

---

**SUPPORTING STATEMENT**

**Part B**

**DRAFT**

Care Coordination Quality Measure for Patients in the Primary Care Setting

**April 2014**

**Agency for Healthcare Research and Quality (AHRQ)**

---

---

# 1TABLE OF CONTENTS

<b>Section</b>	<b>Page</b>
<b>B. Collections of Information Employing Statistical Methods.....</b>	<b>2</b>
1. Potential Respondent Universe and Sampling Methods.....	2
1.1 Sampling Frame Definition.....	3
1.2 Sampling Design.....	6
2. Information Collection Procedures.....	14
3. Methods to Maximize Response Rates.....	15
4. Tests of Procedures.....	16
5. Statistical Consultants.....	16

---

## B. Collections of Information Employing Statistical Methods

### 1. *Potential Respondent Universe and Sampling Methods*

This supporting statement includes information in support of a pilot administration of the Care Coordination Quality Measure for Primary Care (CCQM-PC) – a survey instrument developed according to Consumer Assessments of Healthcare Providers and Systems (CAHPS®) survey design principles. The CCQM-PC is intended to fill a gap in the care coordination measurement space by providing a standardized and detailed measure of the quality of care coordination within primary care practices (PCPs) based on adult patients' reports of their healthcare experiences. To date, the CAHPS® tools have included minimal content on care coordination. This is because they were designed to cover a variety of aspects of care in addition to its coordination. The current project is the third phase of an AHRQ-sponsored and led comprehensive research program focused on the assessment of care coordination.

The goal of the sampling strategy for the pilot survey is not to represent the population of patients who are treated by PCPs or to represent the composition of a primary care practice based network. The goal of our sampling strategy is to obtain a group of respondents that represents the spectrum of care coordination experience from highly coordinated care, to care that is lacking in coordination. Variation in responses to the survey questions is critical to the success of the psychometric analyses and will not be obtained unless respondents do, in fact, differ in their experiences with regard to how coordinated their care has been. To ensure variability in care coordination experience, we will stratify our sample according to patient and practice characteristics. Patients will be stratified according to the intensity of the health care needs for which they seek and receive care. Practices will be stratified according to a four-category classification scheme that considers ownership status, number of practice sites, and whether the practice is solely primary care or part of a multi-specialty group:

- Physician-owned, single-site PCPs
- Physician-owned, multi-site PCPs
- PCPs within an independent multi-specialty group
- PCP units affiliated with an integrated delivery system

A mix of practice sizes (as defined by the number of primary care clinicians in the unit or practice) will be recruited in each category. Maximizing diversity along dimensions such as practice size and ownership status (e.g., physician owned, hospital owned) is important because these factors have been shown in recent studies<sup>1,2</sup> to be correlated with a practice's implementation of process improvement practices that support care coordination and care integration as fostered by the patient-centered medical home model (PCMH). Seeking variation

---

<sup>1</sup> Rittenhouse, D. R., Casalino, L. P., Shortell, S. M., McClellan, S. R., Gillies, R. R., Alexander, J. A., and Drum, M.L. (2011). Small and Medium-size Physician Practices use Few Patient-centered Medical Home Processes. *Health Affairs*, 30 (8), pp. 1575-1584.

<sup>2</sup> Rittenhouse, D.R., Casalino, L.P., Gillies, R.R., Shortell, S.M., and Lau, B. (2008). Measuring the Medical Home Infrastructure in Large Medical Groups. *Health Affairs*, 27 (5), pp. 1246-1258.

---

in both the number of sites (i.e., single-site vs. multi-site) and specialty orientation (i.e., solely primary care or multi-specialty) of included practices also helps ensure that we are including practices that differ in their structural relationships to other entities and health care settings, which may impact the ease or difficulty of coordinating care. In short, we aim to recruit a sufficiently diverse set of participating PCPs, and sufficiently diverse sample of respondents within those practices to enable testing of the CCQM-PC across a range of primary care settings and a spectrum of patients with varying care coordination needs.

## **1.1 Sampling Frame Definition**

The potential respondent universe is operationalized via the sampling frame definition. There are two steps to this activity: 1) define eligible units, and 2) define eligible patients in those units. As noted in Supporting Statement A, the measure development process for the CCQM-PC has been, and will continue to be consistent, with CAHPS® principles (<http://www.cahps.ahrq.gov/about.htm>). In CAHPS surveys, it is essential to define the “accountable unit,” which is the entity from which patients are sampled and for which survey scores will be calculated and reported. However, the design of this pilot survey will be more complex than that used in CAHPS studies because the CCQM-PC is being created for applications in addition to accountability. The CCQM-PC’s primary purpose will be to further our understanding of the variables which act to facilitate or hinder coordinated care so it must be appropriate for use in research conducted within PCPs as well as across PCPs.

### **1.1.1 Eligible Units**

While this survey is being designed for research purposes, we incorporate into the pilot study an “accountable unit” sampling strategy to support the psychometric analyses required to produce scores that ultimately can be used in accountability applications (e.g., public reporting of PCP quality, quality-based reimbursement). That is, we seek to evaluate questions and composite scores according to their ability to reliably detect real differences in care coordination across accountable units. The accountable unit in the case of the CCQM-PC is the *primary care practice (PCP)*. A primary care practice serves as the patient’s first point of entry into the health care system and as the continuing focal point for all needed health care services.

We plan to recruit 30 PCPs nationally, with the assistance of third party entities, for participation in the pilot. Practices will either be identified by professional associations or research networks with which the practices are affiliated (e.g., professional associations of primary care providers and medical group practices, practice-based research networks [PBRN]), or will self-identify their interest in participating in the pilot by responding to recruitment letters distributed via these entities and other listservs/newsletters targeting the primary care provider audience (e.g., AHRQ’s PBRN, Improving Primary Care GovDelivery, and Primary Care Practice Facilitation listservs). We anticipate reaching out with recruitment materials to entities such as the American Academy of Family Physicians (AAFP) and its National Research Network (NRN), the American Medical Association (AMA), the American College of Physicians (ACP), the Society of General Internal Medicine (SGIM), the American Academy of Physician Assistants (AAPA), the American Academy of Physician Assistants in Obstetrics and Gynecology (AAPOG), the American Congress of Obstetricians and Gynecologists (ACOG), the American Association of Nurse Practitioners (AANP), the American Medical Group Association (AMGA), and the Medical Group Management Association (MGMA).

For the psychometric analysis, it is important to select PCPs which represent variation in care coordination. We do not need to select a nationally representative sample of PCPs but we do need to do our best to ensure that the PCPs that participate in the field test are likely to differ in the degree of coordination of the care that they provide. Practices will be stratified according to four categories of primary care practice configurations described above:

- Physician-owned, single-site PCPs
- Physician-owned, multi-site PCPs
- PCPs within an independent multi-specialty group
- PCP units affiliated with an integrated delivery system

We believe we can use information publicly available on the internet (e.g., from scanning practices’ websites or the websites of their affiliated groups or delivery systems) to appropriately classify practices that respond with interest to our recruitment materials. If needed, however, we will develop a brief screener to use with interested practices to ensure they are properly classified. While there will be no other formal stratification variables in the sample design, we will seek variability along other dimensions, as practical, such as:

- Geographic region (e.g., Northeast, South, West, Midwest)
- Urban versus rural catchment area
- Practice size (as defined by number of primary care clinicians in the practice)
- Primary care specialty (i.e., general practitioner, family practice, internal medicine, OB/GYN)

The 2012 AMA Physician Benchmark Survey found that 60% of physicians surveyed worked in practices wholly owned by practice physicians (that is, are individual practices).<sup>3</sup> We will use this as a rough estimate of the distribution of primary care practice configurations and will attempt to recruit a mix of PCPs that reflect this distribution, targeting 18 PCPs in the two wholly physician-owned categories (i.e., 9 practices in each of these two categories), and 12 PCPs in the other two categories (i.e., 6 practices each), as shown in Table 1.

**Table 1. Site Recruitment Table for Primary Care Practices (PCPs)**

<b>Numbers of Practices by Type</b>				
Physician-owned, single-site PCPs	Physician-owned, multi-site PCPs	PCPs within an independent multi-specialty group	PCP units affiliated with an integrated delivery system	<b>TOTAL</b>
<b>9</b>	<b>9</b>	<b>6</b>	<b>6</b>	<b>30</b>

<sup>3</sup> Kane, C. K. & Emmons, D. W. (2013). New Data on Physician Practice Arrangements: Private Practice Remains Strong Despite Shift Toward Hospital Employment. *American Medical Association: Policy Research Perspectives*.

---

### **1.1.2 Eligible Patients**

The target population will include any patients at least 18 years old who are not living in an institutionalized setting. Consistent with criteria established in our formative research, we anticipate limiting the sample to patients who have seen a doctor on at least two occasions in the past year. We will exclude from the sampling frame patients who have not seen a provider in the last 4 months. Our inclusion and exclusion criteria will be informed by preliminary examination of the distribution of patients across these variables and variables related to complexity of health status (e.g., number of visits in past year, diagnoses, referrals to specialists). The complexity of health issues, and associated needs for care coordination, may be thought of on a continuum. The patients included in the sample will represent a range of complexity of health issues to ensure that we represent the range of care coordination needs among our respondents. The range of health issues and health status complexities is defined using three categories: complex health issues requiring ongoing follow up, acute health issues requiring episodic follow up, and routine care needs.

- **Complex health issues** are defined as health conditions that require ongoing care by a combination of health services providers and which require frequent visits, for example, every week to every two months. This includes for example, a person with one chronic disease, such as chronic kidney disease, who visits multiple providers including a primary care physician, nephrologist, nutritionist, and frequent labs. Another example of a person considered to have complex health issues is an individual with multiple health conditions who requires different care providers for different conditions, such as someone with heart disease and diabetes.
- **Acute health issues or injuries** are defined as health conditions that require immediate care by health services provider(s) as limited to a single condition and to a limited number of follow-up or related visits for a duration of time such as would be required for surgical repair of an injury or treatment for pneumonia.
- **Routine care needs** are defined as health conditions that require care by a single health services provider and is limited to a one or two visits, such as for a cold.

Since care coordination needs are highest for those patients with the greatest complexity of health issues, and arguably limited for many patients in our current healthcare system who only have routine care needs, we are faced with a challenge. On the one hand, to fully test the survey it is desirable to represent a range of complexity of health issues to assess how the survey operates for patients across the spectrum of care coordination needs and experiences. Given the project's conceptualization that care coordination needs fall on a continuum, we would want to ensure that the sample drawn from that frame represents the full range of care coordination needs among patients, including those with only routine care needs. Yet we concurrently acknowledge the possibility that some questions may not resonate well with patients with low care coordination needs. The data collection could be inefficient if resources are expended for data collection that results in significant missing data. Fortunately, our experience with the formative research activities generated a good degree of confidence that participants will have something to say about care coordination regardless of their particular health status and needs.

Central to obtaining a sufficient mix of patients with varying care coordination needs is the identification, through proxy indicators, of the complexity of patient health conditions. To

---

eliminate undue burden on the practices to identify eligible patients and provide information needed for segmentation by likely care coordination needs, there is potential value in trading off external validity and generalizability for feasibility of identifying the range of patients' health complexities with low provider burden. This would be achieved by requiring that all participating practices have an electronic health record (EHR) system through which they could run a report of patients meeting the study eligibility criteria defined above (i.e., at least 18 years old, have seen a doctor at least twice in past 12 months and at least once in the past 4 months). Further, the report would include categorization of eligible patients as having "complex health issues", "acute health issues", or "routine health issues", as defined above and based on indicators in the EHR related to complexity of health status (e.g., number of visits in past year, diagnoses, referrals to specialists). By assessing the reports of eligible patients in relation to the sampling goals in table 2, we will determine if a targeted sampling approach will be necessary.

## 1.2 Sampling Design

The psychometric and quality scoring requirements for this effort inform the sampling design and sample size estimates provided in table 2. These requirements are described below:

1. Psychometric methods will be used to test the reliability and validity of selected assessment items, composites, and global ratings. Having a sample sufficient to conduct these analyses is the primary goal of the field test. Standard psychometric practice is to obtain a *minimum* of 10 *complete responses* for each item that will be used in the psychometric analysis (this includes substantive questions, but not screeners or questions designed to determine survey eligibility). This recommendation is grounded in sound measurement theory<sup>4</sup> and practice in the statistical analysis of multivariate data (including factor analyses).<sup>5</sup>
2. Quality scoring involves assessing current performance relative to other entities using the composite scores and overall ratings from the survey and serves as a first step in quality improvement efforts (i.e., knowing where you currently rank). Testing the feasibility of quality scoring is a secondary goal of the pilot test. To make fair comparisons, the scores will be adjusted for patient characteristics such as age, education, overall health rating, and gender (case mix adjusted).
  - a. To assess performance, we will calculate PCP-level estimates of assessment scores (scores for items, composites, and global ratings) and rank-order the PCPs based on their performance.
  - b. For accountability purposes, scores need to be adjusted for case mix in order to facilitate fair comparisons across PCPs<sup>6</sup>. We will use regression models to select a set of case-mix adjusters from a pool of exogenous predictors (e.g., age,

---

<sup>4</sup> Nunnally JC & Bernstein IH (1994). *Psychometric theory (3<sup>rd</sup> Edition)*. New York: McGraw-Hill, Inc.

<sup>5</sup> Stevens J (1992). *Applied multivariate statistics for the social sciences (2<sup>nd</sup> Edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

<sup>6</sup> For research applications in which we seek to understand how individual patient differences might affect the amount and kind of care coordination, we would not want CCQM-PC scores adjusted for case mix so we would use a scoring algorithm that does not include the case mix adjustment.

**Table 2. Sampling goals for patients across PCP recruitment sites\***

<b>Patient N by Type of Care Needs and PCP Configuration</b>												
Physician-owned, single-site PCPs			Physician-owned, multi-site PCPs			PCPs within an independent multi-specialty group			PCP units affiliated with an integrated delivery system			<b>TOTAL</b>
Routine Care (RC)	Acute Care (AC)	Complex Care (CXC)	Routine Care (RC)	Acute Care (AC)	Complex Care (CXC)	Routine Care (RC)	Acute Care (AC)	Complex Care (CXC)	Routine Care (RC)	Acute Care (AC)	Complex Care (CXC)	
844	844	1,687	844	844	1,687	563	563	1,124	563	563	1,124	
<b>3,375 (i.e., 375X9)</b>			<b>3,375 (i.e., 375X9)</b>			<b>2,250 (i.e., 375X6)</b>			<b>2,250 (i.e., 375X6)</b>			<b>11,250</b>

\* Totals by Care Needs are taken from calculations on section 1.2.2.2 of this document, targeting 25% of the sample as patients with routine care needs, 25% of the sample as patients with acute care needs, and 50% of the sample as patients with complex care needs.

---

education, overall health rating, gender, and any other variables that make sense conceptually) and estimate the predictive power and heterogeneity of the adjusters (using regression and variance component models). We will then estimate the impact of case-mix adjustment on the overall rankings of PCPs (based on the star rating assignment produced by the standard, publicly available program for analyzing CAHPS data, and using Kendall's Tau-b to estimate the percentage of PCPs that switch rankings as the result of case mix adjustment).

### **1.2.1 Response Rate Targets**

The response rate for all surveys will be computed as:

$$\frac{\text{Number of completed returned questionnaires}}{\text{Total number of respondents selected} - (\text{deceased} + \text{ineligible})}$$

#### ***Numerator Inclusions:***

- Completed questionnaires. A questionnaire is considered complete if responses are available for 50% or more of a selected list of key survey items (the items that all respondents are eligible to answer).

#### ***Denominator Inclusions:***

- Refusals. The respondent refused in writing or by phone to participate.
- Nonresponse. The respondent is presumed to be eligible but did not complete the survey for some reason (never responded, was unavailable at the time of the survey, was ill or incapable, etc.).
- Bad addresses/phone numbers. In either case, the respondent (or parent or guardian) is presumed to be eligible but was never located.

#### ***Denominator Exclusions:***

- Deceased. In some cases, a household or family member may inform you of the death of the respondent.
- Ineligible – did not meet the eligibility requirements specified in section 1.1.2 above.

Total required sample size is a function of the desired number of completes divided by the estimated overall response rate calculated as described above. Historically, response rates for CAHPS surveys span a fairly wide range. The 2012 Commercial CAHPS response rates are approximately 30% and Medicaid CAHPS response rates are approximately 27%. In the recent field test of the CAHPS survey for Cancer Care, AIR obtained a response rate of 48%; for the Dental CAHPS field test and early implementation, response rates ranged from as low as 40% to as high as 70% in some population segments. Based on AIR's experience with field tests of several different CAHPS® and CAHPS-like instruments, a 40% overall response rate will be targeted.

### **1.2.2 Sample Size Estimates for Initial Sample**

For existing CAHPS surveys where respondents may interact with a number of different individuals, such as with a health plan or hospital, CAHPS recommends obtaining completed

---

questionnaires from at least 300 respondents per reporting entity during implementation (i.e., not a field test). The requirements for a field test, however, are primarily to enable psychometric testing, and secondarily, to enable quality improvement scoring, as described in the beginning of this section. We have proposed that 150 completed surveys per PCP will be sufficient to meet these goals. Assuming a 40% response rate, we would sample 375 patients per PCP, for a total sample size of 11,250.

The details regarding the relationship between these sample sizes and these goals are presented below.

#### 1.2.2.1 Precision and Sample Size

In this section we describe the level of precision a given sample size will be able to achieve for making PCP-level estimates to justify the proposed 150 completes per PCP.

##### 1.2.2.1.1 Psychometric Analysis

This analysis task includes both assessing the measurement properties of the instrument, as well as testing the equivalence of the measurement properties across the mode of administration. The generalizability of the results from the psychometric analysis is obtained by attempting to capture the full range of care coordination experiences, and thus potential response patterns, in the PCP patient population.

The CCQM-PC is provided in [Appendix A](#) to Supporting Statement A. It currently includes 79 assessment items, some of which will drop out of the survey as a result of cognitive testing and further expert review. Assuming we have 79 assessment items in the CCQM-PC pilot survey version, this translates into needing a *minimum* of 790 (i.e., 79 X 10) completed surveys across all pilot sites, assuming that each completed survey contains a non-missing response for each substantive item. However, given that *some* substantive items will be legitimately skipped by respondents to whom the subject matter of the item does not apply, this number will need to be larger. In addition, some completed surveys may still have some degree of item non-response (when a respondent skips an item that he/she should have answered). Moreover, experiences are likely to vary substantially across the 30 PCPs, and so we would want to make sure that each PCP is represented. Thus, we estimate that a minimum number of completes of at least 1,000 would be needed to assure that we meet this goal – assuming an overall response rate of 40%, we would need a sample of at least 2,500 to get 1,000 completes. Our proposed number of completes is 4,500 (i.e., 40% of 11,250) at the national level, which is more than enough to meet this goal.

##### 1.2.2.1.2 Analyses to Compare Scores Across PCPs

A priori estimates of precision (e.g. power calculations) are important to evaluate the CCQM-PC scores for quality reporting and accountability purposes. PCP-level estimates rely on the precision of point estimates for the survey measures (composites, overall ratings, and single item measures). Precision is defined in terms of the margin of error, which is also known as the “half-width” of the confidence interval (typically a 95% confidence interval). The margin of error for a 95% confidence interval (CI) is equal to the standard error of the point estimate multiplied by 1.96 (the margin of error for a 68% CI would be equal to one standard error; the margin of error for a 99% CI would be equal to 2.58 standard errors). Thus, the margin of error is used to construct the CI around the point estimate and describes the range within which we can be confident the true score lies.

---

We estimated confidence interval precision using PROC POWER in SAS. This approach is analogous to a traditional power analysis, with the margin of error (“CI Half-Width” in SAS) taking the place of *effect size* and the half-width probability (“Prob (Width)” in SAS) taking the place of *power*. Using estimates of a range of variances and standard errors observed from some existing CAHPS surveys (e.g., the field test of the draft CAHPS survey for Cancer Care, the NCQA National Distribution of 2009 Adult Medicaid CAHPS Plan-Level Results, and the 2013 Medicare Part C Report Card results) as inputs, we estimated sample sizes associated with different levels of precision. Note that we have decided *a priori* on a target number of completed surveys. Thus, this analysis is designed to illustrate the level of precision that can be obtained under several scenarios with samples designed to obtain 150 completed surveys<sup>7</sup>.

To anchor the margins of error and variance estimates (expressed as standard deviations) to a meaningful CAHPS scale, we have transformed observed scores for the three different types of measures from the existing CAHPS results mentioned above into a 100-pt scale. This allows us to express the inputs to the power analysis in a scale that is comparable across different types of measures.

To express measures on a 100-pt scale, composites and single item measures are transformed from their original 3-pt or 4-pt scales using a simple linear transformation based on expressing the observed score as a percentage of the distance from the floor to the ceiling of a scale:

$$100 \times (\text{observed score} - \text{scale floor}) / \text{range}$$

For a 4-pt CAHPS scale (1=never, 2=sometimes, 3=usually, 4=always) with a mean of 3.5, the transformation would look like this, for example:

$$\frac{3.5 - 1}{3} \times 100 = 83.3$$

Over half of the assessment questions (51%) in the CCQM-PC include the 4-pt CAHPS response scale; an additional 46% of the assessment questions include the 4-point CAHPS response scale plus a tailored response option that respondents may choose to indicate that the item is not applicable to their experiences of care.

As an example of the proposed approach, consider estimating a sample size assuming a goal of having a half-width probability (power) of 0.80, an alpha of 0.05, and a half-width (margin of error) no greater than 3 points. With these parameters, we would be estimating the number of completed surveys needed to give us an 80% chance of obtaining a 95% CI with +/- 3 point margin of error.

To put this example in more concrete terms, if we observed a score of 83.3 from a sample size calculated using the above inputs, there would be a 95% chance that the true score in the population would be between 80.3 and 86.3, and only a 5% chance that it would be outside of that range.

<sup>7</sup> We used a conditional probability approach (that is, the probability of achieving the desired precision is calculated conditionally given that the true mean is captured by the interval), which is a more conservative approach than the unconditional probability approach.

Table 3 is meant to show the impact of sample size on precision and thus to indicate what level of precision we might be able to obtain with the sample sizes we have proposed. Observed standard deviations from several of the CAHPS sources consulted ranged from approximately 2 to 28 points for measures on a 100-point scale. Observed standard errors ranged from around 0.30 to 3.2, which represent margins of error of approximately 0.60 to 6.3 points (on a 100-pt scale) for a 95% CI.

**Table 3. Precision Associated with Different Sample Sizes and Variances**

Number of Completed Surveys Needed per PCP for 80% Half-Width Probability and a 95% Confidence Interval for a CAHPS-like Measure on a 100-Point Scale						
With a Margin of Error of +/-	And a Standard Deviation of:					
	5	10	15	20	25	30
1	110	410	902	1585	2461	3530
2	32	110	236	410	632	902
3	17	53	110	189	288	410
4	11	32	65	110	167	236
5	8	22	44	73	110	155
6	7	17	32	53	73	110

For shaded cells, the first column of table 3 shows different obtainable margins of error associated with a range of completed survey counts for different variances, and thus shows the range in the level of precision that could be obtained for PCP-level estimates (assuming standard deviations no greater than 30), where the maximum number of completes will be 150 per PCP.

As an illustration, assuming a standard deviation of 25 for an observed mean of 82, we would expect that, in a series of 100 independent random samples drawn from the same population of size sufficient to yield at least 110 completed surveys, the true population score would fall between 77 and 87 ( $82 \pm 5$ ) in 95 of those samples; with 150 completed surveys, the margin of error would be smaller (slightly more than  $\pm 4$ ). For smaller variances, the precision gets better with smaller samples (e.g., with a sample size of 32 and a standard deviation of 5 points, the margin of error would be  $\pm 2$  points).

#### 1.2.2.1.3 Ranking PCPs on Care Coordination Performance

As described above, a secondary objective of the pilot test is to provide reports to the participating sites which rank-order PCPs based on their performance scores (on items, composites, and global ratings). We can consider providing comparative rankings for subsets of practices relative to the average of all practices in a similar grouping of PCPs (e.g., each of the four PCP configurations we are using to segment the recruited practices; all PCPs scoring similarly on the practice-level measure of processes of care [i.e., the Medical Home Index].) For each subset, if a global F-test indicates that scores vary across PCPs in the group, the ranking is then done using a t-test of the difference between each PCP and the overall mean of all PCPs in that group.<sup>8</sup>

<sup>8</sup> For a discussion of reliability and its relationship to sample size, see “Fielding the CAHPS Clinician & Group Surveys: Sampling Guidelines and Protocols (Document No. 1033)”: [https://www.cahps.ahrq.gov/clinician\\_group/](https://www.cahps.ahrq.gov/clinician_group/)

Using variances observed from previous CAHPS field tests, AIR conducted a power analysis based on a two sample *t*-test comparing the mean score on a composite (on a 100-pt scale) from one entity to the pooled mean on that composite from all entities, using a range of variances. This *t*-test is the basis for assigning the rankings to entities when reporting results. The power analysis assumes a balanced design (same number sampled from every entity) and equal variances (single entity variance = to pooled variance).<sup>9</sup>

Tables 4 and 5 illustrate the relationship between sample size, variances and effect sizes to demonstrate the power with which we could detect differences among the 6 or 9 (respectively) PCPs within each practice type noted in Table 1.

**Table 4. Relationship between Sample Size, Variances, and Effect Sizes for PCP Rankings: 6 PCPs**

Number of Completes per PCP (Assuming 6 PCPs)	Variance of 15		Variance of 25	
	Mean Diff	ES	Mean Diff	ES
20	10.4	0.69	17.3	0.69
50	6.5	0.43	10.9	0.44
100	4.6	0.31	7.7	0.31
150	3.8	0.25	6.3	0.25
200	3.3	0.22	5.4	0.22
300	2.7	0.18	4.4	0.18
500	2.1	0.14	3.4	0.14

ES = effect size; Mean Diff = difference in means between a single PCP and the mean of all PCPs

As shown in Table 4, when comparing 6 PCPs, and assuming 150 completes per PCP and a variance of 15 points, we would have 80% power (with an alpha of 0.05) to detect a difference of 3.8 points between a single PCP and the overall mean of PCP scores (e.g., 86.2 versus 90). With a wider variance of 25 points, we could detect a difference of just over 6 points (e.g., 66 versus 72.3). The effect sizes associated with these differences (0.25) are relatively small; that is, 150 completes per PCP with 6 PCPs of a given practice type will be sufficient to detect small effect sizes when comparing the mean of a single PCP to the overall mean of 6 PCPs.

Moderate effect sizes ( $ES > 0.40$ ) could be detected with as few as 50 completes per PCP.

**Table 5. Relationship between Sample Size, Variances, and Effect Sizes for PCP Rankings: 9 PCPs**

Number of Completes per PCP (Assuming 9 PCPs)	Variance of 15		Variance of 25	
	Mean Diff	ES	Mean Diff	ES
20	10.0	0.67	16.7	0.67
50	6.3	0.42	10.5	0.42
100	4.5	0.30	7.4	0.30

<sup>9</sup> In practice, this test is conducted using a Satterthwaite unpooled *t*-test on the mean difference, which accounts for unequal variances across entities. We reproduced the analyses presented in table 4 using this test and specifying different variances for the single entity variance and the pooled variance. When the single entity variance is smaller than the pooled variance, the sample size required to detect mean differences of a particular magnitude tends to decrease. When the single entity variance is larger than the pooled variance, the sample size required tends to increase. However, the sample size requirements are still overwhelmingly determined by upper limit of either variance, regardless of how unequal they are. The impact on the estimated number of completes associated with the mean differences and variances presented in the exhibit was negligible.

Number of Completes per PCP (Assuming 9 PCPs)	Variance of 15		Variance of 25	
	Mean Diff	ES	Mean Diff	ES
150	3.6	0.24	6.1	0.24
200	3.2	0.21	5.3	0.21
300	2.6	0.17	4.3	0.17
500	2.0	0.13	3.3	0.13

ES = effect size; Mean Diff = difference in means between a single PCP and the mean of all PCPs

As shown in Table 5, when comparing 9 PCPs, and assuming 150 completes per PCP and a variance of 15 points, we would have 80% power (with an alpha of 0.05) to detect a difference of 3.6 points between a single PCP and the overall mean of PCP scores (e.g., 86.4 versus 90). With a wider variance of 25 points, we could detect a difference of just over 6 points (e.g., 66 versus 72.1). The effect sizes associated with these differences (0.24) are relatively small; that is, 150 completes per PCP with 9 PCPs of a given practice type will be sufficient to detect small effect sizes when comparing the mean of a single PCP to the overall mean of 9 PCPs.

Moderate effect sizes (ES > 0.40) could be detected with as few as 50 completes per PCP.

#### 1.2.2.2 Drawing the Sample

We cannot know *a priori* what the distribution of health condition complexity will be among patients in the 30 selected PCPs. One study that asked 40 primary care physicians from 12 Massachusetts General Hospital-affiliated practices and community health centers to review a list of their own patients found that these physicians designated about 25% of their patients as complex based on their own experiences with those patients.<sup>10</sup> A recent AHRQ white paper that explored two groups of patients with especially complex health and social support needs (the frail elderly and working age adults with disabilities) cited estimates from the Census Bureau indicating that about 10 percent of adults ages 18 to 64 and 37 percent of adults age 65 and older have a disability.<sup>11</sup> This group would represent only some of the overall population of patients with complex health needs (the subset of persons with disabilities in addition to complex care needs), and thus may be considered an underestimate of all patients for whom the care coordination measure would be most relevant.

Based on the above prevalence estimates, using simple random sampling would probably result in no more than 25% of the sample being comprised of those with complex care needs. We cannot say how the remaining 75% would be distributed across the categories of acute and routine care needs, but with complex care patients providing the most opportunity for care coordination (and having the highest need), we will adjust our sampling strategy to include a higher percentage of patients with complex care needs. To deal with the problems described, we will request that participating practices include in the sampling frame data sufficient information that will allow us to group patients into the three complexity categories. These variables would include diagnosis codes (to identify patients with chronic diseases), frequency and duration of

<sup>10</sup> <http://www.massgeneral.org/about/pressrelease.aspx?id=1424>

<sup>11</sup> Rich E, Lipson D, Libersky J, Parchman M. Coordinating Care for Adults With Complex Care Needs in the Patient-Centered Medical Home: Challenges and Solutions. White Paper (Prepared by Mathematica Policy Research under Contract No. HHS290200900019I/HHS29032005T). AHRQ Publication No. 12-0010-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2012.

---

visits, and the number of different providers seen by the patient. These variables can be used to group patients into the three categories. We can then stratify the sampling frame based on the three categories and oversample in the complex care and acute care categories so that the former comprises 50% of the sample and the latter 25%. The remaining 25% would include those in the routine care category.

A stratified random sample of 375 patients will be drawn from each of the 30 PCPs, which results in a total sample of 11,250 – within each PCP, approximately 187 patients will come from the complex care category, and approximately 94 each will come from the acute care category and the routine care category. Assuming a response rate of 40%, this approach should yield 150 completed surveys per PCP (75 complex, 25 acute, 25 routine), for a total of 4,500 completed surveys (2,250 complex, 1,125 acute, and 1,125 routine). Oversampling complex-care patients and drawing an equal number of patients from each PCP will result in a disproportionate sample and thus unequal sampling weights across strata. This does not matter for the psychometric analysis, but could have a bearing on some comparative analyses. AIR will calculate the appropriate sampling weights where applicable. If a given PCP has 375 patients or fewer who meet the eligibility criteria, we will sample all patients.

## **2. Information Collection Procedures**

The survey will be conducted by mail with phone follow-up for non-respondents.

Survey operations will follow standard CAHPS practice:

- Mail the questionnaire package, including a personalized letter introducing the study and explaining the respondent's rights as a research participant. Include a postage-paid envelope to encourage participation.
- Send a postcard reminder to nonrespondents 10 days after sending the questionnaire.
- Send a second questionnaire with a reminder letter to those still not responding thirty days after the first mailing.
- Begin follow-up by telephone with nonrespondents three weeks after sending the second questionnaire. Interviewers will attempt to locate respondents who have not responded to the mailed survey.
- Telephone numbers for sample respondents will be verified prior to calling.
- A maximum of 9 attempts will be made by phone.

The letters and postcards will include a toll-free number for respondents to call if they have questions about the survey. The firm responsible for fielding the survey will establish a helpdesk to start operating at the first mailing and that will remain open until close of fieldwork. Incoming calls will be answered live during business hours and a recording machine will capture after hours calls. The after-hours calls will be returned next business day.

### **Mode**

As noted in the procedural steps above, surveys may be taken by either paper-and-pencil from the mailed versions of the survey, or by phone interview in phone followup with nonresponders.

### **Recruitment and Consent Materials**

---

Recruitment memos to the professional associations and research networks through which we will outreach to prospective practices, and from these entities to affiliated practices are provided in **Appendix C** to Supporting Statement A. The draft invitation letter to the professional associations and research networks may be tailored to the needs of each organization. Each organization agreeing to support our recruitment efforts will send the invitation to its eligible members on its own letterhead with a strong endorsement of the project. This could be accomplished through use of the organization or network's listserv. Practice managers will be asked to contact AIR directly if they wish to participate. We will also welcome recommendations from the professional organizations for specific practices that might be inclined to participate. If there are no objections from the participating membership organizations, we will include every organization's logo on each letter to demonstrate the broad base of support for the survey and the value of participating in the pilot test. Practices that respond positively will be entered into a practice recruitment database.

### **3. *Methods to Maximize Response Rates***

Every effort will be made to maximize the response rate, while retaining the voluntary nature of the effort. Below are several options recommended by CAHPS for maximizing response rates that may be employed:

- We will set up a toll-free number and publish it in all correspondence with respondents. We will assign a trained project staff member to respond to questions on that line and maintain a log of these calls and review them periodically.
- A persuasive advance letter (included in **Appendix C** to Supporting Statement A) will be sent to spark the interest of practices. The letter to professional organizations and research networks will be printed on AHRQ letterhead. The recruitment memo to practices, supported by the participating professional associations, will be provided on that organization's letterhead with an official logo and include an official signature of the organization. It will include the AHRQ logo as well. Letters to the professional organizations and to any specific practices identified by these organizations will be personalized with the name and address of the intended recipient.
- The envelope to survey respondents will include an official logo of the PCP and include a return address to the survey firm; envelopes will be marked "forwarding and address correction" in order to update records for respondents who have moved and to increase the likelihood that the survey packet will reach the intended respondent.
- Reminder cards, a second survey mailing, and phone followup with nonresponders will be utilized to increase response rate.
- For the telephone interviews:
  - Interviewers will be trained
  - Interviewers will read questions exactly as worded so that all respondents are answering the same question.
  - When a respondent fails to give a complete or adequate answer, interviewer probes will be nondirective.

- 
- Interviewers will maintain a neutral and professional relationship with respondents. The primary goal of the interaction from the respondent’s point of view should be to provide accurate information. The less interviewers communicate about their personal characteristics and, in particular, their personal preferences, the more standardized the interview experience becomes across all interviewers.
  - Interviewers will record only answers that the respondents themselves choose. The instrument is designed to minimize decisions that interviewers might need to make about how to categorize answers.
  - The survey vendor for CCQM-PC will be required to use CATI.

#### **4. Tests of Procedures**

The survey development team conducted in-depth interviews and focus groups with patients and caregivers and held meetings with stakeholders to ascertain dimensions of service and care important to patients and their families, as well as language used by patients and families for care coordination concepts. Using those data, as well as an environmental scan of existing surveys, and the existing Care Coordination Measures Atlas (<http://www.ahrq.gov/professionals/systems/long-term-care/resources/coordination/atlas/care-coordination-measures-atlas.pdf>), the survey development team drafted iterative versions of survey. Cognitive testing occurred in two rounds in December 2013 and February 2014.

#### **5. Statistical Consultants**

This sampling and statistical plan was prepared and reviewed by staff of AHRQ and by AIR. The primary statistical design was provided by Chris Evensen, MS, of AIR at (919) 918-2310; San Keller, PhD, of AIR at (919) 918-2309; and Susan Heil, PhD, of the AIR at (301) 592-2227.