

Methodology for 2008 Physician Performance Scoring

Measures and Measures Specifications

The CMS-Delmarva funded Better Quality Information (BQI) contract required that the CCHRI select performance measures that were:

- Relevant to the Medicare population,
- Endorsed by the Ambulatory Quality Alliance (AQA) or the National Quality Forum (NQF), and
- Could be calculated using the available administrative claims data.

CCHRI selected 18 measures shown in Table 1.0, which are derived from the original starter set of 26 measures endorsed by the AQA and the AMA Physician Consortium on Performance Improvement (PCPI).

Fifteen measures used the Healthcare Effectiveness Data and Information Set (HEDIS) specifications and three measures used the Medicare Care Management Performance (MCMP) Demonstration and Quality Measurement specifications. The HEDIS specifications are copyrighted by NCQA and permission to reproduce these is required by NCQA; consequently, we do not include the details of these specifications in this document. The MCMP measurement specifications were taken from “Medicare Care Management Performance Demonstration and Quality Measurement Specifications,” a report prepared by Jody Blatt of CMS (Dated May 7, 2007; Updated August 3, 2007).

Table 1. Final List of Measures Scored Using 2007 Claims Data

Measure
Screening for Cancer
Breast Cancer Screening (BCS)
Colorectal cancer screening (COL)
Diabetes
LDL testing (LDL)
HbA1c testing (HBA)
Diabetes eye exam screening (CDC_Eye Exam)
Diabetes: Nephropathy testing* (NPH)
Heart Disease
Cardiovascular – LDL testing (CMC)
CAD patients receiving lipid-lowering therapy (CAD)
Persistence of beta blocker therapy – Post MI (PBH)
Annual monitoring for patients on persistent medications: Ace inhibitors, Digoxin, Diuretics, Statins (MPM)
Post MI post discharge beta-blocker therapy* (BBH)
Heart Failure
HF: LVEF testing (HF2)
Warfarin for patients w/ CHF & atrial fibrillation (HF8)
Respiratory Disease
Spirometry testing for COPD patients to confirm dx. (SPR)
Pharmacologic Management of COPD Patients * (PCE)

Measure
Care for Older Adults
Percent of patients 65 or older that had glaucoma eye exam (GSO)
Women with osteoporosis age 67 and older who have had a fracture bone density/ medication (OMW)
Rheumatology/ Orthopedics
Rheumatoid arthritis patients prescribed a DMARD drug (ART)

*Three of the 18 measures selected for testing were dropped from reporting (BBH, PCE, NPH) because of concerns about the measures specifications or data completeness.

Performance scores for each measure are calculated as a ratio in which the denominator represents all patients who should have received a particular service (i.e., who meet all of the denominator qualification criteria) and the numerator represents the number of the denominator qualifying patients who received the service, based on the information found in the claims data.

Measures are first scored at the patient level (i.e., patients were categorized as either in or out of the numerator and denominator for each measure), and then measure events were attributed to physicians of the relevant specialty for each measure. Patient-level data (numerator and denominator events) were aggregated to generate physician-level scores, using aggregated commercial and Medicare data sources. Measures were computed using 2007 Measurement Year (MY) specifications, relying on claims and eligibility information from 2004, 2005, 2006 and 2007 to support measure calculations.

- **Denominator period:** the time period during which the member had to be enrolled and/or had a clinical condition/event to qualify for the measure.
- **Numerator period:** the time period during which the member was supposed to have received the numerator qualifying service (e.g., mammogram).
- **Attribution period:** the time period during which the E&M visit that is used to attribute a denominator qualifying member to a physician must have occurred (Table 2).

Table 2. Measure Denominator, Numerator and Attribution Periods for Cycle 3

	Measure Name	Denominator Period	Numerator Period[†]	Attribution Period
BCS	Breast Cancer Screening	2006 and 2007	1/1/06 – 12/31/07	2006 and 2007
COL	Colorectal Cancer Screening	2006 and 2007	1/1/04 – 12/31/07	2006 and 2007
HBA	Diabetes: HbA1c Test	2006 and 2007	1/1/07 – 12/31/07	2007
LDL	Diabetes: LDL Screening	2006 and 2007	1/1/07 – 12/31/07	2007
CMC	Cardiovascular: LDL Screening	2006 and 2007	1/1/07 – 12/31/07	2007
PBH	Persistence of Beta Blocker Treatment After a Heart Attack	7/1/06 – 6/30/07	180 day period after discharge date	180 day period after discharge date
MPM	Annual Monitoring for Patients on Persistent Medications	2007	1/1/07 – 12/31/07	2007
CDC_Eye Exam	Diabetes: Eye Exam	2006 and 2007	1/1/07 – 12/31/07	2007
GSO	Glaucoma Screening in	2006 and 2007	1/1/06 – 12/31/07	2006 and 2007

	Measure Name	Denominator Period	Numerator Period [†]	Attribution Period
	Older Adults			
SPR	Use of Spirometry Testing in the Assessment and Diagnosis of COPD	7/1/06 – 6/30/07 (for index start date)	2 years before to 180 days after Index Start Date	180 day period after Index Episode Start Date
ART	Disease Modifying Anti-Rheumatic Drug Therapy for Rheumatoid Arthritis	2007	1/1/07 – 12/31/07	2007
OMW	Osteoporosis Management in Women	7/1/06 – 06/30/07	180 day period after Index Episode Start date	180 day period after Index Episode Start date
CAD	Coronary Artery Disease: LDL Drug Therapy	2007	1/1/07 – 12/31/07	2007
HF2	Heart Failure: Left Ventricular Ejection Fraction Testing	2007	1/1/07 – 12/31/07	2007
HF8	Heart Failure: Warfarin Therapy for Patients with Atrial Fibrillation	2007	1/1/07 – 12/31/07	2007

Measures marked by a cross (†)All dates are inclusive. [EDIT NOTE: CAN'T HAVE A FOOTNOTE REFERRING TO TWO DIFFERENT ISSUES AS TOO CONFUSING, THE MCMP ISSUE ISN'T IMPORTANT HERE]

The 2008 HEDIS specifications were used for MY 2007 scoring with a few exceptions. Measures specification issues and modifications were as follows:

- For the MPM measure, the combined medication rate does not include the count of numerator positive events for anti-convulsants, as called for in the specifications. The CCHRI BQI PAG felt that the clinical evidence supporting annual monitoring of this drug category was relatively weak and removed the anticonvulsant category from the measure. The MPM rate was calculated by adding the numerators and denominators for the three remaining drug categories: digoxin, diuretics, and ACE/ARB.
- The numerator timeframe for COL spans a longer period of time (10 years) than that of the available claims and eligibility data. Physician reports show scores that reflect four years of data (2004 – 2007) that were used to construct this measure.
- The relevant specialties for the COL and NPH measures were modified to include Colorectal Surgery and Nephrology, respectively. This decision was made in consultation with the BQI PAG and Steering Committee.
- The Beta Blocker Treatment After a Heart Attack measure (BBH) was retired by NCQA in 2007: MY 2007 or 2008 specifications were not available. To score the measure, we used MY 2006 HEDIS specifications.
- The HBA and LDL HEDIS measure specifications excluded Medicare beneficiaries identified as having End Stage Renal Disease (ESRD) (through an ESRD flag in the Medicare data) from the measures' denominators. This exclusion was not possible with the commercial data, as it did not contain an ESRD indicator.

Only one modification was made to a MCMP measure; Endocrinology was added as a relevant specialty for the CAD measure, based on input from the California Physician Performance Initiative (CPPI) Physician Advisory Group.

Data used to Construct Performance Measures

For CPPI, CCHRI combined Medicare fee-for-service (FFS) claims data with PPO claims data from two commercial California health plans (Anthem Blue Cross of California and Blue Shield of California) to generate physician-level quality performance scores. Data files that were used to generate the measures included:

- Eligibility – member benefit enrollment information and member demographics
- Professional Claims – medical services (non-facility claims)
- Facility Claims – inpatient and outpatient facility services claims
- Pharmacy Claims – outpatient pharmaceutical claims
- Provider – physician and/or provider identifiers, demographics, and address.

Members had to meet continuous enrollment requirements within each data source to qualify for the measure. For example, if a member switched from a PPO plan to Medicare FFS during the continuous enrollment period, s/he would not meet continuous enrollment requirements. Members also had to meet any pharmacy coverage and anchor date requirements. The process then identified any contraindications/exclusions and numerator services for all denominator-qualifying members. For example, the Medicare eligibility file was pre-limited prior to the application of the measure programming in the following ways:

- Only Medicare beneficiaries with FFS coverage were included.
- Only Medicare beneficiaries with both Part A and Part B coverage for at least 11 of 12 months in the calendar year, including the last (anchor) month of the calendar year, were included. Medicare eligibility data included monthly enrollment flags (rather than the start and end date of enrollment segments). Consistent with the HEDIS specifications for Medicaid products (where monthly enrollment flags are commonly used), members were required to be continuously enrolled for at least 11 of 12 months in each calendar year to qualify for inclusion in any measure as required by measure specifications.
- The Medicare pharmacy benefit flag was created by Thomson Reuters by using the Part D enrollment file to determine the pool of FFS Medicare members that met measure-specific continuous enrollment and pharmacy coverage requirements.

The continuous enrollment requirement means that some patients who a physician is currently seeing will not be included in their measure rate calculation because they fail the continuous enrollment criteria. For example, the continuous enrollment requirement for Breast Cancer Screening is two years (2006 and 2007 for Cycle 3). If a member joined one of the participating plans/Medicare in June 2007, they would not be included in the measure or in a physician's score even if they saw the physician (after June 2007) and had the required mammogram.

Attributing Patient Denominator Events to Physicians

The attribution rule assigns a patient event to any physician, of a relevant specialty for a given measure, who had one or more Evaluation & Management (E&M) visits during the measurement period. The relevant specialists for each measure are shown in Table 3.

This attribution rule means that a patient denominator event can be assigned to one or more physicians. For example, primary care physicians and OB/GYN's are relevant specialists for the mammography measure. A woman who is eligible for this screening event would be assigned to any and all primary care and OB/GYN's who saw this patient for an E&M visit during the measurement period.

As shown in Table 2 above, the numerator and attribution periods are aligned so that a physician must have seen the member for an E&M visit during the time period in which they were to have received the numerator service.

Many physicians have multiple specialty designations across the commercial health plans and Medicare. A single, primary specialty was assigned to each physician and this specialty was used in the patient attribution step.

CORRECT TABLE 3 TO INCLUDE ALL OF THE MEASURE NAMES IN RIGHT ROWS

Table 3. Measure Descriptions and Designated Relevant Physician Specialties

Measure	Measure Name	Population and Description	Designed Relevant Specialties for the Measure
BCS	Breast Cancer Screening	Women, age 42-69, who had mammogram in 2006 or 2007.	Family practice, internal medicine, and OB/GYN
COL	Colorectal Cancer Screening	Patients, age 51-80, who had a FOBT in 2007, sigmoidoscopy during 2004-2007, DCBE during 2004-2007, or colonoscopy during 2004-2007.	Family practice, internal medicine, gastroenterology, and colorectal surgery
HBA	Diabetes: HbA1c Test	Diabetics, age 18-75, who had an HbA1c screening test during 2007.	Family practice, internal medicine, and endocrinology
LDL	Diabetes: LDL Screening	Diabetics, age 18-75, who had an LDL-C screening test during 2007.	Family practice, internal medicine, endocrinology, and cardiology
NPH	Cardiovascular: LDL Screening	Diabetics, age 18-75, who had a nephropathy screening test or evidence of nephropathy in 2007.	Family practice, internal medicine, endocrinology, and nephrology
CMC	Persistence of Beta Blocker Treatment After a Heart Attack	Patients, age 18-75, who were hospitalized in 2006 for an AMI, CABG, or PTCA, <i>or</i> were diagnosed with IVD in 2006 or 2007, and who had an LDL test in 2007.	Family practice, internal medicine, and cardiology
BBH	Annual Monitoring for Patients on Persistent Medications	Patients aged 35+, who were hospitalized in 2007 for an AMI and received beta-blocker therapy during hospitalization or within 7 days after discharge.	Family practice, internal medicine, and cardiology
PBH	Diabetes: Eye Exam	Patients, age 35+, who were hospitalized in 2007 for an AMI and received beta-blocker therapy for the 6 months after discharge.	Family practice, internal medicine, and cardiology
CAD	Glaucoma Screening in Older Adults	Coronary artery disease patients, age 18+, who were prescribed a lipid-lowering therapy.	Family practice, internal medicine, cardiology, and endocrinology
HF2	Use of Spirometry Testing in the	Heart failure patients aged 18+, who	Family practice, internal

Measure	Measure Name	Population and Description	Designed Relevant Specialties for the Measure
	Assessment and Diagnosis of COPD	were hospitalized during 2006 and had a LVF test.	medicine, and cardiology
HF8	Disease Modifying Anti-Rheumatic Drug Therapy for Rheumatoid Arthritis	Heart failure patients, age 18+ who were hospitalized with paroxysmal or chronic atrial fibrillation during 2006 and were prescribed warfarin therapy.	Family practice, internal medicine, and cardiology
MPM	Osteoporosis Management in Women	Combined rate for patients aged 18+ who were prescribed at least a 180-days supply of ambulatory medication therapy for one of the following medications: (1) ACE inhibitors or ARBs; (2) Digoxin; or (3) Diuretics.	Family practice, internal medicine, and cardiology
CDC Eye	Coronary Artery Disease: LDL Drug Therapy	Diabetics, age 18-75, who had a retinal or dilated eye exam in 2007..	Family practice, internal medicine, endocrinology, ophthalmology, and optometry
COPD: Pharmacotherapy Bronchodilators	Heart Failure: Left Ventricular Ejection Fraction Testing	Patients age 40+, with a COPD exacerbation, with an inpatient discharge or ED encounter between 1/1/2007 – 11/30/2007, who received a bronchodilator within 30 days.	Family practice, internal medicine, allergy and immunology, and pulmonology
COPD: Pharmacotherapy Systemic Corticosteroids	Heart Failure: Warfarin Therapy for Patients with Atrial Fibrillation	Patients age 40+, with a COPD exacerbation, with an acute inpatient discharge or ED encounter between 1/1/2007 – 11/30/2007, who received a systemic corticosteroid within 14 days of the event.	Family practice, internal medicine, allergy and immunology, and pulmonology
COPD: Spirometry Testing		Patients age 42+, with a new or newly active COPD diagnosis between 7/1/2006 – 6/30/2007, who received spirometry testing two years prior to diagnosis or within 6 months of diagnosis.	Family practice, internal medicine, allergy and immunology, and pulmonology
GSO		Patients, age 67+ without history of glaucoma, who received a glaucoma screening in 2006 or 2007.	Family practice, internal medicine, geriatric medicine, ophthalmology, and optometry
ART		Patients, age 18+, diagnosed with rheumatoid arthritis who received at least one ambulatory prescription for a disease modifying anti-rheumatic drug during 2007.	Family practice, internal medicine, orthopedics, rheumatology
OMW		Women, age 67+ with a fracture occurring between 7/1/2006 – 6/30/2007, who received a bone mineral density (BMD) test or prescription to treat/prevent osteoporosis within six months of the injury.	Family practice, internal medicine, OB/GYN, orthopedics, geriatric medicine, rheumatology, endocrinology, and nephrology

*As indicated in the NCQA HEDIS 2008 Technical Specifications for Physician Measurement.

E&M visits were defined as professional encounters/claims with CPT codes identifying E&M services. The E&M visit does not have to have a diagnosis code related to the measure for which the member qualifies. Only office-based visits, hospital outpatient visits, and independent clinic visits were counted. These visits were identified by the Place of Service code on the claim/encounter. The following Place of Service codes were used:

- 11 (Office Visit), 22 (Outpatient Hospital), and 49 (Independent Clinic).

Visits with other Places of Service were excluded: Hospital Inpatient, Emergency Room, and Ambulatory Surgery Centers were among the Places of Service exclusions.

Eligible E&M visits were defined per the following list of codes: CPT Code = 59400, 59410, 59425-59426, 59430, 59510, 59515, 59610, 59614, 59618, 59622, 90750-90764, 90801-90899, 92002-92014, 95105, 99201-99499

or

- HCPCS Code = G0380-G0384.

This attribution rule was chosen based on the results from testing three attribution methods and consultation with the CCHRI BQI Steering Committee) and Physician Advisory Group. The three methods tested were: 1) Method #1, which assigned events to any physician with whom the patient had one or more E&M visits during the measurement period (includes measure relevant and non-relevant specialties); 2) Method #2, which assigned patient events to any measure-relevant physician with whom the patient had one or more E&M visits during the measurement period; and 3) Method #3, which assigned patient events to any measure-relevant physician that represented the majority of the patient's E&M visits during the measurement period. As discussed in more detail in the validation section of this document, the three alternative methods were assessed by examining:

- Number of Denominator Patients Attributed
- Number of Physicians Attributed
- Physician Performance Measure Rates
- Ability to Score Physicians Reliably

Method #3 constitutes a narrower set of accountabilities while Method #1 establishes a broader set of accountabilities for patient care. While narrow accountabilities may be more consistent with physician perceptions of the care they provide to patients—which was confirmed in our physician testing of the various approaches—broad accountabilities (i.e., multi-provider) align with the concept of physicians having shared responsibility for ensuring the proper management of the patient.

There were cases when we could not attribute denominator qualifying patients to a physician; these events are excluded from the performance results. The key reasons that patient events could not be assigned were: 1) member had an E&M visit but not with a physician of a relevant specialty for the given measure, 2) member had an E&M visit but the rendering physician could not be identified (i.e., rendering provider field on claim was blank or did not identify an individual physician), or 3) member did not have an E&M visit during the measurement period. Almost half of the time a patient event was unassigned because the patient had a visit with a physician of a non-relevant specialty.

Constructing Physician Performance Scores

After patients are attributed to physicians, physician-level scores are calculated by aggregating attributed members across all data sources. This produces a set of physician-level scores. For each measure the numerator patient response is either 0 (did not have service) or 1 (had service) for each patient in the denominator. Each physician is attributed certain patients and the physician's performance is the aggregate of those numerator patient responses. Physician performance rates, or scores were calculated for each of the measures. Physician p 's true performance rate, π_k , is unknown. It was estimated from the data as:

$$r_p = \frac{1}{n_p} \sum_{i=1}^{n_p} X_{ip}$$

where physician p has a sample of n_p (denominator) patients and $X_i = 1$ if patient i received the service and $X_i = 0$ if patient i did not receive the service. That is, the performance rate was calculated as the numerator for the measure (i.e., the number of eligible members who had the service, such as breast cancer screening) divided by the denominator for the measure (i.e., the number of eligible members for the service).

Table 4 shows the statistical properties (mean, standard deviation, interquartile range) of the estimated physician-level performance rates assessed for each measure. These statistics reveal considerable variation in performance across physicians and also illustrate that performance rates are below optimal levels.

Table 4. Overall Physician Performance Rates, by Measure

Measure	Cycle 3		
	Mean	Standard Deviation	Interquartile Range
BCS	67.41%	14.14%	59%-78%
CMC	78.06%	12.01%	71%-87%
COL	48.79%	13.69%	40%-56%
HBA	76.51%	12.68%	70%-86%
LDL	75.04%	13.13%	67%-85%
BBH	54.29%	11.79%	43%-64%
CAD	49.64%	11.49%	42%-58%
HF2	47.50%	13.50%	38%-57%
HF8	76.30%	10.56%	70%-83%
MPM	91.43%	9.19%	88%-98%
CDC_Eye	58.28%	17.78%	47%-65%
GSO	25.59%	14.09%	16%-32%
SPR	25.59%	14.09%	16%-32%
ART	78.48%	13.00%	74%-86%
OMW	20.46%	11.69%	14%-25%

Constructing a Master Physician List

Part of the work of aggregating data from different payers involved consolidating the four data contributors' provider files to identify a single set of unique physicians. Physicians were

identified in numerous different ways within and across the commercial and Medicare data contributors. Data contributors have physicians stored in their systems with varying representations of the physicians' name, address, specialty, and identifier numbers. For instance, some data contributors use different health plan identifier numbers for the same physician within their own system. Across data contributors, there is even further variation in the way in which the same physician is represented. The most common identifier, tax identification number (Tax ID), is of limited value for identifying unique physicians because multiple physicians can use the same Tax ID and physicians can bill under more than one Tax ID. Further, those identifiers that are typically unique, such as the Unique Physician Identification Number (UPIN) and Drug Enforcement Administration (DEA) number, are often either entirely or intermittently absent from a given data contributors' files.

Given the absence of a common identifier across the four key sources of data, probabilistic linking of the claims records (professional, facility, and pharmacy) was used in this project to identify the universe of unique physicians practicing in California. Probabilistic linking does not pre-determine a point-value (or weight) for the various identifying variables, but instead allows the data itself to determine what those weights should be, based on the relative importance of an agreement on that variable. Probabilistic linking also allows for a variable's weight to be adjusted (or scaled) up or down to reflect the relative frequency of a particular value for the variable. For instance, in data where there are many last names of "Smith," then the weight for the last name variable would be scaled downward to reflect that agreement on this variable is less important; similarly, in data where there are very few last names of "Abercombe," then the weight for last name would be scaled upward to reflect that agreement on this variable is more important. Finally, probabilistic linking considers not only agreements but also disagreements in determining whether two records are linked (while most deterministic methods only include agreement measures). For instance, agreement on gender does not provide much information about whether two records are linked, but disagreement on gender does provide a lot of information that two records are not linked.

1.

This process resulted in the creation of a Master Physician List (MPL) which contains the names, addresses, and specialties for all physicians that could be successfully identified within the data. A total of 61,984 unique physicians are contained in the MPL; 22,683 of whom (37%) were adult primary care (generalists, family physicians, internal medicine). In addition, CCHRI used external reference source data files to validate and correct physician specialty and fill in missing address and identifier information.

Assessing the Reliability of Performance Scores

CPPI adopted a minimum reliability threshold score of 0.70 to indicate that there was sufficient signal to discriminate performance. Reliability is a key measure of the scientific soundness of a measure and is a ratio of signal-to-noise. The signal in this case is the proportion of the variability that can be explained by real differences in performance. A reliability of zero implies that all of the variability in a measure is attributable to measurement error. A reliability of one implies that all of the variability is attributable to real differences in performance. High

reliability does not mean that performance on a given measure is good, but rather that one can confidently distinguish the performance of one physician from another.

Measures of physician clinical quality, patient experience, peer review, medical errors, and utilization have been evaluated for their reliability (Hofer et al. 1999; Hayward and Hofer 2001; NCQA 2006; Hays and Revicki 2005). The primary motivation for using reliability as a metric is reliability's relationship to misclassification. If a provider is flagged as higher quality than his/her peers this classification is more likely to be correct the higher the reliability. The reliability of a HEDIS rate based on n denominator patients can be calculated from the Spearman-Brown prophecy formula:

$$R_i = \frac{\sigma_{MD}^2}{\sigma_{MD}^2 + \frac{\sigma_i^2}{n_i}}$$

In this formula, σ_{MD}^2 is the between-physician variance, the variance among physician rates, σ_i^2 is the binomial variance associated with the rate for physician i , and n_i is the rate denominator (sample size) for physician i . Consequently, R_i is the proportion of the total variance that is attributable to the variance among physicians. As is evident from the formula for R_i , the reliability of each physician's HEDIS rate increases with 1) the variation in rates among physicians, and 2) the physician's sample size (denominator).

We estimated the physician-to-physician variance, σ_{MD}^2 , using a beta-binomial model, which assumed that the underlying true physician rates were distributed according to a beta distribution and that the observed rate for physician i was distributed according to a binomial distribution with sample size n_i . To estimate σ_{MD}^2 , we used a SAS macro developed for this purpose (Wakeling, 2004). To estimate the binomial sampling variance, σ_i^2 , for each physician, we used the property of the binomial distribution, $\sigma_i^2 = p_i(1-p_i)$, where p_i is the physician's true rate. For moderate to large sample sizes this variance can be estimated by substituting the physician's observed rate $\hat{p}_i = (x_i / n_i)$ for p_i into the formula for σ_i^2 , where x_i is the number of successes for the sample of n_i patients treated by physician i . However, many physicians have small denominators, resulting in unstable estimates of their true rates. Therefore, for the purpose of estimating the reliability, R_i , for physician i , we estimated each physician's rate, p_i , using a shrinkage estimator (explained in the next section) to compensate for the instability of estimated rates for small-denominator physicians. For example, without this adjustment, physicians with a denominator sample of size $n = 1$ would always have $p_i = 0$ or $p_i = 1$, resulting in an estimated reliability of 100 percent, which is clearly inappropriate.

We sought to establish a denominator sample size threshold that would achieve a certain level of reliability. However, for a given sample size, n_i , the reliability is different for different values of σ_i^2 , which depends on the physician's HEDIS rate. Therefore, we established the sample size threshold as the sample size for which 90 percent of physicians had at least 70 percent reliability; we computed the observed 10th percentile of estimated reliability for each sample size n , and

identified the value of n for which the 10th percentile of reliability was 0.70. This value of n was employed as a sample size threshold to categorize physicians as reliable.¹

A physician’s performance rate for a given measure was deemed to be reliable if his or her denominator sample size was at least as large as the sample size for which 90 percent of physicians had at least a 70 percent estimated reliability. Note that some physicians will have slightly less than 70 percent estimated reliability using this sample size threshold, but the vast majority of physicians exceeding this sample size threshold will exceed or nearly exceed an estimated 70 percent reliability. It is also possible for a physician with a sample size below the threshold to exceed 70 percent estimated reliability. We addressed this by designating a physician’s rate as reliable if his or her sample size fell below the sample size threshold, but his or her estimated reliability was at least 70 percent.

The patient denominator counts per physician to achieve the threshold of 0.70 or higher varies moderately measure-to-measure. For a number of measures the reliability threshold was achieved with 25-35 patient denominators. From some measures, such as the Osteoporosis Management and the Heart Failure indicators, denominator sizes of 50+ patients are required to achieve a reliable estimate of performance. For measures that focused on less common conditions (e.g., heart failure, heart attack), achieving reliable denominator sizes at a physician-level is difficult given small patient samples per physician.

Table 5 shows that the proportion of physicians with reliable results ranged considerably across measures. A majority of eligible physicians had reliable results for the preventive care measures while, typically, about one-third of physicians eligible for scoring had reliable results for the diabetes measures. At the extreme, almost none of the physicians had reliable results for the heart failure measures. Table 4 also shows the mean performance rate, for both reliably and unreliably scored physicians and the scores at various percentile cut points.

Table 5. Physician Performance Rates and Denominators By Measure And Reliability Status

			Performance Rate (%)							
			MDs	% MDs	Mean	P10	P25	P50	P75	P90
Measure	Den. Threshold	Reliable?								
ACE	25	No	10,291	62.2	83.3	50.0	75.0	90.9	100.0	100.0
		Yes	6,245	37.8	93.6	83.3	90.8	95.7	100.0	100.0
ART	21	No	8,723	98.4	65.6	0.0	50.0	75.0	100.0	100.0
		Yes	142	1.6	79.4	67.4	74.5	81.2	86.7	92.7
BCS	28	No	9,286	44.9	63.9	33.3	50.0	66.7	80.0	100.0
		Yes	11,387	55.1	67.9	48.2	59.5	69.7	78.1	84.1

¹ For some measures, the 10th percentile of observed reliability never reached 70 percent. In those cases, we calculated the sample size threshold by solving the Spearman-Brown prophecy formula for n, and substituted the overall observed mean rate to estimate the binomial variance.

			Performance Rate (%)							
			MDs	% MDs	Mean	P10	P25	P50	P75	P90
CAD	75	No	14,710	93.7	51.3	0.0	36.4	50.0	66.7	100.0
		Yes	983	6.3	50.6	37.4	43.8	51.3	57.7	63.5
CDC_EYE	35 *	No	7,866	62.9	54.5	25.9	41.7	53.8	66.7	85.7
		Yes	4,641	37.1	60.4	41.0	48.0	56.0	65.8	100.0
CMC	37	No	12,229	83.4	73.5	36.0	61.5	80.0	100.0	100.0
		Yes	2,429	16.6	81.0	66.7	74.4	82.5	88.9	93.7
COL	22	No	4,465	23.3	43.2	0.0	25.0	42.9	60.0	84.2
		Yes	14,696	76.7	48.9	33.0	40.6	47.9	55.6	66.0
GSO	19 *	No	1,598	11.7	65.1	25.0	50.0	66.7	88.2	100.0
		Yes	12,050	88.3	72.0	58.9	65.2	71.2	77.0	85.7
HBA	33	No	10,961	66.7	73.4	42.9	62.5	78.1	92.3	100.0
		Yes	5,467	33.3	77.4	61.9	70.4	79.0	86.2	91.7
HF2	47	No	9,847	99.0	44.6	0.0	20.0	45.0	66.7	100.0
		Yes	102	1.0	49.6	28.3	40.4	48.6	60.7	68.8
HF8	95	No	10,717	99.9	75.7	33.3	63.2	83.3	100.0	100.0
		Yes	7	0.1	86.7	75.0	81.0	85.1	95.8	95.9
LDL	29	No	10,757	62.9	71.4	33.3	58.8	75.0	92.3	100.0
		Yes	6,354	37.1	76.1	58.6	67.7	77.6	85.7	91.9
MPM	23	No	8,935	52.9	82.0	50.0	73.7	88.9	100.0	100.0
		Yes	7,967	47.1	93.0	82.0	89.8	95.5	100.0	100.0
OMW	60	No	10,060	100.0	23.4	0.0	0.0	12.5	35.3	66.7
		Yes	1	0.0	2.9	2.9	2.9	2.9	2.9	2.9
PBH	267	No	1,724	100.0	73.5	0.0	50.0	100.0	100.0	100.0
SPR	28	No	10,748	70.1	29.9	0.0	0.0	25.0	45.5	70.0
		Yes	4,583	29.9	24.0	9.5	14.5	21.4	30.6	41.0

Validation Work Associated with Methods Used for Scoring Performance

CPPI conducted several validation steps to inform its measurement and scoring work. Using 2005 data, we tested the impact of three alternative approaches (see Attribution section) for assigning patient denominator events to physicians. The validation work comprised both empirical analyses of the data and a test with physicians. Additionally, we compared performance rates to national benchmarks and conducted a small test with physicians to examine how frequently our designation of a numerator negative event (i.e., patient did not receive service) would have been overturned had we used medical chart data or data from an electronic medical record.

Empirical Examination of the Three Attribution Methods: The empirical analyses concerned the impact of the three approaches on the:

1. ***Number of Denominator Patients Attributed:*** Under all attribution methods, a substantial percentage of members were not attributable to any physician. Although approximately 25-30% of patient events were lost when assigning under Method #1, that number jumped to 31-48% under Methods #2 and #3. Method #1 saw a smaller loss of patient events because attribution was not limited to relevant specialties in this approach.
2. ***Number of Physicians Attributed:*** Moving from Method 1 to Methods 2 or 3 results in a loss of about half of the population of physicians assigned one or more events– the bulk of whom are designated “non-relevant physicians.”
3. ***Physician Performance Measure Rates:*** Physician-level measure rates were calculated under each of the three patient attribution methods. Measure rates at the physician level were highly correlated across the 3 attribution methods; there was little variation in a physician’s score relative to the method of attribution.
4. ***Ability to Score Physicians Reliably:*** To assess the number of physicians who could be reliably scored, we examined the minimum patient sample size needed to achieve 0.70 physician-level reliability; a commonly used threshold. The minimum patient sample varied by attribution method and measure. The minimum number of denominator patients (i.e., sample size) to achieve a 0.70 reliability is lowest under method 3 with the exception of BSC where method 2 yielded the lowest number of required patients. Under method 3 there were considerably fewer physicians that could be reliably scored and a smaller percentage of patients were attributed. Method 2 resulted in the largest percentage of patients being attributed to physicians with 0.70 reliability or higher and provided for a larger population of physicians with scores achieving 0.70 or higher reliability as compared to method 3.

Physician Review of Attribution Approaches: The physician testing provided an opportunity to get direct feedback from physicians about the patient population being assigned to them as a test of the “face validity” of the three approaches. This work was performed using five measures (BCS, COL, HBA, CMC, and LCL) constructed from the MY2005 Y data.

Five California physician groups, with physicians that treat PPO patients, volunteered to participate in these validation tasks. We used a convenience sample of physicians, initially targeting 50 physicians; as Table 6 shows, the final sample contained 38 physicians who reviewed attribution assignments for 3,478 patient events (1,850 unique patients). This test helped us understand and take into consideration how doctors view their patient relationship responsibilities. We were also able to explore the “patient accountability” concept with pairs of physicians who treated the same patient within the measurement period. Physicians from the five physician groups who were attributed patients under one or more of the three methods were asked to indicate if they felt responsible for the patient’s medical care.

Table 6. Total Number of Patients Validated for the Patient Attribution Validation

PO	BCS	CMC	COL	HBA	LDL	Total Patients Validated Across Measures		Total Unique Patients Validated	
						n	%	n	%
PO 1	192	37	394	160	160	943	99.9%	449	99.8%
PO 2	82	3	518	52	52	707	97.0%	449	97.4%
PO 3	174	67	553	128	128	1,050	99.3%	535	99.3%
PO 4	155	11	460	76	76	778	95.3%	417	95.0%
PO 5	--	--	--	--	--	--	--	--	--
Total	603	118	1,925	416	416	3,478	78.2%	1,850	76.2%

Note: PO 5 (Physician Organization 5) was excluded after learning that the EMR used for validation purposes did not cover the entire measurement period thus the chart review was incomplete.

Physicians were asked two key questions related to their perceptions of responsibility for the care of each patient listed: a) degree of responsibility for the patient’s overall health care, and b) degree of responsibility for the patient’s measure-specific care (e.g., diabetes). Physicians were also asked to provide a reason why they did not feel responsible or felt only partially responsible for a patient’s care, if they did not indicate that they felt fully responsible for the patient.

Table 7 shows how physicians view their responsibility for the patients they were asked to review.

Table 7. Physician Reports of Responsibility for Patients’ Overall Care, by Specialty Relevancy and Attribution Method

Measure	Non-Relevant Physicians (Method 1 Only)	Relevant Physicians Only	
		Not Plurality E&M Visits (Method 2 Only)	Plurality E&M Visits (Method 3)
	(1)	(3)	(4)
BCS	0.0%	18.2%	61.6%
COL	0.0%	26.5%	66.2%
CMC	0.0%	41.3%	74.5%
LDL	0.0%	30.7%	80.4%
HBA	0.0%	33.0%	81.1%
OVERALL	0.0%	26.5%	70.2%

Note: Column 3 (Method 2 Only) refers to a portion of the physicians assigned under Method 2 who are the physicians who did not account for a plurality of the E&M visits. Column (4) captures physicians who represented the plurality of E&M visits with the assigned patient, and represents the other portion of physicians that would be included under attribution Method 2.

[Edit note: column 2 is missing from this table]

For non-relevant specialties (under Attribution Method 1), virtually none of the time did physicians indicate that they were responsible for the patient’s care, whereas relevant-specialty physicians claimed responsibility for the patient’s overall care approximately half of the time. However, relevant specialty physicians were much more likely to claim responsibility for the patient’s overall care where they had the most E&M visits with the patient (Attribution Method 3) than when they did not (Attribution Method 2 only, not Method 3). Indeed, fully three-fourths

of relevant specialty physicians claimed responsibility for the patient’s overall care under Attribution Method 3, whereas only about one-third of relevant specialty physicians claimed responsibility for the patient under Attribution Method 2 only. The pattern of results did not differ when physicians were asked about their feelings of responsibility for the patient’s measure-relevant care, such as diabetes.

CPPI adopted an attribution rule that is a blend of columns 3 and 4 in Table 7: the final rule includes any relevant-specialty physician who the eligible patient saw at least once. As such, the rule captures a mix of attribution situations in which the physician had the plurality of visits with that patient and other situations in which the physician saw the patient but another physician accounted for the plurality of visits.

When physicians did not claim responsibility for the patient’s care, a variety of reasons were provided (Table 8).

Table 8. Reasons Why Physicians Claim Non-Responsibility for Patient’s Care

Measure	Not Primary Care Giver	Treated Patient Through Single Consult Only	Patient Changed Status; No Longer Able to See Patient	Never Treated This Patient	Other
BCS	74.7%	3.6%	0.3%	20.1%	1.3%
COL	79.4%	8.2%	7.8%	9.2%	13.4%
CMC	84.3%	7.8%	0.0%	7.8%	0.0%
LDL	69.2%	9.2%	1.5%	17.7%	2.3%
HBA	63.8%	13.4%	2.4%	18.1%	2.4%

Across measures, the most predominant reason why physicians are not claiming responsibility for patients’ care is that they do not feel that they are the primary care giver for those patients.

Comparison to External Benchmarks: The population-level (pre-attribution) results for each measure were compared to appropriate benchmarks. This was one means used to gauge the reasonableness of the measure scores for each of the four data contributors. Comparisons were made to:

- Commercial HEDIS 2007 audit means for PPOs (NCQA)
- Commercial HEDIS 2007 audit means for HMOs (NCQA)
- Medicare HEDIS 2007 audit means for PPOs (NCQA)
- Medicare HEDIS 2007 audit means for HMOs (NCQA)

For some measures, the measures are too new to have external benchmarks to compare to—such as for the Medicare Care Management Performance Measures. With the exception of a few measures that were set aside for this cycle of reporting and the colorectal screening measure which has a 10 year lookback period, the claims-derived performance rates computed for CPPI were within 1 to 8 percentage points excluding the BBH measure which along with several other measures was removed from the reportable results (Table 9).

Table 9. Comparison of Commercial and Medicare Population-level Rates to External Benchmarks

Measure	Commercial			Medicare		
	Rate	Benchmark	Difference to Benchmark	Rate	Benchmark	Difference to Benchmark
BCS	64.1%	62.7% (1)	1%	54.2%	60.6% (3)	-6%
COL	40.9%	41.8% (1)	-1%	38.3%	53.2% (4)	-15%
CMC	61.3%	68.1% (1)	-7%	82.1%	88.0% (4)	-6%
LDL	60.8%	67.4% (1)	-7%	76.1%	79.0% (3)	-3%
HBA	63.6%	71.7% (1)	-8%	76.3%	83.1% (3)	-7%
NPH	53.2%	60.5% (1)	-7%	66.7%	85.3% (4)	-19%
BBH	43.8%	97.7% (2)	-54%	46.0%	93.7% (4)	-48%
PBH	69.5%	72.5% (2)	-3%	NA	NA	NA
MPM	66.4%	65.6% (1)	1%	89.6%	83.6% (3)	6%
CAD	59.7%	*	*	48.8%	*	*
HF-2	44.4%	*	*	40.7%	*	*
HF-8	68.7%	*	*	73.8%	*	*
CDC - Eye Exams	28.8%	35.0% (1)	-6%	50.4%	53.1% (3)	-3%
GSO	NA	NA	NA	62.7%	63.3% (3)	-1%
SPR	29.7%	33.6% (1)	-4%	23.6%	26.2% (4)	-3%
PCE - Systemic Corticosteroid	43.2%	**	**	21.3%	**	**
PCE - Bronchodilator	57.2%	**	**	30.3%	**	**
ART	81.9%	82.4% (1)	0%	63.2%	68.2% (4)	-5%
OMW	NA	NA	NA	23.7%	21.8% (4)	2%

Note: (1) Commercial HEDIS 2007 audit means for PPOs (NCQA). (2) Commercial HEDIS 2007 audit means for HMOs (NCQA). (3) Medicare HEDIS 2007 audit means for PPOs (NCQA). (4) Medicare HEDIS 2007 audit means for HMOs (NCQA) * No benchmarks available for Medicare Care Management Performance measures. ** No benchmarks available: new HEDIS measures starting with measurement year 2008. In general, HMOs use hybrid methods for relevant measures and PPOs use only administrative data methods.

The CPPI scores compared to benchmark are an important element of the claims-based information approach but more work is needed to understand the extent to which incomplete data affects a physician score. The Cycle 3 results reporting to physicians is an opportunity to further understand the validity of claims-based measures as physicians self-validate their scores.

Physician Reporting

The CPPI project is reporting all available results to physicians, including results that don't meet the reliability threshold, to share whatever information is available about their patients. Physician scores are displayed graphically only for measures with a patient count of 10 or more. If too few patients were attributed to a physician to provide a reliable score, the results are displayed with a 'no fill' bar and an alert that the physician's true ranking likely differs from this result. All scores, regardless of reliability or patient count, are listed in a summary table with scores displayed separately for commercial, Medicare and the combined results.

In some cases, per the measures specifications, the measure is only relevant to the Medicare population; thus, the measure only is scored using Medicare data. In other cases, the project team felt there were data completeness issues that indicated the measure should be scored using only the Medicare or Commercial data. An example of this is the diabetes eye exam measure, where there was concern that often commercial vision care coverage is handled separately from medical coverage and the vision care services may not be captured in the claims data.

Table 13. Data Sources used in Measure Reporting

	Measures		Data Sources used for Measure Reporting
1	BCS	Breast Cancer Screening	Commercial and Medicare
2	COL	Colorectal Cancer Screening	Commercial and Medicare
3	CMC	Cholesterol Management for Pts With Cardiovascular Conditions - LDL Screening	Commercial and Medicare
4	LDL	Comprehensive Diabetes Care - LDL Screening	Commercial and Medicare
5	HBA	Comprehensive Diabetes Care - HbA1c testing	Commercial and Medicare
6	NPH	Comprehensive Diabetes Care - Nephropathy	Not Included in Physician Reports
7	BBH	Beta-Blocker Treatment After a Heart Attack	Not Included in Physician Reports
8	PBH	Persistence of Beta-Blocker Treatment After a Heart Attack	Commercial Only
9	MPM	Annual Monitoring for Patients on Persistent Medications – Total	Commercial and Medicare
10	CAD	Coronary Artery Disease	Commercial and Medicare
11	HF2	Heart Failure: Left Ventricular Ejection Fraction Testing	Commercial and Medicare
12	HF8	Heart Failure: Warfarin Therapy for Patients with Atrial Fibrillation	Commercial and Medicare
13	CDC_Eye Exam	Comprehensive Diabetes Care - Eye Exams	Medicare Only
14	GSO	Glaucoma Screening in Older Adults	Medicare Only (Medicare Only measure per HEDIS)
15	SPR	Use of Spirometry Testing in the Assessment and Diagnosis of COPD	Commercial and Medicare
16	PCE_Cortic	Pharmacotherapy Management of COPD Exacerbation - Systemic Corticosteroid	Not Included in Physician Reports
17	PCE_Bronch	Pharmacotherapy Management of COPD Exacerbation – Bronchodilator	Not Included in Physician Reports
18	ART	Disease Modifying Anti-Rheumatic Drug Therapy for Rheumatoid Arthritis	Commercial and Medicare

	Measures		Data Sources used for Measure Reporting
19	OMW	Osteoporosis Management in Women Who Had a Fracture	Medicare Only (Medicare Only measure per HEDIS)

Note: The measures shaded in grey are excluded from reporting due to measures specification problems, data incompleteness, and one of the measures has been retired by NCQA.

Pilot Testing of Physician Reports: CCHRI worked closely with physicians to engage them in the review and modification of the performance report that would share results. The first step in this process was sharing performance results data with the CPPI's Physician Advisory Group and Steering Committee. In addition to review by the two physician committees, CCHRI engaged practicing physicians to assist with reviewing data results and materials for use during the pilot. The project prepared and disseminated physician-level scores to 74 medical directors of physician groups who routinely saw patients. The medical directors were asked to provide feedback regarding the content of the report. CCHRI received feedback from 12 physicians located in the following regions: Southern California (5), San Diego (2), Santa Cruz (2), Sacramento (1), San Francisco (1), and Humboldt (1). Overall, the feedback was quite positive: noting that the information was clearly presented and helpful. CCHRI also provided the sample Report, Frequently Asked Questions (FAQ), and cover letter to the California Medical Association (CMA) for their review and feedback. The report was also reviewed and approved by the Centers for Medicare & Medicaid Services (CMS). The final report format and content incorporated feedback from these various parties.

To support action on the part of physicians for quality improvement and to support requests from physicians about the project, CCHRI worked with its PAG and SC to develop the Physician Comment and Request for Information Period. This process, accessed via www.cchri.org/cppi provides physicians the opportunity to do the following:

- 1- Update or change their demographic information, specialty designation
- 2- Request a duplicate report
- 3- Request a patient list for their commercial patients
- 4- Submit additional data, if needed
- 5- Ask questions
- 6- Provide feedback

The site will also contain a detailed document that explains the underlying methodology used to generate the performance scores. Thomson Reuters will track requests for information and modifications supplied by physicians; these findings will be reported back for review and consideration of next steps to CCHRI and its PAG and the SC. To comply with HIPAA and the California Civil Code, The Physician Comment and Request for Information Period process includes a consent form that a physician is required to sign to authorize the release of the patient list derived from the aggregated commercial data.

Version: 10/13/2008