

SUPPORTING STATEMENT B

MEDICARE HEALTH OUTCOMES SURVEY

B. Collection of Information Employing Statistical Methods

B.1 Respondent universe and sample

1. Description of sample selection and universe

A. Potential respondent universe.

The target universe is current Medicare beneficiaries enrolled in Medicare managed care health plans (i.e., Medicare Advantage (MA) contracts), residing in the 50 states, the Virgin Islands, and Puerto Rico. Aged and disabled Medicare beneficiaries enrolled in an MA contract with more than 500 members are eligible. Institutionalized members are eligible.

B. Sample Sizes.

(i) Target sample sizes.

The targeted response rate for the baseline survey is 60 percent. The targeted response rates for the follow-up survey and HOS-M survey are 75 percent. The initial sample size and targeted baseline and follow-up response rates are designed to achieve an analytic sample size with adequate statistical power to detect significant variation between MA health plans on physical and mental health outcomes. A post hoc analysis exploring the power and minimum detectable effect associated with the HOS focused on “net effects,” or the mean change in PCS or MCS score between baseline and follow-up for enrollees in each contract relative to the mean change of enrollees in all other MA contracts. The results indicated the HOS has adequate power to detect “small” net effects at the contract level, namely 0.2 of a standard deviation between means. Depending on the measure (PCS or MCS) and cohort, the power was greater than 0.93 to 0.99 for half of the contracts and greater than 0.81 to 0.96 for 75% of the contracts. The minimum detectable effects for PCS and MCS score, at 80% power and $\alpha=.05$, ranged from 1.24 to 1.62. As expected, increasing the effective sample size reduced the minimum detectable effect in all cohorts.

(ii) Baseline Survey sampling.

Due to variations in population size of MA contracts, a minimum of two sampling approaches are required.

- 1) In MAOs with Medicare populations greater than or equal to 1,201, 1,200 members are randomly selected for the Baseline Survey.

- 2) In MAOs with populations of 500 to 1,200 members, all members who meet the eligible population criteria are included in the sample.

MAOs with populations less than 500 members are exempt from HOS reporting.

The current (2017) HOS survey sampling protocol includes a third sampling approach. Currently, members of MAOs with populations greater than 3,000 are excluded from the baseline survey sample frame in the current year if they completed and returned a survey the previous year. This was intended to reduce response burden but has created a sample bias by excluding beneficiaries who should legitimately be part of the survey sample frame. It also selectively excludes only beneficiaries in the largest MAOs, who are least likely to be affected by response burden. The probability that any beneficiary in a large MAO will be selected for the HOS survey sample is at most 1,200/3,000 (40%), and declines to 4.9% for an MAO with 25,000 members and <1.0% for the largest MAOs. Because the current protocol creates a bias and is ineffective at accomplishing its goal of reducing response burden, CMS proposes eliminating the large contract exclusion from the HOS sample protocol beginning in 2018.

(iii) Follow-Up Survey sampling.

Eligible members include all Baseline respondents surveyed two years prior for whom a valid physical component summary score or mental component summary score were calculated and who remain alive and enrolled in the same MA contract or the surviving MA contract in the case of an MAO consolidation, merger, or novation.

CMS examined how the small contracts' HOS scores compare to those of larger contracts as the enrollees of small contracts complete the survey more often. Of the 384 MA contracts required to report 2017 Follow-Up HOS, the latest year for which Star Ratings data are available, 27 contracts were identified as small contracts with fewer than 1,200 enrollees. Of these, 5 contracts had fewer than 500 enrollees and were therefore no longer required to report HOS baseline. The remaining 22 contracts are of special interest to CMS because for MA contracts with between 500 and 1,200 enrollees, all eligible members are included in the annual sample. CMS notes this sampling disparity in the HOS prenotification letter and the letters that accompany HOS questionnaires, which state, "You have been randomly selected to receive the "Medicare Health Outcomes Survey. (For some health plans with few enrollees, all members with Medicare are asked to complete it.)" Survey fatigue is a concern for beneficiaries enrolled in small contracts because they are likely to receive, at minimum, both the MA CAHPS and HOS surveys every year and be designated a "double-duty respondent" for HOS by simultaneously serving in the Baseline and Follow-Up samples. Some small MA contracts use outreach such as newsletters and on hold messages in the call centers to encourage their membership to participate. Despite the potential for survey fatigue and a disproportionate number of double-duty respondents, response rates for contracts with fewer than 1,200 enrollees are not significantly different than response rates for contracts with more than 1,200 enrollees. In fact, Cohort 17 Follow-

Up HOS demonstrates that the response rate for the smallest contracts, those with fewer than 1,200 enrollees, was 2 percentage points higher than the response rates for the largest contracts, those with more than 100,000 enrollees.

**Number of Contracts and Response Rates,
by Enrollment Size, at Cohort 17 Follow-Up**

Enrollment Size	# of Contracts	Response Rate (%)
<=1,200	21	66
1,201 – 3,000	24	66
3,001 – 5,000	29	67
5,001 – 10,000	53	64
10,001 – 15,000	40	67
15,001 – 25,000	40	68
25,001 – 50,000	63	65
50,001 – 100,000	52	67
>100,000	39	64

Source: HOS Round 19 Data Set.

Small MA contracts also perform well on HOS measures in the Medicare Part C Star Ratings. On average, contracts with fewer than 1,200 enrollees performed better than contracts with more than 1,200 enrollees on four HOS measures: Improving or Maintaining Physical Health; Improving or Maintaining Mental Health; Monitoring Physical Activity; and Improving Bladder Control. Although small contracts scored slightly lower on Reducing the Risk of Falling, the Part C Summary Score, and Overall Stars, the difference was not significantly different.

2018 Star Ratings Measures	Small Contracts	Large Contracts
	Avg. Stars	Avg. Stars
C04: Improving or Maintaining Physical Health	3.3	2.9
C05: Improving or Maintaining Mental Health	4.1	3.7
C06: Monitoring Physical Activity	3.3	3.0
C18: Reducing the Risk of Falling	2.3	2.5
C19: Improving Bladder Control	3.8	3.2
Summary Ratings		
Part C Summary	3.4	3.6
Part D Summary	3.9	3.9
Overall	3.6	3.7

C. Response rates.

A total survey response rate is calculated for each sample that is the total number of complete surveys divided by all eligible members of the sample. Eligible members of the sample include the entire random sample minus members assigned a survey disposition code of ineligible. In other words, invalid beneficiaries are removed from the denominator in the response rate calculations.

The total survey response rate is calculated as follows:

$$\frac{\text{Number of Complete Surveys}}{\text{Entire random sample} - [\text{Ineligible for any of the following reasons: Deceased} + \text{Not enrolled in MAO} + \text{Language barrier} + \text{Removed from sample} + \text{Duplicate/Listed twice in the sample frame} + \text{Bad address and nonworking and unlisted phone number or person unknown at the dialed phone number}]}$$

Survey response rates from 1998 through 2015 (the latest completed year of data collection) are illustrated in Attachment B. Over the course of time, the refinement of the Medicare HOS measure has resulted in changes to the definition of a “completed survey.” In general, for *Cohorts 1-3 Baseline*, a completed survey is defined as a survey that has at least 80% of the questions answered. However, beginning with *Cohort 1 Follow Up* and *Cohort 4 Baseline*, a completed survey is defined as a survey with calculable physical or mental health summary scores.

Given that the definition of a completed survey has evolved over time, this table should not be utilized for response rate comparisons across the cohorts. (For the Medicare HOS-Modified, a completed survey is defined as a survey that, at a minimum, had responses to all six activities of daily living measures.)

For the baseline cohorts, the reporting units represent the individual contracts sampled for the survey. However for the follow up cohorts, the reporting units have been adjusted to accommodate selected contract consolidations and service area reductions at the time of performance measurement reporting (which typically occurs in the year subsequent to the collection of the follow up data).

B.2 Data collection procedures

This section describes the procedures used for the national survey. It includes a summary of the questionnaire content, and a general discussion of the case-mix and risk-adjustment methodology. Finally, there is a discussion of rules for allowing proxy response.

A. Years 2017 through 2019 data collection procedures.

- (i) Mail phase.

The mail component of the survey uses standardized questionnaires, cover letters, and postcards. Members may receive up to four survey-related items by mail. A Pre-Notification Letter is mailed to all members of the survey sample one week prior to the mailing of the First Questionnaire. Two weeks later, a Reminder Postcard is mailed to all members who have not returned the First Questionnaire. Four weeks after the Reminder Postcard is sent, a Replacement Questionnaire is mailed to all members who have not returned the First Questionnaire.

Because of the challenge in corresponding with a diverse population, it is necessary to support material in multiple languages. The three most common languages encountered with the Medicare population are English, Spanish, and Chinese. Both the standard and modified instruments (Attachments B and C), plus associated supporting correspondence (Attachment D) are available in each of the three identified languages.

To ensure comparability of results, the survey vendors must follow strict adherence to the established survey protocol, specifications manual, and the quality assurance plan.

First and Replacement Questionnaires must include a bar or alphanumeric coded unique identifier for tracking purposes (i.e., tracking returned surveys and determining which respondents are eligible for follow-up). The Cover Letter for First Questionnaire contains English and Spanish text. The Letter for First Questionnaire is double-sided. One side of the letter contains English text and the other side contains Spanish text. The Spanish text invites Spanish-speaking members to request a Spanish version of the questionnaire by calling the survey vendor's toll-free customer support number.

Completed questionnaires can be manually key-entered into the computer or optically scanned. To ensure quality for key-entered data, two separate data entry specialists must independently key answers for each questionnaire. A comparison of the separate entries identifies data entry errors that need adjudication by a supervisor.

(ii) Telephone phase.

The telephone component uses a standardized CATI script and specific design specification. The survey vendor is responsible for programming the scripts and specifications into its existing CATI software. To ensure the comparability of survey results, the survey vendor cannot change the wording of survey questions, the response categories or order of the questions.

Following the mail portion of the process, survey vendors use the scanned barcodes to identify members who did not respond to the mail survey and members who returned a blank or incomplete mail questionnaire (a questionnaire with less than 80 percent of questions complete or one or more Activities of Daily Living [ADL] items unanswered). These members are eligible for telephone interviews.

The survey vendor attempts to contact non-respondents by telephone so that at least six telephone calls are attempted at different times of day, on different days of the week, and in different weeks.

The survey vendor establishes training programs for all personnel involved in the telephone phase of the process. It establishes quality control procedures and monitors staff performance to ensure the integrity of the telephone interviewing process. The survey vendor monitors 10 percent of CATI interviews to evaluate the quality of interviewing and provides feedback and additional training as necessary.

B. Interview content.

There are two survey instruments used by the Medicare HOS. While both are similar, there are some distinct differences. The two instruments are: Medicare HOS 3.0 and Medicare HOS-Modified (HOS-M).

While the Medicare HOS 3.0 is the primary instrument used for data collection for both baseline and follow-up cohorts since 2015, considerations were made for how the Medicare HOS-M is administered to vulnerable Medicare beneficiaries at greatest risk for poor health outcomes. These beneficiaries are enrolled in PACE organizations. The main goal of the Medicare HOS-M is to assess the frailty of the population in these organizations in order to adjust annual Medicare payments.

PACE is a capitated benefit authorized by the Balanced Budget Act of 1997 (BBA) that features a comprehensive service delivery system and integrated Medicare and Medicaid financing. The PACE model was developed to address the needs of long-term care clients, providers, and payers. For most participants, the comprehensive service package permits them to continue living at home while receiving services rather than be institutionalized. Capitated financing allows providers to deliver all services participants need, rather than be limited to those reimbursable under the Medicare and Medicaid fee-for-service systems.

The Medicare HOS-M survey is a modified version of the Medicare HOS and remains unchanged. The instrument assesses the physical and mental health functioning of the Program members to generate information for payment adjustment. It includes 12 physical and mental health status questions, Activities of Daily Living (ADLs), one question about memory loss interfering with daily activities, and one question about urinary incontinence. If the participant received assistance completing the questionnaire, the respondent is asked why a proxy was needed, how the proxy assisted the participant, and the staff position of the proxy.

The Medicare HOS 3.0 is comprised of three major components: 1) the VR-12, as the core component, which feeds the Medicare Part C Star Ratings outcomes measures (HOS questions 1-9); 2) questions to gather information for case-mix and risk-adjustment (HOS questions 10, 20-35, 40, 58-59, 61-62, 65-66, 68); 3) questions to

gather data for the HEDIS measures (HOS questions 42-52); 4) questions required under Section 4302 of the Affordable Care Act (ACA) which established data collection standards for race, ethnicity, sex, primary language, and disability status for all national population health surveys (HOS questions 15-18, 57-60); and 5) additional items (HOS questions 11-14, 19, 36-39, 41, 53-56, 63-64) to gather comparative data for MA health plan quality improvement, new measure development, and to measure health disparities, health behaviors, and health-related quality of life.

(i) The VR-12.

The VR-12 consists of items categorized into seven (7) health domains (general health perceptions, physical functioning, role limitations due to physical and emotional problems, bodily pain, energy/fatigue levels, social functioning, and mental health) and eight (8) concepts of health (Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE), and Mental Health (MH)). Items are scored and summarized into a physical component summary (PCS) and a mental component summary (MCS).

The taxonomy underlying the construction of the VR-12 scales (concepts) and summary measures has three levels: 1) fourteen items; 2) eight scales that aggregate one or two items each; 3) two summary measures that aggregate the eight scales.

Scales covered in the VR-12 are:

- Physical Functioning (PF) consists of two questions that ask respondents to indicate the extent to which their health limits their physical activities.
- Role—Physical (RP) consists of two questions that assess whether respondents' physical health limits them in the kind of work or other usual activities they perform both in terms of time and performance.
- Bodily Pain (BP) consists of one question that determines the extent to which pain interferes with the respondent's normal activities.
- General Health (GH) consists of one question that asks respondents to rate their current health status overall.
- Vitality (VT) consists of one question that asks respondents to rate their well-being by indicating how frequently they experience energy.
- Social Functioning (SF) consists of one question that asks respondents to indicate limitations in social functioning due specifically to health.
- Role—Emotional (RE) consists of two questions that assess whether emotional problems have caused respondents to accomplish less in their work or other usual activities in terms of time and performance.
- Mental Health (MH) consists of two questions that ask respondents how frequently they experience feelings representing the four major mental health dimensions: anxiety; depression; loss of behavioral/emotional control; and psychological well-being.

In addition, a two-item measure of Change in Health asks respondents to rate their general physical health and emotional problems now compared to their health one year ago.

- (ii) Questions to gather information for case-mix and risk-adjustment.

Case-mix and risk-adjustment are essential for meaningful and valid contract-to-contract comparison of health outcomes. By including variables to control for differences in demographics and socioeconomic characteristics, chronic medical conditions, and Medicare HOS study design variables, regression techniques can be used to adjust the scales and summary measures.

- a) Demographics include questions on gender, race, education, marital status, home ownership, and annual household income.
 - b) Six (6) Activities of Daily Living (ADLs) are included to determine self-reported difficulty with performance of daily tasks. ADLs include bathing, dressing, eating, getting in or out of chairs, walking, and using the toilet.
 - c) Fifteen chronic medical conditions are included in the questionnaire. These conditions are: hypertension or high blood pressure; angina pectoris or coronary artery disease; congestive heart failure; myocardial infarction or heart attack; other heart conditions, such as heart valve defects or arrhythmias; stroke; emphysema, asthma, or Chronic Obstructive Pulmonary Disease (COPD); inflammatory bowel disease, including Crohn's disease and ulcerative colitis; arthritis of the hip or knee; arthritis of the hand or wrist; osteoporosis; sciatica; diabetes, hyperglycemia, or glycosuria; depression; any cancer (other than skin cancer) and treatment of cancer.
 - d) General health of member as compared to other people within the same age.
 - e) Study design variables include who completed the survey, CMS region, and the survey vendor.
- (iii) Questions to gather data for the HEDIS measures (HOS questions 42-52).
 - (iv) Questions required under Section 4302 of the Affordable Care Act (ACA) which established data collection standards for race, ethnicity, sex, primary language, and disability status for all national population health surveys (HOS questions 15-18, 57-60).
 - (v) Questions to gather comparative data for MA health plan quality improvement, new measure development, and to measure health disparities, health behaviors, and health-related quality of life.

- a) Six Activities of Daily Living (ADLs) are included to determine self-reported difficulty with performance of daily tasks. ADLs include bathing, dressing, eating, getting in or out of chairs, walking, and using the toilet.
- b) Three Instrumental Activities of Daily Living (IADLs) assessing difficulty preparing meals, managing money, and taking medication as prescribed.
- c) Three Healthy Days questions, which encompass the number of days in the past thirty days that physical health was not good, mental health was not good, and activities were limited due to poor physical or mental health.
- d) One cognitive functioning question to assess memory problems interfering with daily activities.
- e) Three pain questions with a global assessment of pain and pain interference: pain interfering with day-to-day activities, socializing and rating of pain on average.
- f) HEDIS Effectiveness of Care includes three measures on the management of urinary incontinence, two measures on physical activity in older adults, two measures on fall risk management in older adults, and one on osteoporosis testing in older women. One of the fall risk management questions was updated in 2017 HOS (see Attachment A). This revision was approved by NCQA's Committee on Performance Measurement in 2016.
- g) Two depression screening questions include levels of depression that encompass: little interest or pleasure in doing things; or feeling down, depressed or hopeless; and one question to capture a history of depression.
- h) One question to determine level of smoking activity.
- i) Height and weight for the calculation of body mass index (BMI).
- j) Section 4302 of the Affordable Care Act standardized ethnicity, primary language and disability status questions.
- k) Two living arrangement questions to evaluate type of home and with whom the member resides.
- l) Two questions about sleep duration and quality, which are facets of health-related quality of life with well-evidenced relationships to health outcomes.

C. Adjustment methodologies.

- (i) Case-mix adjustment.

The analytic sample for the Medicare HOS is limited to those seniors (age 65 and over) with calculable physical component summary and mental component summary scores. Regression techniques are used to case mix adjust these measures for each beneficiary.

The longitudinal outcomes for the HOS analysis are based on risk-adjusted mortality rates, changes in physical health as measures by the PCS score, and changes in mental

health as measure by the MCS score for each participating MA contract. For reporting purposes, death and PCS outcomes are combined into one overall measure of change in physical health. Thus, there are two primary outcomes: 1) Alive and PCS Better+Same (vs. PCS Worse or Death) and 2) MCS Better+Same (vs. MCS Worse). For the Medicare Part C Star Ratings, the primary outcomes are reported as the percentage of respondents within a MA contract who are “Improving or Maintaining Physical Health” and the percentage in the MA contract who are “Improving or Maintaining Mental Health” over the two-year period, after adjustment for case-mix.

A series of 12 different multivariate logistic regression models (six death models, three PCS models, and three MCS models) are used to case-mix adjust HOS outcomes, and to calculate expected outcomes for each beneficiary. For each of the three types of models (death, PCS, and MCS), the first model (Model A) is used for those beneficiaries with complete data and the other alternative models are used for those respondents with different patterns of missing data for the model outcome. To address the issue of missing data, a series of cascading logistic regression models was developed.

See (Attachment F) for a detailed breakdown of covariates used in the Case Mix Adjustment of PCS and MCS measures for the 2015 *Cohort 18 Baseline* Report. This methodology will remain in practice, although minor refinements could be explored to maintain the accuracy of the measures.

(ii) Risk Adjustment.

In addition to the VR-12 Health Survey, the Medicare HOS questionnaire collects information for purposes of a standardized contract-to-contract risk adjustment. Additional items include variables for morbid conditions, activities of daily living, and socio-demographic characteristics. Risk adjustment accounts for patient-associated factors before comparing outcomes across different health plans or populations. Therefore, risk adjustment enables a fair comparison between health plans.

For each respondent, a change in functional health score will be estimated by subtracting the first (baseline) questionnaire score from the second (follow up) score. Taking into account an expected decline in health and risk-adjustment factors, change in physical and mental health status can be: better, the same as, or worse than expected.

Because outcome is defined as a change score, each respondent serves as his or her own ‘control.’ In addition plan-to-plan risk adjustments are based on morbid conditions at baseline, income, household size, social support, education, race, and gender. Results are aggregated across respondents for each health care plan.

D. Proxy rules.

While sampled members are encouraged to respond directly to the mail or telephone survey, not all elderly or disabled respondents are able to do so. In such cases, proxy responses are acceptable. The survey instrument instructs members who are unable to complete the survey to have a family member or other proxy complete the survey for them.

If a proxy completed the baseline survey, every effort is made to have the same proxy or, if possible, the sampled member, complete the follow-up survey. Having the same proxy complete both questionnaires minimizes bias. If the same proxy or the sampled member is unable to complete the follow-up survey, another proxy may be used.

B.3 Methods to Maximize Response Rates

Medicare HOS is sampling a heterogeneous population that presents a unique challenge for maximizing response rates. The household survey will be approaching two groups – aged and disabled Medicare beneficiaries – who have characteristics that often lead to refusals on surveys. Increasing age, poor health, or poor health of a family member are prevalent reasons for refusal. On the other hand, older persons are the least mobile segment of the population and thus less likely to be lost due to failure to locate. Little is known about the disabled population in terms of survey response rates. It is anticipated that they will be most similar to the oldest old population because of their ill health.

Because this is a longitudinal survey it is essential that we maximize the response rates. In order to do so, CMS employs an extensive annual outreach effort. This includes the notification of government entities (CMS regional offices and hotline, carriers and fiscal intermediaries, and Social Security Offices). These efforts are undertaken to increase the likelihood that respondents would answer the Medicare HOS questions and remain in the survey panel by: 1) informing authoritative sources to whom survey participants are likely to turn if they suspect the legitimacy of the Medicare HOS; 2) giving interviewers resources to which they can refer to reassure respondents of the legitimacy/importance of the survey; and 3) generally making information about the Medicare HOS available through health plans, senior centers, other networks to which survey participants are likely to belong, and through the CMS website.

In addition to the outreach efforts, the following efforts remain in place to maintain a sense of validity and relevance among the survey participants.

- An advance letter is sent to potential respondents from CMS.
- Interviewer training emphasizes the difficulties in communicating with the older population and ways to overcome these difficulties.
- Proxy respondents are sought for survey participants unable to participate for themselves.
- Mail non-respondents are re-contacted by telephone.
- A toll-free number and email address are available to answer respondent questions.

- A website and email address are available at CMS to answer respondent questions.
- A shortened version of the survey instrument is administered to the frailest portion of the sample.
- “Take my name off your list” refusals are flagged for exclusion from all future Medicare HOS activity.

The following disclosure statement is included with the survey instrument:

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0938-0701. The time required to complete this information collection is estimated to average 20 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: CMS, 7500 Security Boulevard, C1-25-05, Baltimore, Maryland 21244-1850.

The OMB control number “0938-0701” appears on every page of the survey instrument.

B.4 Test of Procedures or Methods

At this time there are no plans to conduct field testing of the currently established procedures or methods. From time to time various parts of the questionnaire are modified or augmented to reflect changes to the Medicare Advantage program, capture information on emerging areas of interest, reduce unnecessary burden or to improve the quality of the data. If field testing becomes desirable in the future, it will be submitted for approval separately or in combination with the next main collection of information.

B.5 Statistical and Questionnaire Design Consultants

CMS receives ongoing input from statisticians in developing, designing, conducting, and analyzing the information collected from this survey. Statistical expertise will continue to be available from NCQA, Boston University, HSAG, and RAND.

Ongoing statistical consultation is provided by:

Lewis Kazis, ScD
Boston University
715 Albany Street (T-3W)
Boston, MA 02118
90401

Doug Ritenour, MPH
Health Services Advisory Group
313 E. Camelback Rd., Suite 100
Phoenix, AZ 85015

Marc Elliott, PhD
RAND Health
1776 Main Street
Santa Monica, CA

NCQA, of Washington, DC manages data collection for the Medicare HOS under contract to CMS.

HSAG of Phoenix, AZ conducts data cleaning, analysis and data dissemination for the Medicare HOS under contract to CMS.