1. Conceptual Framework for an Ideal Benchmark

1.1 Primary Considerations for Benchmarking CPC CQMs

CMS will determine the performance of practice sites beginning in performance year 2015. The goal is to compare sites' performance rates on eCQMs against benchmarks at the 25th, 50th, and 75th percentiles of absolute national performance. In identifying possible sources of data for benchmarks, we aimed to match as closely as possible the characteristics of CPC measures, practice sites, patients, and data. We considered the following specific criteria to be most important:

- **Reporting period.** The reporting period for benchmark data and CPC data should be the same, such as a calendar year. Because the eCQMs are annual measures and practice sites report data annually, the ideal benchmark would be based on annual data.
- **Measure specification.** A measure specification provides a precise definition of all elements of a measure, such as the population it applies to, inclusion and exclusion criteria, and time period of measurement. An ideal benchmark data source should specify a measure exactly the same as the corresponding eCQM measure. Measure specifications can vary across programs, over time, and by reporting method (e.g., Medicare claims, data registry, or direct EHR submission). CPC uses the eCQM specifications, so the ideal benchmark dataset likewise should include electronically specified measures.
- **Payers included.** CPC allows practice sites to report on all patients, not just Medicare beneficiaries. The ideal benchmark would include all-payer data.

- **Aggregated reporting.** Because CPC practices report data aggregated to the practice-site level, the ideal benchmark also would be based on site-level data.
- **Primary care specialty.** CPC is a primary care initiative, so an ideal benchmark would be based on the performance of primary care practitioners.
- **Geography.** CPC practices are in a range of distinct geographies, and CMS wishes to set national performance standards, so the ideal benchmark would include practitioners across the nation.
- **Data quality.** The data should be valid and complete. Reported date should accurately reflect the care that was provided according to the measure specification, and without gaps or logical inconsistencies. *Completeness* represents filled data elements.
- **Statistical reliability.** Reliability of each benchmark is essential when policy decisions that affect patient care are at stake. It represents the likelihood that a similar level of performance would be obtained if another sample of data were drawn from a practice site. In a particular sample of data, reliability will depend on sample size, the variation across reporting units, and measurement error.[7] Reliability may be defined as the proportion of total variation in the data attributable to true differences across reporting units rather than to measurement error. The greater the proportion of total variation, the better our ability to distinguish between the performance of any two reporting units. CMS prefers that a measure have at least 70 percent reliability, a widely used standard across many areas of medicine.[8]
- **Choice over reporting.** CPC providers had to report on at least 9 of 13 measures in 2015. Ideally the benchmark data would come from providers who had a similar options as to the number of measures they reported. The more choices that providers have, the more likely it is that they will choose measures on which they perform best.

Section 2. Data Sources

Insurance claims have long been the primary source for performance measurement. Nearly all of the measures that CMS selected for the CPC program, however, require data beyond that normally found in claims. Two are based on laboratory results, four involve information from the patient's medical history, and one pertains to documentation in the medical record.

---

[7] Adams JL, Mehrotra A, McGlynn EA. Estimating Reliability and Misclassification in Physician Profiling. Technical report. Santa Monica, CA: RAND Corporation; 2010. http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR863.pdf.

[8] See, for example, these studies: (1) Staggs VS. Reliability assessment of a hospital quality measure based on rates of adverse outcomes on nursing units. Stat Methods Med Res. Dec. 31, 2015. doi: 10.1177/0962280215618688. (2) Shih T, Dimick JB. Reliability of readmission rates as a hospital quality measure in cardiac surgery. Ann Thorac Surg. 2014;97(4):1214-1218.

From the outset, data submitted electronically to the PQRS program appeared to be the best source for benchmarking. The CPC measures are a subset of the eCQM measures used in PQRS and Meaningful Use, and CMS uses these measures in other demonstration programs. As a result, we gave first consideration to analyzing data submitted electronically to the PQRS program. Appendix A presents information on other potential data sources.

2.1 eCQM Data Reported to the Physician Quality Reporting System

PQRS is a Medicare program through which physicians and certain other providers can report on their performance on selected quality measures.[9] For PY2014 and PY2015, participants were required to report on at least nine measures covering at least three CMS National Quality Strategy (NQS) domains. Doing so for 2014 would earn an incentive payment in 2015 and avoid a downward payment adjustment in 2016 Medicare payments; starting with 2015, only the downward adjustment applied.[10]

The basic unit of PQRS reporting is the EP. Three types of providers are included in the PQRS definition of EPs:

- Doctors: dentists, oral surgeons, and doctors of medicine, osteopathy, podiatry, and optometry, and chiropractic
- Practitioners: physician assistants, nurse-practitioners, clinical nurse specialists, certified registered nurse-anesthetists (including advanced practice registered nurses) and anesthesiologist assistants, certified nurse-midwives, clinical social workers, clinical psychologists, registered dietitians, nutritional professionals, and audiologists
- Therapists: physical, occupational, and qualified speech-language therapists.

In 2013, approximately 51 percent of eligible professionals nationwide participated in PQRS.[11] We cannot know whether the distribution of performance results among these 51 percent is similar to that of all EPs, nor how the geographic distribution compares with the location of the seven CPC regions. These represent limitations of our approach.

Groups represent two or more EPs who report under a single Tax Identification Number (TIN) and who have assigned their billing rights to the TIN. Groups have the option to report through

---

[9] Centers for Medicare & Medicaid Services. Physician Quality Reporting System. [CMS website]. Last modified December 23, 2015. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html.

[10] Centers for Medicare & Medicaid Services. Physician Quality Reporting System, Analysis and Payment. [CMS website]. Last modified March 4, 2016. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/AnalysisAndPayment.html.

[11] Centers for Medicare & Medicaid Services. Participation Continues to Rise in Medicare Physician Quality Reporting System and Electronic Prescribing Incentive Program. CMS Fact Sheet; April 23, 2015. https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-04-23-1.html.

the Group Practice Reporting Option (GPRO). An EP who reports data to PQRS through GPRO cannot simultaneously submit data as an individual under the same TIN.

GPRO itself has several reporting options[12]:

- through a qualified registry
- electronic health record (EHR) reporting, through direct submission or an EHR data submission vendor (DSV), using the QRDA III standard
- through the GPRO Web Interface, a web-based tool.

The EHR reporting option was the best fit as a potential source of CPC benchmarks. Registries may not report most or all eCQMs selected for CPC, which would limit the sample size available for benchmarking. Data submitted through the GPRO Web Interface were unsuitable because they use a sampling methodology to report performance, represent only Medicare FFS patients, and because only practices with 25 or more EPs could use the Interface.

2.1.1 Data Submission Standard

Data submitted electronically to the PQRS program can be submitted under two technical standards for the exchange of clinical data: Quality Reporting Document Architecture Category I (QRDA-I) and QRDA Category III (QRDA III).

Individual EPs are represented by unique combinations of National Provider Identification (NPI) and Tax Identification Number (TIN). For group practice reporting, data are aggregated for each reported measure for all EPs under the TIN.

Under QRDA-I, a single report (file) is submitted for each patient who meets the Initial Patient Population criterion of a measure. Multiple measures for the same patient can be reported in a single submission. Under QRDA III, measure performance for individual patients is aggregated into summary data for each measure. For individual reporting, the aggregation is over all relevant patients seen by the EP. For group reporting, the aggregation is over all relevant patients seen by all EPs in the group under the TIN. We do not expect different levels of average performance between individual EPs reporting under the QRDA-I and QRDA III standards.

Section 3. Plan for Selecting a Source

3.1 Overview

Our goal was to create absolute benchmarks using external data sources. CMS will then compare CPC practice site performance to these benchmarks. CPC practices earn an increasing

[12] Centers for Medicare & Medicaid Services. Physician Quality Reporting System (PQRS) Group Practice Reporting Option (GPRO) 2015 Criteria. October 2015. https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/pqrs/downloads/2015_pqrs_gpro_criteria.pdf.

number of points at the 25th, 50th, and 75th percentiles, which are considered *gates* of absolute national performance.

Our first task was to analyze potential data sources to determine which was most appropriate for benchmarking. Then we determined the distribution of scores and calculated threshold values to serve as the benchmarks.

In the next section, we present our criteria for choosing the most appropriate data source.

3.2 Criteria

We prioritized the criteria named above, which reflect conceptual and statistical aspects of data quality and sufficiency as well as needs of the CPC initiative. The benchmarks for these quality measures represent the standard against which the sites are compared. To make accurate comparisons, CMS needs high-quality, reliable benchmarks that are based on data that can support the same measure specifications.

Table 3: Criteria for Judging Potential Data Sources

Characteristic	Criterion	Assessment Method
Primary Criteria		
Quality	Have known and minimal levels of reporting error	Statistical analysis
Specification	Match those of CPC eCQM measures, including version number	Program rules and data source features
Reliability	Achieve at least 70% statistical reliability	Statistical analysis
Secondary Criteria		
Reporting Period	Derive from the same time period as CPC data	Data source
Geography	Are nationally representative	Program rules
Reporting level	Are reported at the group level	Program rules and data source
Primary care specialty	Contains data from practitioners who provide primary care	Program rules and data source
Measure choice	Allow little or no choice regarding which measures to report	Program rules
Payers	Include all payer types	Program rules

Abbreviations: CPC, Comprehensive Primary Care Initiative; eCQM, electronic clinical quality measure

CMS considered the most important characteristics to be matching specifications, quality, and reliability. *Specifications match* when the measures in the benchmark source are specified exactly as in the CPC program. For purposes of our analysis, *quality* was limited to what could be ascertained analytically through assessment of incorrectly calculated or missing values and of performance measure distributions. *Reliability* is a statistical assessment of the extent to which observed variation across reporting units reflects actual variation. CMS prefers to use a

minimum threshold of 70 percent reliability.[8] The reliability formula appears in section 3.4.3 below.

There were several possible causes of invalid values: (1) the wrong information was entered into the EHR or into the CPC web application during the attestation process, (2) the data were entered into or stored in a location within the EHR that is not searched according to the measure specification, or (3) the data were corrupted during electronic transmission. The first type of error can be noticed only when the mistake is obvious, such as breast cancer screening reported for an infant or a ratio numerator that exceeds its denominator. We cannot observe the second type of error—problems with EHR data storage—although the EHR vendor and the CPC practice site could locate them. The third type of error—data corruption—would be noticeable if many figures were invalid or missing.

We also identified five characteristics of secondary importance, listed in Table 3. They assess the similarity of the practices and populations underlying the potential data source to those of the CPC demonstration sites.

- *Geography* refers to the national representativeness of the data. We cannot assess that fully, so we considered this criterion to be satisfied if the source data came from all regions of the country.
- *Reporting level* represents the unit of reporting. It can be an individual EP or a set of EPs reporting as a group. Because CPC data are reported at the practice site (i.e., aggregate) level, CMS selected a benchmark data source with group-level reporting.
- *Primary care specialty* refers to using data submitted by primary care practitioners, or at least by practitioners who mostly submit measures within the scope of primary care.
- *Measure choice* reflects the ability of providers to choose whether to report an individual measure. All else being equal, we expect that providers will tend to report measures on which they score well. The more choice allowed, the stronger the potential upward bias in the distribution of reported results.
- *Payers* refers to the insurance mix of reported data, including uninsured/self-pay. We cannot exactly match the payer mix in CPC data because neither reporting method (attestation or PQRS QRDA) reports the proportion of records attributable to each insurance category. Instead we will look for benchmark data sources that feature all insurance types, similar to CPC.

3.3 Process for Selecting Benchmark Data

Using the criteria in Table 3, we developed an approach for selecting the best data source for each measure:

1. Determine how closely the source specification matches the eCQM definition. Retain the source if they are identical.
2. Analyze data quality. If the number of valid responses for a measure was too low across several measures, start over with the next source.

3. Estimate the measure for each site, and determine the statistical reliability of site estimates. If fewer than 80 percent of sites have reliability of 70 percent or more and another source exists, start over with the next source.
4. Report on secondary criteria listed in Table 3. If the results are acceptable to CMS, select the data source for use.

The decisions in steps 2 through 4 were necessarily open to judgment. There are no widely accepted standards for these characteristics to rely upon. As described below, however, the quality and reliability of GPRO QRDA III data were so high, and the secondary criteria so reasonable, that CMS felt confident that the data represented a sound choice.

3.4 Assessing the 2015 Data

The remainder of this section describes how we assessed each characteristic in 2015 CPC and PQRS GPRO QRDA III data. There were three sources of information: (1) CMS program regulations and program documentation, (2) earlier findings on the quality of CMS datasets, and (3) original data analysis.

3.4.1 Data Quality

We judged data quality by the following factors:

- For each measure and across all measures, the percentage of records with missing values of the numerator, denominator, or both;
- Percentage of instances in which the numerator exceeded the denominator;
- Exclusion and exception rates[13]: for each measure whose definition features an exclusion or an exception, the proportion of CPC practice sites or PQRS groups who excluded at least five percent of cases, and the proportion who excepted at least five percent of cases.

Appendix B provides additional detail.

An ideal dataset should have no invalid data, whether missing or out of range. Considerably different exclusion or exception rates between CPC and data submitted electronically to the PQRS program could indicate differences in patient characteristics, practice patterns, or data quality.

3.4.2 Ability to Match eCQM Measure Specification

CPC uses the same eCQM measure specification as PQRS. This criterion would have been a differentiator only if we had needed to consider some other benchmark data source.

[13] A clinical quality measure is a mechanism for assessing the degree to which a provider competently, safely, and timely delivers clinical services that are appropriate for the patient. *Exclusions* represent individuals who should not receive the indicated service or who are not at risk for the clinical outcome. *Exceptions* are allowable reasons for not meeting the standard of the quality measure. Not all measures have exceptions and exclusions.

3.4.3 Statistical Reliability

We used a standard approach to determine the reliability of performance on each measure.[14,15] The unit of measurement is the practice group for PQRS and the practice site for CPC. Following the notation of Adams, Mehrotra, Thomas, and Glynn (2010), we signify the variance between reporting units as $\sigma^2_{\text{provider-to-provider}}$ and the variance of reporting-unit-specific error as $\sigma^2_{\text{provider-specific-error}}$. Reliability is defined as follows:

$$\text{Reliability} = \sigma^2_{\text{provider-to-provider}} / (\sigma^2_{\text{provider-to-provider}} + \sigma^2_{\text{provider-specific-error}}).$$

The two variance figures are derived from a hierarchal linear model. We used the SAS® macro *betabin* to fit a beta-binomial distribution to the provider-level calculated rate for each measure. This method enables calculation of the elements in the formula above.

Each reporting unit had a reliability value for each measure. These values were aggregated to form an overall reliability assessment for the measure and data source. We chose as a standard the percentage of reporting units that achieved at least 70 percent reliability. If at least 80 percent of the reporting units for a measure have 70 percent reliability or greater, we concluded that the measure has sufficient reliability in that data source.

3.4.4 Secondary Characteristics

The secondary data characteristics noted in Table 3 can be determined from program rules published on the CMS website and in federal regulations.

3.5 CPC Practices Reporting to PQRS

Electronic data reported to the PQRS program includes data submitted by CPC practice sites. Although it would appear unusual to use the same data to create a benchmark that will be judged by that benchmark, we deemed it reasonable in this case. First, CPC practice sites constitute only a very small proportion of all groups reporting data electronically to PQRS. Second, CPC practice sites that did not take a waiver from PQRS reporting are submitting data to PQRS separately.[16] Even if desired, we could not remove them from the PQRS data because the TINs used for PQRS practice identification do not exactly overlap with the definition of CPC practices. Finally, if the program were expanded and made permanent the data that a practice reports would be included in the national benchmark. Thus including them in the benchmark at this stage was appropriate.

---

[14] Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling – reliability and risk of misclassification. N Engl J Med. 2010;362(11):1014-1021.

[15] Adams JL, Mehrotra A, McGlynn EA. Estimating reliability and misclassification in physician profiling. Technical report. Santa Monica, CA: RAND Corporation; 2010. http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR863.pdf.

[16] CPC practice sites that elected a PQRS waiver could, by meeting certain reporting requirements, satisfy both CPC and PQRS programs without separately reporting data to PQRS.

3.6 Aggregating Data

3.6.1 Options

CPC collects data at the practice-site level, similar to group reporting in PQRS, so it made sense to develop benchmarks using group-level data. There were several options:

- Use only the group-level data
- Supplement the group-level data with EP-level data that has been aggregated to the group level
- Aggregate group-level and EP-level data, and ignore the difference in aggregation

The appropriate approach was to use QRDA III group-level data, because they are the most similar to CPC practice data. Two potential drawbacks are small sample sizes and the possibility that the distribution of performance within group-level data does not match that of all reporting groups and EPs. A bias could arise if, for example, better-performing providers are more likely to report as EPs than as groups. We do not have reason a priori to expect such a correlation, but one could exist nonetheless.

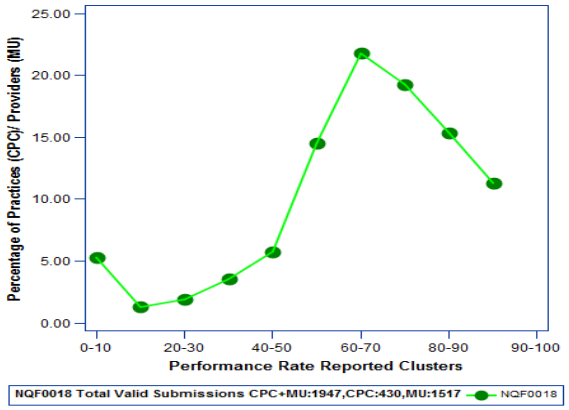
3.7 Calculating the Benchmark Gates

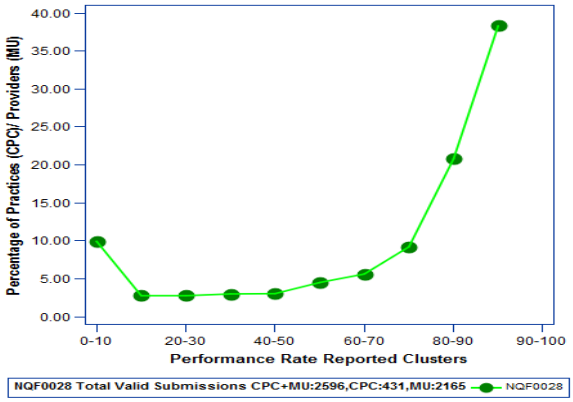
CPC practice sites were scored on their performance on claims-based measures and patient-experience measures for the 2014 performance year. Each site received points for a measure on the basis of how it fared relative to three *benchmark gates*. For claims-based measures, the gates were the 25th, 50th, and 75th percentiles of the benchmark distribution; for patient-experience measures, the gates were 2 standard deviations (SDs) below the mean, the mean, and 2 SDs above the mean. The 2015 benchmark gates for the eCQMs likewise consist of the 25th, 50th, and 75th percentiles of the observed distribution. CMS assigns positive point values to the three performance zones above the 25th percentile (25th-50th, 50th-75th, and above 75th).[17]

The distribution of measure performance could have several shapes, such as those illustrated in Figure 1 (derived from 2014 CPC data). If the observed distribution is heavily skewed with a large mass of observations near the 0 percent or 100 percent marks, it may be difficult to distinguish performance. CMS could address this scenario by changing its methodology to assign similar points to the zones on either side of a gate. For example, suppose that the 75th percentile falls in the middle of a large mass of observations. CMS could respond by assigning the same, or nearly the same, level of points to scores above the 75th percentile and between the 50th and 75th percentiles.

[17] For more information, see https://innovation.cms.gov/files/x/comprehensive-primary-care-initiative-shared-savings-methodology-pdf.pdf.

Figure 1: Examples of Measure Distributions





A. Normal Distribution with Uneven Tails B. U-Shaped Distribution

C. Large Mass at or Near Zero

3.7.1 Problems with Three eCQMs in 2015

CMS decided not to compare the performance of CPC sites to benchmarks for three eCQMs for the 2015 project year: NQF 0043 (Pneumonia Vaccination Status for Older Adults), the second performance rate of NQF 0075 (Ischemic Vascular Disease (IVD): Complete Lipid Panel & LDL Control), and NQF 0059 (Diabetes: Hemoglobin A1c Poor Control). For NQF 0043, the guidelines changed while CMS was using the measure for CPC and we did not want to set a benchmark for a measure with outdated guidelines. For NQF 0059 and for the second performance rate of NQF 0075 ("patients whose most recent LDL-C level performed during the measurement period is <100 mg/dL"), problems with how the measures had to be coded in the EHR yielded inaccurate performance rates. We did not want to benchmark measures that were inaccurate. Thus, we neither analyzed nor created benchmarks for NQF 0043, NQF 0059, and the second performance rate for NQF 0075. We did, however, analyze and create a benchmark

for the first performance rate of NQF 0075 ("patients with a complete lipid profile performed during the measurement period").

Section 4. Results

4.1 Data Quality and Counts of Records

This chapter discusses results of the assessment of data quality for 2015 CPC practice sites and 2015 PQRS GPRO QRDA III data. It also presents the final 2015 benchmarks derived from the GPRO data.

4.1.1 CPC Data

Sites Reporting Valid Values

A requirement for participation in CPC was either the ability to submit quality measure results at the practice level through a direct export from the practice's EHR system or the ability to produce the measure numerator, denominator, and any other required measure characteristics automatically from the EHR and input it into the CPC Web Application Attestation Module. Further, practice sites must have installed the 2014 specification for each measure. For these reasons, a substantial amount of quality control was built into the EHR.

For the 2015 performance year, 450 CPC sites submitted 4,055 records through attestation and 587 records through QRDA III. A record represents performance results for a single measure. The quality of data was very high: for 4,109 attestation records (99.1 percent) and 587 QRDA III records (100 percent), the numerator and denominator were non-missing and positive, and the numerator was less than or equal to the denominator.

Table 4 shows the number and percentage of CPC sites reporting each measure in 2015 with non-zero numerator and denominator. Figures combine attested and QRDA III submissions. Only two measures were reported by fewer than two-thirds of sites: NQF 0083 (Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction) and NQF 0418 (Screening for Clinical Depression and Follow-Up Plan).

Table 4: Number and Percent of CPC Sites Reporting Valid Values, by Measure, 2015[a]

CMS Measure Number & Version	NQF Measure Number	Measure Name	Number	Percent
165v3	0018	Controlling High Blood Pressure	441	98.0
138v3	0028	Tobacco Use; Screening, and Cessation Intervention	408	90.7
125v3	0031[b]	Breast Cancer Screening	412	91.6
130v3	0034	Colorectal Cancer Screening	429	95.3
147v4	0041	Preventive Care and Screening: Influenza Immunization	374	83.1
163v3	0064[b]	Diabetes: Low Density Lipoprotein (LDL) Management	409	90.9
182v4	0075[b]	Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control (first performance rate)	362	80.4
144v3	0083	Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction	119	26.4
139v3	0101	Falls: Screening for Future Fall Risk	318	70.7
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	208	46.2
68v4	0419	Documentation of Current Medications in the Medical Record	313	69.6

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care; NQF, National Quality Forum

[a] Excludes practice sites that reported zero for both the numerator and denominator.

[b] No longer endorsed by NQF

Exclusion and Exception Rates

Four CPC eCQMs had exclusions in their specifications, and six had exceptions. Appendix C presents a brief verbal description of the exclusions and exceptions. Here we describe their frequency.

In the CPC data, fewer than five percent of sites reporting measures 0018, 0031, and 0034 excluded more than five percent of cases. Of the 208 sites that reported measure 0418, however, one-third (33.3 percent) excluded at least five percent of cases. Exclusion rates were uniformly higher among sites that reported data by attestation rather than electronically through PQRS.

Fewer than three percent of sites reporting measures 0028, 0148, 0101, and 0419 had exception rates of five percent or more. For 0041 more than 18 percent of sites excepted at least five percent of cases, while for 0083 it was more than 25 percent. There was no clear pattern distinguishing sites that reported by attestation from those that reported electronically.

4.1.2. 2015 PQRS GPRO QRDA III Data

Sites Reporting Valid Values

In 2015 GPRO group practices submitted 5,633 records for the selected eCQMs. The fourth column of Table 5 shows the number of groups (TINs) reporting each measure in 2015, excluding those with zero numerator and zero denominator. We do not calculate a percentage because the total set group practices reporting by GPRO includes many who did not report any eCQMs used in CPC.

Software used by many group practices to report measures through GPRO will submit zero-valued numerators for selected measures even when the group practice did not intend to report on them. That is, PQRS requires nine measures to be submitted though many GPRO reporters configure software such that all available measures are submitted. Their zero values most likely do not represent the actual performance of the group. Unfortunately, the PQRS system cannot identify which measure reports fall in this category. We therefore dropped zero-valued measure reports from GPRO data if the group reported at least nine additional eCQMs with non-zero numerators. Although this could inadvertently delete some valid zero performance rates, we expect that such errors will be rare.

The final column of Table 5 shows the number of records remaining after eliminating these records as well as records with zero for both the numerator and denominator. A total of 495 records were dropped, ranging from 2 (2.4 percent) for NQF 0083 (Beta-Blocker Therapy for Left Ventricular Systolic) to 134 (60.1 percent) for NQF 0418 (Screening for Clinical Depression and Follow-Up Plan).

Table 5: GPRO Practices Reporting Valid Values, by Measure, 2015

CMS Measure Number	NQF Measure Number	Measure Name	Number of Practices Reporting	
			With Selected Zeros [a]	Without Selected Zeros [b]
165v3	0018	Controlling High Blood Pressure	602	597
138v3	0028	Tobacco Use; Screening, and Cessation Intervention	783	780
125v3	0031[c]	Breast Cancer Screening	442	418
130v3	0034	Colorectal Cancer Screening	476	463
147v4	0041	Preventive Care and Screening: Influenza Immunization	477	447
163v3	0064[c]	Diabetes: Low Density Lipoprotein (LDL) Management	273	223
182v4	0075[c]	Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control (first performance rate)	239	198
144v3	0083	Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	84	82
139v3	0101	Falls: Screening for Future Fall Risk	280	171
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	223	89
68v4	0419	Documentation of Current Medications in the Medical Record	848	844

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care; NQF, National Quality Forum

[a] Excludes practice sites that reported zero for both the numerator and denominator.

[b] Excludes practice sites that reported zero for both the numerator and denominator and sites with zero-valued measure reports from GPRO data if the group reported at least nine additional eCQMs with non-zero numerators.

[c] No longer endorsed by NQF

Exclusion and Exception Rates

The exclusion rates did not indicate any problems with data quality. In the PQRS GPRO data, fewer than three percent of groups excluded more than five percent of cases for measures 0031 and 0034. Nearly six percent of groups excluded more than five percent of cases for measure 0018, and more than 14 percent did so for measure 0418. There was no clear pattern relative to exclusion rates among CPC sites.

Results were broadly similar for denominator exceptions. Fewer than one percent of GPRO groups excepted more than five percent of cases for measures 0028, 0101, and 0419. There were no exceptions at all for measure 0418. A higher proportion of GPRO groups than CPC practice sites excepted more than five percent of cases for the two remaining measures: over 27 percent for 0041 (versus 18 percent for CPC) and over 44 percent for 0083 (versus 25 percent for CPC). These could reflect differences in patient population, however. We do not see patterns clearly indicative of a data quality issue

4.2 Ability to Match eCQM Measure Specifications

As noted earlier, data submitted electronically to the PQRS program through GPRO use the same eCQM measure specifications as the CPC practices.

4.3 Statistical Reliability

We assessed the statistical reliability of 2015 CPC practice data and 2015 GPRO QRDA III PQRS data using the method described in Section 3.4.3 above. Results are summarized here; full tables of results appear in Appendix D.

We found extremely high reliability for every measure among 2015 CPC practice sites. Between 98 percent and 100 percent of sites achieved reliability of 0.70 or greater on every measure except NQF 0083 (Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction). More than 80 percent of sites met the 0.70 threshold for that measure, however, our predetermined standard for reliability.

We found similar results in the 2015 PQRS data. For every measure between 92 and 100 percent of groups had reliability of 0.70 or above. We conclude that the data are sufficiently reliable to use for benchmarking purposes.

4.4 Secondary Characteristics

Table 6 summarizes the secondary characteristics of CPC and PQRS GPRO QRDA III data in 2015. Reporting period is not listed because we are limiting our attention to 2015 data, which matches the year for CPC data. All provider types are included who meet the respective reporting requirements of CPC and PQRS; the data are not limited to physicians, or (in PQRS) to primary care providers or providers working in primary care practices.

Table 6: Secondary Characteristics of 2015 Data Sources

Data Source	Geography	Reporting Level[a]	Primary Care Specialty	Measure Choice	Payers Included
CPC	7 regions	Practice site[b]	Primary Care	9 measures of 13[c]	All[d]
PQRS GPRO QRDA III	National	group (TIN)	Primary care, specialty care	9 measures over 3 domains[e]	All[d]

Abbreviations: CPC: Comprehensive Primary Care Initiative; GPRO: Group Practice Reporting Option; PQRS: Physician Quality Reporting System; TIN: Tax Identification Number

[a] The CPC practice site is defined differently from the PQRS GPRO physician practice, which may cover multiple physical locations.

[b] *Practice site* refers to all patients seen at the physical location during the year who qualify for the initial population inclusion criteria of the respective measure.

[c] Of the 64 available measures, CMS selected 13 that appeared to match to a primary care population and aligned with goals of the model. Sites were required to report 9 of the 13 measures in 2015.

[d] All payer types among patients meeting the initial population inclusion criteria of the respective measure.

[e] If fewer than nine eCQMs across three domains applied to the group, or if the group's reporting method did not support reporting nine measures across three domains, then the group had to report the measures for which there was data. At least one measure had to include Medicare data.

Geography: As noted earlier, nationwide over 50 percent of eligible professionals participated in PQRS. We could not determine the geographic distribution of practices reporting through GPRO QRDA III data, but undoubtedly like CPC it represents provider groups across many states.

Reporting Level: CPC data are reported at the practice level. By definition a CPC practice is limited to a single physical location. GPRO data are reported by group, defined as two or more EPs who reassign their billing rights to the TIN. The definitions of CPC practice and GPRO practice overlap but are not identical; for example, a GPRO practice may have multiple physical locations.

Primary Care Specialty: CPC sites are predominantly primary care practices, whereas eligible providers under PQRS span all medical specialties. Practices that specialize in primary care could differ in quality from those for which primary care is relatively uncommon. Nevertheless, CMS wants all provider specialties to be held accountable to the same standard. For this reason, we used data from all provider and practice specialties when creating the benchmarks.

Measure Choice: CPC practices were required to report a minimum of 9 measures out of 13 in 2015. As shown above in Table 5, six measures were reported by 400 or more practice sites, four by 300 to 399 sites, and three by fewer than 300 sites. On average 351 sites reported a measure.

PQRS reporting through GPRO likewise required nine measures in 2014 and 2015, spanning three NQS domains, or fewer if the group did not have data for nine. The groups could choose

any eCQMs that fulfilled those requirements, and so the range of available measures was much broader than for CPC. The overall number reporting the CPC eCQMs was similar, however. As reported in Table 5 above, once zero measure reports were excluded there were eight measures reported by 400 or more groups and five by fewer than 300 groups. On average 395 sites reported a measure, 12 percent more than for CPC.

Payers Included: CPC explicitly includes all patients in its reporting regardless of insurance status. Data from PQRS reported through GPRO likewise represents all payers of patients who qualified for the measure. We cannot be sure that every reporting group or practice provided care to people with all types of insurance, including those without insurance; rather, we can have confidence that all insurance types, including uninsured status, are represented in the aggregate.

4.5 Benchmark Values

Having concluded that PQRS GPRO QRDA III data had sufficient quality and reliability, we developed the 25th, 50th, and 75th percentiles of the distribution of each. We used percentile definition 1 in the PROC UNIVARIATE command in SAS® to develop the figures.[18]

Table 7 presents the elements of the percentile formula in the left column. The right column works through an example of calculating the 25th percentile for a measure when there are 425 valid observations and particular values for the 106th and 107th ordered scores.

Table 7: Percentile Formula and Example

Percentile Formula Element	Example: Finding the 25th Percentile
n is the number of valid values for a measure	425 observations
$x_1, x_2, \ldots, x_n$ represent the ordered values of the variable, from lowest to highest	106th value is a score of 50.0% 107th value is a score of 51.0%
y *is* the t^{th} percentile	For the 25th percentile, $y = 25$
$p = t/100$	$p = 25/100 = 0.25$
$n*p = j + g$ where j is the integer part of $n*p$, and g is the fractional part of $n*p$	$n*p = 425*.25$ $= 106.25$ $j = 106$ $g = 0.25$
percentile $y = (1-g)*x_j + (g*x_{j+1})$ where x_0 is taken to be x_1	$y = [(1-0.25)*50.0\%] + (0.25*51.0\%)]$ $= 50.25\%$

[18] See http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_univariate_sect028.htm.

There are alternative formulas for the percentile, as noted in the SAS documentation (footnote 18). We selected the default approach in SAS because it represents a commonly used formula. An alternative formula could have caused some sites to move from below to above a threshold (25^{th}, 50^{th}, or 75^{th}) for any particular measure, or conversely to move from above to below it.

We selected an unweighted percentile formula for two reasons. First, the PQRS GPRO QRDA III data are at the group level, and a PQRS practice group approximates a CPC practice better than would an individual Eligible Provider. Thus, there was no need to weight the data to represent the appropriate level. Second, CMS wished to give equal weight to all GPRO groups, regardless of size, as it did to all CPC practices.

Table 8 presents the benchmark values. There was considerable variation across measures in the values of each percentile threshold. The 25^{th} percentile values ranged from 0.0 to 78.3 percent, the 50^{th} percentile values from 1.0 to 90.9 percent, and the 75^{th} percentile values from 11.0 to 100.0 percent. These broad and overlapping regions imply that there is substantial variability in the quality of care provided by groups who report to PQRS through GPRO QRDA III.

If the interquartile range of each measure—the distance between the 25^{th} and 75^{th} percentiles—is relatively small, it implies that most providers have similar rates. We found that the interquartile range varied from a low of 11.0 percent (Screening for Clinical Depression and Follow-Up Plan) to a high of 70.6 percent (Screening for Future Fall Risk). Higher values could reflect wide variation in practice, but they could also point to problems with data capture and reporting.

Table 8: 2015 Benchmarks by Measure Based on PQRS GPRO QRDA III Data

CMS Measure Number & Version	NQF Measure Number	Measure Name	Benchmark Values by Percentile		
			25th	50th	75^{th}
165v3	0018	Controlling High Blood Pressure	54.24	62.61	69.48
138v3	0028	Tobacco Use; Screening, and Cessation Intervention	70.10	85.22	92.58
125v3	0031[a]	Breast Cancer Screening	3.42	32.82	56.70
130v3	0034	Colorectal Cancer Screening	2.19	29.61	57.42
147v4	0041	Preventive Care and Screening: Influenza Immunization	5.78	25.92	43.14
163v3	0064[a]	Diabetes: Low Density Lipoprotein (LDL) Management	23.93	36.51	46.86

CMS Measure Number & Version	NQF Measure Number	Measure Name	Benchmark Values by Percentile		
			25th	50th	75th
182v4	0075	Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control (perf. rate 1)	35.71	54.52	68.26
144v3	0083	Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	78.26	90.83	100.00
139v3	0101	Falls: Screening for Future Fall Risk	1.18	28.57	71.79
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	0.00	1.00	11.00
68v4	0419	Documentation of Current Medications in the Medical Record	72.96	90.88	97.00

Abbreviations: CMS, Centers for Medicare & Medicaid Services; NQF, National Quality Forum

[a] No longer endorsed by NQF

4.6 Performance of CPC Sites against Benchmarks

We calculated the proportion of CPC practice sites that fell into each of the four performance zones created by the three benchmarks. Results appear in Table 9.

CPC practice sites performed better than GPRO groups overall. The benchmark median (50th percentile) represents the score relative to which half of GPRO groups scored higher and half scored lower. For 10 of 11 measures, fewer than 50 percent of CPC sites scored below the median GPRO score. For 9 measures, fewer than 40 percent fell below the benchmark median. At the same time, for 8 measures at least 40 percent of CPC sites scored in the top 25 percent of GPRO values.

For most measures the number of CPC sites generally rose from the lowest to the highest benchmark quartile. The smoking cessation measure (NQF 0028) and the heart failure measure (NQF 0083) were exceptions: both had a pronounced U-shaped distribution, with the largest proportions of CPC sites falling in the lower quartile and the highest quartile. Nevertheless, the results did not suggest a need to adopt alternative benchmarks.

Table 9: Performance of CPC Practice Sites Relative to Benchmarks in 2015, by Measure

CMS Measure Number & Version	NQF Measure Number	Measure Name	0th - 25th Percentile of Benchmark		25th - 50th Percentile of Benchmark		50th - 75th Percentile of Benchmark		75th - 100th Percentile of Benchmark	
			Number of CPC Sites	Percent of CPC Sites	Number of CPC Sites	Percent of CPC Sites	Number of CPC Sites	Percent of CPC Sites	Number of CPC Sites	Percent of CPC Sites
165v3	0018	Controlling High Blood Pressure	47	11	84	19	111	25	198	45
138v3	0028	Tobacco Use; Screening, and Cessation Intervention	140	34	47	12	73	18	147	36
125v3	0031[a]	Breast Cancer Screening	14	3	74	18	141	34	182	44
130v3	0034	Colorectal Cancer Screening	4	1	85	20	173	40	166	39
147v4	0041	Preventive Care and Screening: Influenza Immunization	5	1	64	17	142	38	161	43
163v3	0064[a]	Diabetes: Low Density Lipoprotein (LDL) Management	30	7	43	11	119	29	216	53
182v4	0075 [a,b]	Ischemic Vascular Disease (IVD): Complete Lipid Panel and LDL Control (perf. rate 1)	32	9	47	13	109	30	173	48
144v3	0083	Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	49	43	9	8	9	8	48	42
139v3	0101	Falls: Screening for Future Fall Risk	40	13	83	26	128	41	65	21
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	0	0	53	26	42	20	112	54
68v4	0419	Documentation of Current Medications in the Medical Record	44	14	77	25	64	21	125	40

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care Initiative; NQF, National Quality Forum

[a] No longer endorsed by NQF

[b] First numerator

Section 5. Conclusion and Limitations

We developed and applied a method to develop benchmarks for the eCQMs used in the CPC Initiative in 2015. Our search for the best available benchmarking data source was guided by primary and secondary criteria pertaining to similarity to CPC data, quality, and statistical reliability. We also analyzed the quality and reliability of 2015 CPC data.

PQRS GPRO QRDA III data had the best characteristics for use as a benchmark. We determined that they and the 2015 CPC data both have a high level of quality. Rates of missing numerators and denominators were low. The exception and exclusion rates were relatively high for a few measures, which indicates meaningful variability in the clinical status of patients across practice sites and groups.

Statistical reliability for these 13 measures was very high. For every measure, more than 80 percent of GPRO groups and CPC practices in our data had reliability of 0.7 or greater. For most groups and measures the reliability was 0.9 or greater. On this basis we concluded that the PQRS GPRO QRDA III data are sufficient for forming 2015 benchmarks, and that the performance rates for CPC sites are also reliably measured.

Correct classification of CPC sites would be hampered by floor or ceiling effects, indicated by threshold values of 0 or 100 percent. In practice, however, we found that only a single measure had a threshold of 0 percent, the 25th percentile for NQF 0418 (Screening for Clinical Depression and Follow-Up Plan). Likewise, only one measure had a threshold of 100 percent, the 75th percentile for NQF 0083 (Heart Failure: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction). We conclude that floor and ceiling effects were not important in these 2015 data.

CPC sites generally performed better on these eCQMs than did the GPRO groups who reported to PQRS through QRDA III. Relatively few scored in the lowest quartile, and for 8 of 11 measures at least 40 percent scored in the top quartile.

We note several potential limitations to our analytic approach. Although data submitted eletronically to PQRS represent the best available source for benchmarks, they fall short of being ideal in every respect. For example, the distribution of performance scores in PQRS may not be representative of all physicians in the United States in terms of performance or geographic distribution. We could not judge the accuracy of data entered into individual patient records. Although we have no reason to suspect widespread or consistent errors, we recognize that inaccurate data in individual records could lead to incorrect reporting of performance rates of related eCQMs. That being said, this type of limitation is not unique to eCQM data; data could be incorrectly recorded into patient records for other methods of quality reporting.

The distinction between individual and group reporting could matter to benchmark development because there should be less variation in average performance rates among groups than among the individuals who constitute those groups. If EPs with outlier performance values (such as 0 percent or 100 percent on a measure) report in a group with EPs who are not outliers, then the average performance of the group will be closer to the median than will the performance of the

EP with outlier values. Group reporting should thereby affect the calculated values of the 25th, 50th, and 75th percentiles, although by what amount cannot be known.

We expect that, given a choice, GPRO groups and CPC practices would choose to report data on measures on which they perform best. If so, the observed distribution of scores will be higher than the actual distribution among all providers. As noted earlier, groups reporting electronically to PQRS had considerable choice among measures: the minimum was just 9 of 64 eCQMs covering at least three NQS domains. However, because most EHRs do not support the full set of 64 eCQMs, there could be substantially less choice in practice. As a result, we do not expect meaningful bias to arise.

Appendix A. Additional Data Sources and Analyses

This appendix describes additional data sources that we considered for use in benchmarking and specific analyses performed on data from the EHR Incentive Program (Meaningful Use).

Data Sources

Meaningful Use Data

Meaningful Use provides financial incentives to providers and hospitals for Meaningful Use of "interoperable health information technology (HIT) and qualified electronic health records."[19] The reporting unit for providers is the individual EP, defined as the categories of doctors listed under PQRS above. Practitioners and therapists do not participate separately in Meaningful Use.

In 2014 and 2015, there were several approaches to Meaningful Use reporting: (1) registration and attestation, (2) reporting through the PQRS web portal via direct submission by CEHRT, and (3) CEHRT submission through a data submission vendor (DSV CEHRT). An EP who submits data by attestation manually enters the numerator, denominator, exclusions, exceptions, and performance rate for each measure on the basis of results obtained from his or her certified EHR. Groups of 25 or more EPs had the additional option of submitting their Meaningful Use data to PQRS through the GPRO web interface, which we did not consider for benchmarking because the web interface does not use electronic specifications.

Data submitted by either CEHRT method reflect a 12-month reporting period, but data submitted by registration and attestation can represent a 90-day or a 12-month reporting period. We were interested only in 12-month data for several reasons. First, CPC data also have a 12-month reporting period, so it makes sense to use the same reporting period in the benchmark data. Second, the measure specifications assume a 12-month period; if reported over 90 days, the data may miss follow-up care necessary to satisfy the measure. Third, some measures, such as influenza vaccination, have pronounced seasonality. A 90-day reporting period that missed the most relevant period would misrepresent performance.

Between 40 percent and 90 percent of EPs by state attested to Meaningful Use in the 2014 reporting year.[20] Among EPs submitting data by attestation in 2014, 72 percent attested to Stage 1 and 28 percent attested to Stage 2.[21] These figures indicate that Meaningful Use data should be capable of producing nationally representative performance rates.

---

[19] Centers for Medicare & Medicaid Services. Electronic Health Records (EHR) Incentive Program. [CMS website]. Last modified February 26, 2016. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html.

[20] Office of the National Coordinator for Health Information Technology. Office-Based Health Care Professional Participation in the CMS EHR Incentive Programs. December 2015. http://dashboard.healthit.gov/quickstats/pages/FIG-Health-Care-Professionals-EHR-Incentive-Programs.php.

[21] Authors' tabulations.

Electronic Health Record Data

A few private health systems, such as Kaiser Permanente and Geisinger, have amassed large amounts of EHR data on their patients. The Veterans Health Administration (VHA) and the Department of Defense health care system have extensive EHR systems, but they typically do not export data for external research. Moreover, many patients in VHA also are seen by practitioners outside of the system, making it difficult to know whether particular measures have been met. There currently are no publicly available, all-payer EHR databases. The State of North Carolina is beginning to create a multipayer EHR database, but a state official reported in late 2015 that they were not ready to share the data with researchers. We therefore conclude that large-scale EHR databases are not feasible for use in developing eCQM benchmarks at this time.

CMS Claims Databases

CMS maintains databases of Medicare and Medicaid claims. Although the databases are very large and have good quality, we deemed them inappropriate as CPC benchmark data sources for several reasons. As noted earlier, most of the 13 CPC measures require data elements not found in claims. As a result, the claims specifications of those measures are different enough from the electronic specifications to cause difficulties in comparing performance on an electronically reported measure with a claims-based benchmark. Moreover, CPC data include all payer types and patients with the entire range of demographics. Medicare and Medicaid claims each represent a single payer type and cover specific subgroups: older adults and individuals who are chronically disabled for Medicare, and individuals with low income for Medicaid. This limitation does not affect data submitted electronically to PQRS because those data reflect individuals with all types of insurance, including those with no insurance.

Quality of 2014 Meaningful Use Data

Before 2015 data were available, we assessed the quality of 2014 Meaningful Use data using the approach applied to CPC and GPRO QRDA III GPRO data. In addition, we studied the number of TIN and NPI combinations, as follows:

- The number of unique TIN and NPI combinations
- The number of TINs that have multiple NPIs
- The number of NPIs that have multiple TINs

The 2014 Meaningful Use data file contains provider information, business ZIP Codes, and other information relevant to electronic health record (EHR) software used for submission. The file has one row for each provider.

The columns represent the following: National Provider Identification (NPI), Provider Tax Identification Number (TIN), Provider Name, Business State/Territory, Business ZIP Code, EHR Certification Number, EHR Reporting Period, Clinical Quality Measure (CQM) eReporting Election, and Certified EHR Technology (CEHRT) Edition Year Number.

The NPI acts as a foreign key in this data set. Each NPI measure is arranged in rows. For example, if Provider 123 has submitted 10 measures, there should be 10 rows for Provider 123 in the data set. This data file has 28,898 rows.

The columns of this dataset represent the following: NPI, Program Year, Payment Year Number, Provider Stage Number Category, Category Compliance Status, Objective Identifier/Title, Numerator Value, Denominator Value, Performance Rate Number, Exclusion Attestation Indicator, Exclusion Description, Exclusion Value, and Exemption Indicator.

Results

There were 2,478 unique providers in the 2014 Meaningful Use data set. Every NPI was associated with a unique TIN, a surprising finding. The providers submitted a total of 25,308 measure records with valid data. The CPC eCQMs are all proportions; to be valid, the numerator and denominator had to be non-missing, and both could not be zero. The numerator never exceeded the associated denominator. The median exclusion rates were 0.0 for the same nine measures as CPC data, 0.07 for measure NQF 0018, and 0.17 for measure NQF 0034. The latter two rates were slightly lower than those for CPC data.

Exclusion rates were often greater in Meaningful Use data than in CPC data. For 8 of 13 measures, the percentage of Meaningful Use providers who excluded at least five percent of cases was at least 50 percent greater than for CPC practices. We cannot tell whether this reflects different interpretation of, or accuracy in applying, the measure definitions, or instead whether it is attributable to differences in the clinical status of patients.

We did not calculate exception rates for Meaningful Use data.

Secondary Characteristics of Meaningful Use Data

Table 10 summarizes the secondary characteristics of Meaningful Use data in 2014 and 2015. The reporting period is clear by construction and so it omitted from the table. All reporting methods for Meaningful Use had similar characteristics.

Geography: At least 40 percent of potentially eligible providers participated in Meaningful Use by state. This is a sufficiently large sample to enable creation of credible national benchmarks.

Reporting Level: Meaningful Use data are reported at the level of an eligible provider.

Primary Care Specialty: Eligible providers under Meaningful Use span all medical specialties.

Measure Choice: The number of measures varied, depending on the data submission method. If fewer than nine measures across three domains applied to the EP, or if the EP's reporting method did not support reporting nine measures across three domains, then the EP could report fewer than nine measures, report measures across fewer than three domains, or both.

Payers Included: Meaningful Use data are based on all payer types among patients meeting a measure's denominator.

Table 10: Secondary Characteristics of 2014 and 2015 Meaningful Use Data

Data Source	Geography	Reporting Level	Primary Care Specialty	Measure Choice	Payers Included
Meaningful Use: EHR attestation	National	EP	Primary care, specialty care	9 measures over 3 domains	All
Meaningful Use: e-reporting by PQRS portal	National	EP	Primary care, specialty care	9 measures over 3 domains	All
Meaningful Use: direct CEHRT	National	EP	Primary care, specialty care	9 measures over 3 domains	All
Meaningful Use: DSV CEHRT	National	EP	Primary care, specialty care	9 measures over 3 domains	All

Abbreviations: CEHRT: certified electronic health record technology; DSV, data submission vendor; EHR, electronic health record; EP, eligible professional

Statistical Reliability of CPC Measures in 2014 Meaningful Use Data

We assessed the statistical reliability of the eleven 2014 CPC measures in 2014 Meaningful Use data. Table 11 presents the results. There were 8,927 Meaningful Use records submitted for the 11 CPC measures by attestation, of which 8,080 records (90.5 percent) had valid measure data. Between 94 percent and 100 percent of sites achieved reliability of 0.70 or greater on every measure, and average reliability ranged from 93 percent to 99 percent across measures.

Table 11: Reliability by Measure for 2014 Meaningful Use Attestation Data

CMS Measure Number	NQF Measure Number	No. of Meaning-ful Use Providers	Avg	Std Dev	Sites with Reliability ≥0.70, %	Min	Percentile			Max
							25th	50th	75th	
165v2	0018	1,517	0.94	0.11	95	0.32	0.95	0.99	0.99	1
138v2	0028	2,165	0.98	0.05	99	0.44	0.99	1	1	1
125v2	0031	679	0.98	0.04	100	0.55	0.98	0.99	1	1
130v2	0034	552	0.99	0.02	100	0.77	0.99	1	1	1
147v2	0041	728	0.99	0.05	99	0.43	0.99	1	1	1
163v2	0064	357	0.97	0.07	98	0.57	0.97	0.99	1	1
182v3	0075	294	0.97	0.06	99	0.58	0.97	0.99	1	1
144v2	0083	35	0.93	0.13	94	0.35	0.87	1	1	1
139v2	0101	577	0.99	0.03	100	0.48	1	1	1	1
2v3	0418	313	0.99	0.02	100	0.83	1	1	1	1

Abbreviations: Avg, average; CMS, Centers for Medicare & Medicaid Services; , Max, maximum, Min, minimum, NQF, National Quality Forum; Std Dev, standard deviation

Appendix B. Data Quality Analyses

This appendix describes in more detail the data quality analyses performed on Comprehensive Primary Care (CPC) Initiative and PQRS GPRO QRDA III databases.

Methodology

We assessed the appropriateness of PQRS GPRO data for benchmarking through analyses of program requirements and data completeness. We performed similar analyses of CPC data for comparison.

Program Requirements Analysis

This analysis assessed the data sets with respect to compliance with program requirements. We performed the analyses listed here and described below:

- Number of providers reporting
- TIN and NPI combinations
- Incomplete data
- Numerator greater than denominator
- Fewer than nine measures
- Providers reporting by measures

Number of Providers Reporting

This is a simple count of the number of practice sites or groups in each data set.

Data Completeness

We evaluated data completeness with respect to the following parameters:

- Denominator incomplete: proportion of provider measures with numerator available but denominator missing (Note: 10 measures for 1 provider = 10 provider measures)
- Numerator incomplete: proportion of provider measures with denominator available but numerator missing
- Dual incomplete: proportion of provider measures with performance rate available but both numerator and denominator missing
- Total incomplete: proportion of provider measures with denominator, numerator, and dual incompletes.

If all values for a measure—numerator, denominator, exceptions, exclusions, and performance rate—were missing or zero, we did not interpret it as an error. We could not know how many measures each CPC practice site or GPRO practice group intended to submit, except that all CPC sites were required to submit at least nine. (Although PQRS also required submitting at least nine measures, they could have included eCQMs that were not among those used in the CPC Initiative. There was no effective minimum number of CPC eCQMs expected in GPRO data as a result.) We assume that when all values are missing or zero, the site did not intend to report the measure.

Numerator Greater Than Denominator

This is the proportion of provider measures for which the numerator exceeds the denominator.

Fewer Than Nine Measures

CPC provider sites should report at least 9 of the 13 eCQMs. For each data set, this is the number and proportion of providers reporting fewer than nine measures.

Providers Reporting by Measures

This is the number and proportion of providers reporting for each of the 13 eCQMs, and the average number and percentage of providers reporting for the entire eCQM measure set.

Performance-Based Analysis

This analysis assessed the data sets with respect to three factors:

- Exclusions
- Exceptions
- Performance

Exclusions and Exceptions

For each data set, for measures defined with an exclusion or an exception, we assessed the rates of exclusion and exception with respect to two parameters:

- Rate: for each measure and in the total across all measures, the average rate of exclusion or exception (where the rate is the number of exclusions/exceptions divided by the denominator)
- Proportion of high outliers: for each measure and in the total across all measures, the proportion of providers with an exclusion/exception rate exceeding 5 percent

We did not analyze the proportion with no exclusions or exceptions, as it would not be surprising or problematic to find a practice that had none.

Performance

For each data set, performance rates were assessed with respect to two parameters:

- Performance rate: for each measure and the total across all measures, the average performance rate (the numerator divided by the denominator)
- Performance high outliers: for each measure and in the total across all measures, the proportion of providers with a performance rate exceeding 95 percent

We examined high outliers because very high performance rates, if accurate, imply that there is little room for clinical improvement. If many sites are scoring very high then it would give CMS justification in retiring the measure from use in CPC.

Appendix C. Measure Exclusions and Exceptions

Tables 12 and 13 present the exclusions and exceptions, respectively, in the eCQMs used in CPC. They reflect the measure specifications that applied during the 2015 performance year.

Table 12: Measure Exclusions

CMS Measure Number & Version	NQF Measure Number	Measure Name	Exclusions
165v3	0018	Controlling High Blood Pressure	Patients with evidence of end stage renal disease (ESRD), dialysis or renal transplant before or during the measurement period; patients with a diagnosis of pregnancy during the measurement period.
125v3	0031	Breast Cancer Screening	Women who had a bilateral mastectomy or for whom there is evidence of two unilateral mastectomies
130v3	0034	Colorectal Cancer Screening	Patients with a diagnosis or past history of total colectomy or colorectal cancer
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	Patients with an active diagnosis for depression or a diagnosis of bipolar disorder

Abbreviations: CMS, Centers for Medicare & Medicaid Services; NQF, National Quality Forum

Table 13: Measure Exceptions

CMS Measure Number & Version	NQF Measure Number	Measure Name	Exceptions
138v3	0028	Tobacco Use Screening and Cessation Intervention	Documentation of medical reason(s) for not screening for tobacco use (e.g., limited life expectancy, other medical reason)
147v4	0041	Influenza Immunization	Documentation of medical reason(s) for not receiving influenza immunization (e.g., patient allergy, other medical reasons); documentation of patient reason(s) for not receiving influenza immunization (e.g., patient declined, other patient reasons); documentation of system reason(s) for not receiving influenza immunization (e.g., vaccine not available, other system reasons)
144v3	0083	Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	Documentation of medical reason(s) for not prescribing beta-blocker therapy (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons); documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons); documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the healthcare system)
139v3	0101	Falls: Screening for Future Fall Risk	Documentation of medical reason(s) for not completing a risk assessment for fall risk (i.e., patient is not ambulatory)
2v4	0418	Screening for Clinical Depression and Follow-Up Plan	Patient refuses to participate; patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status; situations where the patient's functional capacity or motivation to improve may impact the accuracy of results of standardized depression assessment tools (for example: certain court appointed cases or cases of delirium).
68v4	0419	Documentation of Current Medications in the Medical Record	Patient is in an urgent or emergent medical situation where time is of the essence and to delay treatment would jeopardize the patient's health status.

Abbreviations: CMS, Centers for Medicare & Medicaid Services; NQF, National Quality Forum

Appendix D. Statistical Reliability Tables for 2015 CPC and PQRS GPRO QRDA III Data

Table 14 presents the total count of records and the count of valid performance records. Records were deemed invalid if the numerator and denominator were both 0, the number of numerator records exceeded the number of denominator records, the performance rate exceeded 100 percent, or the number of denominator records minus the number of exceptions and exclusions was zero or less.

Table 14: Count of Records by Data Source

Provider Measure	No. of Records	No. of Records With Valid Performance Measure Data
2015 CPC Attested Data	4,055	4,019
2015 CPC QRDA III Data	587	587
2015 PQRS GPRO QRDA III Data (CPC Measures Only)	5,979	5,872

Abbreviations: CPC, Comprehensive Primary Care; GPRO, Group Practice Reporting Option; QRDA, Quality Reporting Document Architecture

Tables 15, 16, and 17 present reliability statistics by data source and measure.

Table 15: Reliability by Measure for 2015 CPC Attestation Data

CMS Measure Number & Version	NQF Measure Number	No. of CPC Sites	Sites With Reliability ≥0.70 (%)	Minimum Value	Percentile			Maximum Value
					25th	50th	75th	
165v3	0018	389	388 (99.7)	0.65	0.98	0.99	0.99	1.00
138v3	0028	367	367 (100.0)	0.96	1.00	1.00	1.00	1.00
125v3	0031	360	359 (99.7)	0.63	0.99	1.00	1.00	1.00
130v3	0034	378	378 (100.0)	0.94	1.00	1.00	1.00	1.00
147v4	0041	323	323 (100.0)	0.79	0.99	1.00	1.00	1.00
163v3	0064	358	358 (100.0)	0.70	0.97	0.98	0.99	1.00
182v4	0075	318	317 (99.7)	0.68	0.98	0.99	0.99	1.00
144v3	0083	115	99 (86.1)	0.44	0.85	0.97	1.00	1.00
139v3	0101	285	285 (100.0)	0.97	1.00	1.00	1.00	1.00
2v4	0418	175	175 (100.0)	0.99	1.00	1.00	1.00	1.00
68v4	0419	277	277 (100.0)	0.98	1.00	1.00	1.00	1.00

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care Initiative; Max, maximum; Min, minimum; NQF: National Quality Forum

Table 16: Reliability by Measure for 2015 CPC QRDA III Data

CMS Measure Number & Version	NQF Measure Number	No. of CPC Sites	Sites with Reliability ≥0.70 (%)	Min	Percentile			Max
					25th	50th	75th	
165v3	0018	52	52 (100.0)	0.87	0.95	0.97	0.98	0.99
138v3	0028	41	41 (100.0)	0.98	1.00	1.00	1.00	1.00
125v3	0031	52	52 (100.0)	0.97	0.99	1.00	1.00	1.00
130v3	0034	51	51 (98.0)	0.96	0.99	0.99	0.99	1.00
147v4	0041	51	50 (98.0)	0.96	0.99	0.99	0.99	1.00
163v3	0064	51	50 (100.0)	0.90	0.96	0.97	0.99	1.00
182v4	0075	44	44 (100.0)	0.87	0.94	0.96	0.98	1.00
144v3	0083	4	1 (25.0)	0.00	0.00	0.00	0.00	1.00
139v3	0101	33	32 (97.0)	0.98	0.99	1.00	1.00	1.00
2v4	0418	33	33 (100.0)	0.99	1.00	1.00	1.00	1.00
68v4	0419	36	34 (94.4)	1.00	1.00	1.00	1.00	1.00

Abbreviations: CMS, Centers for Medicare & Medicaid Services; CPC, Comprehensive Primary Care Initiative; Max, maximum; Min, minimum; NQF: National Quality Forum; QRDA: Quality Reporting Document Architecture

Table 17: Reliability by Measure for 2015 PQRS GPRO QRDA III Data

CMS Measure Number & Version	NQF Measure Number	No. of Groups	Groups With Reliability ≥0.70 (%)	Min	Percentile			Max
					25th	50th	75th	
165v3	0018	602	582 (96.7)	0.17	0.98	0.99	1.00	1.00
138v3	0028	783	783 (100.0)	0.79	1.00	1.00	1.00	1.00
125v3	0031	442	442 (100.0)	0.74	1.00	1.00	1.00	1.00
130v3	0034	476	475 (99.8)	0.42	1.00	1.00	1.00	1.00
147v4	0041	477	476 (99.8)	0.54	1.00	1.00	1.00	1.00
163v3	0064	273	272 (99.6)	0.57	0.99	1.00	1.00	1.00
182v4	0075	239	239 (100.0)	0.83	0.99	1.00	1.00	1.00
144v3	0083	84	78 (92.9)	0.22	0.92	0.98	1.00	1.00
139v3	0101	280	280 (100.0)	0.81	1.00	1.00	1.00	1.00
2v4	0418	223	223 (100.0)	0.94	1.00	1.00	1.00	1.00
68v4	0419	848	847 (99.9)	0.51	1.00	1.00	1.00	1.00

Abbreviations: CMS, Centers for Medicare & Medicaid Services; GPRO: Group Practice Reporting Option; NQF, National Quality Forum; PQRS: Physician Quality Reporting System; QRDA: Quality Reporting Document Architecture