Physician Compare Quality Measurement Technical Expert Panel (TEP) Summary Report

October 2014

Thomas MaCurdy
Keziah Cook
Stephen McKean
Rob Warnock
Sumit Verma

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010

[This page is intentionally left blank.]

TABLE OF CONTENTS

LIST OF TABLES

1 ABOUT THE TEP

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC (henceforth "Acumen") to assist in the selection of physician quality measures for public reporting via the Physician Compare website. As part of the measures selection process, Acumen has convened the Physician Compare Quality Measurement Technical Expert Panel (henceforth "the TEP") to obtain expert input on physician quality measures that CMS has proposed for public reporting and seek recommendations regarding future quality measures for public reporting. Specifically, the TEP will help meet the Affordable Care Act (ACA) requirement of ensuring that data reported on Physician Compare provide an accurate and robust portrayal of physician performance. The TEP composition meets the CMS Measures Management Blueprint criteria, including the involvement of individuals who represent the perspectives of patient/caregiver and purchasers, and technical experts who can provide a broad range of technical experience and expertise in public reporting of performance measures, health care quality improvement, and quality measure development and testing.

Acumen first convened with the TEP via teleconference on September 30-October 1, 2013 to discuss the final selection of candidate quality measures for public reporting in early 2014. A summary of this meeting is available on the CMS website.[1] Acumen and the TEP met again on August 18, 2014 to review the measures available for public reporting in late 2014. The remainder of this report summarizes the discussions and conclusions from this meeting. Section 2 summarizes Acumen's analysis findings for the candidate measures and Section 3 summarizes the TEP's input. Table 1.1 on the following page lists the 15 individuals who comprise the TEP, nine of whom were present on the teleconference.

[1] Physician Compare Quality Measurement Technical Expert Panel (TEP) Summary Report, December 2013.

Table 1.1: TEP Members

TEP Member	Position(s),Organization	Location
A.J. Yates, MD	Associate Professor, Department of Orthopedic Surgery/University of Pittsburgh School of Medicine	Pittsburgh, PA
Bettina Berman, RN, MPH	Project Director for Quality Improvement, Jefferson School of Population Health/Thomas Jefferson University	Philadelphia, PA
Dale Shaller, MPA **(TEP Chair)**	Principal, Shaller Consulting Group	Stillwater, MN
David Baker, MD, MPH	Michael A. Gertz Professor in Medicine, Chief of the Division of General Internal Medicine and Geriatrics, and Deputy Director of Institute for Public Health and Medicine at Feinberg School of Medicine, Northwestern University	Chicago, IL
Dave deBronkart*	Co-Founder, Society for Participatory Medicine (ePatient Dave)	Nashua, NH
Emma Kopleff, MPH*	Senior Policy Advisor, National Partnership for Women & Families	Washington, DC
Eric Holmboe, MD*	Chief Medical Officer and Senior Vice President, American Board of Internal Medicine	Philadelphia, PA
Gregory Dehmer, MD	Cardiologist, American College of Cardiology	Temple, TX
Jeffrey P. Jacobs, MD	Professor of Cardiac Surgery, Johns Hopkins University	St. Petersburg, FL
Michael Mihlbauer, MS*	Practice Administrator, Anesthesiology Associates of Wisconsin	Milwaukee, WI
Robert Krughoff, JD	Founder and President, Center for the Study of Services/Consumers' Checkbook	Washington, DC
Sara Schoelle, DrPH	Vice President, Research & Analysis/National Committee for Quality Assurance	Washington, DC
Sherrie Kaplan, PhD, MSPH, MPH*	Professor of Medicine and Assistant Vice Chancellor, Healthcare Evaluation and Measurement Executive Co-Director, Health Policy Research Institute School of Medicine/ University of California, Irvine	Irvine, CA
Ted von Glahn, MS	Senior Director, Performance Information and Consumer Engagement/Pacific Business Group on Health	San Francisco, CA
Thomas Smith, MD, MS*	Vice President of Research & Analysis, New York State Office of Mental Health / Columbia University Medical Center	New York, NY

**TEP member was unable to participate in the teleconference, but received and reviewed all meeting materials and was invited to provide written feedback.*

2 SUMMARY OF MEASURES SELECTION ANALYSES

Per the 2013 Medicare Physician Fee Schedule (PFS) Final Rules, CMS intends to report in late 2014 a select set of the 2013 Physician Quality Reporting System (PQRS) - Group Practice Reporting Option (GPRO) Web Interface quality measures for patients with Diabetes Mellitus (DM) or Coronary Artery Disease (CAD). CMS still has to select the specific DM and CAD quality measures for public reporting; up to *six* DM measures and *two* CAD measures may be selected. Table 2.1 summarizes these eight candidate measures.[2] The first two columns list the PQRS number and title of each measure. The third column describes the patient outcome or clinical care process that each measure assesses. The fourth column identifies the measure developer, which includes the National Committee for Quality Assurance (NCQA), the Minnesota Community Measurement (MNCM), and the American Medical Association-Physician Consortium for Performance Improvement (AMA-PCPI). The final column indicates the measure type (i.e., outcome or process).

Table 2.1: Eight Candidate Measures for Public Reporting

Measure Number	Measure Title	Measure Description	Measure Developer	Measure Type
GPRO DM-2	Hemoglobin A1c Poor Control	Percentage of patients aged 18 through 75 years with diabetes mellitus who had most recent hemoglobin A1c greater than 9.0%	NCQA	Outcome
GPRO DM-13	High Blood Pressure Control	Percentage of patients aged 18 through 75 years with diabetes mellitus who had most recent blood pressure in control (less than 140/90 mmHg)	NCQA	Outcome
GPRO DM-14	Low Density Lipoprotein (LDL-C) Control	Percentage of patients aged 18 through 75 years with diabetes mellitus who had most recent LDL-C level in control (less than 100 mg/dL)	NCQA	Outcome
GPRO DM-15	Hemoglobin A1c Control (< 8%)	Percentage of patients aged 18 through 75 years with a diagnosis of diabetes (type 1 or type 2) who had HbA1c < 8%	NCQA	Outcome
GPRO DM-17	Tobacco Non-Use	Percentage of patients aged 18 to 75 years with a diagnosis of diabetes who indicated they were tobacco non-users	MNCM	Outcome

[2] The measure specifications can be downloaded from: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/GPRO_Web_Interface.html

Measure Number	Measure Title	Measure Description	Measure Developer	Measure Type
GPRO DM-16	Daily Aspirin or Antiplatelet Medication Use for Patients with Diabetes and Ischemic Vascular Disease	Percentage of patients aged 18 to 75 years with diabetes mellitus and ischemic vascular disease with documented daily aspirin or antiplatelet medication use during the measurement year unless contraindicated	MNCM	Process
GPRO CAD-2	Lipid Control	Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who have a LDL-C result < 100 mg/dL OR patients who have a LDL-C result ≥ 100 mg/dL and have a documented plan of care to achieve LDL-C < 100 mg/dL, including at a minimum the prescription of a statin	AMA-PCPI	Process
GPRO CAD-7	ACE-I/ARB Therapy – Diabetes or LVSD	Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have diabetes OR a current or prior Left Ventricular Ejection Fraction (LVEF) < 40% who were prescribed ACE inhibitor or ARB therapy	AMA-PCPI	Process

Amongst other requirements, Section 10331 of the 2010 Patient Protection and Affordable Care Act (ACA) requires that the data made public via Physician Compare are statistically valid and reliable, including any risk adjustment mechanisms as appropriate. To support CMS in the final selection of quality measures for public reporting in late 2014, Acumen analyzed the empirical and statistical properties of group practices' data collected via the 2013 GPRO Web Interface and presented our findings to the TEP for consideration. The remainder of this section summarizes these findings.

2.1 Data Quality Assurance

To ensure the reliability and accuracy of the quality measures reported through Physician Compare, Acumen performed two sets of QA checks. First, Acumen verified that the group-level performance rates in the group-level data files are consistent with the beneficiary-level data for each DM and CAD quality measure. Beneficiary-level data were aggregated according to the group practices responsible for recording each measure, and the resulting group-level rates matched those of the group-level files.

Second, Acumen confirmed that group practices have correctly adhered to CMS criteria in excluding (or "skipping") patients from their DM and CAD samples. Patients could be

skipped from a group practice sample for a number of reasons, including them being deceased, in hospice care, residing out of the country, or insured by a private insurer.[3] Where possible, Acumen leveraged Medicare claims/enrollment data to verify these skips and found that 79% of them were valid. In total, the average skip rate among all group practices was 10%. The single most common skip reason was that the group practice was unable to confirm that the sampled patient had the DM or CAD outcome. This accounted for 89% of DM skips and 74% of CAD skips.

2.2 Measure Performance Rates

For each quality measure, the performance rate refers to the percentage of a group practice's sampled beneficiaries who meet that measure's numerator criteria, as outlined in the third column of Table 2.1. Table 2.2 shows the distribution of group practice performance rates in 2013 for each DM and CAD measure. For all measures except **DM-2,** meeting the measure is indicative of good heath or good care (e.g., patients meeting measure **DM-15** are successfully managing blood glucose levels better than those who do not). Therefore, higher rates indicate better group practice performance. For **DM-2**, however, patients who meet this measure have recorded blood glucose levels exceeding a healthy threshold; therefore not meeting the measure represents a better patient outcome and low rates indicate better group practice performance.

Table 2.2: Distribution of Group Practice Performance Rates across Measures

Measure		Minimum	25th Percentile	Median	75th Percentile	Maximum	Mean	Std. Deviation
Outcome	**DM-2**	7.3	11.9	16.1	22.1	100	18.4	10.8
	DM-13	44.8	65.9	70.5	76.2	91.6	70.8	8.5
	DM-14	27.8	54.0	60.6	65.3	80.6	59.6	9.5
	DM-15	0.0	68.4	74.1	78.0	90.2	71.2	10.6
	DM-17	8.7	79.3	84.2	87.7	97.9	80.6	14.8
Process	**DM-16**	25.0	80.9	89.3	94.5	100	85.8	12.9
	CAD-2	41.4	75.6	83.4	89.9	99.5	81.7	10.7
	CAD-7	42.9	71.2	81.5	88.3	97.8	79.2	11.7

Since the Physician Compare website aims to provide actionable information for users to choose physicians and other health care professionals based on the quality of care, CMS should only publicly report performance rates that enable users to make "meaningful" comparisons across group practices. To assess whether the comparison of these rates is meaningful, Acumen examined their statistical significance by comparing the rate of each group practice with the average rate across all group practices. Acumen conducted a two-sided test of the null hypothesis that a group practice's performance is not different from the mean performance of all

[3] Other skip reasons that could not be verified using claims data are: the patient's diagnostic condition could not be confirmed or other medical and non-medical "CMS-approved reasons."

group practices with at least 20 measure-eligible cases at the 5% significance level. Across all measures, more than half of the group practices' rates differ statistically from the mean, implying that there is enough variation between groups to make meaningful comparisons.

2.3 Measure Reliability

Measure reliability refers to the extent to which differences in performance rates for each quality measure are due to actual differences in group practice performance versus variation that arises from measurement error. Statistically, reliability depends on performance variation for a measure across group practices, the random variation in performance for a measure within a provider's panel of attributed beneficiaries, and the number of beneficiaries attributed to the provider. High reliability for a measure suggests that comparisons of relative performance across group practices are likely to be stable over different performance periods and that the performance of one group practice on the quality measure can be confidently distinguished from another. Potential reliability values range from zero to one, where one (highest possible reliability) means that all variation in the measure's rates is the result of variation in differences in performance across group practices, while zero (lowest possible reliability) means that all variation is a result of measurement error.[4]

Acumen fit a beta-binomial model to calculate reliability scores for each measure. Acumen concluded that reliability was high across all measures; the 25th percentile ranged from 0.91 to 0.97, which is well above the range considered acceptable for drawing inferences about group practices (i.e., 0.70 – 0.80). Additionally, to measure between-group practice variation and within-group practice variation, Acumen calculated the test-retest reliability using the intra-class correlation coefficient (ICC). ICC values that approach 1 indicate that the fraction of the total variance due to between-group variation is high. The ICC values across all measures range from 0.83 to 0.94, indicating that most of the total variation is due to between-group practice variation. Table 2.3 shows the reliability coefficients and ICC values for each measure.

Table 2.3: Reliability of Group Practice Performance Rates

Measure		Reliability Coefficient					ICC
		Minimum	25th Percentile	Median	75th Percentile	Maximum	
Outcome	DM-2	0.52	0.96	0.97	0.98	1.00	0.91
	DM-13	0.42	0.92	0.93	0.94	0.97	0.86
	DM-14	0.43	0.93	0.93	0.94	0.96	0.84

[4] For more information about reliability testing for physician performance measurement, as well as the methodology for constructing the reliability score reported on Table 6, see "Reliability of Provider Profiling: A Tutorial" by John Adams, RAND. http://www.rand.org/pubs/technical_reports/TR653.html

Measure		Reliability Coefficient					ICC
		Minimum	25th Percentile	Median	75th Percentile	Maximum	
Outcome	**DM-15**	0.54	0.96	0.97	0.97	1.00	0.88
	DM-17	0.80	0.97	0.98	0.99	1.00	0.94
Process	**DM-16**	0.69	0.91	0.95	0.97	1.00	0.88
	CAD-2	0.74	0.95	0.97	0.98	1.00	0.93
	CAD-7	0.56	0.91	0.94	0.96	1.00	0.83

2.4 Impact of Case-Mix on DM Outcome Measures

Case-mix adjustment refers to the statistical process of identifying and adjusting for differences in population characteristics (i.e., risk factors) before comparing outcomes of care. While case-mix adjustment is generally not applied to certain structure and process measures, it may be necessary for outcome measures that are not fully within a providers' control (e.g., blood sugar or cholesterol levels).

The DM outcome measures (i.e., **DM-2, DM-13, DM-14, DM-15,** and **DM-17**) do not include case-mix adjustment as part of their specifications. To determine the impact of case-mix on group practice performance rates across measures, Acumen adjusted the performance rates for certain patient characteristics that are outside the control of a group practice (e.g., demographic characteristics and pre-existing health conditions) and evaluated how group-level performance rates were affected.

To compare the impact of different sets of case-mix factors, Acumen constructed two predictive models; Model 1 includes selected basic demographic characteristics and health status variables (e.g., age, sex, reason for Medicare eligibility) that Medicare commonly uses as part of case-mix adjustment for other publicly reported measures. However, group practice performance rates may vary systematically based on racial and regional attributes that Medicare does not typically use for case-mix adjustment; Model 2 is an expanded model that includes these additional characteristics (e.g., race, income, and urban or rural region type). Based on these models, Acumen reached the following conclusions about the impact of case-mix adjustment on the DM outcome measures:

(1) In Model 1, the case-mix consisting of the basic demographic indicators (i.e., age, and sex) and health status variables (i.e., Medicare eligibility, Medicaid eligibility,

and the presence of assorted health conditions) does predict, on the patient-level, the probability of meeting a measure. However, this case-mix does not differ significantly across the group practices included in the sample. By comparing actual performance rates with performance rates predicted by this case-mix, Acumen found that case-mix adjustment has only a small impact on the predictive ability of Model 1.

(2) In Model 2, when the case-mix was expanded to include additional demographic and regional characteristics (i.e., race, region, region type, typical household income, and typical home value), predicted performance rates differed from actual performance rates on the group practice level to some extent.

3 TEP MEASURE RECOMMENDATIONS

During the August teleconference, the TEP reviewed the measure specifications and analysis findings summarized above to assess the appropriateness of each measure for public reporting. Since five of the eight candidate measures are currently reported on Physician Compare, the TEP discussed whether or not these measures are still suitable for reporting. The TEP unanimously agreed that the measures have been treated appropriately, with the exception of **DM-17**, the tobacco non-use measure. Table 3.1 below shows the TEP's recommendations and the remainder of this section summarizes the TEP's conclusions.

Table 3.1: Measures Available for Public Reporting for Program Year 2013

Available Measures		Reported for Program Year 2012	Recommendation for Program Year 2013
Outcome	**DM-2:** Hemoglobin A1c Poor Control	No	No
	DM-13: High Blood Pressure Control	Yes	Yes
	DM-14: Low Density Lipoprotein (LDL-C) Control	No	No
	DM-15: Hemoglobin A1c Control (<8%)	Yes	Yes
	DM-17: Tobacco Non-Use	Yes	Some Concerns
Process	**DM-16:** Daily Aspirin or Antiplatelet Medication Use	Yes	Yes
	CAD-2: Lipid Control	No	No
	CAD-7: ACE-I/ARB Therapy	Yes	Yes

3.1 TEP Input on DM Outcome Measures

Listed below are the TEP's recommendations for the five DM outcomes measures being considered for public reporting in late 2014: **DM-2, DM-13, DM-14, DM-15,** and **DM-17.**

3.1.1 DM-2: Hemoglobin A1c Poor Control

DM-2 was not selected for public reporting in early 2014. The TEP believes that CMS should continue to not report this measure because the website already displays results from an inverse glucose control measure, **DM-15.** In addition to the redundancy of having two outcome measures related to blood glucose levels, the interpretation of **DM-2** is opposite of the other publicly reported measures. A high performance rate indicates poor patient health, which may cause unnecessary confusion among the website's users.

3.1.2 DM-13: High Blood Pressure Control

DM-13 is currently publicly reported on the Physician Compare website. After reviewing the results of Acumen's analyses, the TEP agrees that it is a valuable quality measure and recommends that it continues to be reported.

3.1.3 DM-14: Low Density Lipoprotein (LDL-C) Control

DM-14 was not selected for public reporting in early 2014. The TEP believes that CMS should continue to not report this measure because its current specifications are no longer consistent with clinical guidelines for lipid control.

3.1.4 DM-15: Hemoglobin A1c Control (<8%)

DM-15 is currently publicly reported on the Physician Compare website. Although similar in purpose to **DM-2, DM-15** has a more logical interpretation: higher performance rates indicate good patient health. After reviewing the results of Acumen's analyses, the TEP agrees that **DM-15** is a valuable quality measure and recommends that it continues to be reported.

3.1.5 DM-17: Tobacco Non-Use

DM-17 is currently publicly reported on the Physician Compare website. However, the TEP was not unanimously in favor of continuing its reporting. Therefore, Acumen instructed members of the TEP to rate the performance measure on a scale from one to five, where one corresponds to "not appropriate for public reporting" and five corresponds to "very appropriate for public reporting." Acumen also asked everyone to include a comment justifying his or her opinion. Nine TEP members responded to this request and gave the measure an average score of 3.0 out of 5. They were split on the measure's usefulness; those in favor of reporting the measure generally felt tobacco has a major impact on public health and is a topic patients identify with, while those opposed argued that measure performance is largely driven by regional variation of the prevalence of tobacco use. Some of the dissenting opinions recommend that instead the measure be replaced with a process measure identifying whether or not patients who do smoke are receiving guidance from their physician to quit. The following comments were provided:

Comments in favor of reporting **DM-17:**

(1) "Despite the big variation in tobacco use by geography, ethnicity, and other factors, this is a measure providers can impact and they will be able to focus on developing strategies targeted on diabetics (and possibly other conditions). Data users can be helped to understand that the difficulty of scoring well varies among providers. The incentive the measure provides can be important."

(2) "I find that the measure is easy for providers to document in electronic health records. It should be measured for diabetic as well as non-diabetic patients."

(3) "I can imagine substantial concern might have been raised for a measure heavily dependent on changes in patient behavior. Such behavioral change is also multi-factorial. However, given the individual and public health implications of tobacco use in diabetic patients, I think this tips the balance in favor of public reporting. I think this type of measure might also be better suited for normative comparisons (such as achievable benchmarking)."

(4) "This is an important measure that the public can identify with. It's not clear how this classifies as an outcome measure when aspirin use is a process measure, but both should be reported as important indicators of good diabetes management. Differences in rates of tobacco use by region and patient characteristics do not negate the validity of this measure for public reporting."

(5) "Smoking prevalence differences by area is one marker of SES/other influences on health but the extension of that argument is to discount performance measures for a host of other issues like variations in BMI, alcohol consumption etc. I assume the Tobacco Non-Use measure is predicated on evidence that health professionals can influence tobacco use and if that is so, it should be retained given tobacco's impact on health. This raises a larger issue about benchmarking techniques and use of regional vs. national benchmarks that would be good for the TEP to consider."

Comments opposed to the reporting of **DM-17:**

(1) "[The measure is] more reflective of regional tobacco use rates than physician care. Consider using the tobacco screening and follow up measure."

(2) "Tobacco use varies markedly by state/region, race/ethnicity, and education. Thus, physician "performance" on this measure is highly influenced by factors beyond their control. Even with optimal care, only a minority of patients will stop smoking."

(3) "There are several potential problems with this measure. It is an absolute number and thus affected by the incidence of tobacco use in at baseline. This is just an example, but consider how this could be misleading – among all of the diabetic patients treated by Provider A, at baseline 50% are tobacco users. Provider A really doesn't care much about telling patients to stop, so after 1 year the incidence of diabetic smokers decreases from 50% to 45% (5% decrease). In contrast, provider B starts with 75% of diabetics smoking and works really hard to get them to quit. After one year the provider gets 25% to quit so now only 50% are still smokers. Since this reporting is for the public and if I were a diabetic smoker who needed help and encouragement, who would I go to see? The provider with the lowest percent of smokers (provider A who could care less if one quits) or provider B who has a higher rate, but works hard with patients and gets 25% to quit. This would be more meaningful to report the decrease in tobacco users over time rather than an absolute number without much context. Furthermore, the unintended consequence of this is providers will shun new diabetic smokers from their practice and only see patients who don't use tobacco. Thus, they could easily stack the deck to make their metric look good."

(4) "As it is, it is an "outcome" over which the MD has limited control depending on culture and patient. I would recommend changing it to a process measure (smoking cessation program discussed or ordered for patients that smoke)."

3.2 TEP Input on DM and CAD Process Measures

Listed below are the TEP's recommendations for the three DM and CAD process measures being considered for public reporting in late 2014: **DM-16, CAD-2, *and* CAD-7.**

3.2.1 DM-16: Daily Aspirin or Antiplatelet Use

DM-16 is currently publicly reported on the Physician Compare website. After reviewing the results of Acumen's analyses, the TEP agrees that it is a valuable quality measure and recommends that it continues to be reported.

3.2.2 CAD-2: Lipid Control

CAD-2 was not selected for public reporting in early 2014. The TEP believes that CMS should continue to not report this measure because its current specifications are no longer consistent with clinical guidelines for lipid control.

3.2.3 CAD-7: ACE-I/ARB Therapy

CAD-7 is currently publicly reported on the Physician Compare website. After reviewing the results of Acumen's analyses, the TEP agrees that it is a valuable quality measure and recommends that it continues to be reported.

4 GENERAL TEP COMMENTS

In addition to comments specific to the DM and CAD candidate measures described in the previous section, the TEP offered general comments and questions that spanned several topic areas, including future growth of Physician Compare's reporting, CAHPS measures, benchmarking, and concept testing. The following sections summarize the main discussion points covered during the teleconference, as well as comments and questions offered by TEP members. CMS will take the TEP's input into consideration to inform future Physician Compare measures selection activities.

4.1 Physician Compare as per the 2012 Public Reporting Plan

- Since last year's TEP teleconference, Physician Compare has begun publicly reporting performance data at the group practice level from the 2012 PQRS GPRO Web Interface measures. Each measure's performance score is displayed by a series of five stars, each representing 20 percentage points.
- A TEP member asked if the same display methodology will be used for the future reporting of measures in late 2014.
    - Yes.
- The same TEP member then asked if CMS has received any feedback on the usability of these measure displays, or if any questions or issues have surfaced since public reporting began.
    - In response, yes. CMS has conducted both usability and concept testing prior to public reporting. These are the measures and display option that resonated well with customers. The star-percent display option was also tested on other measure types, and consistently users were able to interpret them clearly. The addition of stars has been helpful in streamlining information; users are not getting lost trying to overanalyze numbers as they did when just a percentage display was tested.
- Another TEP member asked if partial stars were tested in addition to partially-filled stars.
    - No, partial stars were not considered.

4.2 Physician Compare as per the 2013 Public Reporting Plan

- Whereas the 2012 rule specified that 2012 PQRS GPRO data would be publicly reported in early 2014, the 2013 rule specifies that 2013 data will be publicly reporting in late 2014.
- This year we also have ACO measures collected via the web interface.
- A big addition for 2013 data is the inclusion of CAHPS measures, available for both PQRS GPROs and ACOs. These measures are for groups of 100 or more eligible professionals reporting via the web interface.

- However, CAHPS for PQRS measures will not be available in time for publication in late 2014, due to a slight delay in data collection. CAHPS for ACOs, on the other hand, will be available for reporting on the ACO pages.
- This year includes a jump in group practices reported from 66 to 169. One TEP member asked to put this into some context and explain the overall trajectory for what is expected in the next couple years. CMS noted that this is a very small percentage of the groups that are a participating in GPRO, much less all of Medicare. In the proposed reporting plans for 2014 and beyond, there will begin a significant expansion to not only the web interface, but also group-level measures selected via registry and electronic health records, as well as individual physician-level measures. This year (2013) is the last of the very limited set of public reporting, and part of a phased approach to public reporting. But yes, as of this year, it consists an incredible small percentage of group practices that have data in 2012 and 2013. What we will discuss today is simply those who reported data on at least 20 patients via the web interface.

4.3 2013 CAHPS Measures

- A TEP member asked for a brief description of how the CAHPS data collection process.
    - A survey development contractor, RAND, had a single vendor collect the CAHPS data for all group practices and ACOs. Data was not collected by the group practices themselves.
- The TEP member then asked if the same 5-stay display methodology will be used for reporting CAHPS scores.
    - As it turns out, the CAHPS data will be presented as percentages for the ACOs, as their web pages are separate from the consumer-driven Physician Compare profile pages. As for the CAHPS PQRS, the plan is to publicly report using the stars plus percentages display.
- Another TEP member asked what the burden of collection was for the individual group practice. His assumption was that the group somehow collects its own data and then uploads it to RAND.
    - For 2013/2014 there is none. Currently the CMS does the sampling and provides the sample to the single, CMS-chosen vendor. Following 2014, the only addition burden will be that groups will select their own CMS-approved vendor, after which CMS will again collect and provide the sample to that vendor.
- The TEP member then asked if the collection method is via mail.
    - The survey administration is a mixed model – it is mailed and then followed up via telephone. Ultimately it consists of a pre-notification letter, two survey mailings, and then up to six telephone calls if the mail is not returned.
- The TEP then asked if CMS has done any preliminary look at differences in results that may exist between people that mail in the data on time versus those who take six phone calls to respond.
    - As of yet, no, there have been no such analyses.

- The TEP member's concern is that there may be a risk of patients feeling like they will lose coverage or be thought less of if they do not reply to the survey.
    - It is clearly outlined in the pre-notification letter that participation in the survey does not affect their Medicare benefits, that they are free to participate or not participate, and that the results are not shared with their provider.
- A final question from the TEP member is how CMS handles patients who cannot read, or are blind, or have some other barrier to communication.
    - CMS has a few things in place. One, although not ideal, is for that patient to assign a proxy permission to answer the survey. The survey can also be administered in a number of languages.

4.4 Current State of Measure Reporting on Physician Compare

- Acumen reviewed where public reporting on Physician Compare stands today. In October 2013, the TEP voted on 11 candidate quality measures using data from the 2012 GPRO web interface. In February 2014, CMS began publicly reporting the results of five of those measures on the Physician Compare website.
- The following measures are currently reported: DM High Blood Pressure Control, DM Hemoglobin A1c Control (<8%), DM Daily Aspirin or Antiplatelet Use for those with IVD, DM Tobacco Non-Use, and CAD ACE-I/ARB Therapy.
- The following measures were not selected for public reporting: CAD Lipid Control, DM LDL-C Control, DM Hemoglobin A1c Control (>9%), CAD Antiplatelet Therapy, DM Eye Exam, and DM Foot Exam.
- The next step is to review the newer 2013 data for the measures currently reported, so the analyses will be very similar to those of last year.
- A TEP member pointed out that in the 2015 proposed rule, eye exam, foot exam, and HbA1c >9% are all set to be included, whereas HbA1c <8% is to be excluded. This seems like the opposite of what we recommended during the last TEP Panel.
    - The ultimate decisions involve a very large set of stakeholders and a significant amount of input. But as you said, the rule is still in the proposed form so we will see where things end up.

4.5 Overview of Acumen Analyses

- Acumen briefly explains to the TEP their quality assurance analyses (explained in detail in Section 2.1 of this report).
- A TEP member asked how the results of the 2013 skip rate analysis compare to those of 2012?
    - Overall they were similar. Being unable to confirm the diagnostic condition was the most common skip reason in both years.
- Acumen briefly explains to the TEP their descriptive statistics, reliability analyses, and case-mix analyses for each measure (explained in Sections 2.2, 2.3, and 2.4 of this report, respectively).

- Another TEP member asked how the case-mix design and results compare between this year and last year.
    - Again, the design and results were comparable. One model included patient health characteristics, the other also included race, region, and income controls. There was limited movement in the first model, but more movement in the second when controlling for region, race, and income.
- A TEP member asked about how the data are reported, specifically if they will include the quartiles described in the case-mix analyses.
- The data are going to be reported using the star and percentage displays. This year there will no benchmarks or information about relative performance.
- Next Acumen displayed a map showing the geographic areas served by the group practices reporting via the web interface. Most areas of the country are served by only one group practice, so for the most part, at this stage in limited public reporting, beneficiaries will not be making many choices about seeing one practice over another. Displaying relative performance for comparison purposes will be useful in the future, but for now it will not be included in the public display.
- A TEP member asked if Acumen was able to control for obesity as a case-mix risk factor.
    - No, and that is a challenge when working with Medicare claims data. Ideally we could have height and weight recorded for every Medicare enrollee, but we do not. In some cases there are indicators for obesity but unfortunately we cannot assume that everyone without that indicator is not also obese.
- The TEP member then commented to CMS that there is actually a way of recording BMI as a continuous variable using an L code. It is used for the outpatient setting, and may be worthwhile to explore for inpatient purposes.
- Another suggestion from the TEP member is to one day collect information about a beneficiary's level of education, which is a good indicator of people ability to practice compliance.
- Acumen reiterated that only the unadjusted percentages will be presented for these measures. Applying the adjustment methodology for public reporting is not really on the table; the goal was just to determine whether or not the data would present a misleading picture of group practice performance by reporting rates in their unadjusted form.
- A TEP member asked if it will be publicly made clear that results are unadjusted. Something acknowledging that may ameliorate some commentary later.
- A TEP member commented that it may be useful to document the robustness of the case-mix model.
    - The challenge is that these measures were developed without risk adjustments. So some risk factors that we may want to collect are not available since adjustment was not the original intention. In order to get a fully robust model to compare these results to, hypothetically you would need to start again at the data collection stage and find group practices willing to provide data on all

possible risk factors. This is good to document for future endeavors, but would not be feasible after the fact.

4.6 Tobacco Non-Use Measure DM-17

- For this measure Acumen found the largest impact of case-mix adjustment, of which a main driver is geographic region. So it is worth considering to what extent we are measuring the quality of care versus some characteristic of the region where the beneficiary resides.
- A TEP member pointed out that for this measure, group practices in states that have high smoking rates will appear to do poorly on the measure. According to the CDC website, there is more than a two-fold difference in smoking rates across states.
- Acumen mentioned that there is also another PQRS measure that was not included in the 2013 web interface, which first documents a patient's smoking status, and then checks whether or not smokers are receiving a proper cessation intervention. That measure, strictly speaking, is a process measure, whereas DM-17 had a pretty large outcome component.

4.7 Measure Recommendations

- Last year TEP members we asked to rate and comment each candidate measure; this year the TEP can first see if there is consensus moving forward, and if not, individual score cards can be distributed to measures in question.
- A TEP member commented that he has always had concerns about the Tobacco Non-Use measure, and now this data backs up his concerns.
- A TEP member commented that there is value in continuity – as more group practices are added it is worthwhile to have the same measures reported. However he has no qualms with dropping the tobacco measure if there is consensus.
- TEP members agreed on continuing to not report the two cholesterol control measures, which are no longer consistent with clinical guidelines.
- Acumen suggested that it would be useful to circulate the measure scorecard for the tobacco use measure, and everyone agreed.

4.8 Physician Compare Going Forward (2014 Reporting Plan)

- 2014 data is scheduled for public reporting in late 2015 if technically feasible. This marks the beginning of what should be a significant expansion of the public reporting plan.
- Before, we presented PQRS GPRO measures collected via the web interface for just diabetes and coronary artery disease. This will be expanded to all possible web interface measures.
- Additionally, a subset of PQRS GPRO measures collected via registry (16 measures) and EHR (13 measures) will be available.
- For ACOs, the expansion includes all web interface measures, three claims-based measures, and one administrative claims-based measure.

- CAHPS for PQRS and ACO measures will be available if technically feasible.
- 2014 will also mark the first year that there will be publicly reported measures at the individual physician level. These 20 measures are in-line with those currently available through the GPRO web interface.
- Additionally, six cardiovascular prevention measures will be made available for public reporting in support of the Million Hearts Initiative.
- For some measures, if an individual chooses not to report data, this will be disclosed and explained publicly. A TEP member stressed that it would be very important to have strong explanations of why someone may choose not to report a particular measure.

4.9 Physician Compare Going Forward (2015 Reporting Plan)

- As for 2015, measures will be available for public reporting in late 2016. In the proposed rule, not all web interface PQRS GRPO measures are available but all registry and EHR measures are. As for ACOs, it would basically be all available measures, including benchmarks.
- Also proposed is to have CAHPS for PQRS and ACO measures available for groups so two or more eligible professionals and all ACOs who report via a pre-approved CMS survey vendor.
- One thing to mention is that the rule does not intend on publicly reporting every single measure on a physician or group practices' Physician Compare profile page, but rather have them all available in a downloadable database. What actually does get presented on the profile page is up for discussion.
- Only data from measures deemed valid and reliable would be included in this downloadable packet. Although there is no limitations on the way people may use this data, a TEP member stated that it would be nice for the TEP to have some say in determining what data is considered to be high enough quality for inclusion in the download.
- Another TEP member commented that if something in the PQRS universe turns out to have little reliability or validity and is not worth reporting of being downloadable, to should be removed entirely so that it does not keep populating.
- A TEP member asked whether or not CMS plans to risk adjust these measures. Hospital Compare, for instance, is running risk adjusted reporting.
    - Risk adjustment varies by PQRS measure, so that would be outlined in the measure specifications by the measure developer. At this point, most PQRS measures are not risk adjusted, although some are. Since we have to work with what we get from PQRS, it may not be something that Physician Compare can directly add to the measures, so to speak.
- The TEP member then suggested that as PQRS measures are reviewed for endorsement by the NQF, they ought to disclose whether or not they want to consider risk adjustment.
- Composites are available for comment but not something CMS plans on finalizing in this rule. Benchmarking, on the other hand, is something that is being proposed.

- Regarding composites, a TEP member stressed that if Physician Compare presents all PQRS measures, it is in a sense, creating a composite presentation to patients. And as with composites, things are correlated so that one piece may drive the other in a similar direction. So it is import that we are careful not to choose to report things that are so in line with one another that a group may do well in one area because they did well in a driving part of that composite.

4.10 Preference Elicitation

- Another TEP member suggests that as the number of quality measures expand, it may be useful to group measures into composites or summary categories that can be layered in a hierarchical fashion.
    - CMS explained that one limitation to this is that unfortunately we do not have composite benchmarks available to use for 2014 data. This is a problem that would be great to work through as a solution would help the consumers work through the data.
- A TEP member urged for the elicitation of preferences from the consumers. For example, rather than exposing someone to a long list of measures, some of which may not be relevant to that person, you would want to be able to bring the measures that are most relevant to the forefront.
- This suggestion sounds to the TEP like a good work around, so a TEP member asked Westat or Acumen about the feasibility of adding some kind of consumer preference functionality on the Physician Compare site.
    - Nothing like that currently exists.
- To elaborate on how this might work, a TEP member gives an example: measures are grouped by logical, intuitively appealing categories, such as diabetes, cancer screening, and so forth. That way you elicit from the user if they have a particular interest in A, B, C, or D and zeroing in on what is relevant to that user. Another aspect of this is explaining the relationship of performance across measures within a group, for instance if there is a set of measures that are highly related.

4.11 Benchmarking

- CMS is very interested in the TEP's opinion on the proposed benchmarking methodology as well as possible alternatives. The specific methodology is included as an excerpt to the rule in the packet of materials shared earlier. It is in line with the Shared Savings Program methodology currently in use.
- This methodology would create a system in which group practices earn quality points on a sliding scale based on their performance. Performance below the minimum (30^{th} percentile) receives no points; performance above the maximum (90^{th} percentile) receives the maximum number of points. Then the scoring system assigns stars that are presented for each group practice.
- There are concerns about this methodology and one thing to consider is the use of value based modifiers, particular for their use of weighted averages.

4.12 Concept Testing

- A round of concept testing looked at CAHPS for PQRS measures and PQRS measures at the individual physician level, and both were received incredibly well by participants.
- Also looked at open payments data and utilization data, which unlike PQRS measures, are not yet on the official radar for public reporting on Physician Compare. That being said, CMS wants the TEP's input on how to make these date useful to consumers, if possible.
- Open payments data are being collected by CMS, and represent the payment to clinicians from industry. The idea is to promote transparency and openness about who is providing money and how much.
- Patients states that they would not base their healthcare decisions on this data, but appreciated it and were willing to engage with it. What became clear was that the amount and origin of the money mattered.
- A TEP member pointed out that a recent experiment made payment data public, and despite some staggering results that were previously undisclosed, it had little effect on consumer choices. He would expect there to be a long time lag before the publication of this data results in consumer change.
- As for utilization data, these data are currently available on CMS.gov and display how physicians are billing their time. This did not resonate well as raw data with consumers, which is not surprising. The information was often misinterpreted, and quickly it became clear that it would be challenging to make this data meaningful.