

# 2016 | METHODOLOGY REPORT



Centers for Medicare & Medicaid Services (CMS)  
Office of Enterprise Data and Analytics (OEDA)

November 16, 2018

## TABLE OF CONTENTS

<b>1. Introduction to MCBS .....</b>	<b>1</b>
1.1 Purpose and Goals .....	1
1.2 Survey Overview and History .....	1
1.3 Key Data Products and Analyses.....	3
<b>2. Changes Unique to 2016.....</b>	<b>5</b>
2.1 Sampling .....	5
2.2 Data Collection.....	5
2.3 Questionnaires .....	5
2.4 Data Processing .....	6
<b>3. Sample Design for the MCBS 2016 Panel.....</b>	<b>7</b>
3.1 Overview of MCBS Sample Design .....	7
3.2 Selection of MCBS PSUs .....	11
3.3 Selection of MCBS SSUs.....	11
3.4 Selection of Beneficiaries for the 2016 MCBS Panel .....	12
Sample Size Determination .....	14
2016 Sampling Frame.....	19
Sample Selection for the 2016 Panel.....	21
Sampling Results.....	24
Coverage Analysis of the 2016 Sampling Frame.....	26
3.5 Continuing Sample (2013-2015 Panels).....	28
3.6 Fielded Sample Sizes by Panel and Round .....	29
<b>4. Instrument and Materials Design .....</b>	<b>30</b>
4.1 Community Questionnaire Content .....	31
Interview Type.....	31
Community Questionnaire Flow .....	32
Core Section Content.....	35
Topical Section Content .....	43
Changes to the Community Questionnaire for 2016 .....	44
4.2 Facility Instrument Content .....	46
Interview Type.....	46
Facility Screener.....	47

Facility Instrument Flow .....	47
Core Section Content.....	49
Changes to the Facility Questionnaire for 2016.....	53
4.3 CAPI and Case Management System Programming and Testing .....	53
Community Questionnaire.....	54
Facility Screener and Instrument.....	54
Case Management System .....	54
4.4 Letters and Other Respondent Materials.....	55
<b>5. Interviewer Recruitment and Training .....</b>	<b>56</b>
5.1 Interviewer Recruitment and Staffing.....	56
5.2 Interviewer Training Programs for 2016.....	56
<b>6. Data Collection.....</b>	<b>56</b>
6.1. Clearance .....	57
OMB Approval.....	57
IRB Approval .....	57
6.2. Data Collection Process and Procedures .....	57
Data Collection Schedule and Timeline.....	57
Sample Releases and Preloads .....	58
Beneficiary Eligibility for MCBS Survey .....	59
MCBS Data Collection Protocols.....	61
Crossover Definitions and Procedures .....	63
Proxy Interviews and Assistants .....	64
Interviewing Languages.....	65
Questionnaire Breakoffs.....	65
6.3 Data Collection Results .....	66
6.4 Data Collection and Quality Control.....	66
<b>7. Data Processing and Data Delivery.....</b>	<b>68</b>
7.1 Data Processing Overview.....	68
Process Description .....	68
7.2 Preload Editing and File Production.....	70
Community and Facility Preload Timeline .....	71
Community and Facility Preload Process Description .....	71
7.3 MCBS 2016 Survey Data File .....	72
File Eligibility Criteria .....	72
File Contents.....	72

7.4 MCBS 2016 Cost Supplement File .....	73
Reference Period.....	73
Insurance Timeline.....	74
Event Cost Consolidation .....	74
Prescription Medicine File.....	74
Facility Stay File .....	75
<b>8. Weighting and Imputation.....</b>	<b>76</b>
8.1 Overview .....	76
8.2 MCBS Weighting Procedures .....	76
Overview .....	76
Process .....	76
2016 Initial Weighting .....	77
2016 Survey File Weights .....	82
2016 Cost Supplement Weights.....	87
2016 Topical Module Weights.....	89
8.3 MCBS Imputation Processes .....	91
Overview .....	91
Income and Asset Imputation .....	92
Prescription Medicine and Non Prescription Medicine Imputation .....	93
<b>9. Response Rates and Nonresponse.....</b>	<b>98</b>
9.1 Response Rates.....	98
2016 Cost Supplement File Response Rates .....	98
2016 Survey File Response Rates .....	100
9.2 Nonresponse Bias Analysis .....	103
<b>10. Using MCBS Data Files .....</b>	<b>104</b>
10.1 MCBS Data User's Guide .....	104
10.2 MCBS Public Use Data File.....	104
10.3 MCBS Limited Data Sets.....	104
<b>11. Glossary .....</b>	<b>105</b>
<b>Appendix A: Technical notes.....</b>	<b>110</b>

## TABLE OF EXHIBITS

<b>Exhibit 1.2.1:</b> MCBS Participation Timeline .....	2
<b>Exhibit 1.3.1:</b> 2016 Contents of Data Releases .....	4
<b>Exhibit 3.1.1:</b> 2011-2016 MCBS Rotating Panel Design.....	8
<b>Exhibit 3.1.2:</b> 2016 MCBS Sampling Strata .....	10
<b>Exhibit 3.4.1:</b> Assumed Rates (in Percent) Used in Determining Sample Sizes for the MCBS 2016 Panel, by Age Group .....	16
<b>Exhibit 3.4.2:</b> 2016 Fall Planning Sample Sizes and Resulting Projected Survey File and Cost Supplement File Estimates, by Age Group, Including Recent Enrollees and Hispanic Oversample .....	18
<b>Exhibit 3.4.4:</b> Number of Beneficiaries in 2016 Enrollment Data Subsample Extracts (Combined) and 2016 Sampling Frame, by Stratum .....	20
<b>Exhibit 3.4.5:</b> 2016 MCBS Panel, Final Sampling Fractions by Stratum .....	23
<b>Exhibit 3.4.6:</b> 2016 MCBS Panel, Number of Beneficiaries Selected by Age Group.....	24
<b>Exhibit 3.4.7:</b> 2016 MCBS Panel, Number of Beneficiaries Selected by Stratum.....	25
<b>Exhibit 3.4.8:</b> 2016 MCBS Panel, Number of Current-Year Enrollees Selected by Age Group .....	25
<b>Exhibit 3.4.9:</b> Number of Beneficiaries in 2016 Enrollment Data Subsample and Estimated 2016 MCBS Population, by Stratum.....	27
<b>Exhibit 3.6.1:</b> 2016 Fielded Sample Sizes by Round for Each Panel .....	29
<b>Exhibit 4.1:</b> MCBS Questionnaire Overview.....	30
<b>Exhibit 4.1.1:</b> Community Questionnaire Interview Types .....	32
<b>Exhibit 4.1.2:</b> 2016 MCBS Community Questionnaire Flow for Baseline Interview .....	33
<b>Exhibit 4.1.3:</b> 2016 MCBS Community Questionnaire Flow for Continuing Interview.....	34
<b>Exhibit 4.1.4:</b> 2016 MCBS Community Core Sections by Data File and Administration Schedule .....	36
<b>Exhibit 4.1.5:</b> Utilization and Cost Section Flow .....	41
<b>Exhibit 4.1.6:</b> 2016 MCBS Community Topical Sections by Data File and Administration Schedule .....	43
<b>Exhibit 4.2.1:</b> Facility Instrument Interview Types .....	47
<b>Exhibit 4.2.2:</b> 2016 MCBS Facility Instrument Flow for Baseline Interview .....	48

<b>Exhibit 4.2.3:</b> 2016 MCBS Facility Instrument Flow for Continuing and Crossover Interviews....	49
<b>Exhibit 4.2.4:</b> Facility Core Sections by Data File and Administration Schedule.....	50
<b>Exhibit 6.2.2:</b> MCBS Contacting Guidelines by Interview Round .....	59
<b>Exhibit 6.2.3:</b> Interviewer-Reported Abstraction Rates by Facility Instrument Section .....	63
<b>Exhibit 6.3.1:</b> 2016 Completed Interviews by Component .....	66
<b>Exhibit 7.1.1:</b> Data Review and Editing Process .....	69
<b>Exhibit 8.2.1:</b> 2016 MCBS Data Files Summary of Weights .....	78
<b>Exhibit 8.2.2:</b> Control Totals for Ever Enrolled Weight Raking, Dimension 1: Age Group.....	86
<b>Exhibit 8.2.3:</b> Control Totals for Ever Enrolled Weight Raking, Dimension 2: Enrollment Year...	87
<b>Exhibit 8.2.4:</b> 2016 Data Year Topical Module Survey Weights Datasets and Contents .....	89
<b>Exhibit 8.3.1:</b> Payers and Payment Amounts Missing Together, Total Charge Known .....	94
<b>Exhibit 8.3.2:</b> Payers and Payment Amounts Missing Differentially, Total Charge Known .....	94
<b>Exhibit 8.3.3:</b> Total Charge Unknown.....	95
<b>Exhibit 9.1.1:</b> 2016 MCBS Annual Cost Supplement File Unconditional Response Rates .....	100
<b>Exhibit 9.1.2:</b> 2016 MCBS Annual Cost Supplement File Conditional Response Rates.....	100
<b>Exhibit 9.1.3:</b> 2016 MCBS Annual Survey File Unconditional Response Rates for Ever Enrolled Beneficiaries.....	102
<b>Exhibit 9.1.4:</b> 2016 MCBS Annual Survey File Unconditional Response Rates for Continuously Enrolled Beneficiaries.....	102
<b>Exhibit 9.1.5:</b> 2016 MCBS Annual Survey File Conditional Response Rates for Ever Enrolled Beneficiaries.....	103
<b>Exhibit 9.1.6:</b> 2016 MCBS Annual Survey File Conditional Response Rates for Continuously Enrolled Beneficiaries.....	103

# 1. INTRODUCTION TO MCBS

## 1.1 Purpose and Goals

Medicare is the nation's health insurance program for persons 65 years and older and for persons younger than 65 years who have a qualifying disability. The Medicare Current Beneficiary Survey (MCBS) is a representative national sample of the Medicare population sponsored by the Centers for Medicare & Medicaid Services (CMS). The MCBS is designed to aid CMS in administering, monitoring, and evaluating the Medicare program. A leading source of information on Medicare and its impact on beneficiaries, the MCBS provides important information on beneficiaries that is not available in CMS administrative data and plays an essential role in monitoring and evaluating beneficiary health status and health care policy.

The MCBS is a continuous, in-person, multi-purpose longitudinal survey covering a representative national sample of the Medicare population. It covers the population of Medicare beneficiaries aged 65 and over and beneficiaries aged 64 and below with disabilities, residing in the United States and Puerto Rico. CMS contracts with NORC at the University of Chicago (NORC) to conduct the MCBS. Fieldwork for the first round of data collection began in September 1991; since then, the survey has continued to collect and provide essential data on the costs, use, and health care status of Medicare beneficiaries. Recently celebrating its 25<sup>th</sup> anniversary of continuous data collection, the MCBS has completed more than one million interviews provided by thousands of respondents.

The MCBS primarily focuses on economic and beneficiary topics including health care use and health care access barriers, health care expenditures, and factors that affect health care utilization. As a part of this focus, the MCBS collects a variety of information about the beneficiary, including demographic characteristics, health status and functioning, access to care, insurance coverage and out of pocket expenses, financial resources, and potential family support. The MCBS collects this information in three data collection periods, or rounds, per year. Over the years, data from the MCBS have been used to inform many advancements, including the creation of new benefits, such as Medicare's Part D prescription drug benefit.

## 1.2 Survey Overview and History

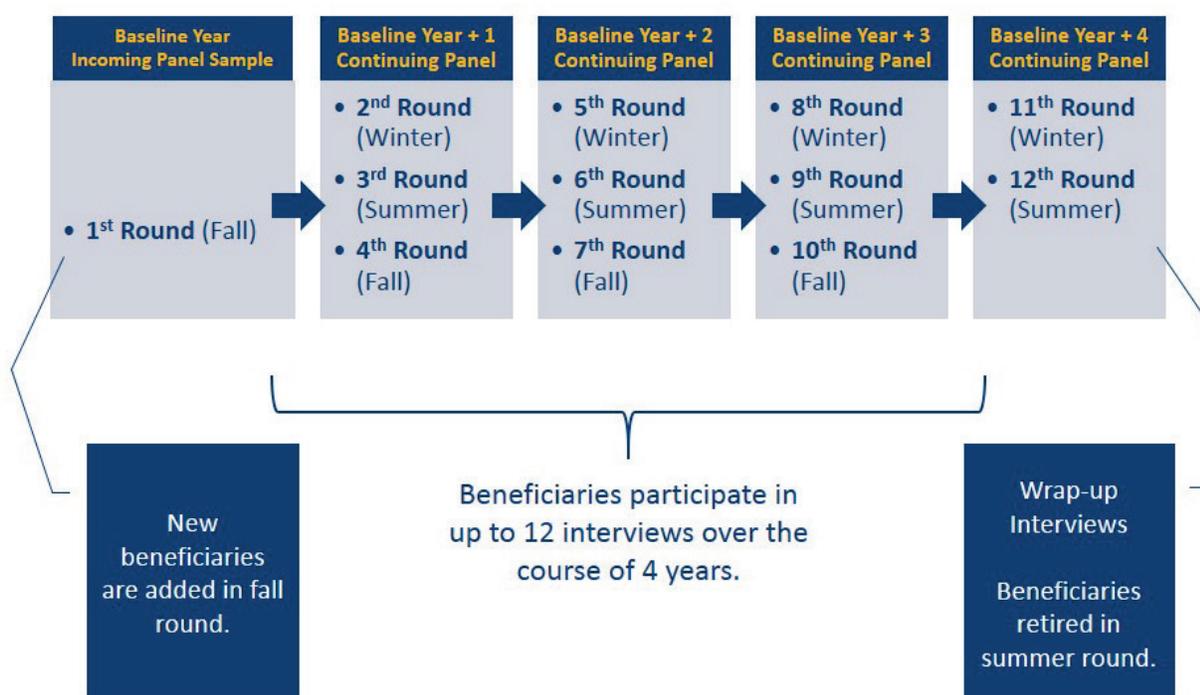
In its initial design, the MCBS was to serve as a traditional longitudinal survey of the Medicare population, with no predetermined limit to the duration of participation. However, beginning in 1994, participation of beneficiaries in the MCBS was limited to no more than four years.

Initial interviews of newly-selected respondents take place once per year in the fall data collection period (i.e., round). Often the fall round begins early (i.e., late July or early August) to allow more time to conduct outreach and collect information from the new survey respondents who are selected to participate in the MCBS. That is, the early start of the fall round overlaps with the final weeks of data collection for the summer round. These small overlap periods that occur as one round ends and another begins are acceptable design features of the survey. For example, the fall

round usually extends into early January to allow for the completion of interviews that may have been postponed due to the holiday period.

Subsequent rounds, which occur every four months, involve re-interviewing of the same respondent (or appropriate proxy respondents) until they have completed four years of participation (up to 12 interviews in total). Interviews are conducted regardless of whether the respondent resides at home or in a long-term care facility, using a questionnaire version appropriate to the setting. Exhibit 1.2.1 depicts the timeline of participation for beneficiaries selected to be in the MCBS sample.

**Exhibit 1.2.1:** MCBS Participation Timeline



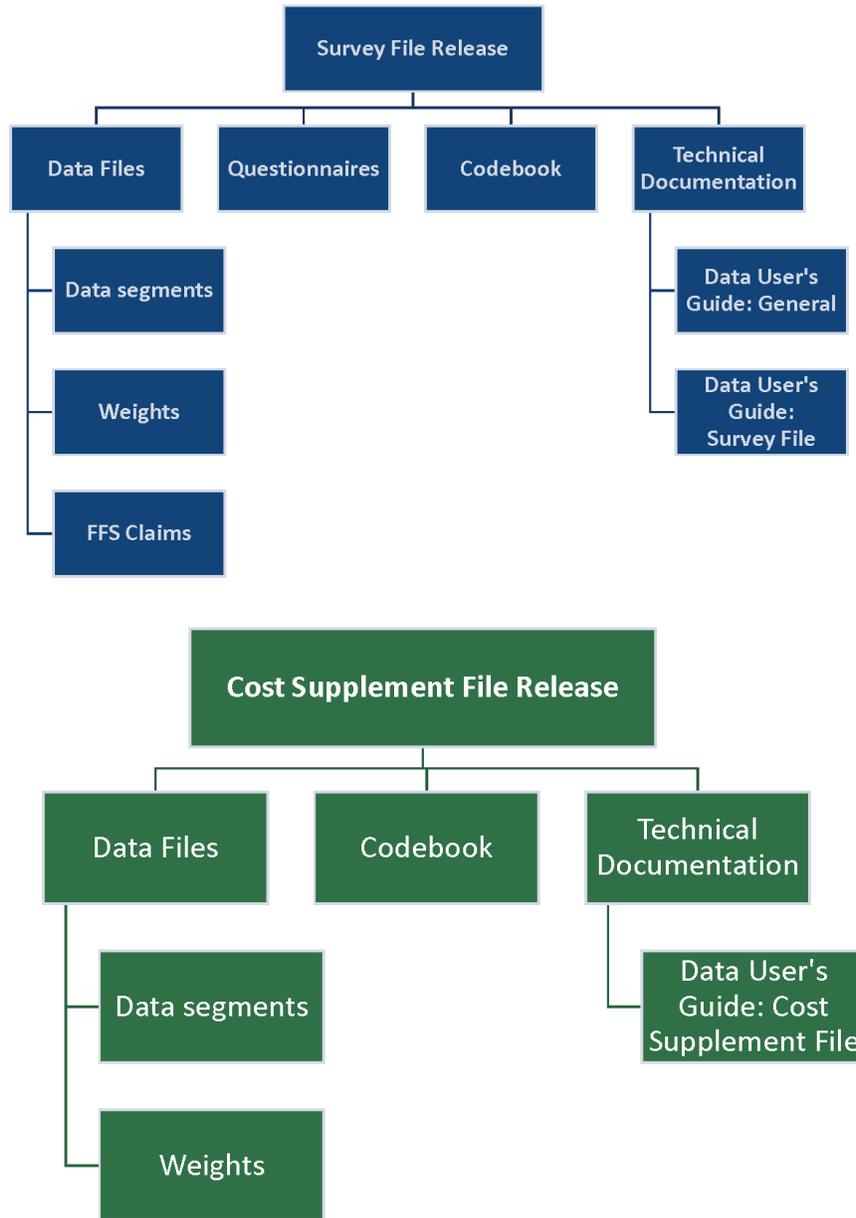
Detailed information on the sampling design can be found in Chapter 3 of this report. Chapter 6 describes the data collection fielding procedures, including eligibility for each round of the interview, and Chapter 9 summarizes the results of data collection, including response rates. Please see the 2015 MCBS Methodology Report to obtain the results of the most recent non-response bias analysis.

### 1.3 Key Data Products and Analyses

MCBS data are made available via releases of annual files. For 2016, two annual Limited Data Set (LDS) releases, the Survey File and the Cost Supplement File, and one Public Use File (PUF) (based on the Survey File data only) are released. The LDS releases each contain multiple files, called segments, which are easily linkable through a common beneficiary key ID. Detailed descriptions of each segment, including the core contents of each segment, key variable definitions, and special notes on new variables, recodes, and administrative sources for select variables can be found in the data release-specific chapters of the MCBS Data User's Guide (see MCBS Data User's Guide: Survey File and MCBS Data User's Guide: Cost Supplement File).

Exhibit 1.3.1 displays the components of each LDS release. Both the Survey File and Cost Supplement File contain data segments, codebooks, questionnaires and technical documentation. The Survey File release contains the Fee-for-Service claims data, which provide CMS administrative information on medical services and payments paid by Medicare claims; claims data for Medicare Advantage beneficiaries are not available. While users can conduct analyses with the Survey File alone, users interested in the Cost Supplement File data will need both LDS files in order to link cost and utilization variables with demographic or health insurance coverage variables.

**Exhibit 1.3.1:** 2016 Contents of Data Releases



Chapter 4 of this report provides information on the specific questionnaire sections associated with each data file. Chapter 7 describes the creation of these data files and Chapter 8 provides an overview of weighting and imputation procedures. Detailed descriptions of each file, including the contents of the files, file structure, information on new variables, key recodes, and administrative sources for select variables can be found in the data file-specific chapters of the MCBS Data User’s Guide (see MCBS Data User’s Guide: Survey File and MCBS Data User’s Guide: Cost Supplement File).

## 2. CHANGES UNIQUE TO 2016

Several key changes were made to the MCBS during 2016, affecting the areas of data collection, questionnaire design, and data processing, including weighting and imputation procedures. These changes are highlighted below and described later in this report.

### 2.1 Sampling

**Sample eligibility:** Beginning in 2015, beneficiaries who became eligible for Medicare Part A or B and enrolled anytime during the sampling year were eligible to be sampled as part of the annual MCBS panel. Prior to 2015, only beneficiaries who became eligible on or before January 1 of the sampling year were eligible to be sampled. Including the current-year eligible beneficiaries allows for the release of data products up to one year earlier. Additional information can be found in Section 3.4.

**Census tracts replaced ZIP Code areas for Secondary Sampling Units (SSUs):** Beginning in Fall 2014, Census tracts or groups of tracts, replaced ZIP Code areas as SSUs for the new panel selected and fielded each fall. See Section 3.3 for more information.

**Hispanic oversample:** Beginning in 2015, Hispanic beneficiaries living outside of Puerto Rico were oversampled for the MCBS. The main goals of the oversampling were to increase the number of Hispanic beneficiaries to allow for more precise estimates of health disparities experienced by these populations and to increase the proportion of MCBS Hispanic beneficiaries from outside Puerto Rico. The goal is to achieve the full target sample of 1,500 U.S. Hispanic beneficiaries by approximately 2020.

### 2.2 Data Collection

In 2016, data collection returned to its regular cycle of three rounds per year (i.e., winter, summer, and fall).

### 2.3 Questionnaires

In 2016, MCBS implemented changes to the Community Questionnaire question text, response options, programming logic, and text fills. Questions about preventive care from the Health Status and Functioning (HFQ) section were moved to a new section, Preventive Care (PVQ). Within the HFQ, existing mental health questions were replaced with standardized scales. In addition, a new section on nicotine products and alcohol use, Nicotine and Alcohol Use (NAQ), was added to the fall interview. Finally, the content of the Usual Source of Care (USQ) section was reverted back to the version fielded in Fall 2014. Additional details about questionnaire content and changes made in 2016 can be found in Section 4.1.

## 2.4 Data Processing

**Imputation methods changes:** Due to the substantial changes in the Income and Assets Questionnaire (IAQ), the 2015 IAQ processing imputed a new set of IAQ variables, which carried through to 2016. New imputation methods were tested and used for this new set of variables. For the 2016 non-prescription medicine imputation, methodological changes were made that improve the efficiency and accuracy of the imputation. See Section 8.3 for more detail about imputation procedures.

## 3. SAMPLE DESIGN FOR THE MCBS 2016 PANEL

### 3.1 Overview of MCBS Sample Design

The MCBS employs a three-stage cluster sample design. In 2016, the survey continued to use the set of 107 primary sampling units (PSUs) that have been employed for sampling for the MCBS since 2001 as well as the census tract-based secondary sampling units (SSUs) that were selected in 2014.<sup>1</sup> At the third stage, Medicare beneficiaries, the ultimate sampling units (USUs), were selected from within the selected tract-based SSUs.<sup>2</sup> In 2015, the sample design underwent a few significant changes, including the oversampling of Hispanic beneficiaries and the inclusion of current-year enrollees in the sampling frame; these changes are carried forward in 2016.

In 2016, the MCBS continued to use the sample rotation pattern used historically. In particular, the newly selected 2016 Panel, and the panels selected in 2013, 2014, and 2015, continued into Fall 2016 and beyond according to their established rotation schedules.<sup>3</sup> The 2012 Panel (which was first fielded in Fall 2012) exited in Summer 2016 and was replaced with a new sample of beneficiaries in Fall 2016. Exhibit 3.1.1 displays the MCBS rotating panel design from 2011 to the present panel.

---

<sup>1</sup> These SSUs were sized to last 20 years; therefore, there is no current plan to reselect the SSUs.

<sup>2</sup> In late 2000, the current set of PSUs was selected. In 2014, SSUs were reconstructed using Census tracts and a new sample was drawn. While the MCBS PSUs and SSUs do not align directly with other surveys, they may overlap in some areas with PSUs and/or SSUs used for other surveys.

<sup>3</sup> A new panel is added each fall and retains the year of its entry as its sampling panel designation for projections and response rate analysis. Once a panel is selected, it remains in the MCBS for 4 years, participating in 12 total rounds.

**Exhibit 3.1.1:** 2011-2016 MCBS Rotating Panel Design

Data Collection Schedule			Panel					
Data Year	Season	Round#	2011	2012	2013	2014	2015	2016
2012	Winter	62	■	■	■	■	■	■
	Summer	63						
	Fall	64						
2013	Winter	65	■	■	■	■	■	■
	Summer	66						
	Fall	67						
2014	Winter	68	■	■	■	■	■	■
	Summer	69						
	Fall	70						
2015	Winter/Summer	71/72	■	■	■	■	■	■
	Fall	73						
2016	Winter	74	■	■	■	■	■	■
	Summer	75						
	Fall	76						

This section documents the procedures used to select the new sample for Fall 2016 (i.e., the 2016 Panel). The 2016 Panel will be retained in the study for the four years specified under the MCBS sample rotation scheme and is designed to: (a) replace approximately one-third<sup>4</sup> of the respondents in the existing MCBS sample; and (b) extend survey coverage to persons added to the Medicare roles during the current year (see Section 3.4 for details).

**PSUs.** In 2016, an investigation into the continued representativeness of the PSUs was undertaken. If the Medicare population had experienced substantial geographic shifts over time, reselection of the PSUs may have been required to control sampling variability. The analyses suggested that while there has been a noticeable increase in the 65 and older population both in the U.S. overall and within the current set of PSUs as well as some geographic movement, only modest precision gains would be expected from a reselection of the PSUs, and the gains are not expected to outweigh the cost of implementation. Thus, the set of PSUs for the 2016 Panel consisted of the 107 previously constructed PSUs used for MCBS sampling since 2001, including 29 certainty PSUs and 78 non-certainty PSUs.

**SSUs.** Prior to 2014, MCBS SSUs were made up of ZIP Codes and ZIP Code fragments. Beginning in 2014, the MCBS implemented the use of census tracts, or clusters of adjacent tracts, as SSUs. A total sample of 703 SSUs was selected within the 107 PSUs, consisting of a proportional allocation of 242 SSUs to the 29 certainty PSUs and an equal allocation of approximately 6 SSUs to each of the 78 non-certainty PSUs. These SSUs are sized to last for a full 20 years of use for MCBS

<sup>4</sup> Due to the cumulative effects of attrition over time as well as cost-related sample cuts from past years, the number of MCBS respondents varies by panel, with fewer respondents in the older panels than in newer ones. Thus, while the newly-selected panel replaces one of four existing panels, the net effect has been to replace about one-third of the existing MCBS respondents. Furthermore, because attrition has been higher than expected in recent years, some of the newer panels may be required to replace more than one-third of the respondents.

sampling. 2016 is the third year in which the annual MCBS panel sample is selected from within the new SSUs; with the selection of the 2017 Panel, the MCBS sample will have fully transitioned from the ZIP-based to the census tract-based SSUs.

**USUs.** The third stage of sampling was the selection of Medicare beneficiaries from within each SSU. Previously, to be eligible for sample selection, beneficiaries had to be eligible for Medicare and enrolled by the first day of the sampling year. Beginning with the 2015 Panel, however, all current-year enrollees are also eligible to be sampled as part of a new sampling approach. As a result, all beneficiaries who are enrolled in Medicare before January 1, 2017 are eligible to be sampled as part of the 2016 Panel. Sampling of current-year enrollees is discussed in detail in Section 3.4.

Hispanic beneficiaries were first oversampled beginning in 2015 and continue to be oversampled in the 2016 Panel. An additional 75 completed interviews from Hispanic beneficiaries are targeted per year with the aim of producing 1,500 Hispanic beneficiaries in the 2020 Survey File. Oversampling of Hispanic beneficiaries is discussed in detail in Section 3.4.

The sampling frame for the Medicare beneficiaries begins with Medicare administrative enrollment data. To avoid duplication in the various panels of MCBS beneficiaries, a unique and disjoint 5-percent sample of the enrollment data is specified annually by CMS for the MCBS. The most recent 5-percent file was used as a basis for selecting the sample for the 2016 MCBS Panel. A first extract of the enrollment data 5-percent file was provided in March 2016, and the bulk of the 2016 Panel sample was selected from that extract. Two additional extracts of the enrollment data 5-percent file, containing only new enrollees who were not included in the initial extract, were also needed to support sampling of current-year enrollees.<sup>5</sup> The combination of these extracts constitutes the full frame from which the 2016 Panel was selected. Details about the sampling frame construction can be found in Section 3.4.

The MCBS enrollment data 5-percent file extracts were subset based on eligibility and other criteria (described in detail later in this section) and then geocoded to the tract level. The set of all records that geocoded to the selected SSUs constituted the MCBS sampling frame of beneficiaries. A random sample of beneficiaries residing in the selected SSUs was then selected within defined age group and ethnicity (Hispanic/non-Hispanic) strata within the U.S., and within age group strata in Puerto Rico. An ethnicity flag (see Section 3.4 for a full description) was used to classify beneficiaries into the Hispanic strata; a value of “yes” indicates that the beneficiary is expected to be Hispanic; a value of “no” indicates that the beneficiary is not expected to be Hispanic. (Actual, or self-reported, Hispanic origin status may differ from the ethnicity flag.) Thus, the sample was selected within the strata displayed in Exhibit 3.1.2.

---

<sup>5</sup> Note that while all new enrollees added to the enrollment data since the previous extract(s) are received, only new *current-year* enrollees are sampled from these additional two extracts.

**Exhibit 3.1.2:** 2016 MCBS Sampling Strata

<b>U.S. Hispanic</b>	<b>U.S. Non-Hispanic</b>	<b>Puerto Rican Resident</b>
<b>Under 45 U.S. Hispanic</b>	Under 45 U.S. non-Hispanic	Under 45 Puerto Rican resident
<b>45 - 64 U.S. Hispanic</b>	45 - 64 U.S. non-Hispanic	45 - 64 Puerto Rican resident
<b>65 - 69 U.S. Hispanic</b>	65 - 69 U.S. non-Hispanic	65 - 69 Puerto Rican resident
<b>70 - 74 U.S. Hispanic</b>	70 - 74 U.S. non-Hispanic	70 - 74 Puerto Rican resident
<b>75 - 79 U.S. Hispanic</b>	75 - 79 U.S. non-Hispanic	75 - 79 Puerto Rican resident
<b>80 - 84 U.S. Hispanic</b>	80 - 84 U.S. non-Hispanic	80 - 84 Puerto Rican resident
<b>85 and over U.S. Hispanic</b>	85 and over U.S. non-Hispanic	85 and over Puerto Rican resident

Sampling rates varied by stratum, with the strata containing younger beneficiaries with disabilities (under 45) and elderly beneficiaries (85 and over) being oversampled. Hispanics are also oversampled relative to their non-Hispanic age stratum counterparts.<sup>6</sup> The MCBS sampling design for an annual panel provides nearly self-weighting (i.e., equal probabilities of selection) samples of beneficiaries within each of the 21 sampling strata.

For the 2016 Panel, the historical target of 11,500 responding beneficiaries across all panels was reduced to 10,112<sup>7</sup> responding beneficiaries across all panels that would comprise the 2019<sup>8</sup> Cost Supplement File. It should be noted that the Hispanic oversample is based on the Survey File counts because it is an annual addition to the fall interview.<sup>9</sup> The reduced targeted annual completed sample size of 10,112 responding beneficiaries with Cost Supplement data is expected to be comprised of approximately 900-1,000 beneficiaries from each of the under 65 (disability) age groups and approximately 1,500-2,000 beneficiaries from each of the remaining age groups. The traditional targeted annual completed sample size of 11,500 responding beneficiaries with Cost Supplement data is typically expected to be comprised of approximately 1,000 beneficiaries from each of the under 65 (disability) age groups and approximately 1,800-2,000 beneficiaries from each of the remaining age groups.

<sup>6</sup> The Hispanic oversample is relatively small and does not impact the sampling rates as much as the age group targets do.

<sup>7</sup> The historical target of 11,500 responding beneficiaries across all panels was reduced to 10,112 in 2016, which was the maximum number of completed interviews achievable within budget.

<sup>8</sup> The target corresponds to the 2019, rather than the 2016, Cost Supplement File because 2019 is the final year that the 2016 Panel beneficiaries will contribute to a Cost Supplement File. The goal is to start with a large enough sample to achieve, after attrition and deaths, the required number of completes in the panel's final Cost Supplement year.

<sup>9</sup> While the base sample sizes are calculated with a Cost Supplement File target in mind, the oversamples were calculated based on a Survey File goal. This is because the oversamples are built over time via additional completes from the newest panel each year, which is easiest to enumerate in terms of Survey File yield.

### 3.2 Selection of MCBS PSUs

The original MCBS PSU sample was selected in 1991 using a sampling frame that was developed using 1980 Census data. In 2001, the set of PSUs was redesigned and reselected. For the redesign, those PSUs in the continental U.S. with at least 224,000 Medicare beneficiaries, according to the June 2000 enrollment data 5-percent file, were included in the sample with certainty. The cutoff of 224,000 corresponds roughly to a probability of selection of 0.75 under a PPS (probability-proportionate-to-size) sample design. The use of the specified cutoff resulted in designating the 28 largest PSUs in the continental U.S. as certainty PSUs. Of these, 27 were also certainties in the original design, and one had been selected as a noncertainty PSU in the original design. In addition, the largest PSU in Puerto Rico (the central part of the San Juan Metropolitan Statistical Area (MSA)) was included in the sample with certainty.

The remaining noncertainty PSUs were grouped by census region and MSA status (where Puerto Rico was treated as a separate "region" for sampling purposes). Within each major group of PSUs, detailed sampling strata were formed by sorting PSUs by the percentage of Medicare beneficiaries enrolled in Health Maintenance Organization (HMO) plans (and in some cases also by the percentage of Medicare beneficiaries who were minorities) and then forming strata of roughly equal size from this sorted list. Thirty-eight noncertainty strata were formed within the continental U.S., and one was formed in Puerto Rico. Two PSUs were then selected from each stratum with probabilities proportionate to size using procedures designed to maximize overlap with the existing MCBS sample. The procedure developed by Ernst (1986)<sup>10</sup> was used for this purpose. In the Ernst approach, each stratum in the new design is treated as a separate linear programming problem, and the optimization of the linear programming problem determines a set of coefficients that are used to select the new sample.

The PSUs are examined periodically for representativeness to the national Medicare population. The most recent analysis was conducted in 2016, and it was determined that a reselection of PSUs was not necessary at that time.

### 3.3 Selection of MCBS SSUs

Prior to 2014, MCBS SSUs were defined as clusters of ZIP Codes and ZIP Code fragments. In 2014, the MCBS SSUs were reselected using census tracts. This change reduces the need for maintenance of SSUs previously required due to the nature of ZIP Code boundary changes and allows for easier merging of MCBS data with U.S. Census Bureau data and other aggregate level geographic or environmental extant data.

---

<sup>10</sup> Ernst, L. (1986). Maximizing the overlap between surveys when information is incomplete, *European Journal of Operational Research*, 27, 192-20.

The creation and selection of SSUs involved several steps. First, the enrollment data<sup>11</sup> were geocoded to the tract level. Next, enrollment data counts by tract were used to calculate SSU measures of size. Small tracts were combined to create SSUs that met a minimum measure of size. (The SSUs were sized to provide for a full twenty years of sampling for the MCBS.) Finally, a core sample of 703 SSUs, plus an additional reserve sample of 339 SSUs, were selected from a frame of 24,212 SSUs within the 107 PSUs using systematic probability proportional to size selection. A fixed number of SSUs was selected from each of the 78 noncertainty PSUs (subject to a maximum of 6), and the remainder were proportionally allocated to the certainty PSUs (subject to a minimum of 6 and constrained to an even number).

A total of 703 core SSUs, comprised of 242 SSUs from the certainty PSUs and 461 SSUs from the noncertainty PSUs, were selected in 2014. An additional reserve sample of 339 SSUs (122 from certainty PSUs and 217 from the noncertainty PSUs) was also selected to provide CMS the possibilities to expand the sample or to study special rare populations in future years. In 2016, the set of 703 core SSUs was used again for sample selection.

### 3.4 Selection of Beneficiaries for the 2016 MCBS Panel

The third stage of sampling is the selection of Medicare beneficiaries from within each SSU. In 2015, two major design innovations were introduced at this stage and have been carried forward in the 2016 sample selection. First, current-year Medicare enrollees are now eligible to be sampled as part of each new annual panel. Their inclusion will allow for the release of data files up to one year earlier than previously possible.<sup>12</sup> Second, Hispanic beneficiaries were oversampled. These innovations and their implementation in the 2016 sample selection process are discussed in detail below.

**Current-Year Enrollee Sample.** Historically, to be eligible for sample selection, beneficiaries had to be eligible for Medicare and enrolled by January 1<sup>st</sup> of the sampling year ( $t$ ), which resulted in the release of data products containing information about the cost and use of health care services (i.e., the Cost Supplement File) in reference year  $t$  during the middle of the year two years later (year  $t + 2$ ).<sup>13</sup> Such late release arose because the year  $t$  cohort of beneficiaries, which contributes to the cost and use of health care services in reference year  $t$ , was not even sampled until year  $t + 1$  and not initially interviewed until the fall round of year  $t + 1$ .

---

<sup>11</sup> Because the enrollment data are so plentiful and the geocoding process so laborious, only enrollment data records falling into ZIP Codes that overlapped with one of the 107 PSUs were geocoded to the tract level. Then, only those falling into one of the 107 PSUs were kept for the selection of SSUs.

<sup>12</sup> Persons who became eligible for Medicare during 2016, for example, could have incurred health care costs in 2016. By including such persons in the sampling process up to a year earlier than was done previously, they can be appropriately represented in the 2016 Cost Supplement File up to a year earlier. The same is true for all years beyond 2016 as well.

<sup>13</sup> Final versions of Medicare claims for a calendar year are available six months after the start of the subsequent calendar year.

Beginning in 2015, the year  $t$  cohort<sup>14</sup> of beneficiaries (i.e., the set of current-year enrollees) was included in the sampling frame of beneficiaries from which the year  $t$  panel<sup>15</sup> was selected and resulted in the collection of health care cost and utilization (Cost Supplement File) data for these cases one year earlier than previously possible. This allows for processing of these data to occur one year earlier than under the previous design, and may therefore allow for release of data files up to one year earlier.

The inclusion of the current-year enrollees introduced the need for multiple enrollment data extracts for sampling and multiple sample draws. Because not all 2016 enrollees are included in the enrollment data by the time the initial sampling needs to occur, additional extracts, or “updates” to the original enrollment data extract for the 2016 Panel, were required. The first, and largest, extract, which contained the bulk of the 2016 sampling frame, was delivered by CMS in two parts, in March and May of 2016.<sup>16</sup> The majority of the 2016 Panel was selected from this initial extract. Additional enrollment data extracts of 2016 enrollees were delivered in early August and early October 2016, and additional samples of 2016 enrollees were drawn from these extracts. The sampling frame for the 2016 Panel is made up of the beneficiaries in the three extracts falling into the MCBS PSUs and SSUs. A fourth and final extract was delivered in mid-January 2017 and used to fully enumerate the 2016 population of Medicare enrollees. Because data collection had already ended for Fall 2016, no sample was drawn from the January extract; however, the information was used for weights calibration. Please see the Coverage Analysis section below for a detailed description of this extract and the results of the coverage analysis.

*Timing of the Interview.* Members of the year  $t$  cohort of beneficiaries sampled under the new design will all be enrolled in Medicare sometime during sampling year  $t$ . Because these individuals may be more cooperative after they become eligible and have a connection to Medicare, and because the interview is geared toward those who are already enrolled, these sampled individuals are interviewed only after they are enrolled. The majority become eligible and enroll before fall interviewing begins; for those not enrolled until after interviewing begins, an interview is conducted with the sampled beneficiary after he or she enrolls in Medicare (i.e., on or after their enrollment date in the enrollment data).

**Hispanic Oversample.** Also beginning in 2015, Hispanic beneficiaries living outside of Puerto Rico are oversampled in new panels. Hispanics are identified, and the sampling frame is stratified using a flag provided by CMS based on Census records of Hispanic surnames and other enrollment information, such as language preference. The Hispanic stratum was oversampled relative to the non-Hispanic stratum. The main goal of the oversampling is to increase the number of

---

<sup>14</sup> An annual cohort is the set of beneficiaries that are enrolled in Medicare and appear in the Medicare enrollment data within a given year.

<sup>15</sup> An annual panel is the set of beneficiaries sampled in a given year and initially interviewed in the fall round of that year.

<sup>16</sup> Note that normally the first extract is delivered as one file; in 2016, a small part of the first extract was delivered a few months later than the majority of the extract. This was an anomaly that is not expected to recur in future years. We will refer to the first extract in 2016 as one extract throughout this report, even though it was delivered in two parts, and extract, frame, and sample sizes from that combined extract will be reported as the sum of the two respective parts.

beneficiaries of Hispanic, Latino/a, or Spanish origin in the MCBS to allow for precise estimates of health disparities experienced by these populations; this is accomplished by increasing the proportion of MCBS Hispanic beneficiaries from outside Puerto Rico. An additional 75 completed interviews with Hispanic beneficiaries are targeted annually with the goal of achieving 1,500 annual Hispanic completes across all panels by 2020.

**Sample Selection Overview.** The sample of MCBS beneficiaries was selected using systematic sampling within each PSU, and specifically only within the 703 core SSUs selected within the PSUs. In May 2016, the majority of the 2016 Panel was selected. In August and October, additional small samples of 2016 enrollees were selected using the same sampling rates as for the initial sample. The sample sizes for the 2016 Panel were determined in early Spring 2016 based on the most up-to-date response rates available at that point in time. Because the sample size for the 2016 Panel was larger than the sizes of past panels and some strata could not support substantial reserves, a reserve sample was not selected for the 2016 Panel.

For the 2016 MCBS Panel, an initial sample of 11,928 beneficiaries (including the Hispanic oversample) was selected in May. In August, an additional 171 current-year enrollees were selected using the sampling rates computed for the first extract and were added to the 2016 MCBS Panel. In October, a further 72 current-year enrollees were selected, again using the sampling rates computed for the first extract, and added to the 2016 MCBS Panel. As of October, the 2016 MCBS Panel was complete, with a total of 12,171 beneficiaries (including the Hispanic oversample).

Details of the determination of the sample size, the construction of the sampling frame, and the selection of the sample of beneficiaries for the 2016 MCBS Panel are given below.

### *Sample Size Determination*

The sample size requirements for the 2016 Panel were derived using estimated sample losses due to “immortals,” deaths, and nonresponse. Immortals are defined as:

- a) Persons in the CMS sampling frame who enrolled prior to the year preceding the sampling year, and are determined to be deceased at the first or second interview, have a date of death confirmed by a proxy to be prior to the sampling year, and whose death is not recorded in CMS administrative updates;
- b) Persons in the CMS sampling frame who enrolled prior to the year preceding the sampling year, and are determined to be ineligible for Medicare in the first or second interview, have a loss of entitlement date confirmed by the respondent or a proxy to be prior to the sampling year, and who have no record of having lost eligibility in CMS administrative updates; or
- c) Persons who enrolled prior to the year preceding the sampling year and died or lost Medicare eligibility prior to the sampling year based on CMS administrative updates.

These three types of immortals all share the characteristic that they would never have been sampled if up-to-date and accurate information on death and eligibility status had been available in the CMS sampling frame.<sup>17</sup> Sampled beneficiaries who were deceased at the first or second interview and for whom a date of death **after** January 1 of the sampling year (or after the enrollment date, in the case of current-year enrollees) is recorded in CMS administrative updates or obtained from a proxy are “true” deaths, and, unlike the immortals, were alive and eligible for Medicare at the beginning of the sampling year (or as of their enrollment date, for current-year enrollees).<sup>18</sup> The essential difference is that the immortals are not eligible for inclusion in the MCBS since by definition they could not have incurred any health care costs in the year in which they were sampled.

For sample size determination purposes, death rates,<sup>19</sup> response rates, and immortal rates were computed within each age group.<sup>20</sup> The immortal and death rates used were an average of historical rates and actual rates from Fall 2014 and Fall 2015. The immortal rates apply to losses in the first fall interview round only. Similarly, the initial losses due to deaths in the sample selection year apply only to the first fall interview round. On the other hand, persons who completed one or more rounds of interviews but who later died in year  $t$  are eligible for inclusion in the Cost Supplement File covering year  $t$ . In other words, these later deaths do not necessarily result in a reduction in sample size in the Cost Supplement File corresponding to the year in which the beneficiary died, but do represent losses in the *subsequent* Cost Supplement Files. Thus, the “first-” and “second-year” death rates that were computed for sample design purposes are used to estimate losses in the second and third Cost Supplement Files, respectively, in which a particular panel can appear. Exhibit 3.4.1 below displays the assumed rates used in determining the sample sizes for the MCBS 2016 Panel. These rates were used in each of the Hispanic, non-Hispanic, and Puerto Rico sampling strata within age group.

---

<sup>17</sup> Note that members of the 2016 cohort (i.e., 2016 sampled panel members who first became eligible for Medicare during 2016) who died or lost eligibility during the sampling year (i.e. sometime during 2016 after becoming eligible) are not immortals and should still be sampled. These cases contribute to the 2016 Cost Supplement File.

<sup>18</sup> Data for beneficiaries in this group who were newly enrolled (i.e., enrolled during the sampling year) are, in fact, pursued, and proxy interviews are attempted. Their data will be used to aid in imputation of their cost and use data.

<sup>19</sup> Included in the calculation of death rates is a small number of persons who lost Medicare eligibility.

<sup>20</sup> Note that during Fall 2014 (Round 70), a decision was made by CMS to replace any newly sampled (Incoming Panel) beneficiaries found to be incarcerated in the first interview because they would not be eligible for benefits. These numbers are quite small and are currently not significant enough to warrant inclusion in the calculation of the sample size for the annual panel.

**Exhibit 3.4.1:** Assumed Rates (in Percent) Used in Determining Sample Sizes for the MCBS 2016 Panel, by Age Group

Sampling Rate	Age Group (as of 12/31/2016)							
	<45	45-64	65-69	70-74	75-79	80-84	85+	Total
Estimated "immortal" rate	0.5	0.5	0.2	0.3	0.4	0.4	1.2	0.5
Estimated selection year death rate	2.2	2.1	1.5	1.8	3.1	4.4	10.5	4.0
Selection year response rate: 2016 Panel	49.7	59.5	49.2	49.0	53.1	49.9	47.7	50.4
Selection year response rate: 2017 and 2018 Panels	54.3	65.0	53.8	53.5	58.0	54.5	52.1	55.0
Post-fall round death/loss rate	1.5	0.7	0.4	0.5	0.5	0.7	1.2	0.7
First year response rate	64.3	72.5	74.8	73.8	71.8	72.6	71.0	72.0
Estimated first year death rate	2.9	2.3	1.7	2.1	3.8	6.0	11.1	4.4
Second year response rate	75.2	76.0	78.5	80.6	79.4	78.8	69.8	77.2
Estimated second year death rate	3.0	3.4	1.7	2.8	3.7	7.3	13.1	5.0
Third year response rate	83.2	82.0	85.1	86.5	84.9	81.1	73.6	82.8
Year 1 Retention rate <sup>1</sup> : 2016 Panel	30.6	41.7	36.1	35.3	36.6	34.3	29.5	34.4
Year 1 Retention rate <sup>1</sup> : 2017 and 2018 Panels	33.4	45.5	39.4	38.5	40.0	37.4	32.2	37.6
Year 2 Retention rate <sup>2</sup>	73.0	74.3	77.2	78.9	76.4	74.1	62.0	73.8
Year 3 Retention rate <sup>3</sup>	80.7	79.2	83.7	84.1	81.7	75.2	64.0	78.7

<sup>1</sup> The Year 1 Retention rate takes into account the immortal rate, selection year death and response rates, post fall round death/lost entitlement rate, and first year response rate. Year 1 refers to the first year after the selection year.

<sup>2</sup> The Year 2 Retention rate takes into account the Year 1 death rate and the Year 2 response rate. Year 2 refers to the second year after the selection year.

<sup>3</sup> The Year 3 Retention rate takes into account the Year 2 death rate and the Year 3 response rate. Year 3 refers to the third year after the selection year.

The response rate for the selection year used in the sample size calculations (i.e., the proportion of sampled beneficiaries, excluding deaths and immortals, who complete the initial fall interview) was assumed to be 50 percent for the 2016 Panel and 55 percent for the 2017 and 2018 Panels. This assumption is based on Fall 2015 achievement as well as historical rates, and includes a slightly lower rate assumption for the 2016 presidential election year. The response rate for the first year in the survey (i.e., the proportion of persons completing the initial fall interview who provide substantially complete data for the first Cost Supplement File to which they contribute) was computed based on the response rate for the 2014 Panel. The response rate for the second year in the survey (i.e., the proportion of living respondents in the first Cost Supplement File who also provide substantially complete data for the second Cost Supplement File) was computed based on the response rate for the 2013 Panel. Finally, the response rate for the third year in the survey (i.e., the proportion of living respondents in the second Cost Supplement File who also provide substantially complete data for the third Cost Supplement File) was based on the response rate for the 2012 Panel.

The sample size projections also included adjustments to account for movement of beneficiaries from one age category to the next over the course of three years in the study. This adjustment affects primarily the youngest age category (44 years or younger), the oldest age category (85 years or older), and the 65 to 69 year-old age category. As the panel ages, the oldest beneficiaries in the 44 years or younger age category will move to the next age category, and there will be no migration into the 44 years or younger age category. On the other hand, there will not be any migration out of the oldest age category (85 years or older), while about 19 to 20 percent of the beneficiaries from the 80 to 84 year-old age group will move into this age group after one year. The 65 to 69 year-old age category will also be affected as the migration into this category from the 45 to 64 year-old age category will be less (about 6 to 8 percent) than the migration out of this category (about 20 to 29 percent) every year. The remaining age categories (45 to 64, 70 to 74, 75 to 79, and 80 to 84) are not affected as much since the migration in and out of these categories occurs at approximately the same rate.

The sample size target (including the Hispanic oversample) for the 2016 Panel was determined to be 12,044 beneficiaries. Exhibit 3.4.2 displays the initial sample size planning numbers and resulting projected Survey File completed interviews in 2016 and projected Cost Supplement File completed interviews in 2017, 2018, and 2019.

**Exhibit 3.4.2:** 2016 Fall Planning Sample Sizes and Resulting Projected Survey File and Cost Supplement File Estimates, by Age Group, Including Recent Enrollees and Hispanic Oversample

<b>Age Category</b>	<b>Number Sampled for 2016 Panel: Fall 2016</b>	<b>Expected Total Survey File Completes from All Panels: 2016 Annual Estimates</b>
<45	1,325	1,344
45-64	827	1,069
65-69	2,394	2,628
70-74	1,426	1,872
75-79	1,799	2,218
80-84	1,893	2,218
85+	2,535	2,469
Total	12,200	13,818

<b>Age Category</b>	<b>Projected Total Cost Supplement File Completes from All Panels: 2016 Annual Estimates</b>	<b>Projected Total Cost Supplement File Completes from All Panels: 2017 Annual Estimates</b>
<45	727	732
45-64	683	702
65-69	1,602	1,621
70-74	1,159	1,199
75-79	1,318	1,294
80-84	1,272	1,261
85+	1,202	1,313
Total	7,963	8,122

<b>Age Category</b>	<b>Projected Total Cost Supplement File Completes from All Panels: 2018 Annual Estimates</b>	<b>Projected Total Cost Supplement File Completes from All Panels: 2019 Annual Estimates</b>
<45	870	995
45-64	814	902
65-69	1,838	2,076
70-74	1,287	1,374
75-79	1,432	1,535
80-84	1,423	1,505
85+	1,572	1,726
Total	9,235	10,112

## 2016 Sampling Frame

1. As described earlier, the inclusion of current-year enrollees in the sampling frame requires additional steps to be taken in the building of the frame and sampling of beneficiaries from the frame because not all year  $t$  enrollees are included in the enrollment data by the spring of year  $t$ , when initial sampling operations occur. Instead, year  $t$  enrollees are added to the enrollment data in two distinct manners. First, beneficiaries who will be automatically enrolled in Medicare appear in the enrollment data up to four months prior to their automatic enrollment. Second, beneficiaries can self-enroll within a seven-month window comprised of the three months prior to their 65th birthday month, their 65th birthday month, and the three months following their 65th birthday month. Those who self-enroll typically appear in the enrollment data within a month after their enrollment in Medicare. Thus, someone enrolling in December 2016 may not appear in the enrollment data until January 2017.

In March 2016, when the first enrollment data extract was pulled to facilitate sampling for the fall round, only a portion of the current-year enrollees were included in the enrollment data. Beneficiaries who enrolled prior to March 1 of year  $t$  or who were automatically enrolled within four months of March (i.e., by July 1 of year  $t$ ) were included in the enrollment data extract. However, any beneficiary who self-enrolled on or after March 1 or was automatically enrolled after July 1 of year  $t$  did not yet appear in the enrollment data. Thus, multiple enrollment data extracts were required to facilitate sampling of the full year  $t$  cohort.

Two additional enrollment data extracts are pulled each year and contribute to the year  $t$  sampling frame: (1) an extract in August, which contains additional self-enrollees through August 1 of year  $t$  and scheduled automatic enrollees through December 1 of year  $t$ ; and (2) an extract in early October, which contains additional self-enrollees through October 1 of year  $t$  and scheduled automatic enrollees through December 31 of year  $t$ . The October extract is scheduled for the latest date possible to facilitate sampling and fielding in year  $t$ ; however, it leaves a slight under-coverage of any self-enrollees between October 2 and December 31 of year  $t$ . A final extract is pulled in mid-January of year  $t+1$  to identify this under-coverage and account for it in weighting adjustments.

For the 2016 MCBS Panel, the first, or initial, extract of the enrollment data, delivered in March, included:

- Beneficiaries who were first eligible for Medicare before January 1, 2016 and still alive and eligible on January 1, 2016; and
- Beneficiaries who were first eligible for Medicare between January 1, 2016 and March 1, 2016 (inclusive), or who would be automatically enrolled in Medicare between March 2, 2016 and July 1, 2016 (inclusive), regardless of vital status.

To avoid duplication across the various panels of MCBS beneficiaries, a unique and disjoint 5-percent sample of the enrollment data<sup>21</sup> is specified annually by CMS, and a subset (based on the eligibility and mortality selection criteria listed above, as well as other data quality checks) is

---

<sup>21</sup> The enrollment data include over 100,000,000 beneficiaries.

specified for the MCBS for use in sampling beneficiaries for the annual panels. This is referred to as the 2016 enrollment data subsample.

CMS subset each of its enrollment data extracts as described above, keeping only beneficiaries meeting the criteria for the 2016 enrollment data subsample. These enrollment data subsample extracts are further subset to include only beneficiaries falling within the 703 selected MCBS SSUs. Exhibit 3.4.4 shows the number of beneficiaries by sampling stratum (age group by ethnicity or Puerto Rican residency) in the three 2016 enrollment data subsample extracts and the resulting 2016 sampling frame. Of the 2,941,050 beneficiaries in the combined 2016 enrollment data subsample extracts, a total of 46,835 beneficiaries fell within the selected MCBS PSUs and SSUs and were eligible for sampling in 2016.

**Exhibit 3.4.4:** Number of Beneficiaries in 2016 Enrollment Data Subsample Extracts (Combined) and 2016 Sampling Frame, by Stratum

Stratum	Age Group/ Ethnicity	Three Extracts Combined	2016 Sampling Frame
1	<45, U.S. Hispanic	11,227	135
2	45-64, U.S. Hispanic	37,460	427
3	65-69, U.S. Hispanic	60,065	760
4	70-74, U.S. Hispanic	50,813	601
5	75-79, U.S. Hispanic	34,992	430
6	80-84, U.S. Hispanic	23,804	261
7	85+, U.S. Hispanic	23,523	249
8	<45, U.S. non-Hispanic	82,794	1,241
9	45-64, U.S. non-Hispanic	317,415	4,892
10	65-69, U.S. non-Hispanic	722,371	11,195
11	70-74, U.S. non-Hispanic	565,836	8,838
12	75-79, U.S. non-Hispanic	392,701	5,978
13	80-84, U.S. non-Hispanic	277,731	4,220
14	85+, U.S. non-Hispanic	323,270	4,820
15	<45, Puerto Rican	419	77
16	45-64, Puerto Rican	2,633	476
17	65-69, Puerto Rican	3,947	674
18	70-74, Puerto Rican	3,407	565
19	75-79, Puerto Rican	2,662	427
20	80-84, Puerto Rican	1,931	293
21	85+, Puerto Rican	2,049	276
Total		2,941,050	46,835

Using the initial 2016 enrollment data subsample extract in combination with previous annual enrollment data subsamples, the size of the total 2016 enrollment data subsample (containing all projected 2016 Medicare enrollees, through December 31, 2016) could be forecast at the time of initial sampling (May 2016). This forecast was used to determine how much of the current-year enrollee sample was expected to be selected from the first extract and how much would be expected to be drawn from future extracts, and to determine the sampling fractions for beneficiaries.

A final enrollment data subsample extract was provided in mid-January 2017 and used to fully enumerate the 2016 cohort to (a) inform undercoverage of the 2016 sampling frame, and (b) contribute to weighting adjustments to account for this undercoverage. Results of these analyses are provided in the Coverage Analysis section below.

### *Sample Selection for the 2016 Panel*

The goal for the 2016 Panel was to select a sample of 12,200 beneficiaries with targeted oversamples in the 64 and younger and 85 and over age groups; this sample also includes an oversample of Hispanic beneficiaries (from the U.S. portion of sampling frame) targeted to yield interviews with 75 additional true (self-reported) Hispanic beneficiaries in 2016.

**Sampling Fractions.** As discussed earlier, the Hispanic flag provided by CMS and used for sampling is imperfect. Some cases flagged as Hispanic do not self-report as Hispanic, and others flagged as non-Hispanic actually self-report as Hispanic. The sampling fractions for the Hispanic and non-Hispanic strata, therefore, were jointly determined to compensate for the misclassification errors inherent in the Hispanic flag to achieve the required sample sizes of self-reported Hispanic and non-Hispanic beneficiaries.

The calculation of the sampling fractions for 2016 was accomplished in two phases. First, the sampling fractions for the U.S. (not including Puerto Rico) portion of the sample were completed at the national level within the 14 strata (seven age groups by the Hispanic/non-Hispanic flag). Then, the sampling fractions for the Puerto Rico portion of the sample were calculated within the seven age group strata.

**Probabilities of Selection.** The probabilities of selection for beneficiaries were then computed. Let  $f_{1a}$  be the national sampling fraction for the Hispanic stratum in age group  $a$ , let  $f_{-1a}$  be the national sampling fraction for the non-Hispanic stratum in age group  $a$ , and let  $f_{2a}$  be the sampling fraction for the Puerto Rican stratum in age group  $a$ . The inclusion probability for the  $i$ -th PSU is denoted by  $\pi_i$  and the conditional inclusion probability for the  $j$ -th SSU given the  $i$ -th PSU is  $\pi_{j|i}$ . Thus, the conditional probability of selection for beneficiary  $k$  in the Hispanic stratum in age group  $a$  given PSU  $i$  and SSU  $j$  is

$$\rho_{1ak|ij} = \min\left(1, \frac{f_{1a}}{\pi_i \pi_{j|i}}\right), \quad a = 1, \dots, 7,$$

and for non-Hispanic beneficiary  $k$  in the non-Hispanic stratum in age group  $a$  given PSU  $i$  and

SSU  $j$  is

$$\rho_{-1ak|ij} = \min\left(1, \frac{f_{-1a}}{\pi_i \pi_j |i}\right), \quad a = 1, \dots, 7.$$

The probabilities of selection for beneficiaries in the Puerto Rico PSUs are calculated as follows:

$$\rho_{2ak|ij} = \min\left(1, \frac{f_{2a}}{\pi_i \pi_j |i}\right), \quad a = 1, \dots, 7.$$

Actual sample sizes can fall short of expectations when SSUs actually contain fewer beneficiaries in the enrollment data subsample extract than what is called for by the initial national sampling fractions. To avoid a shortfall, the initial sampling fractions must be adjusted and the conditional probabilities  $\rho_{1ak|ij}$ ,  $\rho_{-1ak|ij}$ , and  $\rho_{2ak|ij}$  recomputed. Within each stratum, the cumulative sums of the probabilities of selection were formed. In an iterative process, the initial national sampling fractions were repeatedly adjusted until the cumulative sums were as close as possible to the final targeted sample sizes. Exhibit 3.4.5 displays the final sampling fractions used for calculating probabilities of selection, by stratum, for the 2016 Panel.

**Exhibit 3.4.5:** 2016 MCBS Panel, Final Sampling Fractions by Stratum

Stratum	Age Group/ Ethnicity	Final Sampling Fraction, in Percent
1	<45, U.S. Hispanic	0.0180
2	45-64, U.S. Hispanic	0.0024
3	65-69, U.S. Hispanic	0.0037
4	70-74, U.S. Hispanic	0.0028
5	75-79, U.S. Hispanic	0.0046
6	80-84, U.S. Hispanic	0.0101
7	85+, U.S. Hispanic	0.0156
8	<45, U.S. non-Hispanic	0.0220
9	45-64, U.S. non-Hispanic	0.0023
10	65-69, U.S. non-Hispanic	0.0029
11	70-74, U.S. non-Hispanic	0.0022
12	75-79, U.S. non-Hispanic	0.0043
13	80-84, U.S. non-Hispanic	0.0064
14	85+, U.S. non-Hispanic	0.0074
15	<45, Puerto Rican	0.0317
16	45-64, Puerto Rican	0.0032
17	65-69, Puerto Rican	0.0074
18	70-74, Puerto Rican	0.0054
19	75-79, Puerto Rican	0.0083
20	80-84, Puerto Rican	0.0116
21	85+, Puerto Rican	0.0145
Total		0.0045

**Selection of the 2016 Panel.** The 2016 Panel was drawn by systematic random sampling with probability proportional to the conditional probabilities of selection with an independently selected random start within each PSU. For the sample drawn from the U.S. (outside Puerto Rico), the beneficiaries were ordered within each PSU by age group, SSU (to approximate geographic serpentine sorting), ethnicity flag, and extract.<sup>22</sup> There were 1,250 beneficiaries with a conditional

<sup>22</sup> The second extract was added to the end of the first extract, in the same sort order, and the systematic selection was continued into the range of newly enrolled beneficiaries. The same process was used for the third extract.

probability of selection equal to 1.<sup>23</sup> These beneficiaries were selected with certainty, given the selection of their PSUs and SSUs.

The same process was used for the sample drawn from Puerto Rico. There were no beneficiaries in Puerto Rico with a conditional probability of selection equal to 1.

### *Sampling Results*

Exhibit 3.4.6 below shows the number of selected beneficiaries within each age group, and Exhibit 3.4.7 shows the number of selected beneficiaries within each stratum. These tables present the total number of beneficiaries in the 2016 Panel, including the Hispanic oversample.

**Exhibit 3.4.6:** 2016 MCBS Panel, Number of Beneficiaries Selected by Age Group

<b>Age Group (as of December 31, 2016)</b>	<b>Total Selected Beneficiaries</b>
<45	1,310
45-64	838
65-69	2,395
70-74	1,428
75-79	1,800
80-84	1,889
85+	2,511
<b>Total</b>	<b>12,171</b>

<sup>23</sup> This is a large increase compared to 2015 due to the fact that the sample size increased considerably from 2015 to 2016. As a result, there are more SSUs from which *all* eligible beneficiaries are selected in 2016 compared to 2015, which in turn means more beneficiaries with conditional probabilities of selection equal to 1.

**Exhibit 3.4.7:** 2016 MCBS Panel, Number of Beneficiaries Selected by Stratum

Stratum	Age Group/Ethnicity	Total Selected Beneficiaries
1	<45, Hispanic	117
2	45-64, Hispanic	67
3	65-69, Hispanic	192
4	70-74, Hispanic	111
5	75-79, Hispanic	134
6	80-84, Hispanic	159
7	85+, Hispanic	208
8	<45, non-Hispanic	1,180
9	45-64, non-Hispanic	760
10	65-69, non-Hispanic	2,174
11	70-74, non-Hispanic	1,298
12	75-79, non-Hispanic	1,641
13	80-84, non-Hispanic	1,710
14	85+, non-Hispanic	2,277
15	<45, Puerto Rican	13
16	45-64, Puerto Rican	11
17	65-69, Puerto Rican	29
18	70-74, Puerto Rican	19
19	75-79, Puerto Rican	25
20	80-84, Puerto Rican	20
21	85+, Puerto Rican	26
Total		12,171

The number of current-year enrollees (those who enrolled in 2016) selected into the 2016 Panel (including the Hispanic oversample) is displayed in Exhibit 3.4.8 below.

**Exhibit 3.4.8:** 2016 MCBS Panel, Number of Current-Year Enrollees Selected by Age Group

Age Group	Initial Extract	Three Extracts Combined
<45	47	70
45-64	29	57
65-69	154	346
70+	0	0
Total	230	473

Several quality checks were performed after sample selection. These included the comparison of the weighted 2016 enrollment data subsample extract counts (combining all three extracts) with the corresponding weighted counts for the selected sample as well as the distributions of selected beneficiaries by PSU and SSU.

### *Coverage Analysis of the 2016 Sampling Frame*

As discussed above, a final enrollment data 5-percent file extract was provided in mid-January 2017. This extract was used to fully enumerate the 2016 cohort to (a) inform undercoverage of the 2016 sampling frame, and (b) contribute to weighting adjustments to account for this undercoverage. The results of the analysis of this extract are given in this section.

**Coverage Analysis.** The fourth enrollment data subsample extract, along with the first three extracts, was used to fully enumerate both the 2016 enrollment data subsample and the 2016 MCBS population. In order to construct the full 2016 enrollment data subsample, all records of eligible beneficiaries enrolled through December 31, 2016, from the four extracts were combined. From that universe, the 2016 MCBS population was constructed by retaining only beneficiaries falling into the MCBS PSUs and SSUs. Including the fourth extract, which contains beneficiaries who were automatically enrolled or self-enrolled through the end of 2016, ensures that all eligible beneficiaries, particularly current-year enrollees who were not included in the first three extracts, are included in the final population. Thus, the final 2016 MCBS population includes all beneficiaries who were enrolled in Medicare in 2016 and reside in the MCBS PSUs and SSUs.

Exhibit 3.4.9 displays the full 2016 enrollment data subsample and the estimated 2016 eligible U.S. Medicare population, by stratum. This table builds on Exhibit 3.4.4, which displayed the 2016 enrollment data subsample file through the third extract.

The fourth enrollment data subsample extract is similar in size to the third extract, at approximately 20,000 beneficiaries overall. Further, the number of cases from the fourth extract falling into the MCBS PSUs and SSUs is also very similar to those in the third extract (333 in the fourth extract versus 317 in the third extract). Overall, the fourth extract accounts for 0.7 percent of the total 2016 MCBS population. While the cases included in the fourth extract consist exclusively of new enrollees, the exclusion of this extract from the frame could lead to imbalances in the representativeness of the sample. However, because the final extract accounts for such a small proportion of the overall population, it was expected to have minimal impact on the representativeness of the 2016 Panel. Any imbalance will be accounted for in adjustments made to the weights, discussed in Chapter 8.

**Exhibit 3.4.9:** Number of Beneficiaries in 2016 Enrollment Data Subsample and Estimated 2016 MCBS Population, by Stratum

Stratum	Age Group/ Ethnicity	Beneficiaries in Four Enrollment Data Extracts Combined	Estimated Beneficiaries in Full U.S. Medicare Population
1	<45, Hispanic	11,299	225,980
2	45-64, Hispanic	37,596	751,920
3	65-69, Hispanic	61,281	1,225,620
4	70-74, Hispanic	50,907	1,018,140
5	75-79, Hispanic	35,041	700,820
6	80-84, Hispanic	23,829	476,580
7	85+, Hispanic	23,537	470,740
8	<45, non-Hispanic	83,128	1,662,560
9	45-64, non-Hispanic	318,442	6,368,840
10	65-69, non-Hispanic	739,218	14,784,360
11	70-74, non-Hispanic	566,180	11,323,600
12	75-79, non-Hispanic	392,797	7,855,940
13	80-84, non-Hispanic	277,762	5,555,240
14	85+, non-Hispanic	323,293	6,465,860
15	<45, Puerto Rican	421	8,420
16	45-64, Puerto Rican	2,637	52,740
17	65-69, Puerto Rican	3,981	79,620
18	70-74, Puerto Rican	3,408	68,160
19	75-79, Puerto Rican	2,662	53,240
20	80-84, Puerto Rican	1,932	38,640
21	85+, Puerto Rican	2,049	40,980
Total		2,961,400	59,228,000

Exhibit 3.4.10 compares the original forecast of the full 2016 enrollment data subsample, including all cases expected to be in the enrollment data through the end of the 2016, to the actual count of beneficiaries in the combined four enrollment data subsample extracts. As described above, the forecast was used to develop sampling fractions for use in the selection of the 2016 MCBS Panel sample. The comparisons in Exhibit 3.4.10 are given by stratum and overall. The counts are quite close; the total actual overall count is only slightly higher than the forecast (2,961,400 actual versus 2,947,410 forecast beneficiaries), and the differences by stratum are very small.

**Exhibit 3.4.10:** Forecast Compared to Actual Beneficiaries in Full 2016 Enrollment Data Subsample, by Stratum

Stratum	Age Group/Ethnicity	Forecast <sup>1</sup> of Beneficiaries in Full 2016 Enrollment Data Subsample	Actual <sup>2</sup> Beneficiaries in Full 2016 Enrollment Data Subsample
1	<45, Hispanic	11,332	11,299
2	45-64, Hispanic	37,930	37,596
3	65-69, Hispanic	62,449	61,281
4	70-74, Hispanic	51,494	50,907
5	75-79, Hispanic	35,431	35,041
6	80-84, Hispanic	24,066	23,829
7	85+, Hispanic	23,800	23,537
8	<45, non-Hispanic	83,613	83,128
9	45-64, non-Hispanic	319,640	318,442
10	65-69, non-Hispanic	721,198	739,218
11	70-74, non-Hispanic	565,649	566,180
12	75-79, non-Hispanic	392,602	392,797
13	80-84, non-Hispanic	277,682	277,762
14	85+, non-Hispanic	323,252	323,293
15	<45, Puerto Rican	434	421
16	45-64, Puerto Rican	2,666	2,637
17	65-69, Puerto Rican	4,122	3,981
18	70-74, Puerto Rican	3,407	3,408
19	75-79, Puerto Rican	2,662	2,662
20	80-84, Puerto Rican	1,931	1,932
21	85+, Puerto Rican	2,049	2,049
<b>Total</b>		<b>2,947,410</b>	<b>2,961,400</b>

NOTE: The Full 2016 Enrollment Data Subsample in this table includes all current-year enrollees through December 31, 2016.

<sup>1</sup>Forecast was calculated at the time of sampling (May, 2016).

<sup>2</sup>Actual counts based on all enrollment data records received for 2016, including those in final extract delivered in January, 2017.

### 3.5 Continuing Sample (2013-2015 Panels)

Each Continuing panel is fielded, along with the newly selected Incoming Panel, according to its rotation schedule. Panels are fielded for a total of 12 rounds, starting in the fall round of the year the panel is selected. In Summer 2016, the 2012 Panel completed its rotation schedule and was replaced by the 2016 Panel in Fall 2016. The 2013 Panel was in its 10<sup>th</sup> round of participation in Fall 2016, the 2014 Panel was in its 7<sup>th</sup> round, and the 2015 Panel was in its 4<sup>th</sup> round.

### 3.6 Fielded Sample Sizes by Panel and Round

During 2016, sampled beneficiaries were interviewed during three rounds: a winter round, a summer round, and a fall round. During Summer 2016, the 2012 Panel was interviewed for its final time, and in Fall 2016, the 2016 Panel was interviewed for its first time. The fielded sample sizes,<sup>24</sup> by panel, for each round are given in Exhibit 3.6.1.

**Exhibit 3.6.1:** 2016 Fielded Sample Sizes by Round for Each Panel

Round	Panel	Fielded Sample Sizes
<b>Winter 2016</b>	2012	2,294
	2013	2,616
	2014	4,495
	2015	4,349
	All	13,754
<b>Summer 2016</b>	2012	1,998
	2013	1,717
	2014	2,911
	2015	3,122
	All	9,748
<b>Fall 2016</b>	2013	2,274
	2014	3,670
	2015	3,345
	2016	12,145
	All	21,434

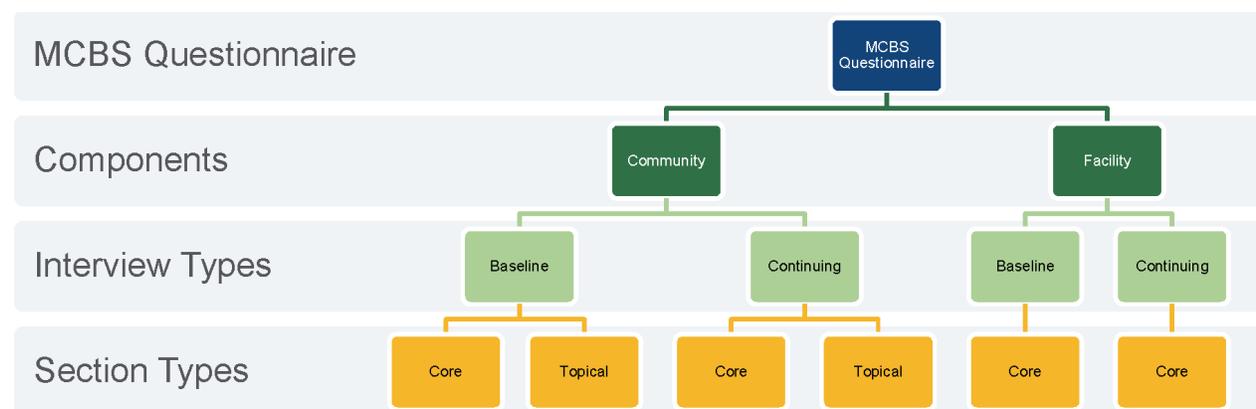
<sup>24</sup> Note that these are not the original sample sizes when the panel was selected (except in the case of the 2016 Panel), but the sample remaining in the round, less attrited beneficiaries and other sample losses, that are fielded in that round. For the summer round, the fielded sample size counts do not include deferred cases (cases that were intentionally not fielded in the summer round and instead deferred to the fall round).

## 4. INSTRUMENT AND MATERIALS DESIGN

The MCBS Questionnaire structure features two components (Community and Facility), administered based on the beneficiary's residence status. Within each component, the flow and content of the questionnaire varies by interview type and data collection season (fall, winter, or summer). There are two types of interviews (Baseline and Continuing) containing two types of questionnaire sections (Core and Topical). The beneficiary's residence status determines which questionnaire component is used and how it is administered. See Exhibit 4.1 for a depiction of the MCBS Questionnaire structure.

- **Community Component:** Survey of beneficiaries residing in the community at the time of the interview (i.e., their residence or a household). Interview may be conducted with the beneficiary or a proxy.
- **Facility Component:** Survey of beneficiaries residing in facilities such as long-term care nursing homes or other institutions at the time of the interview. Interviewers do not conduct the Facility component with the beneficiary, but with staff members located at the facility (i.e., facility respondents). This is a key difference between the Community and Facility components.

**Exhibit 4.1:** MCBS Questionnaire Overview



Interviews are conducted in one or both components in a given data collection round, depending on the beneficiary's living situation. Procedures for these "crossover" interviews (where the beneficiary moves from one component to another) are described in Section 6.2.

Within each component, there are two types of interviews – an initial (Baseline) interview administered to new beneficiaries, and an interview administered to repeat (Continuing) beneficiaries as they progress through the study.

- ▶ **Baseline:** The initial questionnaire administered to beneficiaries new to the study; administered in the fall of the year they are selected into the sample (interview #1).

- ▶ Continuing: The questionnaire administered to beneficiaries as they progress through the study (interviews #2-12).

MCBS uses dependent interviewing to ensure that the flow of the interview takes into account known and previously reported information, such as beneficiary sex, health insurance coverage, health status and conditions. Dependent interviewing based on preloaded data is especially important for the design and flow of the Continuing questionnaire. This allows for a more streamlined interview by prompting the respondent for confirmation of previously-reported information, and for more complex queries to be crafted that address a beneficiary's particular situation. Section 7.2 describes the role of preloads in dependent interviewing in more detail.

Depending on the interview type and data collection season (fall, winter, or summer), the MCBS Questionnaire includes Core and Topical sections. See Exhibits 4.1.4 and 4.1.6 for tables of the 2016 Core and Topical sections.

- ▶ Core: These sections are of critical purpose and policy relevance to the MCBS, regardless of season of administration. Core sections collect information on beneficiaries' health insurance coverage, health care utilization and costs, and operational management data such as locating information.
- ▶ Topical: These sections collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

Data collected by the Community and Facility interviews are released to users via two primary limited data sets (LDS) – the Survey File and the Cost Supplement File. The Survey File includes data collected via Core and Topical sections related to beneficiaries' access to care, health status, and other information regarding beneficiaries' knowledge, attitudes towards, and satisfaction with their health care. This file also contains demographic data and information on all types of health insurance coverage. The Cost Supplement File delivers information collected via Core sections on the use and costs of health care services as well as information on supplementary health insurance, living arrangements, income, health status, and physical functioning.

## 4.1 Community Questionnaire Content

The section that follows provides an overview of the Community component of the MCBS questionnaire. The actual content administered varies based upon several factors, including the questionnaire administration season or round, the type of interview which reflects the length of time the respondent has been in the MCBS, and the component of the most recent interview.

### *Interview Type*

As MCBS is a panel survey, the type of interview a given beneficiary is eligible for depends on his or her status in the most recent round of data collection. Interview type (also referred to in this report by its Community Questionnaire variable name, INTTYPE) is a key determinant of the path followed through the Community Questionnaire. For example, the Baseline interview is an abbreviated interview that includes many Core and Topical sections but does not include questionnaire sections that collect health care utilization and cost information. For the purposes of

administering the Community Questionnaire, there are ten interview types, summarized in Exhibit 4.1.1 below. Several of these interview types are applicable only in a certain season. For example, the Baseline interview (INTTYPE C003) is always conducted in the fall.

**Exhibit 4.1.1:** Community Questionnaire Interview Types

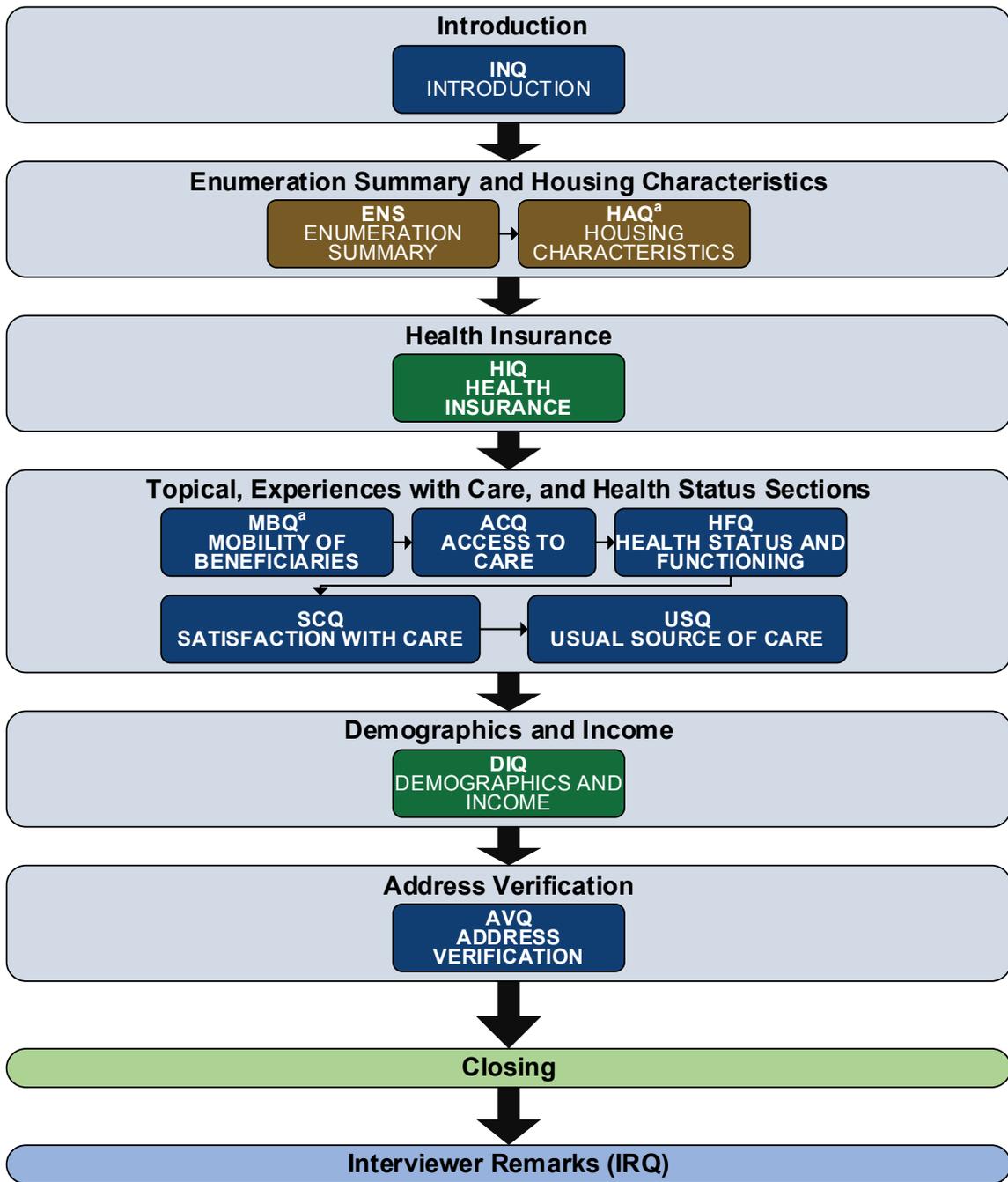
INTTYPE	Description	Seasons
C001	Standard Continuing interview, meaning the most recent interview was in the community during the last round.	All
C002	New from facility, meaning the most recent interview was in a facility. No prior community interview.	All
C003	Baseline interview. First round in the sample.	Fall
C004	Standard community "holdover," meaning the last round interview was skipped. Most recent interview was in the community.	All
C005	Facility "crossover," meaning the most recent interview was in a facility. Last community interview was two rounds ago.	All
C006	Facility "crossover," meaning the most recent interview was in a facility. Last community interview was three or more rounds ago.	All
C007	Second round interview. Most recent interview was the fall Baseline interview. The second round interview is the first time utilization and cost data are collected.	Winter
C008	Standard exit interview. Most recent interview was in the community during the last round. Final round in the survey.	Summer
C009	Exit interview "holdover," meaning the last round interview was skipped. Final round in the survey.	Summer
C010	Second round "holdover," meaning the winter interview was skipped. Most recent interview was the fall Baseline interview. The third round interview is the first time in which utilization and cost data are collected.	Summer

### *Community Questionnaire Flow*

Interview type and data collection season (fall, winter, or summer) are the two main factors that determine the specific sections included in a given interview. Further factors include whether the interview is conducted with the beneficiary or with a proxy and, for proxy interviews, whether the beneficiary is living or deceased.<sup>25</sup> The Baseline interview contains an abbreviated flow which does not include the utilization or cost sections of the questionnaire. Exhibit 4.1.2 shows the flow for the Baseline interview.

<sup>25</sup> Only one final interview is conducted with proxy respondents for deceased beneficiaries. See Section 6.2 for more details on fielding procedures for deceased beneficiaries.

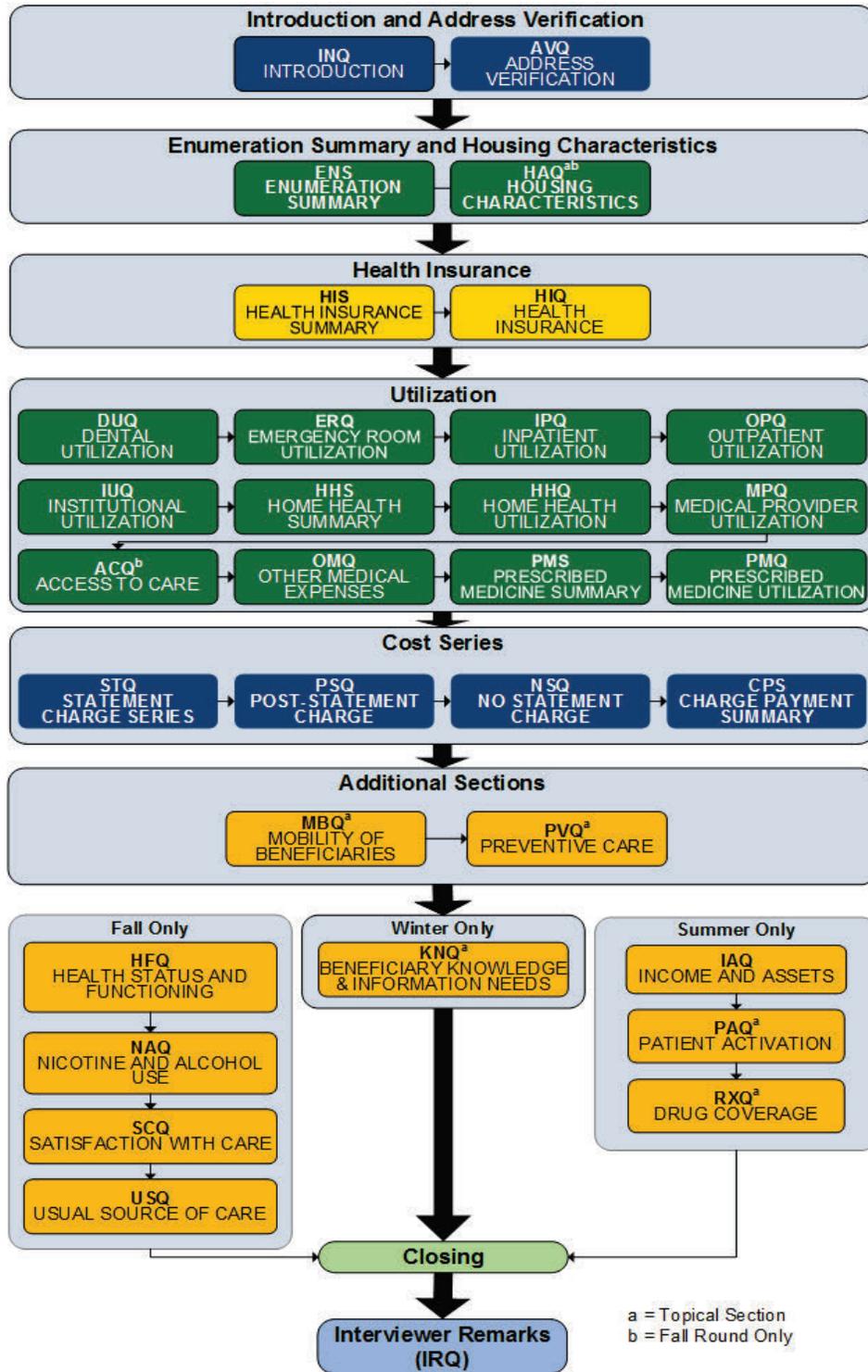
**Exhibit 4.1.2:** 2016 MCBS Community Questionnaire Flow for Baseline Interview



a = Topical Section

Exhibit 4.1.3 shows the most common Community Questionnaire flow for standard Continuing community sample.

**Exhibit 4.1.3:** 2016 MCBS Community Questionnaire Flow for Continuing Interview



## *Core Section Content*

Core survey content is grouped into questionnaire sections that collect data central to the policy goals of the MCBS. These sections collect information related socio-demographics, health insurance coverage, health care utilization and costs, beneficiary health status and experiences with care, as well as operational and procedural data. Data from these questionnaire sections are found on the Survey File and Cost Supplement File data releases. Many of the core sections are administered each round. The following pages describe core sections of the Community Questionnaire, organized by topic of information collected. Exhibit 4.1.4 lists the core sections of the Community Questionnaire along with the data file they inform and the seasons in which they were administered.

### **Socio-Demographics**

Two sections in the Community Questionnaire capture key socio-demographic characteristics of the beneficiary. The Demographics and Income section is administered for each Community beneficiary once during the Baseline interview. Income and Assets is administered to all Continuing beneficiaries once per year.

The **Demographics and Income (DIQ)** section includes traditional demographic items such as Hispanic origin, race, English proficiency, education, and a total household income. This section is administered during the Baseline interview.

**Income and Assets (IAQ)** collects detailed information about income and assets of the beneficiary and spouse or partner (if applicable). IAQ covers beneficiary (and spouse/partner) income from employment, Social Security, Veteran's Administration, and pensions. The respondent is also asked to indicate the value of the beneficiary's (and spouse's/partner's) assets including retirement accounts, stocks, bonds, mutual funds, savings accounts, businesses, land or rental properties, and automobiles. Also included is homeownership or rental status, and food security items. The Income and Assets section is asked in the summer round to collect income and asset information about the previous calendar year. The Income section asked in the Summer of 2016 asked about income and assets for the 2015 calendar year.

**Exhibit 4.1.4:** 2016 MCBS Community Core Sections by Data File and Administration Schedule

Section Group	Abbr.	Section Name	Survey File	Cost Supplement File	Administrative Season
<b>Socio-Demographics</b>	IAQ	Income and Assets	X		Summer**
	DIQ	Demographics/Income	X		Fall, Baseline Interview
<b>Health Insurance</b>	HIS	Health Insurance Summary***	X		All Seasons
	HIQ	Health Insurance	X		All Seasons
<b>Utilization</b>	DUQ	Dental Utilization		X	All Seasons
	ERQ	Emergency Room Utilization		X	All Seasons
	IPQ	Inpatient Hospital Utilization		X	All Seasons
	OPQ	Outpatient Hospital Utilization		X	All Seasons
	IUQ	Institutional Utilization		X	All Seasons
	HHS	Home Health Summary**		X	All Seasons
	HHQ	Home Health Utilization		X	All Seasons
	MPQ	Medical Provider Utilization		X	All Seasons
	OMQ	Other Medical Expenses Utilization		X	All Seasons
	PMS	Prescribed Medicine Summary***		X	All Seasons
	PMQ	Prescribed Medicine Utilization		X	All Seasons
<b>Cost</b>	STQ	Statement Cost Series		X	All Seasons
	PSQ	Post-Statement Charge		X	All Seasons
	NSQ	No Statement Charge		X	All Seasons
	CPS	Charge Payment Summary***		X	All Seasons
<b>Experiences with Care</b>	ACQ	Access to Care	X		Fall
	SCQ	Satisfaction with Care	X		Fall
	USQ	Usual Source of Care	X		Fall
<b>Health Status</b>	HFQ	Health Status and Functioning	X		Fall

**SOURCE:** 2016 MCBS Community Questionnaire

\*Certain procedural or operational management sections are collected specifically to manage the data collection process. These sections are not directly included in the LDS files (e.g., Introduction (INQ), Address Verification (AVQ), Enumeration (ENS), Closing (CLQ), and Interview Remarks (IRQ)).

\*\*The IAQ is administered in the Summer round following the current data year.

\*\*\*Summary sections: Updates and corrections are collected through the summary sections. The respondent is asked to verify summary information gathered in previous interviews. Changes are recorded if the respondent reports information that differs from what was previously recorded.

## Health Insurance

Two sections of the Community Questionnaire capture health insurance information.

**Health Insurance Summary (HIS)** reviews information about health insurance plans that the beneficiary had at the time of the last interview. Plans reported in the prior round may be deleted, or edited in this section, if the respondent indicates they are not correct for the previous reference period. Additional plans may be added if they are active during the prior round reference period but are not reported at that time. HIS is administered for all beneficiaries that had an interview in the prior round or those who skipped the prior round but have always remained in the Community component of the MCBS. Therefore, newly sampled beneficiaries in the Baseline interview, beneficiaries in their final interview who did not skip the prior round interview, and those that crossed over from the Facility Questionnaire do not receive this section.

**Health Insurance (HIQ)** records all health insurance plans that the beneficiary has had since the beginning of the reference period. The survey prompts for coverage under each of the following types of plans: Medicare Advantage, Medicaid, Tricare, non-Medicare public plans, Medicare Prescription Drug Plans, and private (Medigap or supplemental) insurance plans. Detailed questions about coverage, costs, and payment are included for Medicare Advantage, Medicare Prescription Drug, and private insurance plans.

## Utilization

The utilization sections of the questionnaire capture health care use by category. Generally, four types of health care utilization are recorded: provider service visits, home health care, other medical expenses, and prescribed medicines. Provider service visits includes visits to dental providers, emergency rooms, inpatient and outpatient hospital departments, institutional stays, and medical providers. In these sections, visits are reported as unique events by date, although in cases where there are more than five visits to a single provider during the reference period, the events are entered by month with the number of visits specified. A slightly different reporting structure is used for home health care, other medical expenses, and prescribed medicines.

All utilization sections are administered in all Continuing interviews; these sections are not part of the Incoming Panel's Baseline interview. Additional detail is provided on each of the four types of health care utilization collected by the community survey below.

### Provider Service Visits

The utilization sections collecting provider service dates are as follows.

**Dental Utilization (DUQ)** collects information about dental visits during the reference period. DUQ collects the name and type of dental providers, dates of visits, services performed, and medicines prescribed during the visits.

**Emergency Room Utilization (ERQ)** records visits to hospital emergency rooms during the reference period. ERQ collects the names of the hospitals, dates of visits, whether the visit was associated with a particular condition, and medicines prescribed during the visits. If a reported

emergency department visit resulted in hospital admission, an inpatient visit event is created, with follow up questions asked in the Inpatient Utilization section.

**Inpatient Utilization (IPQ)** collects information about inpatient stays during the reference period. IPQ collects the names of the hospitals, beginning and end dates of the stays, whether surgery was performed, whether the visit was associated with a particular condition, and medicines prescribed to be filled upon discharge from the hospital (medicines administered during the stay are not listed separately). Inpatient stays resulting from emergency room admissions are also covered.

**Outpatient Hospital Utilization (OPQ)** prompts for visits that the beneficiary may have made to hospital outpatient departments or clinics during the reference period. OPQ collects the name of the outpatient facility, dates of visits, whether surgery was performed, whether the visit was associated with a particular condition, and medicines prescribed during the visits.

**Institutional Utilization (IUQ)** collects information about stays in nursing homes or any similar facility during the reference period. IUQ collects the name of the institution(s) and the dates the beneficiary was admitted and discharged from the institution(s).

**Medical Provider Utilization (MPQ)** collects information about medical provider visits during the reference period. In addition to physicians and primary care providers, this includes visits with health practitioners that are not medical doctors (acupuncturists, audiologists, optometrists, chiropractors, podiatrists, homeopaths, naturopaths), mental health professionals, therapists (including speech, respiratory, occupational, and physical therapists), and other medical persons (nurses, nurse practitioners, paramedics, and physician's assistants). MPQ collects names and types of providers, dates, whether the visit is associated with a particular condition, and medicines prescribed during the visit.

### Home Health Care Visits

A second type of health care utilization captured by the community survey are home health care visits. For Continuing beneficiaries that reported home health events during the prior round, **Home Health Summary (HHS)** reviews those providers and confirms whether the same providers were visited during the current round. These visits are recorded not by date, but by the number of visits. In addition, the length of visits and services performed are recorded. **Home Health Utilization (HHQ)** then collects information about home health provider visits, both professional and non-professional, during the reference period. HHQ collects names and types of home health providers, number and length of visits, and services performed during visits.

### Other Medical Expenses

The community survey also records other medical expenses. These expenses are reported using a slightly different reporting structure within the questionnaire.

**Other Medical Expenses Utilization (OMQ)** collects information about medical equipment and other items (excluding prescriptions) that the beneficiary purchased, rented, or repaired during the reference period. Other medical expenses includes glasses, hearing devices, orthopedic items (wheelchairs, canes, etc.), diabetic equipment and supplies, dialysis equipment, prosthetics, oxygen-related equipment and supplies, ambulance services, other medical equipment (beds, chairs, disposable items, etc.) and alterations to the home or car. For each item the date(s) of rental, purchase, or repair are recorded. For disposable medical items (e.g., bandages), the number of purchases is collected, rather than a date.

## Prescribed Medicines

For Continuing interviews with at least one medicine reported in the prior round, the **Prescribed Medicine Summary (PMS)** presents the list of medicines reported during the prior round interview and asks the respondent to verify that those medicines are correct as of the date of the prior round interview. This allows the interviewer to add, delete, or edit medicines applicable to the prior round reference period.

The **Prescribed Medicine Utilization (PMQ)** section collects details about prescribed medicines obtained during the reference period. For medicines recorded in the provider service visit sections (in the context of those visits), PMQ collects the medicine strength, form, quantity, and number of purchases. Medicines that are not previously reported during the course of the provider service visit utilization sections, including those that are refilled or called in by phone, are also collected in this section. Unlike for provider service visits, event dates are not collected for prescribed medicines. Instead, the interviewer records the number of purchases or refills. Information about non-prescription medicines and prescriptions that are not filled are not recorded.

## Cost Series

Once all utilization sections are completed, the questionnaire flows to the cost series, wherein the costs of all reported visits and purchases are recorded, along with the amount paid by various sources. Importantly, additional visits and purchases not reported in the utilization sections of the questionnaire could be recorded within the cost series, and all corresponding data for those events are collected within the cost series.

The cost series consists of four sections: Statement, Post-Statement, No Statement, and Charge Payment Summary. Each is described below.

The **Statement section (STQ)** collects medical cost information directly from Medicare Summary Notices (MSNs), insurance explanations of benefits (EOB), Prescription Drug Plan statements, and TRICARE or other insurance statements. In cases where the beneficiary had more than one payer (e.g., Medicare and private insurance), interviewers organize statements into charge bundles, which are driven by the claim total on a MSN or EOB and may include one or more utilization events (visits, medicines, or purchases). Each charge bundle is entered separately, and all previously-reported events associated with the charge bundle are linked to the cost record. Payment details are entered from the statements and any remaining amount not accounted for is confirmed with the respondent. This process is repeated for all available, not

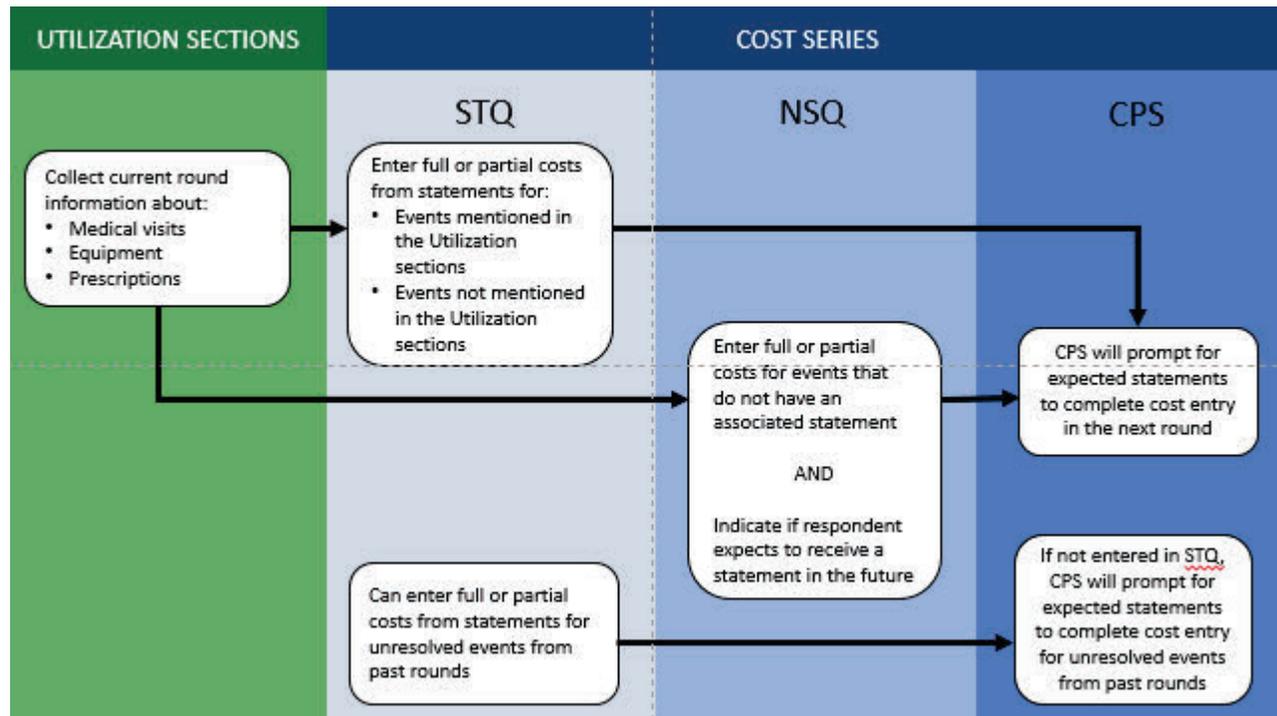
previously recorded insurance statements containing events that occurred within the survey reference period (roughly the past year).

The **Post-Statement section (PSQ)** facilitates cost data collection for rental items that span multiple rounds of interviews (such as a long term wheelchair rental) and for which cost data has not yet been reported.

The **No Statement section (NSQ)** prompts for cost data for all events that do not have a Medicare, insurance, or TRICARE statement reported in the current round. This section attempts to capture cost data even in absence of insurance statements. The respondent may refer to non-statement paperwork such as bills or receipts to help collect accurate cost information. NSQ loops through a series of cost verification items for each event or purchase reported during the current round utilization but not already linked to a cost record via the Statement section. If respondents indicate a statement for the event is expected, then the NSQ items are bypassed.

The final cost series section, the **Charge Payment Summary (CPS)** reviews outstanding cost information reported from previous rounds. For example, if the respondent reported in the previous interview that he/she expected to receive an insurance statement for a particular event, then this event is carried forward to the next round CPS. Any charge bundle for which costs are not fully resolved is asked about in the next round CPS section. There are a variety of reasons a cost record might qualify to be asked about in CPS (referred to as "CPS Reasons"). For example, a respondent may have been expecting to receive a statement related to the event or may have reported payments that account for only part of the total charge. The amount of information collected in CPS and the path through the section is determined by the CPS reason for the cost record. One case can have multiple cost records flagged for CPS with a variety of CPS reasons. The questionnaire loops through each eligible cost record in an attempt to collect further cost data.

The flow of sections and questions within the Cost series varies depending on data collected in the current round (e.g., whether the beneficiary had a health insurance statement for a visit reported in the current round) and data collected in prior rounds (i.e., whether there was outstanding cost information reported from a prior round). Exhibit 4.1.5 illustrates how paths through these sections may vary depending on health care utilization and cost information collected in the current and previous rounds.

**Exhibit 4.1.5:** Utilization and Cost Section Flow

## Experiences with Care

Three sections cover the beneficiary's experience with care in various medical settings. All three Experiences with Care sections are part of the fall round interview for Incoming Panel and Continuing respondents.

**Access to Care (ACQ)** focuses on the beneficiary's experience with particular types of medical encounters (hospital emergency room, hospital clinic or outpatient department, long-term care facility, or medical doctor visits) during the reference period. If the beneficiary had one or more of a particular type of medical encounter, additional items collect information about services received and waiting times associated with the most recent encounter.

**Satisfaction with Care (SCQ)** collects the respondent's opinions about the health care that the beneficiary had received. The questions refer to medical care received from all medical providers, including both doctors and hospitals.

In Fall 2016, the Usual Source of Care/Patient Perceptions of Integrated Care (USQ/PPIC) section fielded in 2015 was reverted back to the originally fielded **Usual Source of Care (USQ)** section. The USQ obtained specific information about the usual source of health care for the beneficiary as well as any specialists seen during the reference period.

## Health Status

**Health Status and Functioning (HFQ)** collects information on the beneficiary's general health status and needs. This includes specific health areas such as disabilities, vision, hearing, and preventive health measures. HFQ includes measures of the beneficiary's ability to perform physical activities, moderate and vigorous exercise, health care maintenance and needs, and standard measures of Instrumental Activities of Daily Living (using the telephone, preparing meals, etc.), and Activities of Daily Living (bathing, walking, etc.). In addition, HFQ asks about medical diagnoses for common conditions (cancer, arthritis, hypertension, etc.). Finally, the section covers mental health conditions, falls, urine loss, and a more extensive series of questions for beneficiaries with high blood pressure and diabetes.

## Operational and Procedural

These sections help guide the interviewer through the interview, providing scripts for introducing and ending the interview. They also facilitate collection of address and household information to augment sample information for the purposes of locating respondents for follow-up interviews. Data collected in these sections are not included in the Survey or Cost Supplement data files.

**Introduction (INQ)** introduces the survey and records whether the interview was completed by the beneficiary or a proxy. For interviews completed by a proxy, the introduction collects the proxy's name and relationship to the beneficiary and determines if the proxy is a member of the beneficiary's household. The introduction is part of every community interview.

**Address Verification (AVQ), Closing (CLQ) and Exit (EXQ)** sections obtain contact information necessary to locate the respondent for future rounds of interviewing. AVQ collects the beneficiary's contact information (address, phone number, alternate address, etc.), and CLQ collects contact information for the proxy (if applicable), two additional individuals who may serve as proxies in the future, and two additional contact persons who do not live in the beneficiary's household. AVQ and CLQ are administered in all rounds except the final exit interview. EXQ is administered in place of CLQ during the exit interview and contains additional scripts to thank the respondent for participation over the four years of the MCBS.

**Enumeration (ENS)** collects household information and a roster of persons living in the household. For each household member added to the roster, his/her relationship to the beneficiary, sex, date of birth, age and employment status are collected. ENS is administered in all rounds except the final exit interview.

The **Interviewer Remarks Questionnaire (IRQ)** captures additional metadata about the interview, as recorded by the interviewer. This includes the length of the interview, assistance the respondent may have received, perceived reliability of the information provided during the interview, and comments the interviewer had about the interviewing situation. IRQ is administered after every interview, but is generally completed after leaving the respondent's home, as none of the questions are directed to the respondent.

### Topical Section Content

In addition to the core content, there are several topical questionnaire sections that capture data on a variety of key topics that relate to the beneficiary's housing characteristics, health behaviors, knowledge about Medicare, and health-related decision making. All data from the topical sections are included in the Survey File data release. Each topical section is described below, organized by information collected. Exhibit 4.1.6 lists the topical sections and administration schedule.

**Exhibit 4.1.6:** 2016 MCBS Community Topical Sections by Data File and Administration Schedule

Section Group	Abbr.	Section Name	Traditional Season	Administrative Season
<b>Housing Characteristics</b>	HAQ	Housing Characteristics	Fall only	Fall 2016
<b>Health Behaviors</b>	MBQ	Mobility of Beneficiaries	All seasons	Winter 2016, Summer 2016, Fall 2016
	NAQ	Nicotine and Alcohol Use	Fall only	Fall 2016
	PVQ	Preventive Care	All seasons	Fall 2016, Winter 2017, Summer 2017
<b>Knowledge and Decision Making</b>	IAQ	Food Insecurity	Summer only	Summer 2017
	KNQ	Beneficiary Knowledge and Information Needs	Winter only	Winter 2017
	PAQ	Patient Activation	Summer only	Summer 2017
	RXQ	Drug Coverage	Summer only	Summer 2017

### Housing Characteristics

**Housing Characteristics (HAQ)** collects information on the beneficiary's housing situation. This includes the type of dwelling, facilities available in the household (e.g., kitchen and bathrooms), accessibility, and modifications to the home (e.g., ramps, railings, and bathroom modifications). This section also records if the beneficiary lives in an independent or assisted living community (distinct from a nursing or long-term care facility) where services like meals, transportation, and laundry may be provided. HAQ is administered in the fall for all beneficiaries in the Community component.

## Health Behaviors

Three questionnaire sections record additional information about health behaviors, specifically mobility, preventive care, and nicotine and alcohol use.

**Mobility of Beneficiaries (MBQ)** determines the beneficiary's use of available transportation options, with a focus on reduced mobility and increased reliance on others for transportation.

The **Preventive Care (PVQ)** section collects information about beneficiaries' preventive health behaviors. Questions administered in this section vary by data collection season. In the winter round, the PVQ focuses on the influenza vaccine while in the summer round, the PVQ asks about the shingles and pneumonia vaccines. In the fall round, the PVQ asks whether the beneficiary has received various types of applicable preventive screenings or tests, such as a mammogram, Pap smear, or digital rectum exam.

**Nicotine and Alcohol Use (NAQ)** collects information on beneficiaries' smoking and drinking behavior, including past and current use of cigarettes, cigars, "smokeless" tobacco, and e-cigarettes. It also asks about past and current drinking behavior.

## Knowledge and Decision-Making

Respondent knowledge of Medicare and health-related decision making is captured in three topical sections.

The **Beneficiary Knowledge and Information Needs (KNQ)** section is administered in the winter round. These items measure the respondent's self-reported understanding of Medicare and common sources of information about health care and Medicare.

The **Patient Activation (PAQ)** section is administered during the summer round. It covers items such as the beneficiary's interaction with health care providers and ability to find and understand information from those providers.

The **Drug Coverage (RXQ)** section is a summer round section that focuses on the Medicare Prescription Drug benefit, including respondent knowledge of the benefit, and opinions of the beneficiary's drug coverage, whether through a Medicare Prescription Drug Plan, a Medicare Advantage plan with prescription drug coverage, or a private insurance plan that covers prescription drugs.

### *Changes to the Community Questionnaire for 2016*

Questionnaire changes implemented for 2016 generally included updates to question text, response options, programming logic, text fills, and the addition of new questionnaire items.

## Summary of Global Questionnaire Updates Made for 2016

Global questionnaire updates included aligning the MCBS Community Questionnaire specifications with terminology used by other federally-funded surveys when referring to doctors. In any

question asking about a specific health care visit, diagnosis, prescription, or other medical encounter, the word “doctor” was replaced with the term “doctor or other health professional.”

## Summary of Item- and Section-Level Questionnaire Revisions

In Fall 2016, questions were revised to align the MCBS Community Questionnaire specifications with the Department of Health and Human Services (DHHS) Data Council Recommended Guidance for Survey Questions and Preferred Practices in HHS Surveys for Selected Estimates.<sup>26</sup> This effort resulted in the replacement of existing mental health questions with more standardized scales in the HFQ section and a new section on the use of nicotine products and alcohol. Questions about preventive care were moved from the HFQ section to a new PVQ section to enhance the flow of the questionnaire. Additional changes were made in Fall 2016 to the USQ section, to revert the section to the version fielded in the Fall 2014.

### Demographics and Income (DIQ)

In Fall 2016, a Limited English Proficiency (LEP) measure about reading was added to the questionnaire.

#### Income and Assets (IAQ)

In Summer 2016, a five question set was added to the Income and Assets Questionnaire (IAQ). These items were adapted from the Health and Retirement Survey (HRS) to collect information about inheritances and other lump sum assets.

### Beneficiary Knowledge and Information Needs (KNQ)

In Winter 2016, three items were added to the Beneficiary Knowledge and Information Needs (KNQ) section to identify some of the reasons beneficiaries find it difficult to compare plans and make plan choices.

#### Nicotine and Alcohol Use (NAQ)

In Fall 2016, existing items about tobacco and alcohol use were replaced with items based on DHHS recommendations. Previously, tobacco and alcohol use questions had been asked as part of the Health Status and Functioning (HFQ) section. The new nicotine and alcohol use items were placed in a new section, Nicotine and Alcohol Use (NAQ), to enhance the flow of the questionnaire.

---

<sup>26</sup> Department of Health and Human Services (DHHS) Data Council. 2015. *Draft: 12-14-2015 HHS data council recommended guidance for survey questions and preferred practices in HHS surveys for selected estimates: Improving utility, comparability and alignment*. In author's possession.

## Preventive Care (PVQ)

Existing items about the seasonal flu vaccine were moved from the HFQ section and added to a new section – Preventive Care (PVQ)—beginning in Winter 2016. In Summer 2016, two items were added to the PVQ section. One item, regarding the shingles vaccine, is a new item for MCBS, while the other item, about the pneumonia vaccine, had previously been part of the HFQ section. In Fall 2016, screening questions about blood pressure, cholesterol, mammogram, Pap smear, digital rectal exam, prostate cancer that were previously asked in the HFQ section were migrated to the PVQ section to enhance the flow of the questionnaire.

## Usual Source of Care (USQ)

Changes were made in Fall 2016 to the Usual Sources of Care (USQ) questionnaire section to remove the Patient Perceptions of Integrated Care (PPIC) items that were added in 2015. This change essentially reverted the USQ back to the version fielded in Fall 2014. The decision to remove the PPIC items in Fall 2016 was informed by feedback from interviewers about the challenges associated with administering these items to beneficiaries. The motivation for removing PPIC items in Fall 2016 was to allow NORC and CMS researchers time to reevaluate the wording, placement, and relevance of PPIC items before refielding a subset of the items in Winter 2018.

## 4.2 Facility Instrument Content

The following section provides an overview of the content of the Facility component of the MCBS questionnaire. The content of the Facility Instrument varies based upon several factors, including the season of data collection, the type of interview (which reflects the length of time the beneficiary has been in the facility), and the component of the most recent interview.

### *Interview Type*

Similar to the Community Questionnaire, the Facility Instrument uses interview type as a key determinant of which questionnaire sections to administer during a facility interview.

The MCBS uses five interview types, also known as sample types, to describe MCBS beneficiaries who reside in a facility, summarized in Exhibit 4.2.1.

**Exhibit 4.2.1:** Facility Instrument Interview Types

<b>INTTYPE</b>	<b>Description</b>	<b>Season</b>
CFR	Continuing Facility Resident. Beneficiary whose previous round interview was a facility interview and who currently resides at the same facility.	Any
CFC	Community-Facility-Crossover. Beneficiary who interviewed in the community previously and has now moved to a long-term care facility.	Any
FFC	Facility-Facility-Crossover. Beneficiary who was previously interviewed in a long-term care facility and has now moved to a different facility.	Any
FCF	Facility-Community-Facility Crossover. Beneficiary whose last interview was in the community and for whom a facility interview has been conducted in a previous round, and who has been admitted to a new facility or readmitted to a facility where the beneficiary had a previous stay. This sample type is rarely encountered.	Any
IPR	Incoming Panel Respondent. Beneficiary who was just added to the MCBS sample (fall round only) and currently resides in a facility.	Fall

NOTE: Interview type (INTTYPE) is typically referred to as Sample Type in the Facility Instrument section specifications.

*Facility Screener*

The Facility screener is administered to a facility staff member when a beneficiary moves to a new facility setting. The Facility screener confirms whether the beneficiary is currently living at the facility (or lived at the facility at some point during the reference period) and determines whether the facility is a public or private residence.

*Facility Instrument Flow*

The Facility Instrument collects similar data to the Community Questionnaire. However, the Facility Instrument is administered to facility staff and not to the beneficiary; that is, the beneficiary does not answer questions during a Facility interview – instead, facility administrators and staff answer questions on behalf of the beneficiary.

Just like the Community Questionnaire, the sections administered in a given facility interview vary by interview type and data collection season (fall, winter, or summer). The Baseline interview, administered to Incoming Panel Respondents, contains an abbreviated flow which does not include the utilization or cost sections of the questionnaire. Exhibit 4.2.2 shows the flow for the Baseline interview.

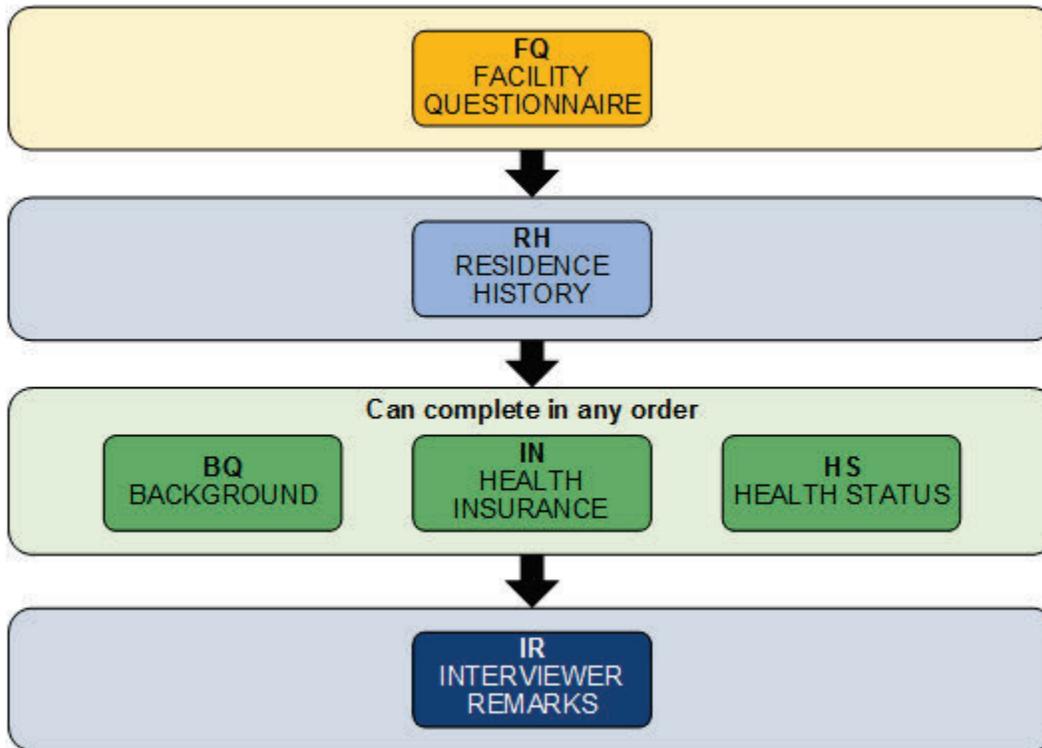
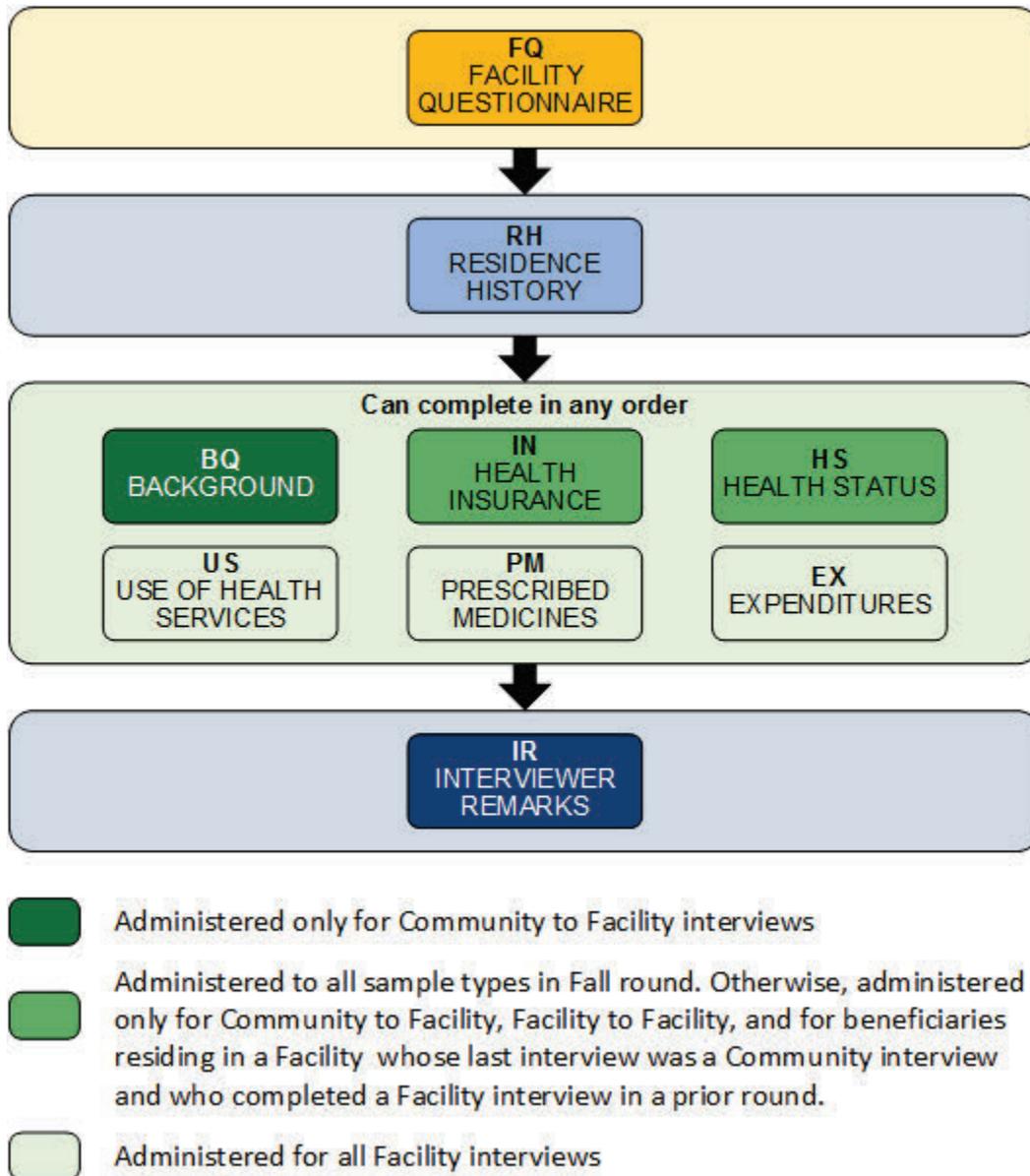
**Exhibit 4.2.2:** 2016 MCBS Facility Instrument Flow for Baseline Interview

Exhibit 4.2.3 shows the flow for the Continuing and crossover interview types.

Because the Facility Instrument is administered to facility staff and not directly to the beneficiary, the Facility Instrument is designed to have a modular, flexible flow. The interviewer first completes the Facility Questionnaire (FQ) section. Next, the interviewer administers the Residence History (RH) section. The remaining sections may be completed in any order. Interviewers are instructed to conduct the sections in the order most suitable to the facility structure and the availability of facility staff. For example, the interviewer may conduct three sections with the head nurse and then visit the billing office to complete the remaining sections. Interviewers complete the Interviewer Remarks (IR) section at the end of the interview.

**Exhibit 4.2.3:** 2016 MCBS Facility Instrument Flow for Continuing and Crossover Interviews



*Core Section Content*

The Facility Instrument consists of only core sections with no topical content. The following pages describe core sections of the Facility Instrument, organized by topic of information collected. Exhibit 4.2.4 shows the core sections of the Facility Instrument along with the data file they inform and the seasons in which they are administered.

**Exhibit 4.2.4:** Facility Core Sections by Data File and Administration Schedule

Section Group	Abbrev	Section Name	Survey File	Cost Supplement File	Administrative Season
<b>Facility Characteristics</b>	FQ	Facility Questionnaire	X		All seasons
<b>Socio-Demographics</b>	RH	Residence History	X		All seasons
	BQ	Background	X		Fall**
<b>Health Insurance</b>	IN	Health Insurance	X		Fall***
<b>Utilization</b>	US	Use of Health Services		X	All seasons
	PM	Prescribed Medicines		X	All seasons
<b>Cost</b>	EX	Expenditures		X	All seasons
<b>Health Status</b>	HS	Health Status	X		Fall***

SOURCE: 2016 MCBS Facility Instrument

NOTE: Certain procedural or operational management sections are collected specifically to manage the data collection process. These sections are not directly included in the LDS files (e.g., Interview Remarks (IR)).

\*The BQ section is also administered to Community-to-Facility crossover cases each season.

\*\*The IN and HS sections are also administered to Community-to-Facility and Facility-to-Facility crossover cases each season.

## Facility Characteristics

The Facility Characteristics core section contains the **Facility Questionnaire (FQ)** section of the Facility Instrument. The FQ section collects information on the number, classification, and certification status of beds within the facility; sources of payment for facility residents; and facility rates. Interviewers typically conduct the FQ with the facility administrator. Interviewers are not allowed to abstract this section of the interview; it must be conducted with a facility staff member.

## Socio-Demographics

The Socio-Demographics core sections capture key characteristics of the interview and the beneficiary. These include residence history and demographics.

The **Residence History (RH)** section collects information about all of the places that the beneficiary stayed during the reference period. Information is collected about where the beneficiary was just before entering the facility and where he/she went if they had been discharged. For each stay, the interviewer collects the name of the place of residence, the type of place it is, and the start and end date for the period the beneficiary was living there.

The RH section creates a timeline of the beneficiary's whereabouts from the date the beneficiary entered the facility or the date of the last interview, through the date of interview, date of

discharge, or date of death. The goal is to obtain a complete picture of the beneficiary's stays during the reference period, including any stays of one night or more in hospitals, other facilities, or any other place.

The **Background Questionnaire (BQ)** collects background information about the beneficiary such as use of long-term care before admission to the facility, level of education, race, ethnicity, service in the Armed Forces, marital status, spouse's health status, living children, and income. The BQ is completed only once for each beneficiary during their first interview in the Facility.

## Health Insurance

The Health Insurance core section contains the **Health Insurance (IN)** section of the Facility Instrument. The IN section collects information about the beneficiary's type(s) of health insurance coverage. This includes questions about all types of health insurance coverage the beneficiary had in addition to Medicare: private insurance, long-term care insurance, Department of Veterans Affairs eligibility, and TRICARE or CHAMPVA.

## Utilization

The Utilization sections collect data on the beneficiary's use of health care and prescribed medicines. These sections are administered to all sample types except for the Incoming Panel.

The **Use of Health Care Services (US)** section collects information on the beneficiary's use of health care services while a resident of the facility. This includes visits with a range of providers including medical doctors, dentists, and specialists; visits to the hospital emergency room; and other medical supplies, equipment, and other types of medical services provided to the beneficiary.

The best facility respondent for this questionnaire section is usually someone directly involved with the beneficiary's care or someone who is familiar with the medical records.

The **Prescribed Medicines (PM)** section collects data on the beneficiary's monthly use of prescribed medicines while in the facility. The data collected includes the medicine name, form, strength, dosage, how often it is administered, and the total times per month that the medicine is taken. The data for this section may be obtained by abstracting from the Medication Administration Record (MAR), which is the report that serves as a legal record of the medications administered to a patient at a facility by a health care professional. MARs are commonly referred to as drug charts. Unlike other parts of the questionnaire, the PM section collects data by month, asking for each month separately.

In the Facility Instrument, prescribed medicines are defined as medicines ordered by a physician through a written or verbal order for a pharmacist to fill. In the long-term care setting, doctors often order medications that do not require a prescription in the community setting, such as vitamins. This is a key difference between the Facility and Community Instruments. As a result, medications that are not considered prescribed medicines in the Community Instrument are collected as prescribed medicines in the Facility Instrument. Prescribed medicine data collected via

the Facility interview are not incorporated into data products that are shared with users. Prescribed medicine data from Medicare Part D claims for facility respondents are incorporated into the Cost Supplement file.

## Cost

The Facility Cost component consists of the **Expenditures (EX)** section. The EX section collects information about bills for the beneficiary's care at a facility and payments by source for those charges. Data are only collected for the time period when the beneficiary was a resident of the facility at which the interview takes place. The EX section collects information by billing period (e.g., monthly semi-monthly, quarterly, etc.).

Unlike the Community Questionnaire, which collects information for each service, the EX section collects information on the fees the facility bills for the beneficiary's care. The EX section collects information on the amount billed for the beneficiary's basic care and for any health related ancillary services. Typically, the EX section is administered to facility staff located in the billing office.

## Health Status

The **Health Status (HS)** section collects information on the beneficiary's general health status, ability to perform various physical activities, general health conditions, instrumental activities of daily living, and activities of daily living.

Most of the information needed to conduct the HS section may be found in a medical chart. The Federal Government requires that all nursing facilities certified by Medicaid or Medicare conduct comprehensive and standardized assessments of each resident's health status when the resident is admitted to the nursing home and at regular intervals thereafter.<sup>27</sup> These assessments are captured by the Long-Term Care Minimum Data Set (MDS),<sup>28</sup> which contains a set of key items measuring a resident's capacity to function independently. Nursing homes use this information to assess each resident's health status, identify problem areas and, where problems exist, formulate care plans to address them.

The HS section is designed to mirror the flow and wording of the MDS items; it contains a subset of the MDS items. In addition, the HS section contains some questions that are not found on the MDS. Interviewers ask these questions of someone knowledgeable about the beneficiary's care or find the information in the medical chart.

---

<sup>27</sup> Centers for Medicare & Medicaid Services. *Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual v. 1.16*. Baltimore, MD, October 2018.

<sup>28</sup> <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/Minimum-Data-Set-3-0-Public-Reports/index.html>

## Operational and Procedural

The **Interviewer Remarks (IR)** section captures additional metadata about the interview, as recorded by the interviewer. This includes comments the interviewer may have about the interviewing situation and notes to themselves for use in gaining cooperation in the future. Data from this section are not included in the Survey File or the Cost Supplement File.

## Missing Data Sections

There are three additional sections, called missing data sections, which are activated when essential survey information is coded as “don’t know” or “refused” in the Facility Questionnaire (FQ), Residence History (RH), or Background (BQ) sections. The missing data sections prompt the interviewer for the specific piece of information that is missing. There are no new questions in the missing data sections, just repeats of questions initially asked in the FQ, RH, or BQ. Examples of the type of missing information that activate the missing data sections are the name of the facility or date of death.

The purpose of the missing data sections is to reduce item non-response for key variables in a highly modular, flexible format. If the interviewer is able to obtain the missing information from another facility staff member or from a different medical document, then the interviewer uses the missing data section to capture a non-missing response for the key questionnaire item without modifying responses for the other already-completed items in the FQ, RH, and BQ sections. If the interviewer is unable to obtain the missing information, either “don’t know” or “refused” is entered in the missing data sections.

The missing data sections are:

- Facility Questionnaire Missing Data (FQ\_MD): collects data missing from the FQ section of the interview;
- Residence History Questionnaire Missing Data (RH\_MD): collects data missing from the RH section;
- Background Questionnaire Missing Data (BQ\_MD): collects data missing from the BQ section.

### *Changes to the Facility Questionnaire for 2016*

There were no substantive changes to the Facility Instrument for 2016.

## 4.3 CAPI and Case Management System Programming and Testing

MCBS interviewers receive project laptops with computer assisted personal interviewing (CAPI) software and an electronic case management system to facilitate data collection activities and questionnaire administration. Interviewers conduct the MCBS interviews using the CAPI software on the laptops and organize their cases and workload using the case management system. This section describes the CAPI and case management systems.

## *Community Questionnaire*

The MCBS Community Questionnaire used in 2016 was programmed in UNICOM® Intelligence data collection software (formerly IBM® SPSS® Data Collection or mrInterview). The software allows for full control of interviewer routing through the complex questionnaire. It uses built-in data quality measures, such as range and logic checks, dynamic text fills, and respondent exit and re-entry management. Several lookup tools are also included within the questionnaire to allow for more effective identification of some types of health insurance plans (Medicare Advantage and Prescription Drug plans), medical providers, and prescribed medicines. Throughout the questionnaire, specially formatted grid screens allow interviewers to easily reference providers, health care events, and medicines added in the current round, as well as those added in prior rounds (and preloaded into the questionnaire). In addition, screen-by-screen help text is available to assist interviewers with definitions and additional instruction.

## *Facility Screener and Instrument*

The MCBS Facility Instrument is programmed in Blaise® interview software. Unlike the Community Questionnaire, the Facility Instrument is modular, meaning the software allows the interviewer to select sections based on the interviewing situation, rather than on a set order (with some restrictions, see Section 4.2 for more information). Like the Community Questionnaire, the Facility interview includes built-in data quality checks such as range and logic checks, dynamic text fills, respondent exit and re-entry, and a prescribed medicine lookup utility. The Facility Instrument also features a facility stay history timeline.

The Facility Screener is a separate instrument programmed in UNICOM® Intelligence. This module allows for basic information about a facility to be recorded electronically and transferred to an interviewer certified to complete the facility interview.<sup>29</sup>

## *Case Management System*

The case management system facilitates management of interviewer case assignments and questionnaire administration. It is a web-based application that provides interviewers and other project staff with a consistent way to access, update, and organize case information (e.g., contact names, addresses, telephone numbers, date and location of the last interview, and optimal contact time). The system includes a portal-based case management view and a laptop-based interviewing module. Field managers and other project staff use the management portal to monitor interviewer workload and productivity. Interviewers use the laptop-based module to view their MCBS case assignments, record attempts to locate and contact respondents, update respondents' personal contact information, schedule appointments, and record case status information. The case management system is the gateway for interviewers to access the Community Questionnaire, the Facility Instrument, and the Facility Screener. Interviews are

---

<sup>29</sup> Not all interviewers may complete Facility interviews – additional training and certification is required beyond the standard Community interview training.

conducted off-line and case management and survey data are synchronized between the laptop database and the central office servers over a secure, encrypted internet connection.

Paradata elements captured within the case management system include contact level information, mode of contact attempt, source of contact information referenced (phone, address, email, etc.), and the result of the contact attempt. The case management system integrates questionnaire and case management data both within and across rounds, allowing interviewers to identify the best or most recent telephone numbers and locations for expedited contacting.

The case management system also includes the Automated Crossover Process (ACP), which automatically transfers cases from the Community component to the Facility component. The ACP creates case management updates and questionnaire preloads for these cases through a set of stored procedures, allowing interviewers to conduct an interview with the facility as quickly as one day after they located and screened the facility. Prior to Fall 2016, only Continuing cases were eligible for automated transfer from the Community component to the Facility component. Transfers for all other types of cases were completed manually. In Fall 2016, the ACP was expanded to include the automated transfer of Incoming Panel Respondents found to be residing in a facility setting.

#### **4.4 Letters and Other Respondent Materials**

A series of materials and other resources provides respondents with information about the MCBS and requests their cooperation and participation in the survey. Respondents receive letters by mail and additional materials from interviewers during their visits. In addition, a MCBS respondent website, a project toll-free number, and project email address are available for respondent communication.

Respondent materials include a variety of standard letters, such as advance letters mailed prior to the Baseline interviews and a community authority letter. This letter is sent to communicate survey legitimacy of the survey to entities such as state resources for senior citizens. Materials are tailored to whether respondents reside in the community or in facilities. In addition to the standard letter mailings, a set of contacting and refusal conversion letters are used to address common contacting problems and respondent concerns about participating in the study.

Other materials include brochures about the survey, NORC, and the income and assets questionnaire section; greeting cards; a frequently asked questions (FAQ) document; and an annual newsletter. The brochures, cards, and FAQs may be used by interviewers or managers at their discretion to assist in gaining cooperation. The MCBS respondent newsletter was provided to all beneficiaries residing in the community during the Winter 2016 interview. Finally, to assist in recording health care visits, purchases, and costs, all respondents to the Community interview in the fall round receive a planner booklet with a calendar that allows them to record health care events and other appointments and costs.

## 5. INTERVIEWER RECRUITMENT AND TRAINING

### 5.1 Interviewer Recruitment and Staffing

A professional interviewer staff is required to complete in-person interviews throughout the year. In 2016, most MCBS interviewers were experienced and had conducted MCBS interviews for at least a year or more. Some new-to-MCBS interviewers were recruited to replace those who had left the project; hiring is targeted based on local staffing needs and MCBS-specific skill requirements. The set of preferred skills included experience with financial data and complex surveys; language skills; working with individuals who have hearing, visual, or cognitive challenges; and experience interviewing people with disabilities and the elderly. In 2016, over 200 interviewers worked on the MCBS.

### 5.2 Interviewer Training Programs for 2016

The 2016 MCBS Training Program included in-person and remote trainings, customized by level of interviewer experience (new-to-MCBS or MCBS-experienced), interview component (Community or Facility), sample type (Incoming Panel or Continuing), and season-specific requirements (new questionnaire modules or data collection protocols). The program was structured to expose all field staff to the same training content, ensuring that the performance of data collection responsibilities was standardized, methodical, and measurable.

Remote trainings targeted MCBS-experienced interviewers in advance of each round of data collection. In-person trainings educated new-to-MCBS interviewers on the project's background and purpose, preparations for gaining cooperation under various circumstances, and proper administration of the MCBS questionnaires to both Incoming Panel and Continuing respondents.

In-person interviewer training programs were implemented to instruct two different types of interviewer staff: a subset of experienced MCBS interviewers identified to be trained on the Facility interview and new interviewer staff recruited to administer the Community Questionnaire.

## 6. DATA COLLECTION

NORC and CMS are committed to protecting respondent confidentiality and privacy, and both organizations diligently uphold provisions established under the Privacy Act of 1974, the NORC Institutional Review Board (IRB), the Office of Management and Budget (OMB), and the Federal Information Security Management Act of 2002. As such, MCBS data collection activities include a set of approved procedures designed to guide outreach and questionnaire administration with beneficiaries across three rounds of continuous data collection each year. In-person data collection is facilitated through a series of protocols that define eligibility for the survey, provide instruction for questionnaire administration by round and component (Community and Facility), and establish rules for how to conduct the interview within a given round. Quality control procedures are also instituted to ensure high quality data are collected.

## 6.1. Clearance

### *OMB Approval*

The Office of Management and Budget (OMB) regularly re-authorizes data collection for the MCBS. In January 2016, in anticipation of changes required for Summer 2016, CMS submitted a non-substantive change to the approved collection (OMB control number 0938-0568, expiration date 7/31/17) for questionnaire revisions, including the addition of items to the Preventive Care (PVQ) section on the shingles vaccination and pneumonia vaccination; the revision of some items in the Income and Assets (IAQ) section to accommodate deceased respondents and/or partners; and the addition of items to the IAQ section on lump sum payments. OMB clearance was received on February 10, 2016.

In spring 2016, CMS submitted a revision to the approved collection. This included modifying the sample design to achieve additional Hispanic completed cases beginning in Fall 2016. The questionnaire revisions included adding a new section on Use of Nicotine and Alcohol (NAQ) to incorporate current measures of nicotine and alcohol use; updating the Health Status and Functioning (HFQ) section to use more current terminology and improve mental health measures as specified in the DHHS standards; reducing the number of items asked in the Patient Perceptions of Integrated Care/Usual Source of Care (PPIC/USQ) section; moving some items from HFQ into Preventive Care (PVQ) and updating terminology; and adding a question to the Demographics and Income (DIQ) section that measures English literacy among respondents with Limited English Proficiency. The clearance was obtained on June 27, 2016, with an expiration date of June 30, 2019.

### *IRB Approval*

The NORC Institutional Review Board (IRB) reviews and approves all MCBS data collection protocols, questionnaires, and respondent materials to ensure human subject protections are properly addressed before field data collection began. For 2016 data collection, the MCBS research protocol and consent procedures were approved by NORC's IRB in July 2014, with subsequent changes to the protocol approved through amendments and annual renewal for 2016.

## 6.2. Data Collection Process and Procedures

The MCBS data collection process includes a timeline to fulfill three continuous rounds of annual data collection. MCBS data collection procedures define how beneficiaries are contacted, determine when a MCBS beneficiary is eligible to participate, and include protocols designed to facilitate longitudinal data collection, establish contacting rules, and maintain beneficiary participation throughout twelve rounds over a four year period.

### *Data Collection Schedule and Timeline*

The annual MCBS fielding schedule includes three rounds of data collection, with the winter and summer rounds typically lasting 16 to 17 weeks and a slightly longer fall data collection round of

24 weeks. The fall round is scheduled as a longer data collection period to accommodate contacting and interviewing efforts with the Incoming Panel.

In 2016, Winter 2016 (Round 74) data collection started January 13, 2016 and concluded April 30, 2016; Summer 2016 (Round 75) data collection started May 3, 2016 and concluded August 31, 2016. Fall 2016 data collection started July 25, 2016 and concluded January 8, 2017. Thus, data collection represented in 2016 annual files includes data collected from January 13, 2016 through January 8, 2017, with a reference period start date of January 1, 2016 for health care event data.

### *Sample Releases and Preloads*

For a given round, MCBS data collection is structured around several case releases. This is primarily due to the cyclical nature of fielding the MCBS as a continuous longitudinal survey. For members of Continuing panels, questionnaire data from the prior round need to be cleaned using structure, logic and reasonableness checks, edited, and preloaded before a case is released into production for the next round (see Chapter 7: Data Processing and Data Delivery for more information). Continuing cases are staged and released in batches scheduled throughout the data collection round.

### **Contacting Efforts and Outreach Rules**

Given the longitudinal panel design of the MCBS, it is imperative that sampled beneficiaries engage with the study throughout the 12 rounds of data collection to minimize non-response bias and the impact of sample attrition over time. After 2015, recall that the MCBS data collection design no longer follows a beneficiary who misses two consecutive rounds of data collection. While beneficiaries can miss a single round, non-completion of an interview in a previous round can lead to long recall periods and less complete information collected. Various data collection strategies are used to limit respondent burden, strengthen the beneficiary's commitment to the survey and maximize response rates across rounds.

### **Contacting Protocols**

During each case release, interviewers receive case assignments for contacting and questionnaire administration. Interviewers are trained to establish contact with respondents using guidelines on the frequency and type of contact, typically starting with initial contacts to introduce the survey and gain cooperation, schedule an interview and administer the questionnaire.

Following CMS guidance, and shown in Exhibit 6.2.2, interviewers use contacting strategies that promote efficiency and ensure continuity in contacts across all beneficiaries actively fielded during a given round. The contacting effort required often corresponds to the number of rounds a beneficiary has previously participated. For example, greater effort, in terms of the number and types of contacts made, is invested in contacting the Incoming Panel beneficiaries in the first-interview fall and second-interview winter rounds as activities, such as locating, gaining cooperation and establishing familiarity with the MCBS, are often required. Contacting efforts for the 3<sup>rd</sup> through 12<sup>th</sup> interviews typically require a reduced number of attempts necessary to make contact with respondents by phone and schedule appointments to interview respondents in person.

**Exhibit 6.2.2:** MCBS Contacting Guidelines by Interview Round

	<b>Interview 1</b>	<b>Interview 2</b>	<b>Interviews 3-12</b>
Attempts made prior to initial contact with beneficiary, designated proxy, or facility staff	10 attempts	8 attempts	4 attempts
Contact attempts after initial contact to secure appointment	8 attempts	8 attempts	4 attempts
Visits to complete interview	2 visits	2 visits	2 visits

**Case Management**

Interviewers access their case assignments using a case management system. This system collects and displays primary contact information, contacting histories and key elements that describe case status which interviewers use to facilitate efficient outreach and questionnaire administration in a secure and standardized manner. They also use the case management system to update contact information, describe and classify outcomes of contact attempts and launch the CAPI questionnaires. This information is synchronized with central office databases for reporting and data processing tasks. See Section 4.3 for more information about the case management system.

The case management system also houses historical summaries of previously reported utilization and cost records captured during past interviews. These summaries are produced for all Continuing Community cases and are used by interviewers to prepare for the interview. They include information such as previously reported medicines, previously entered insurance statements, previously reported utilization without associated costs collected, and summaries of utilization events reported during the last interview.

*Beneficiary Eligibility for MCBS Survey*

Eligibility to participate in the survey depends upon a number of factors encountered throughout the four years of panel participation. Changes in survey eligibility are generally identified either by the interviewer while attempting to contact the beneficiary in a given round, or from Medicare program eligibility updates reported by CMS on a regular basis throughout the year. Factors that impact whether future interviews will be conducted include whether beneficiaries are deceased, have lost Medicare entitlement, have relocated outside of PSU boundaries, or are no longer fielded due to *Not-in-Round* case finalization rules.

**Recently Deceased.** Sampled beneficiaries reported as deceased during data collection are finalized as *Complete-Deceased* at the end of the round. The standard data collection procedure for a beneficiary reported as having died at any point between the 2<sup>nd</sup> and 12<sup>th</sup> interview is to attempt an interview with a proxy in order to collect utilization and cost data between the date of the last interview and the beneficiary's date of death. A proxy completes the questionnaire in the

Community setting or a final interview is completed at a facility before the case is finalized and no longer contacted in future rounds.

Fielding procedures are also in place to handle Incoming Panel beneficiaries reported as deceased. The date of death reported and the beneficiary's enrollment year are key drivers for determining when an interviewer pursues a proxy interview during the first and second interviews. Any Incoming Panel beneficiary reported as deceased who became eligible for Medicare prior to the Incoming Panel year (e.g., for 2016, any Incoming Panel beneficiary who enrolled in Medicare prior to 2016) is finalized as deceased without pursuing a proxy interview. Any Incoming Panel beneficiary reported as deceased who enrolled in Medicare during the same year (e.g., for 2016, any Incoming Panel beneficiary who became eligible for Medicare in 2016) is fielded for a proxy interview before being finalized as deceased. These rules apply to any Incoming Panel beneficiary who is reported as deceased at any point during the Incoming Panel year. This also impacts fielding considerations in the second round winter interview.

**Lost Medicare Entitlement.** Beneficiaries are no longer eligible for participation in MCBS after Medicare entitlement is lost. The CMS uses enrollment records to provide periodic updates for beneficiaries selected to participate in the MCBS who have lost entitlement. These updates are compared with current round case management status to determine fielding procedures. If entitlement is lost while a case is being fielded as part of the Incoming Panel (first round interview), the case status is finalized as "*Ineligible for Contact*." If the beneficiary has lost entitlement during the data collection round for any Continuing interview, an interview attempt is made in order to collect utilization and costs associated with the period of time when the beneficiary still maintained coverage. At the end of the Continuing round, the case is finalized as "*Lost Entitlement*" and is no longer fielded in future rounds.

**Beneficiaries Who Move Outside of Sampled PSUs.** Consistent with fielding rules from past MCBS data collection rounds, if a beneficiary permanently moved or relocated more than 30 miles outside of MCBS sampled PSU boundaries, the case is finalized as *Moved out of Area* and not fielded in future rounds.

**Case Finalization and Holdover Consideration for Fielding Next Round.** Each actively fielded case is assigned a final disposition to represent the status of the case at the end of a round. Any case without a completed interview is reviewed by field management and assigned a final disposition to reflect the not-in-round status. Cases assigned not-in-round status such as final refusal or final unlocatable are no longer fielded in future rounds. The majority of beneficiaries finalized as not-in-round are no longer fielded in future rounds.

**Holdover Rules for Participation.** A beneficiary must have data collected for at least two-thirds of a year to be included within the annual delivery files. For data collection purposes, any respondent finalized as not-in-round for two consecutive rounds is no longer considered eligible for participation. However, to ensure participation can continue for beneficiaries unavailable in a present round but likely to participate in the future, a holdover process is used to prepare the case for fielding in the subsequent round. For example, a beneficiary could be away for an extended family visit; a beneficiary could be staying at a second home not in the area; or a beneficiary could have canceled appointments but without seeming to be a hard refusal. Cases

meeting similar criteria are finalized as "*Unavailable this Round*" and are staged for fielding in the following round.

### *MCBS Data Collection Protocols*

A primary objective of the MCBS is to collect complete information about medical care, services, and costs for each beneficiary residing in a community or a facility setting across all twelve data collection rounds. To facilitate collecting a full and complete picture of beneficiary utilization and costs, data collection protocols are used to ensure the proper mode of administration, to conduct the interview in the correct setting, and to identify rules for who respond on behalf of the beneficiary to complete the interview.

### **Community Questionnaire Administration**

The Community Questionnaire is administered in person. Longstanding MCBS protocols have required that Incoming Panel beneficiaries be contacted and interviewed in person. This approach ensures survey legitimacy is established early on and allows the interviewer to establish rapport with the respondent, provide context for future rounds, and introduce materials in support of future rounds.

A key goal of Continuing interviews involves associating health care events with costs and payments. In preparation for the future rounds, interviewers provide respondents with a calendar and instructional aid that reminds them to document medical events and save any Medicare or insurance statements and any other health care-related paperwork received after the date of the current interview. During the subsequent round, interviewers review calendars with respondents, as well as sort and match any hardcopy documentation associated with past reported medical events, such as Medical Summary Notices (MSNs), explanation of benefits (EOBs) and other supplemental insurance forms, and medicine summaries. Interviewers are trained to match these documents into charge bundles to ensure streamlined entry within the questionnaire (see Section 4.1 for more information on how these statements are used during the Cost Series).

The only MCBS questionnaire that is administered by telephone is the shorter 12<sup>th</sup> round interview for Community respondents exiting the survey. This final interview does not collect utilization and cost information, so in-person documentation sorting and matching is not required.

### **Facility Component Interviewing**

If a beneficiary spent time in both the community and a long-term care facility during a given round of data collection, both community and facility interviews may be administered to ensure that continuous records are obtained for the entire reference period. Prior to conducting a facility interview, a potential facility must be screened to ensure the facility meets the MCBS facility definition.

## MCBS Definition of a Facility

For the MCBS, a facility interview is conducted when the beneficiary resided in a long-term care or other residential facility with three or more beds that meets the following conditions.

- Certified by Medicare as a Skilled Nursing Facility (SNF); or
- Certified by Medicaid as a Nursing Facility or an Intermediate Care Facility for the Mentally Challenged; or
- Licensed as a Personal Care Home, Board and Care Home, Assisted Living Facility, Domiciliary Care Home or Rest Home by a state or local government agency; or
- Provides 24 hours a day, 7 days a week supervision by a person willing and able to provide personal care; or
- Provides personal care services to residents (personal care may include assistance with eating, dressing, walking, preparing meals, etc.).

If a facility does not meet the above definition, or if the beneficiary does not reside in the section of the facility that provides long term care, then a Community Questionnaire is administered to collect the data.

Most beneficiaries who reside in a place that meets the MCBS definition of a facility live in a type of nursing home. Other qualifying facilities include institutions for people with mental disabilities, domiciliary or personal care homes, retirement homes, mental health facilities, assisted living, board and care homes, rehabilitation facilities, and group homes.

Institutions such as jails and prisons do not meet the MCBS facility definition. The Facility Screener and the Facility Questionnaire (FQ) section, the first section within the Facility Instrument, are used to confirm that a facility meets the MCBS definition. The Screener and FQ work in tandem to determine whether a case is eligible for the Facility component.

## Facility Screener

When an interviewer learns that a beneficiary who was previously residing in the community has moved into a facility, or a beneficiary who was residing at a facility has moved to a new facility, the interviewer determines whether the new facility meets the MCBS definition of a facility and therefore is eligible for the Facility interview.

As a first step in determining eligibility for the facility interview, the interviewer administers a Facility screener over the phone to a facility contact. The Facility Screener serves to confirm the beneficiary has lived in the facility during the reference period, identifies the current location of the beneficiary, and verifies the location of the facility and relevant contact information.

## Facility Instrument Administration

Unlike in the Community component, interviewers never directly administer the questionnaire to the beneficiary during a Facility interview. Instead, the interviewer administers the questionnaire to staff at the facility, referred to as “facility respondents,” who answer questions about the beneficiary. It is common for field interviewers to interview more than one person at the facility because different staff at the facility have the most complete information for specific sections of the questionnaire.

Much of the content of the Facility interview can be found in medical documentation. Therefore, facility staff may refer to records, such as the beneficiary’s medical chart, during the interview. Further, facility staff may allow the interviewer to abstract responses directly from medical records. The extent of abstraction conducted varies greatly by instrument section, facility structure, and number of events occurring at the facility on the day of the interview. Exhibit 6.2.3 shows the percentage of cases for which interviewers reported abstracting data by section as reported at the conclusion of each questionnaire section for which such data are available.<sup>30</sup>

**Exhibit 6.2.3:** Interviewer-Reported Abstraction Rates by Facility Instrument Section

Questionnaire Section	Reported Abstraction in Winter 2016	Reported Abstraction in Summer 2016	Reported Abstraction in Fall 2016
Background (BQ)	17%	21%	7%
Prescribed Medicine (PM)	53%	49%	52%
Residence History (RH)	6%	7%	4%
Health Status (HS)	24%	27%	17%
Use of Health Services (US)	24%	23%	22%

### *Crossover Definitions and Procedures*

If a beneficiary spends time in both the community and a long-term care facility during a given round of data collection or since the date of the last interview, both Community and Facility interviews are staged for administration to ensure that continuous records are obtained for the entire reference period. Crossovers are cases that have moved into a new setting since the last interview.<sup>31</sup> In a crossover situation, because the beneficiary has spent part of the reference period in more than one setting, interviewers complete two separate questionnaires to collect data from both locations.

<sup>30</sup> Data regarding abstraction in the Expenditures (EX), Insurance (IN), and Facility Questionnaire (FQ) sections are not collected.

<sup>31</sup> Crossovers do not include respondents that have moved, but remained within the Community setting.

Incoming Panel cases in Winter and Summer 2016 had a different policy that depended on when the beneficiary entered the new component and when s/he gained Medicare entitlement. All other crossover cases in their 3<sup>rd</sup>-11<sup>th</sup> interviews follow the crossover procedures outlined below.

**Community-to-Facility Crossover.** When a contact attempt with a Continuing Community beneficiary leads to the discovery that the beneficiary moved into a facility since the last interview, a Community-to-Facility crossover occurs. An interviewer first attempts to administer the Community interview to a proxy followed by administering the facility screener to staff at the facility where the beneficiary is residing. Once the facility screener confirms that the facility meets the MCBS definition, an appointment is scheduled to conduct the facility interview. An automated crossover process for staging a Facility Questionnaire allows both the Community and Facility Questionnaires to be fielded within the same round.

**Facility-to-Community Crossover.** When contact with a facility where a Continuing beneficiary was residing during the last interview indicates that the beneficiary moved back to the community setting, a Facility-to-Community crossover occurs. An interviewer first administers the Facility interview with the original facility to cover utilization and costs from the date of the last interview through the time of the move into the Community. The interviewer also collects information such as the date the beneficiary left the facility as well as the beneficiary's current community residence. The Community interview will be administered in the following round.

**Facility-to-Facility Crossover.** When contact with a facility where a Continuing beneficiary was residing during the last interview indicates that the beneficiary moved to another facility since the date of the last interview, a Facility-to-Facility crossover occurs. An interviewer first administers the Facility interview with the original facility to cover utilization and costs from the date of the last interview through the time of the move into the new facility. The interviewer then collects the required facility screener information for the case to be fielded in the second Facility setting. This second facility will be contacted in the following round.

### *Proxy Interviews and Assistants*

Beneficiaries often require assistance in providing the detailed information needed to accurately respond to survey items. During the course of data collection, the beneficiary may designate a proxy to participate in the interview on his or her behalf or an assistant to provide help when responding to specific survey questions.

### **Proxies and Assistants**

A proxy is a person, generally designated by the beneficiary, who is sufficiently familiar with the beneficiary's health care events and costs and responds on behalf of the beneficiary. In addition, a proxy completes a Community interview when a beneficiary is no longer able to participate, including when a beneficiary died since the date of the last interview, or has entered a Facility setting. In 2016, the percent of interviews completed by a proxy ranged from 10-12% depending on the round.

An assistant helps the beneficiary answer specific questions, but unlike a proxy, an assistant does not answer all questions on behalf of the beneficiary. The assistant is chosen by the beneficiary to help in situations where the beneficiary could respond to the interview as long as he/she received some help from another knowledgeable person. Some examples of this are where a spouse or partner manages the Medicare statements for the household or maintains a calendar of medical visits and appointments. The percent of interviews completed with the help of an assistant in 2016 ranged from 14-19%, again depending on the round.

### Criteria for Proxy Selection

During Community Questionnaire administration, all beneficiaries are asked to identify a person or persons best able to provide information about health care visits and the costs of any health care the beneficiary may receive should the beneficiary not be able to complete a future interview. For Continuing round interviews, the named proxy is in the case management system, along with information indicating if a proxy completed the interview in the prior round. Community interviews conducted with proxies follow a slightly different path than those administered directly to the beneficiary (see Section 4.1 for the Community Questionnaire flow).

When initial contacts with Incoming Panel beneficiaries suggest possible comprehension or physical impairments that would make the interview difficult, interviewers work with their managers to determine if an assistant or proxy is necessary, and who an appropriate person would be to serve as a proxy or assistant.

### *Interviewing Languages*

The Community Questionnaire is programmed for administration in English or Spanish. The Facility Instrument is available for administration in English. Approximately 5 percent of Community interviews were conducted in Spanish in 2016. This includes interviews conducted in Puerto Rico, all of which are completed in Spanish.

Bilingual field interviewers are trained to administer the Community Questionnaire in both English and Spanish. The language of administration is captured within the questionnaire. In rare instances in which the beneficiary speaks a language other than English or Spanish, the interview is conducted in English with an English-speaking proxy or assistant acting as an interpreter for the beneficiary.

### *Questionnaire Breakoffs*

Interviewers are able to suspend the interview prior to completion while administering both the Community and Facility Questionnaires. This break-off feature provides flexibility to address schedule constraints, technical issues, and other extenuating circumstances that prevent completion of the interview in one sitting. Once restarted, the CAPI resumes at the screen of the last question administered. If a questionnaire is broken off, it must be fully administered before the end of the round to count as a completed interview. If the suspended questionnaire is never completed, it is finalized as a *Final Breakoff* at the end of the round (see Chapter 8 for more information on weighting and imputation procedures).

### 6.3 Data Collection Results

An interview is considered to be complete once administration of all questionnaire sections to the respondent has concluded, the Interviewer Remarks Questionnaire (IRQ) is completed, and data are fully transmitted. In 2016, the mean length of winter round community interviews was about 79 minutes, while the mean length of summer round community interviews was about 75 minutes. The mean length of fall round interviews was about 81 minutes for the Incoming Panel and approximately 104 minutes for Continuing beneficiaries.

Exhibit 6.3.1 provides the count of completed interviews by round and component for 2016. Detailed information on response rates can be found in Chapter 9.

**Exhibit 6.3.1:** 2016 Completed Interviews by Component

Round	Component	Completed Interviews	Mean Interview Duration
<b>Winter 2016</b>	Community	10,984	79.0
	Facility	937	60.9
<b>Summer 2016</b>	Community	7,524	75.3
	Facility	726	53.4
<b>Fall 2016</b>	Community	12,952	92.8
	Facility	1,099	62.7

### 6.4 Data Collection and Quality Control

To ensure the collection of high quality data, several quality control procedures are conducted including systematic review of questionnaire data and case management paradata, follow-up contacts with respondents, and ongoing interviewer coaching. Systematic review of interview recordings and ride-along observations of in-person interviews are used to directly observe interviewer interaction with beneficiaries and provide feedback. Verification phone calls and review of survey data are also conducted to validate interviewer performance.

The systematic monitoring and evaluation of interview performance and verification is primarily conducted via digital computer-assisted recorded interview (CARI) recordings. A subset of questionnaire items are recorded with respondent consent. By listening to a random sample of CARI recordings, supervisors identify areas where interviewers require correction in administration, stress the improvement of interviewer techniques to add clarity or minimize potential bias, and emphasize standardization in approach and administration. Any serious deviations from protocol or data quality concerns are reviewed for corrective action in consultation with field management.

Data review procedures are also enacted to identify any systematic CAPI issues resulting from the data collection effort. In 2016, data review procedures consisted of two components: review of survey data within the preload data cleaning process, and review of metadata to assess interviewer performance. Because the Continuing interview by design is highly dependent upon data collected in prior rounds, a multistep cross-team process is used to review questionnaire data prior to preloading for the next data collection round (see Section 7.1). The data cleaning process, including structure, logic and reasonableness checks, informs future questionnaire development as well as additional training and follow-up.

Finally, field managers periodically contact respondents throughout the round to verify the interview was conducted, confirm the interviewer was present, and collect administration information. When necessary, field managers use CARI reports and data review feedback that indicate potential quality issues to prioritize follow-up contacts in order to collect additional information for coaching purposes.

## 7. DATA PROCESSING AND DATA DELIVERY

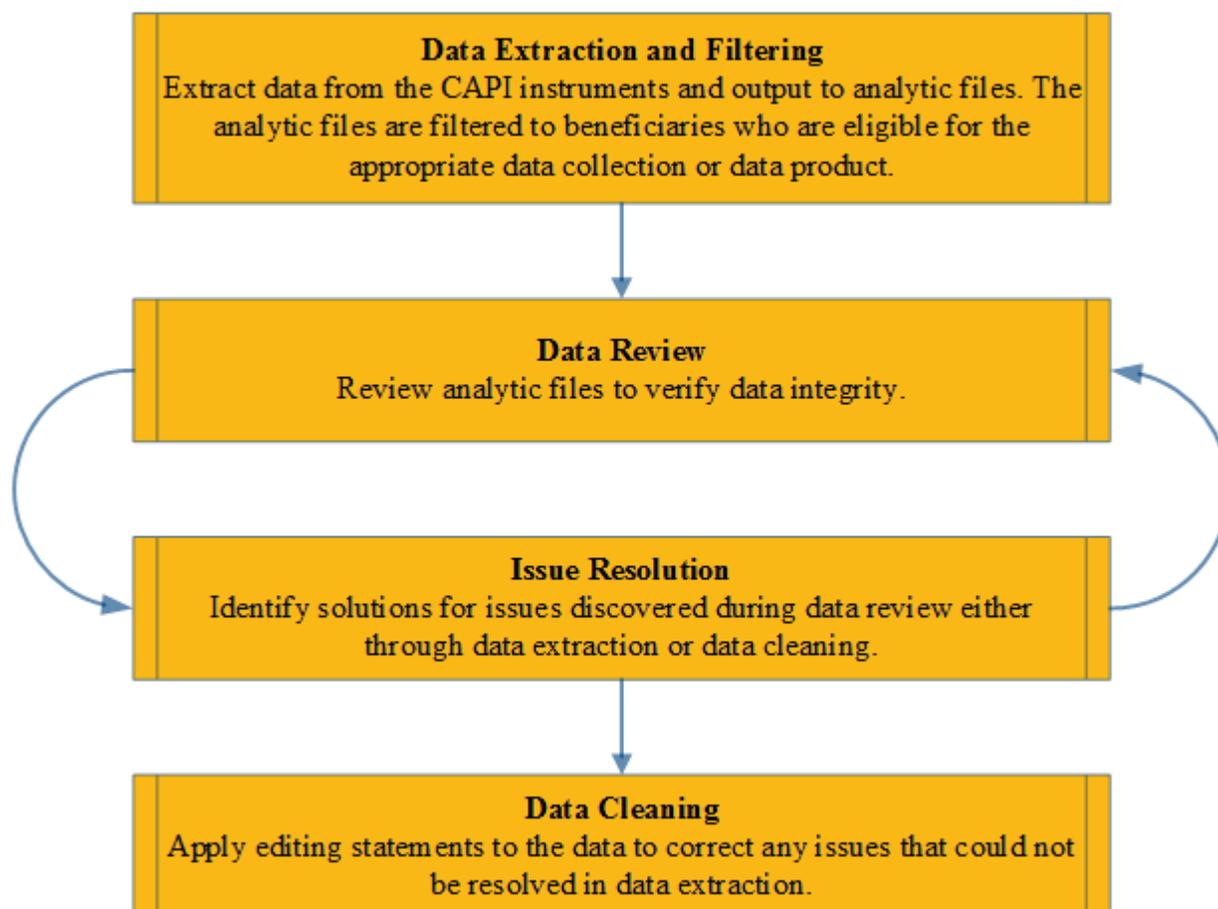
Longitudinal data collection requires both interim and final post-processing of the data in order to prepare them for release. These activities include data editing, for both preloading subsequent round instruments and final file production, data concatenation and reconciliation for the annual data products including the Survey file and Cost Supplement files, and the development of other post-processing inputs to the files. This chapter describes both the data editing process and the annual data concatenation and reconciliation process.

### 7.1 Data Processing Overview

During the interview process, beneficiaries may provide information that is either incomplete or inconsistent (internally or with administrative records). These data require further processing to ensure the highest quality of estimates produced from the MCBS. The processing may involve resolving inconsistencies using logical methods or utilizing imputation techniques, where appropriate, to fill in missing information. Thus, CAPI data are reviewed and processed for three primary purposes: Community and Facility Questionnaire preloads, the 2016 Survey File, and the 2016 Cost Supplement File. The same types of data review and processing protocols are used for each effort, with different source instruments and editing protocols. This section provides an overview of data review and processing procedures. The sections that follow will provide further description of the data cleaning efforts.

#### *Process Description*

Exhibit 7.1.1 illustrates steps in and iterative nature of the data review and editing process.

**Exhibit 7.1.1:** Data Review and Editing Process

**Data Extraction and Filtering.** At the conclusion of data collection in each round, data are extracted from the raw Community and Facility CAPI questionnaires and transformed into SAS analytic files for further processing. This extraction includes the development of appropriate questionnaire metadata. Prior to data review, the individual records and associated analytic files are limited to beneficiaries who are deemed eligible for the appropriate data collection or data product.

**Data Review and Issue Resolution.** Given the complexity of the data structure, the analytic files undergo column and row checks to confirm each individual analytic file is structurally sound. Column checks include confirming that all necessary variables are on the file, checking variable attributes, and identifying high rates of missingness or out of range values. Row checks include confirming the inclusion of expected beneficiary IDs and checking for duplicate or missing linking variables. Structural issues discovered during this process may reinitiate the data extraction process or may be resolved in data cleaning.

Logic and reasonableness checks follow for each analytic file. Logic checks are used to verify that the questionnaire worked as expected, particularly with respect to questionnaire routing and skip logic. The complexity of the event and cost questionnaire sections require particular attention to

the CAPI routing routines that generate analytic files that are specific to these portions of the questionnaire. Errors identified during logic checking result in two types of data edits: flagging values that were incorrectly skipped or setting incorrectly populated values to null to indicate a valid missing.

Furthermore, unreasonable or impossible values are identified and checks for values that are not explicitly disallowed by the questionnaire are identified for reasonableness. For example, in the Community Questionnaire, male beneficiaries should not report female-only conditions, like cervical cancer. Beginning in 2016, continuous variables are reviewed to identify illogical extreme values. For instance, in the Community Questionnaire, the number of living children reported by the beneficiary cannot exceed 20. Based on the results of this data review, edits are developed to either set the unreasonable or impossible value to a logical value or an inadmissible code during data cleaning.

Beginning in 2016, other specify responses for a subset of open-ended variables are reviewed and backcoded into existing codeframes when possible.

**Data Cleaning.** Once the data review and issue resolution steps are complete for each analytic file, data cleaning routines are implemented. During data cleaning, edits are applied to the analytic file and additional quality control (QC) is conducted to ensure that the edits are applied correctly.

## 7.2 Preload Editing and File Production

This section describes Community Questionnaire and Facility Instrument preload production, including the purpose of preloads, examples of preloaded variables, and a general description of timeline and processes. The preload process feeds back questionnaire data from previous rounds' interviews and populates the Community and Facility CAPI Instruments to help drive data collection in the subsequent round. Preloaded data serve to both forestall asking MCBS respondents the same questions in subsequent rounds and to act as the basis for collecting additional information about a medical event, insurer, or associated financial cost or payment. As the data must be loaded into an active CAPI instrument available to interviewers, it requires that the preload data are in a form that is recognized by the case management system, which supplies it to the Community Questionnaire and Facility Instrument in the field. Preloaded information is used to determine questionnaire routing and text fills.

For example, if a beneficiary previously reported having ever smoked cigarettes in his/her lifetime, the questionnaire can then use this information in a subsequent round to probe if the respondent is still smoking. The logic within the questionnaire that determines whether such a question is asked in the next round is driven by preload variables set during the preload process. Examples of preloaded data included information on health plans, medical events, insurance claims, prescription medicines, household members, facility characteristics, and facility stay history.

Preloads generally fall into two categories: direct response data and derived variables. Direct response data are raw questionnaire responses generated in one round that are passed through to the next round. For example, the list of a beneficiary's medical care providers are passed from

one round's Community Questionnaire to the next via the preload process. Similarly, facility name and address are passed from one round's Facility Instrument to the next.

Derived variables require modification of the source data before being preloaded into the next round. Some modifications are quite complex and many derived variables have a significant impact on questionnaire functioning. Examples of derived variables include sample type assignments, Facility and Community Questionnaire reference dates, and the reason a cost is sent through Charge Payment Summary (CPS reason) (see Section 4.1 for more details on this questionnaire section).

### *Community and Facility Preload Timeline*

Preload creation is an ongoing process with a rigid schedule as each round lasts only four months. The MCBS is continuously fielded; thus, the preloads must be ready for the next round of data collection to be conducted.

Each round of data collection in 2016 contained multiple waves of sample release that required preloads. The first wave is the Continuing Community Release 1, which contains beneficiaries who completed the prior round's Community Questionnaire within the first half of the data collection period, while Continuing Community Release 2 beneficiaries completed in the second half. There is one Continuing Facility Release each round. This release includes beneficiaries that either completed the prior round's Facility Instrument or crossed over from the Community Questionnaire to the Facility Instrument because the beneficiary continues to live in a facility or has entered a facility in the new round.

### *Community and Facility Preload Process Description*

The Community Questionnaire and Facility Instrument preload creation processes consist of five steps: data extraction and filtering, data review, issue resolution, data cleaning, and rollover. The first four steps were described in Section 7.1. The final phase of preload creation, which was not described previously, is the rollover process. After data review and editing occur, datasets are constructed with the data required for preloading. Key items set during the rollover process are the derived variables that assign sample type, reference dates, and CPS reason. Sample type assignment is based on previous interview history, including whether respondents missed the previous interview, crossed over from one component to the other, or are in their first year of the MCBS. This information is used to determine which questionnaire sections and items are administered and to set the reference dates for questionnaire items. Reference dates are used in the Community Questionnaire and Facility Instrument to define the time periods about which data will be collected in the upcoming round. There are a number of reference dates that are derived from the dates of the respondents' prior interviews. CPS reason determines which medical costs are collected in the Community Questionnaire based on whether the respondent has a billing statement for that item and whether the total charges were accounted for in previous rounds.

The rollover process, which is designed to ensure that all of the preload data are loaded properly into the questionnaire, occurs before every sample load in a round and between rounds. The

eligible population for each subsequent round is determined by examining case dispositions in the current round.

Thorough quality control steps, including ensuring the data types, dates, and variable definitions are appropriate, are conducted to ensure that preloaded data are successfully created according to the round-based specifications. The preload data needs to be in the specified format acceptable to the case management system, which then makes the preload data available to be called into the Community Questionnaire and/or Facility Instrument for the upcoming round.

### 7.3 MCBS 2016 Survey Data File

The 2016 Survey File release is built from 36 analytic data files encompassing Community and Facility data collection from five rounds of data including Winter 2016, Summer 2016, Fall 2016, Winter 2017, and Summer 2017. These files are input into CMS processes that generate the final data files available to the public. Further detailed descriptions of these final Survey Data files are available in the 2016 Data User Guide. This section describes the eligibility criteria for the analytic files, file preparation, and file contents.

#### *File Eligibility Criteria*

The Survey File LDS contains data collected in Fall 2016 and topical survey data collection from Winter and Summer 2017, which are about experiences during 2016. The criteria for inclusion in these analytic files include: beneficiaries continuously residing in the Community or Facility, beneficiaries who move between a Facility and the Community, data which are collected from proxy respondents for deceased beneficiaries, or data from individuals who lost entitlement to Medicare. A beneficiary only needs to have completed a Community or Facility interview in one of the data collection rounds of interest to be included in these analytic files.

#### *File Contents*

##### **Community**

There are two subcategories of Community analytic files. The Community Continuing Questionnaire analytic files contain data collected in questionnaire sections critical to the purpose of the MCBS. Core data are collected in each round of an annual data collection cycle. The Community Topical Questionnaire analytic files contain data collected in questionnaire sections that cover special interest issues. Topical data are collected only in selected rounds. See Chapter 4 (Exhibits 4.1.4 and 4.1.6) for a list of the Community Questionnaire sections included in each data file.

##### **Community Continuing Questionnaire**

The Community core content analytic files included in the 2016 Survey File contain information about access to medical care, health status and functioning, health insurance plans, medical providers, and income and assets. The Survey File analytic files do not include survey-reported cost, healthcare utilization, or case management data.

## Community Topical Questionnaires

The Community Questionnaire includes sections that are focused on specific topics of interest, such as mobility of beneficiaries and preventive care and drug coverage. The 2016 Survey File contains data from some topical sections that were administered in the Winter 2017 (Round 77) and Summer 2017 (Round 78) round but have reference periods for 2016. These files are processed in combination with the 2016 Survey File deliveries and as a result, topical analytic files are considered part of the MCBS 2016 Survey File.

## Facility

The Facility analytic files contain information about access to medical care, health status and functioning, health insurance plans, facility characteristics, and beneficiary characteristics. The Facility analytic files do not include cost, healthcare utilization, or case management data.

## 7.4 MCBS 2016 Cost Supplement File

The Cost Supplement File data include information on beneficiaries' medical events occurring in 2016, and the cost of those events. The Cost File LDS contains cost and utilization data collected in Winter, Summer, and Fall 2016 about utilization and expenditures occurring in 2016. The MCBS Cost Supplement File is described in the 2016 Data User's Guide.

Substantial post-processing occurs to the questionnaire items that create health care events, the costs and payments associated with those events, and the source of payments. This is the result of the way in which the annual data collection occurs. Five processes are used to create the inputs to the final data files. The five processes build annualized files, define eligibility for the Cost Supplement File, establish insurance coverage across the year and create events that are linked to defined payers and the cost of the services provided. The first four analytic processes are inputs to the claims match process that return matched events for additional post-processing and imputation. The final process, the facility stay file, combines all the steps already described for the Community Questionnaires and adds the claims match into a single step. The facility stay process then generates data files for the production of the Cost Supplement File release.

These five processes (reference period, insurance timeline, event cost consolidation, prescription medicine file, and facility stay file) are described below. The shared goals of all of these interim analytic steps are to combine data across rounds, annualize eligibility for data release, and create analytic products that can be consumed in the context of the final file production. These interim data products are not part of the final Survey File or Cost Supplement File releases. Each interim analytic product is described below.

### *Reference Period*

Reference Period uses case management data to define time periods in 2016 covered by Community and Facility survey data. Reference Period is run for all beneficiaries who had interviews in 2016, and it is used to identify the population of beneficiaries eligible to be included in the Event Cost Consolidation and Prescription Medicine File. The population identified by

Reference Period represents a subset of the population assigned a Survey File Ever Enrolled weight.

Reference Period creates a calendar history of a beneficiary's MCBS interviews. A number of eligibility checks are run against this calendar history to identify beneficiaries who had complete survey data for the calendar year, either because they were interviewed for a full year or interviewed until death or loss of Medicare entitlement. Beneficiaries who pass these eligibility checks become the population eligible for the Event Cost Consolidation and Prescription Medicine File data products.

### *Insurance Timeline*

Insurance Timeline creates a calendar history of a beneficiary's insurance plans and types of insurance coverage. The process pulls together health insurance plan data from the Community Questionnaire, Facility Instrument, and administrative records. Insurance Timeline in 2016 was produced for the same population as was assigned a Survey File Ever Enrolled weight.

Insurance plan timelines are constructed independently across these three data sources. Plans that are identical across data collection periods are collapsed into one record, with each time period identified as having definite or possible coverage by the plan. Plans identified as "Medicare HMO" in the Community Questionnaire data are linked to Medicare Advantage (MA) plans in the administrative and claims data. Finally, the timelines from each of the three data sources are concatenated. The resulting dataset allows these timelines to be examined independently or together to understand insurance coverage in the calendar year for each beneficiary. Plan coverage data from the Insurance Timeline are used downstream to define potential sources of payment in the Event Cost Consolidation process as well as to construct monthly insurance coverage records for each beneficiary.

### *Event Cost Consolidation*

Event Cost Consolidation creates a file containing health care events and their associated costs, payments, provider information, and dates of service for all health care utilization reported by or on behalf of beneficiaries residing in the community. The process matches events to reported periods of insurance coverage as summarized by Insurance Timeline to identify possible and definite sources of payment for each event. Reported charges and payments are matched before being appended to the file of events. The process then applies global editing rules to resolve partial charges and charges with incomplete cost information. Finally, records for recurring events are replicated to represent repeated instances of these events. The resulting dataset of consolidated event and cost information is used to match survey-reported events to Medicare claims. These matched results are the inputs to the Prescription Medicine and non-Prescription Medicine Imputation processes and the final Cost Supplement Files.

### *Prescription Medicine File*

The Prescription Medicine (PMED) file is a list of all prescription medicines that are collected by the MCBS. For 2016, the list included every combination of prescription medicine names, forms, and strengths provided by MCBS respondents during interviews conducted in 2016 (including a

total of five rounds between Fall 2015 and Winter 2017). It includes both medicines that were reported by MCBS respondents for the first time during one of these five rounds and refilled medications that were originally reported earlier, but updated as being currently prescribed during one of these five rounds. It only includes medicines that were reported during the Community Questionnaire administration for beneficiaries who were eligible to be included in the Cost Supplement File.

A number of cleaning steps implemented for the 2015 PMED file have been carried forward for the 2016 PMED file, including fixing common misspellings and other common errors in the verbatim fields entered by interviewers, and standardizing the spacing, punctuation, and other formatting of the prescription medicine information. This simplifies the subsequent CMS process of matching the PMED list to the First Databank list of prescription medicines, and eventually to administrative claims information.

The process of creating the PMED file includes assembling a full list of all beneficiaries' reported prescribed medicines for 2016 from the Community Questionnaire, de-duplicating it, developing and implementing cleaning rules, and then de-duplicating the list again after the cleaning process was complete. For the 2016 data, the original list of prescription medicines included nearly 46,000 records. After cleaning and de-duplicating again, the final file included just over 34,000 unique medicines.

### *Facility Stay File*

The Facility Stay File summarizes data related to facility characteristics, costs and payments, and health care utilization for interviews conducted on behalf of beneficiaries residing in facilities. The process brings in data from the Facility Instrument and reconfigures the data to create one record per facility stay during the calendar year. Medicare Claims data for inpatient hospital visits and skilled nursing facility visits are matched to Facility Instrument data to provide more accurate reporting of Medicare payments. Three imputation routines are applied within the context of the Facility Stay process to remedy missing data issues with payments, as well as edit outliers and other anomalies. The Facility Stay File population in 2016 included any beneficiaries completing one or more Facility Questionnaire interviews covering residence in an MCBS-eligible facility for one or more days in 2016.

## 8. WEIGHTING AND IMPUTATION

### 8.1 Overview

Weighting and imputation are used in surveys to enhance the usability of the data for analysis and increase the accuracy of resulting estimates. Weights are calculated to reduce potential nonresponse and sample coverage bias, ensuring that the sample is representative of the population of interest. They are especially important when particular sampling methods are in place, such as stratification, cluster sampling, and oversampling of particular populations. The MCBS employs all of these sampling methods; weights then account for the resulting differences in probabilities of selection as well as nonresponse, and also calibrate to control totals using post-stratification. Imputation is used to replace missing values of survey variables with admissible complete values and create data where they were not actually collected, allowing for the retention of observations for statistical analysis that would otherwise be excluded. MCBS imputation falls under two umbrellas that focus on imputing monetary amounts: Income and Asset (IA) imputation, and Event, Payer, and Cost imputation, which includes imputation for Prescription Medicine (PM) and Non Prescription Medicine (Non PM) events and costs. The weighting and imputation methods used for the MCBS are described in detail below.

### 8.2 MCBS Weighting Procedures

#### *Overview*

Weighting activities for the 2016 data year consist primarily of four main stages. The first is the initial weighting stage in which the members of the Incoming Panel are given base weights, and these weights are then raked to population control totals and adjusted for nonresponse at the first interview (Fall 2016). The remaining three stages of weighting each lead to delivered weights files. These are the Survey File weights, the Cost Supplement weights, and the weights for topical questionnaire sections. A listing of all of the weights for the MCBS is presented in Exhibit 8.2.1.

#### *Process*

Initial weighting requires receipt of the final combined enrollment data extracts and the finalization of the interview dispositions in the fall round of the data year (i.e., Round 76 for the 2016 data year). Survey File weighting follows initial weighting. Cost Supplement File weighting requires completion of the Survey File weighting process and the Reference Period process. Topical questionnaire modules related to the Survey File and Cost Supplement File are weighted separately as they are fielded in the winter and summer rounds following the data year.

### *2016 Initial Weighting*

In the initial weighting stage, the initial nonresponse adjusted weights for the Incoming Panel of Medicare beneficiaries, which for the 2016 data year is referred to as the “2016 Panel” or the “Incoming Panel” are derived. First, base weights are calculated based on the probabilities of selection for the beneficiaries in the panel and 100 replicate weights for use in variance estimation are created. Then, these weights are raked to population control totals. Finally, the weights are adjusted for nonresponse at the first interview in Fall 2016.

**Exhibit 8.2.1:** 2016 MCBS Data Files Summary of Weights

Limited Data Set	Description	Segment	Full-Sample Weight	Replicate Weights	Population
Survey File	Continuously Enrolled Cross-Sectional Weights	CENWGTS	CS1YRWGT	CS1YR001- CS1YR100	Continuously enrolled from 1/1/2016 through the fall of 2016
Survey File	Ever Enrolled Cross-Sectional Weights	EVRWGTS	EEYRSWGT	EEYRS001- EEYRS100	Ever enrolled for at least one day at any time during 2016
Survey File	Continuously Enrolled Two-Year Longitudinal Weights	LNG2WGTS	L2YRSWGT	L2YRS001- L2YRS100	Continuously enrolled from 1/1/2015 through the fall of 2016
Survey File	Continuously Enrolled Three-Year Longitudinal Weights	Will not be released in 2016	-	-	Continuously enrolled from 1/1/2014 through the fall of 2016
Survey File	Continuously Enrolled Four-Year Longitudinal Weights	LNG4WGTS	L4YRSWGT	L4YRS001- L4YRS100	Continuously enrolled from 1/1/2013 through the fall of 2016
Cost Supplement File	Ever Enrolled Cross-Sectional Weights	CSEVRWGT	CSEVRWGT	CSEVR001- CSEVR100	Ever enrolled for at least one day at any time during 2016
Cost Supplement File	Two-Year Longitudinal Weights	CSL2WGTS	L2YRSWGT	L2YRS001- L2YRS100	Enrolled on or before 1/1/2015 and still enrolled at any time during 2016
Cost Supplement File	Three-Year Longitudinal Weights	Will not be released in 2016	-	-	Enrolled on or before 1/1/2014 and still enrolled at any time during 2016
Cost Supplement File	Four-Year Longitudinal Weights	Will not be released in 2016	-	-	Enrolled on or before 1/1/2013 and still enrolled at any time during 2016

Limited Data Set	Description	Segment	Full-Sample Weight	Replicate Weights	Population
Survey File Topical Section	KNQ Continuously Enrolled	MCREPLNQ	KNCWT	KNC1-KNC100	Continuously enrolled in 2016 and still alive, entitled, and non-institutionalized in Winter 2017
Survey File Topical Section	KNQ Ever Enrolled	MCREPLNQ	KNEWT	KNE1-KNE100	Ever enrolled in 2016 and still alive, entitled, and non-institutionalized in Winter 2017
Survey File Topical Section	IAQ Continuously Enrolled	INCASSET FOODINS	IACWT	IAC1-IAC100	Continuously enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017
Survey File Topical Section	IAQ Ever Enrolled	INCASSET FOODINS	IAEWT	IAE1-IAE100	Ever enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017
Survey File Topical Section	PAQ Continuously Enrolled	PNTACT	PACWT	PAC1-PAC100	Continuously enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017
Survey File Topical Section	PAQ Ever Enrolled	PNTACT	PAEWT	PAE1-PAE100	Ever enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017
Survey File Topical Section	RXQ Continuously Enrolled	RXPARTD	RXCWT	RXC1-RXC100	Continuously enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017
Survey File Topical Section	RXQ Ever Enrolled	RXPARTD	RXEWT	RXE1-RXE100	Ever enrolled in 2016 and still alive, entitled, and non-institutionalized in Summer 2017

## Full-sample and Replicate Raked Base Weights

A full-sample base weight is derived for all beneficiaries in the 2016 Panel. The base weight is equal to the inverse of the beneficiary's overall probability of selection and reflects probabilities at the PSU, SSU, and beneficiary (USU) sampling stages. Let  $\pi_{k|ij}$  be the conditional probability of selection for beneficiary  $k$  given the PSU  $i$  and the SSU  $j$ , such that  $\pi_{k|ij} = \rho_{1ak|ij}$  for beneficiaries in the Hispanic sampling stratum and age group  $a$ , and similarly equals  $\rho_{-1ak|ij}$  and  $\rho_{2ak|ij}$  for beneficiaries in the non-Hispanic and Puerto Rico sampling strata, respectively, as described in Chapter 3. Then, for all selected beneficiaries, the base weights are defined by

$$W_{1ijk} = \frac{20}{\pi_i \pi_{j|i} \pi_{k|ij}}$$

where  $\pi_i$  is the probability of selection for the PSU,  $\pi_{j|i}$  is the conditional probability of selection for the  $j$ -th SSU given the PSU, and  $\pi_{k|ij}$  is the conditional probability of selection for the  $k$ -th beneficiary in the 5-percent enrollment data extract given the PSU and SSU.

Then, one hundred replicate base weights are derived from the full sample base weights, using the variance stratum, and the variance unit of the beneficiary. The variance strata and variance units are derived from the PSUs and SSUs used for sampling. For sampled beneficiary  $ijk$  as described above, the  $\alpha = 1, \dots, 100$  replicate weights for BRR estimation are defined by

$$W_{1ijk\alpha} = \begin{cases} \{\tau(H_{h\alpha} + 1) + (1 - \tau)(1 - H_{h\alpha})\} W_{1ijk} & \text{if in stratum } h \text{ and unit 1} \\ \{\tau(1 - H_{h\alpha}) + (1 - \tau)(H_{h\alpha} + 1)\} W_{1ijk} & \text{if in stratum } h \text{ and unit 2} \end{cases}$$

where  $H_{ha}$  is the associated element in a 100x100 Hadamard matrix. For calculation purposes, this can be written as

$$W_{1ijk\alpha} = 2[\tau\delta_{ja} + (1 - \tau)(1 - \delta_{ja})]W_{1ijk}$$

where  $\tau$  is a compositing factor between zero and one,  $\delta_{ja}$  is a 0-1 indicator of whether the beneficiary is in replicate half-sample  $a$  as determined by the value of  $H_{ha}$ , and  $W_{1ijk}$  is the base sampling weight for the beneficiary. A value of  $\tau = 0.85$  is used, continuing the practice used in prior MCBS years.

The full-sample and replicate base weights are then adjusted in such a way that the sum of the weights for various demographic domains are equal to pre-determined control totals based on the enrollment data extracts, through a process called "raking." The final enrollment data 5-percent extract, received in January 2017, contained additional records for beneficiaries that became eligible near the end of 2016. Due to the timing of this file, these newly-added beneficiaries were not subjected to sampling and could not be included in the 2016 Panel. This small amount of effective population undercoverage is adjusted for in this raking step. Thus, even though those beneficiaries are not eligible for sampling, they are counted in the population totals. This ensures that the weights for the 2016 Panel sum to the correct population total.

The raked full-sample weight is defined by

$$W_{2ijk} = \varphi_{ijk} W_{1ijk}$$

where  $\varphi_{ijk}$  is the raking step adjustment factor for beneficiary  $ijk$ . The raking process calibrates the weights by adjusting them to match the control totals for the first raking dimension, then for the second raking dimension, then for the third dimension, and so on, iterating until the weights perfectly match the control totals in all dimensions. The four dimensions used at this raking step are

2. Age Group (5-level)  $\times$  Sex (2-level)  $\times$  Race (2-level)
3. Census Region (4-level)  $\times$  Age Group (5-level)
4. Metropolitan Status (2-level)  $\times$  Age Group (5-level)
5. Accretion year (6-level; year of enrollment in Medicare)

This adjustment, and all adjustments mentioned in the remainder of this chapter, are made both to the full-sample weights and the 100 replicate weights.

### Initial Nonresponse Adjustments

The raked base weights for the 2016 Panel are then adjusted for nonresponse at the first interview in Fall 2016. The response statuses in Fall 2016 are determined, where a respondent is a beneficiary that is alive and entitled and completed the Fall 2016 interview. Nonresponse adjustment cells are constructed prior to performing the adjustment. First, the beneficiaries are divided into four primary adjustment cells: alive community, deceased community, facility, and Puerto Rico.

Separately within each of these main adjustment cells, response propensity models are fit using logistic regression to model the probability of response at Fall 2016 as a function of covariates derived from multiple sources. These include county-level American Community Survey (ACS) estimates, tract-level ACS estimates, county-level physician fee schedules, rural-urban and Metropolitan Statistical Area (MSA) information, and administrative and claims data at the beneficiary level. Generally, the covariates are selected into the logistic regression model using stepwise selection procedure with an entry  $p$ -value of 0.10 and a stay  $p$ -value of 0.15. Using the predicted response probabilities, beneficiaries are grouped into cells of approximately 100 each. Separately within each of these cells, a ratio adjustment is performed to distribute the weights of the nonrespondents to the respondents, where the adjusted weights are defined by

$$W_{3ijk} = \left( \frac{\sum_{ijk} W_{2ijk}}{\sum_{ijk} I(ijk \in R) W_{2ijk}} \right) W_{2ijk}$$

where  $I(ijk \in R)$  is a 0-1 indicator function indicating whether beneficiary  $ijk$  was a respondent to the first round of interviewing. In other words, the raked weights are adjusted by a factor equal to the ratio of the sum of the weights in the sample in the cell to the sum of the weights among only the respondents in the adjustment cell. The resulting weights are the initial nonresponse-adjusted weights for the 2016 panel.

## *2016 Survey File Weights*

The 2016 Survey File data were collected in Fall 2016 from beneficiaries sampled in the 2013 through 2016 annual panels. To facilitate estimation from the resulting data, five sets of full-sample and replicate weights are derived. These include the 2016 continuously enrolled cross-sectional weights; the 1-year longitudinal weights for analysis of 2015-2016 data, 2-year longitudinal weights for analysis of 2014-2016 data,<sup>32</sup> and 3-year longitudinal weights for analysis of 2013-2016 data; and finally, the 2016 ever enrolled weights. In addition to the weights, the dataset includes the panel (selection year) identifier, and variance strata and variance unit variables for variance estimation. These variance strata and variance unit variables, along with the weights, capture all of the sampling design information necessary to estimate variances and make inferences to the population of Medicare beneficiaries.

### **Composition of Sample and Populations of Interest**

The weights file includes records for beneficiaries that were sampled in the 2013, 2014, 2015, and 2016 Panels. The 2013, 2014, and 2015 panels are referred to as “Continuing panels,” while the 2016 Panel is referred to as the “Incoming Panel” as members of this sample were interviewed for the first time in Fall 2016. The Survey File weights include both continuously enrolled and ever enrolled weights in addition to the longitudinal weights. The continuously enrolled weights represent a population of beneficiaries who were enrolled continuously between January 1<sup>st</sup> of the data year and completion of the fall interview. The ever enrolled weights represent the population of beneficiaries who were ever enrolled in Medicare for at least one day at any time during the data year.

The 2016 Survey File continuously enrolled cross-sectional weights are populated for the subset of records with a completed Fall 2016 interview that are alive and entitled at the time of the interview. The resulting cross-sectional weights represent the population of beneficiaries that were continuously enrolled in Medicare from January 1, 2016, through completion of the Fall 2016 interview.

The one-year longitudinal weights are populated for members of the 2013, 2014, and 2015 panels that were continuously enrolled in both 2015 and 2016. The resulting weights represent the population of Medicare beneficiaries that enrolled on or before January 1, 2015, and are still alive and entitled as of completion of the Fall 2016 interview. The two-year longitudinal weights are populated only for members of the 2013 and 2014 panels who were continuously enrolled in each of the years 2014, 2015, and 2016. The population represented by these weights is the population of beneficiaries enrolled on or before January 1, 2014, and surviving and entitled as of completion of the Fall 2016 interview. Finally, the three-year longitudinal weights are populated only for members of the 2013 panel that were continuously enrolled during all of the years 2013-2016. The resulting weights represent the population of Medicare beneficiaries that enrolled on or before January 1, 2013, and are still alive and entitled as of completion of the Fall 2016 interview.

---

<sup>32</sup> The 2014 Survey File data were not released; as a result the 2-year longitudinal weight cannot be used for the 2016 data year.

The 2016 Survey File ever enrolled weights are populated for all records on the delivered file and include continuously enrolled beneficiaries and beneficiaries who died or lost entitlement prior to completing the Fall 2016 interview. Beneficiaries who first became enrolled in 2016 are also included; these current-year enrollees were sampled and interviewed for the first time in 2016. The resulting weights represent the population of beneficiaries that were enrolled in Medicare on at least one day at any point in 2016.

### Continuously Enrolled Cross-Sectional Weights

The continuously enrolled cross-sectional weights are the traditional Survey File weights and have been provided every year. They represent the population of beneficiaries that were enrolled in Medicare for the entire period between the first of the year through the Fall 2016 interview period.

### Fall 2016 Nonresponse Adjustment

Continuing sample from the 2013, 2014, and 2015 panels are adjusted for nonresponse through Fall 2016. The process begins with weights for these panels that were previously adjusted through Fall 2015. Response status in Winter 2016, Summer 2016, and Fall 2016 is then identified, where a respondent is a beneficiary that was alive and entitled with a complete Fall 2016 interview, or who died or lost entitlement at some time in prior to Fall 2016 but had a completed final interview after death (via proxy) or loss of entitlement.

Nonresponse adjustment cells are constructed prior to performing the adjustment. First, the beneficiaries are divided into five primary adjustment cells: alive community, deceased community, alive facility, deceased facility, and Fall 2015 nonrespondents.

Separately within each of these main adjustment cells, and separately by panel, response propensity models are fit using logistic regression to model the probability of response through Fall 2016 as a function of covariates derived from the Fall 2015 Survey File data. Generally, the covariates are selected into the logistic regression model using stepwise selection with an entry p-value of 0.10 and a stay p-value of 0.15. Using the predicted response probabilities, beneficiaries are grouped into cells of approximately 100 each. Across all panels there are a total of 109 adjustment cells formed following the response modeling process. Separately within each of these cells, a ratio adjustment to distribute the weights of the nonrespondents to the respondents is performed. The resulting weights are the within-panel weights adjusted for response through Fall 2016.

### Derivation of the Continuously Enrolled Weights

The next step takes the weights for Continuing panels that are now adjusted through Fall 2016 and combines them with the weights for the 2016 Panel that were separately adjusted for initial nonresponse at the first interview (Fall 2016) as part of the initial weighting process. Next, the process removed cases that either died or lost entitlement prior to the Fall 2016 interview, or were cases from the 2016 Panel that enrolled after January 1, 2016.

At this stage there is quadruple coverage of beneficiaries who accreted on or before January 1, 2013, triple coverage of beneficiaries who accreted from January 2, 2013 through January 1, 2014, and double coverage of beneficiaries who accreted from January 2, 2014 through December 31, 2015. To account for this overlap, the weights for the four panels are adjusted by compositing factors derived from the number of effective completes by accretion year and age group across the four panels.

The compositing factor applied to beneficiaries from panel  $p$  in accretion year/age group domain  $d$  is

$$\varphi_{pd} = \frac{n_{pd}^{eff}}{\sum_{p \in P} n_{pd}^{eff}}$$

where  $n_{pd}^{eff}$  is the effective number of Fall 2016 completes in panel  $i$  in accretion year/age group domain  $d$ . The subscript  $p$  indexes the four panels in the set of active panels  $P$ . The effective sample sizes are calculated as

$$n_{id}^{eff} = \frac{n_{id}^{act}}{1 + \left( \frac{s_{id}^{76}}{\bar{w}_{id}^{76}} \right)^2}$$

where  $n_{pd}^{act}$  is the actual number of completed interviews,  $\bar{w}_{id}^{76}$  is the average of the Fall 2016 adjusted weights for the panel, and  $s_{id}^{76}$  is the standard deviation of these weights.

The resulting weights are the final continuously enrolled cross-sectional weights for the 2016 Survey File (SF). They represent the 2016 continuously enrolled population.

### Longitudinal Weights<sup>33</sup>

The derivation of two-year longitudinal weights begins with the weights adjusted through Fall 2016 from the 2013, 2014, and 2015 Panels, subset to beneficiaries who were alive and entitled at the Fall 2016 interview. A ratio adjustment accounted for cases that did not have complete Survey File data in both 2015 and 2016. The weights were then further adjusted to account for triple coverage of those accreting on or before January 1, 2013, and double coverage of those accreting from January 2, 2013 through January 1, 2014, using compositing factors derived similarly as described in the previous section. The final resulting weights represent the two-year

---

<sup>33</sup> Beginning with the 2016 LDS, the Survey File longitudinal weight names reflect the number of years the beneficiary was enrolled in Medicare (i.e., LNG2WGTS weights are referred to as 'two-year' rather than 'one-year' as they represent the population continuously enrolled for two years). This change was made to align the names of the longitudinal weights in the Survey File LDS with the naming convention used for the Cost Supplement LDS.

longitudinal population, which is the population of beneficiaries that enrolled on or before January 1, 2015, and were alive and entitled as of the Fall 2016 interview.

The derivation of three-year longitudinal weights begins with the weights adjusted through Fall 2016 from the 2013 and 2014 Panels, subset to beneficiaries that were alive and entitled at the Fall 2016 interview. A ratio adjustment accounted for cases that did not have complete Survey File data in both 2014 and 2016. The weights are then further adjusted to account for double coverage of those accreting on or before January 1, 2013, using compositing factors. The final resulting weights represent the three-year longitudinal population, which is the population of beneficiaries that enrolled on or before January 1, 2014, and were alive and entitled as of the Fall 2016 interview.

The four-year longitudinal weights are comprised of members of the 2013 Panel and are equal to the weights adjusted through Fall 2016 for this panel, subset to beneficiaries that were alive and entitled at the Fall 2016 interview. There is no need for further adjustment by compositing factors because there is only one panel providing four-year data so the weights are equal to the final cross-sectional weights for these beneficiaries. The final weights represent the four-year longitudinal population, which is the population of beneficiaries that enrolled on or before January 1, 2013, and were alive and entitled as of the Fall 2016 interview.

### **Final Ever Enrolled Weights**

Ever enrolled Survey File weights represent the population of Medicare beneficiaries who were ever enrolled at any time during 2016 (i.e., enrolled on at least one day in 2016). The continuously enrolled beneficiaries are a subset of the ever enrolled beneficiaries in two ways, both in terms of the real-world populations they represent and in terms of the sampled and interviewed beneficiaries that appear on the Survey File.

### **Fall 2016 Nonresponse Adjustment**

Continuing sample from the 2013, 2014, and 2015 Panels are adjusted for nonresponse through Fall 2016. As with the continuously enrolled weights, the process begins with weights for these panels that were previously adjusted through Fall 2015. The response status in Winter 2016, Summer 2016, and Fall 2016 is then identified. Under the ever enrolled design, respondents include beneficiaries with a complete Fall 2016 interview, those who lost entitlement prior to Fall 2016 and had a final complete interview, those who died prior to Fall 2016 whether or not a final proxy interview was obtained, and Fall 2016 nonrespondents who were successfully re-fielded in Winter 2017.

Next, the weights are adjusted for nonresponse through Fall 2016, using the same cells that are created for the adjustment of the weights under the continuously enrolled design. Following ratio adjustments within these cells, the resulting weights are the within-panel weights adjusted for response through Fall 2016 for purposes of the ever enrolled weights.

## Derivation of the Ever Enrolled Weights

The next step begins with the weights for the continuing panels adjusted through Fall 2016 in the previous step and combines them with the weights for the 2016 Panel that are separately adjusted for initial nonresponse at the first interview (Fall 2016). Next, the small number of cases that died or lost entitlement prior to January 1, 2016, and hence were never enrolled in 2016, are removed.

At this stage, beneficiaries from the Continuing panels who died or lost entitlement during 2016 are included, as described in the previous section. However, the 2016 Panel cases include only those who were respondents to the Fall 2016 initial interview, and as such they do not include any beneficiaries that died or lost entitlement prior to Fall 2016. Beneficiaries who enrolled before January 1, 2016, who died or lost entitlement are accounted for by the Continuing panels. Enrollees on or after January 1, 2016, who died or lost entitlement are not represented by any other panels, but they are few in number and are accounted for during final poststratification.

As with the continuously enrolled and longitudinal weights, the ever enrolled weights for the four panels are adjusted by compositing factors to account for overlap between the panels. These are derived from the number of effective completes by accretion year and age group. For the ever enrolled weights, beneficiaries from the Continuing panels who died or lost entitlement in 2016 are combined separately to account for the fact that these beneficiaries are not represented by the 2016 Panel.

To finalize the ever enrolled weights, the raking technique to calibrate the weights to known population control totals for the ever enrolled population is used. These are derived from the enrollment data extracts for drawing the 2016 Panel. The raking dimensions used are age category (7-level) and accretion year (6-level). The raking process adjusts the weights to match the control totals for the first raking dimension, then for the second raking dimension, then for the first dimension again, and so on until the weights perfectly match the control totals in both dimensions. The resulting weights are the final ever enrolled weights for 2016. They represent the population of beneficiaries that were enrolled for at least one day at any time in 2016. Exhibit 8.2.2 and 8.2.3 present the control totals used for the raking adjustment step.

**Exhibit 8.2.2:** Control Totals for Ever Enrolled Weight Raking, Dimension 1: Age Group

<b>Age Group</b>	<b>Control Total</b>
< 45 Years	1,880,360
45 -64 Years	7,138,060
65 - 69 Years	15,950,920
70 - 74 Years	12,281,700
75 - 79 Years	8,511,340
80 - 84 Years	5,997,260
85+ Years	6,881,800
<b>Total</b>	<b>58,641,440</b>

**Exhibit 8.2.3:** Control Totals for Ever Enrolled Weight Raking, Dimension 2: Enrollment Year

<b>Enrollment Year</b>	<b>Control Total</b>
< 2012	40,721,100
2012	3,731,780
2013	3,630,180
2014	3,640,000
2015	3,483,620
2016	3,434,760
<b>Total</b>	<b>58,641,440</b>

*2016 Cost Supplement Weights*

Data for the 2016 Cost Supplement File were collected in Winter 2016 through Winter 2017. The weights include beneficiaries sampled in the 2013 through 2015 Panels, plus members of the 2016 Panel who were recently enrolled in Medicare. These Cost Supplement File weights are ever enrolled weights representing the population of beneficiaries who were enrolled for at least one day in 2016. In addition to the weights, the dataset includes panel (selection year) identifier, and variance strata and unit variables for variance estimation.

**Composition of Sample and Populations of Interest**

The 2016 Cost Supplement weights include beneficiaries who were sampled in the 2013, 2014, 2015, and 2016 Panels. The 2013, 2014, and 2015 Panels are referred to as “Continuing panels” and provide survey-reported cost and utilization for 2016 through participation in the MCBS during Winter 2016 through Winter 2017. Members of the 2016 Panel who were first enrolled in 2016 are referred to as “recent enrollees.” They were first interviewed in Fall 2016 and did not provide cost and utilization data for the period of time between enrollment and completion of the Fall 2016 interview; cost and utilization data for the period between the Fall 2016 interview and the end of 2016 were collected in Winter 2017. A combination of the survey-collected data for the end of year and Medicare claims data were used to impute beneficiary-level data for the entire period of enrollment in 2016. The final weights, which include both the Continuing panels and the recent enrollees, represent the population of beneficiaries that whoever enrolled in Medicare at any time during 2016.

**Adjustment Derivation of Cross-Sectional Weights for the Continuing Panels**

The process begins with weights for the 2013, 2014, and 2015 Panels that were previously adjusted through Fall 2016 as part of the 2016 Survey File weights. These weights are further adjusted based on a product of the 2016 reference period process that identifies which beneficiaries contributed enough cost and utilization data to be included in the final data products. To be included, sample members must meet at least one of the following three criteria: (a) the ratio of days covered by interviews to the number of days enrolled in Medicare in 2016 is equal to or greater than 0.66; (b) the difference between the number of days enrolled in Medicare and the number of days covered by interviews is less than or equal to 60 days; or (c) the beneficiary is a

recent enrollee from the 2016 Panel who completed the initial Fall 2016 interview. Beneficiaries who died or lost entitlement prior to January 1, 2016, are ineligible and removed at this stage. Beneficiaries who survived into 2016 but do not meet the above criteria are considered to be nonrespondents for the 2016 Cost Supplement File and are adjusted for in the resulting weights. The adjustment cells used for this ratio adjustment are the same cells that were created during weighting for the 2016 Survey File weights.

Note that at this stage there is triple coverage of beneficiaries who accreted on or before January 1, 2013, in the Continuing panels, and double coverage of beneficiaries who accreted from January 2, 2013 through January 1, 2014. Therefore, the weights for the three panels are adjusted by compositing factors derived from the effective number of completes by panel, accretion year, and age group. The resulting weights are the pre-raked cross-sectional weights for the Continuing panels.

### **Cross-Sectional Weights for the Recent Enrollees**

The “recent enrollees” are those who enrolled between January 1, 2016, and December 31, 2016, inclusive. This step begins with the initial weights for the 2016 Panel, adjusted for nonresponse at the Fall 2016 interview. The subset of all Fall 2016 respondents from the 2016 Panel that are recent enrollees is isolated, and the resulting weights for this subset are the pre-raked cross-sectional weights for the recent enrollees.

### **Final Cross-Sectional Ever Enrolled Weights for the Cost Supplement**

The sum of the combined weights across all four panels (the three Continuing panels plus the recent enrollees from the 2016 Panel), provides an estimate of the ever enrolled population in 2016, but is not exact. To finalize the ever enrolled weights, the raking technique is used to calibrate the weights to known population control totals for the ever enrolled population. The raking dimensions used are age category (7-level) and accretion year (6-level), and the control totals used are the same as those used for the Survey File ever enrolled weights calibration presented in Exhibit 8.3.2. The resulting weights are the final weights for the 2016 Cost Supplement File. They represent the population of beneficiaries that were enrolled for at least one day at any time in 2016.

### **Two-Year Longitudinal Weights for the 2016 Cost Supplement**

The two-year longitudinal weights are populated for members of the 2013, 2014, and 2015 panels who were enrolled in both 2015 and 2016 and provided utilization and cost data for both years. Members of the 2013 and 2014 panels provided data for the 2015 and 2016 data years through participation in the MCBS during Winter 2015 through Winter 2017. Members of the 2015 panel who first enrolled in 2014 or 2015 provided data for the end of 2015 in the Winter 2016 interview, and provided data for the 2016 data year in Winter 2016 through Winter 2017. The final two-year longitudinal weights represent the population of beneficiaries who were ever enrolled in Medicare at any time during both 2015 and 2016.

## 2016 Topical Module Weights

The Beneficiary Knowledge Questionnaire (KNQ) was administered in the Community Questionnaire in Winter 2017 (Round 77). The Summer 2017 (Round 78) Community Questionnaire included the Income and Assets Questionnaire (IAQ), the Patient Activation Questionnaire (PAQ) and the Prescription Medicine Questionnaire (RXQ). To facilitate estimation from the resulting data, two sets of full-sample and replicate weights were derived for each module, one based on the 2016 Survey File continuously enrolled population, and the other based on the 2016 Cost Supplement ever enrolled population. These weights can be used to conduct joint analyses of Topical Module data, Survey File data, and Cost Supplement data. Exhibit 8.2.4 lists the Topical Module weights for these rounds.

Note that counts of cases with positive topical module weights may vary within the data year and may change across years due to response rates, sample sizes, and fielding methods. The topical weights account for these changes.

**Exhibit 8.2.4.** 2016 Data Year Topical Module Survey Weights Datasets and Contents

Dataset Name	Record Count	Variable Count	Full-Sample Weight	Replicate Weights	Description
KNCWT	10,784	102	KN77CWT	KN77C1- KN77C100	R77 KNQ Continuously Enrolled
KNEWT	6,878	102	KN77EWT	KN77E1- KN77E100	R77 KNQ Ever Enrolled
IACWT	9,737	102	IA78CWT	IA78C1-IA78C100	R78 IAQ Continuously Enrolled
IAEWT	6,376	102	IA78EWT	IA78E1-IA78E100	R78 IAQ Ever Enrolled
PACWT	8,690	102	PA78CWT	PA78C1- PA78C100	R78 PAQ Continuously Enrolled
PAEWT	5,678	102	PA78EWT	PA78E1- PA78E100	R78 PAQ Ever Enrolled
RXCWT	7,556	102	RX78CWT	RX78C1- RX78C100	R78 RXQ Continuously Enrolled
RXEWT	4,324	102	RX78EWT	RX78E1- RX78E100	R78 RXQ Ever Enrolled

## Composition of Sample and Populations of Interest

The topical module data were collected from beneficiaries selected in the 2013, 2014, 2015, or 2016 Panels who responded to the Community Questionnaire in the round in which the module was administered, and each of the datasets includes members of these panels who completed the topical module. Each of the eight weights were derived to represent a population that was alive, entitled, and in the community in the given round.

The four weights that began with the 2016 Survey File continuously enrolled weights as a starting point (KNCWT, IACWT, PACWT, and RX7WT) all represent populations that were continuously enrolled from the start of 2016 and were still alive, entitled, and in the community during the round. The KNCWT weights represent the population of beneficiaries that were continuously enrolled from January 1, 2016 and still alive, entitled, and in the community in the Winter of 2017. The IACWT, PACWT, and RXCWT weights represent the population of beneficiaries that were continuously enrolled from January 1, 2016 and still alive, entitled, and in the community in Summer 2017.

The four weights that began with the 2016 Cost Supplement ever enrolled weights as a starting point (KNEWT, IAEWT, PAEWT, and RXEWT) all represent populations that were ever enrolled at any time in 2016 and were still alive, entitled, and in the community during the round. The KNEWT weights represent the population of beneficiaries that were ever enrolled for at least one day in 2016 and were still alive, entitled, and in the community in the winter of 2017. Another way to look at this population is that it is the population of beneficiaries that were continuously enrolled from December 31, 2016 to the winter of 2017. The IAEWT, PAEWT, and RXEWT weights represent the population of beneficiaries that were ever enrolled for at least one day in 2016 and were still alive, entitled, and in the community in the Summer of 2017. Similar to the KNEWT, this population can also be summarized as having been continuously enrolled from December 31, 2016 to the summer of 2017.

## Derivation of Topical Module Weights

Each of the topical module weights is based on a starting weight, which is either the 2016 Survey File continuously enrolled weight (those that end in 'CWT') or the 2016 Cost Supplement ever enrolled weight (those that end in 'EWT'). The choice of starting weight determines the population that the derived topical module weight represents, as described in the previous section. However, once this choice of starting weight is made, the process for each topical module is largely the same.

The weighting adjustments for each delivered weight are carried out in two steps. At each, the existing model-based adjustment cells that were developed for the 2016 Survey File and Cost Supplement weights were used, with collapsing of the cells where necessary to preserve adequate sample sizes.

The first adjustment distributes the weights for cases with unknown eligibility for the module to those with known eligibility. Beneficiaries may have unknown eligibility if they were unlocatable during the round or if they were nonrespondents during the round or earlier rounds and there was

no indication of mortality or residential (community or facility) status. The number of cases with unknown eligibility was small in Winter 2017 because this round immediately followed 2016 and the Fall 2016 Survey File interviews, whereas in Summer 2017 there was an intervening round in which some members of the sample became nonrespondents. In all cases, this first adjustment for unknown eligibility makes the implicit assumption that if eligibility were observable for these cases they would exhibit the same proportions of eligibility as the cases whose eligibility we are able to observe.

Prior to the second adjustment, the set of beneficiaries is limited to those who were eligible to receive the topical module. A beneficiary was considered ineligible if they had died, lost entitlement, or were in the Facility component only during the round. The second adjustment, the nonresponse adjustment, then distributes the weights for the eligible nonrespondents to the eligible respondents. For purposes of the PAQ, alive and entitled beneficiaries in the community for whom the Community Questionnaire was administered by proxy were considered nonrespondents. For the RXQ, the exit panel (2013) was not administered the module. Therefore, the weights are limited only to beneficiaries from the 2014, 2015, and 2016 Panels. To account for the loss of the 2013 Panel, the nonresponse adjustment also included an additional adjustment to increase the weight totals so that they agree with the total weights for the IAQ and PAQ, which were conducted in the same quarter as the RXQ.

## 8.3 MCBS Imputation Processes

### *Overview*

As noted earlier, MCBS imputation falls under two umbrellas that focus on imputing monetary amounts: Income and Asset (IA) imputation, and Event, Payer, and Cost imputation, which includes imputation for Prescription Medicine (PM) and Non Prescription Medicine (Non PM) events and costs. All three imputations focus on imputing a monetary amount. IA imputation completes income and asset information for the beneficiary and spouse, and PM and Non PM imputation complete medical event and cost data. For all three types, two groups of variables are imputed:

- Probes: Yes/no variables indicating whether the type of income, asset, or payer should have a nonzero amount.
- Amounts: The value of the income, asset, or cost paid for a medical event. For IA imputation, amounts are nonzero if the associated probe indicates the income or asset exists and missing otherwise. For PM and Non PM imputation, amounts are nonzero if the associated probe indicates that the payer paid and zero otherwise.

For both probes and amounts, single value imputation is performed sequentially from variables or records with the least to the most item nonresponse.

## *Income and Asset Imputation*

### **Overview**

The 2016 Income and Assets (IA) imputation imputes detailed information about income and assets of the beneficiary and spouse for Community Questionnaire respondents. For Facility Questionnaire respondents, and Community and Facility Questionnaire non-respondents<sup>34</sup>, only total income is imputed due to the lack of detailed asset information.

### **Process**

Respondents are asked about their prior year income and assets during the summer round. The income and asset data first go through data editing to ensure that respondent-reported values are either appropriate or set to missing. Data editing is performed to:

- Ensure consistency with questionnaire skip logic within the Income and Asset Questionnaire (IAQ)
- Set extreme outliers at the tails of the distributions of each IA variable to missing
- Set outliers based on joint distributions of highly-correlated IA variables to missing

Next, probe variables are imputed via a hot deck method. Probes had very low item nonresponse rates. The hot deck method is used because it can impute all of the missing values and is relatively easy to implement. This method takes the non-missing IA value directly from another beneficiary in the same imputation cell to fill in the missing IA value of the recipient beneficiary. If the probe is imputed as “no”, indicating that a beneficiary does not have a particular type of asset, the corresponding amount variable is set to missing.

Amount variables are imputed after probes. While most respondents report whether the beneficiary has an asset type, some respondents refuse to provide or don't know the amount of the asset. As a result, amount variables need more imputation. When respondents report value ranges, the hot deck method is used to impute an exact dollar amount using the given value range as a boundary. When value ranges are not given but prior-year IA information exists, values are imputed using a prior-year carry-forward method with an inflation adjustment. This method uses the non-missing IA variable value for the same beneficiary and variable from the prior year to impute the current-year missing value. This prior-year carry-forward method provides reasonable and consistent imputed values for these respondents. For the rest of the missing amount values, hot deck imputation is used.

Each variable imputed via hot deck imputation has a unique set of imputation cell variables. In the hot deck method, recipient and donor records are segregated into pools of records (“imputation cells”) that have the same values on a set of auxiliary (or explanatory) variables. In general, the

---

<sup>34</sup> The Income and Assets questionnaire section (IAQ) is only administered once per year. Non-response to this section may be due to non-response in the round the questionnaire section is administered, or non-response to questions in the IAQ. For more information on IAQ, see Section 4.1.

auxiliary variables that define imputation cells for probe variables include prior-year probe values, beneficiary's age, indicator of spouse/partner, and other related IA probes. Auxiliary variables that define imputation cells for amount variables include other related IA amounts, poverty indicators, beneficiary's age, and metropolitan status.

## *Prescription Medicine and Non Prescription Medicine Imputation*

### **Overview**

Both the Prescription Medicine (PM) imputation and Non Prescription Medicine (Non PM) imputation fill in missing payer and payment information for beneficiaries' medical events. The imputation procedures used for Prescription Medicine (PM) events versus all other event types (Non PM) are very similar but not identical.

Beginning in 2016, it was determined that payments made by the Veterans Administration (VA) could not be estimated with sufficient accuracy. Therefore, observed payments from the VA have been combined into the 'Other Sources' payer beginning with 2016.

### **Process**

Both PM and Non PM imputation begin with the receipt of the survey-reported events matched against the Medicare claims. Three categories of records are returned: events found in the claims only (claims-only), events found in the survey-reported data only (survey-only), and survey-reported events that were successfully matched to a Medicare claim (survey-matched).

For the PM imputation, only unmatched survey-only events are processed through imputation. Claims-only and survey-matched events are considered complete. For the Non PM imputation, all three claims match statuses are processed through imputation.

First, data preprocessing and editing are performed to identify the total charge for the event and the most likely payers for the event. This procedure is described in detail in the MCBS Data User's Guide: Cost Supplement File. Imputation then proceeds in three steps.

First imputed are events where the total charge is known and the payers and payment amounts are missing together (when a payer is missing, the amount is missing, and vice versa). Exhibit 8.3.1 gives an illustration of the type of record that would be imputed in this group, with a simplified potential payer vector. The donor record is required to be a complete record, and must have at least one of the recipient's missing payers as a payer with a positive payment amount, so that there is at least one amount value to which the difference between the total charge and the sum of the known payments can be allocated. In the example shown in Exhibit 8.3.1, a donor would need to have either "Employment-based private health insurance" or "Out of Pocket" as a payer with a nonzero amount. The payers and payment amounts are pulled from the same donor.

**Exhibit 8.3.1.** Payers and Payment Amounts Missing Together, Total Charge Known

<b>Variable Type</b>	<b>Medicare Fee-for-Service</b>	<b>Medicaid</b>	<b>Employment-based private health insurance</b>	<b>Out of Pocket</b>	<b>Total Charge</b>
Payer Indicator	Yes	No	(null)	(null)	--
Amount	50	0	(null)	(null)	200

Next imputed are events where the total charge is known and the payers and payment amounts have different missing patterns (i.e., there is at least one instance where the payer is known to have paid but the amount is missing). This is illustrated by Exhibit 8.3.2. The payers are imputed first. Donors are required to be complete records. There is no restriction that the donor is a payer for any of the recipient's missing payers because by definition of this group, there is at least one known payer already to which the missing payment amount can be allocated. Payment amounts are imputed next. If the payer is imputed not to have paid, the payment amount is set to zero. If there is only one missing payment amount after the payer imputation, that amount is completed by subtraction. If possible, payment amounts are all pulled from the same donor; if a donor with the required payer pattern does not exist<sup>35</sup>, payment amounts are imputed individually from different donors.

**Exhibit 8.3.2.** Payers and Payment Amounts Missing Differentially, Total Charge Known

<b>Variable Type</b>	<b>Medicare Fee-for-Service</b>	<b>Medicaid</b>	<b>Employment-based private health insurance</b>	<b>Out of Pocket</b>	<b>Total Charge</b>
Payer Indicator	Yes	No	Yes	(null)	--
Amount	50	0	(null)	(null)	200

Finally, events with the total charge unknown are imputed (illustrated by Exhibit 8.3.3). Payers are imputed first and are all taken from the same donor. Payment amounts are imputed next and are

<sup>35</sup> In this group, we impute a vector of missing payers together from the same donor, and have at least one additional payer who is known to have paid but the amount is unknown. Thus, a new payer pattern that did not exist in the original data may be created – the vector of imputed payers, plus the known payer with unknown amount.

taken from the same donor when possible, or are imputed individually if a donor with the required payer pattern does not exist<sup>36</sup>. Total charge is set to the sum of the payment amounts.

### Exhibit 8.3.3. Total Charge Unknown

Variable Type	Medicare Fee-for-Service	Medicaid	Employment-based private health insurance	Out of Pocket	Total Charge
Payer Indicator	Yes	No	Yes	(null)	--
Amount	50	0	(null)	(null)	(null)

In all PM and nearly all Non PM cases, the payment amount is not imputed directly from the donor; it is ratio-adjusted to fit with the recipient's known payment amounts.

The PM and Non PM imputation processes are very similar up to this point but then diverge.

#### PM Imputation

One final step is applied in PM imputation processing. After the general imputation procedure has been run, cases are reviewed and those found to be inconsistent or to have potential imputation issues are reviewed and edited. Records where the payers and payment amount vectors are complete but total charge is less than or more than the sum of the payment amounts, or records that are incomplete but have total charge less than the known payment amounts, are subjected to edits to make the record complete and consistent. Events where an imputed payment amount is less than a penny or a total charge is less than 50 cents are re-imputed from a new donor. The number of records requiring editing or re-imputation is very small (1.5% of records in 2016).

The PM imputation produces one file, an event-level dataset of survey-only events.

#### Non PM Imputation

Beginning in 2015, current-year enrollee sample beneficiaries are included in the Non PM imputation.<sup>37</sup> The current-year enrollees have some portion of the year covered by claims data only, and not by survey data. This may result in biased estimates as some medical events and costs, such as vision and dental health care services, are not covered by the Medicare claims and would be captured only by the survey data that were not collected. Please see the MCBS Data User's Guide: Cost Supplement File for a further discussion of gaps in survey data coverage. A

<sup>36</sup> Similar to when total charge is known, some records with total charge unknown will have payers and payment amounts missing at different rates (i.e., there is at least one instance where the payer is known to have paid but the amount is missing). After the payer imputation, a new payer pattern may be created that did not exist in the original data.

<sup>37</sup> See Section 3.4, "Current-Year Enrollee Sample", for more information on these beneficiaries.

new unit-level imputation procedure was added to address the issue of gaps in survey data coverage for the current-year enrollees. This procedure imputes survey-only events that may not be covered by the claims, adding new event records to the file that did not previously exist.

The time period within which claims-only events are to be imputed varies by individual, ranging from the beneficiary's enrollment date to the first of: the fall interview date (if there was a completed winter interview), the date of death, the date of lost entitlement, or December 31. First, this time period (the "Missing Period") is defined for each current-year enrollee. A donor is selected for each current-year enrollee, and the donor's survey-only records (excluding those with a Medicare and not Medicare Advantage payment, as these would be covered by claim data) that occur within the recipient's Missing Period are then created for the recipient. If the donor has no donation-eligible records of a given event type, no records are created.

All variables populated on the donor record are populated on the newly-created (recipient) record. Variables that relate to the event are pulled along from the donor record. Variables that relate to the beneficiary are retained from the recipient.

As described in the MCBS Data User's Guide: Cost Supplement File, the event types used in the survey differ from the event types in the Medicare claims. For the Non PM events, an administrative event type is imputed from the survey-reported event type. Event type imputation recipients are events found in the survey-only data, and donors are survey-matched events. Recipient records are matched to donors on survey-reported event type and cost, and the donor's administrative event type is assigned to the recipient.

Next, hospice event data are appended to the Non PM events. These data come directly from CMS and are not imputed. More information on hospice data is provided in the MCBS Data User's Guide: Cost Supplement File.

Finally, the Non PM data are aggregated to the service and person level. The Non PM imputation produces three files: at the event level (most disaggregate), at the service level (one record per beneficiary and event type), and at the person level (one record per beneficiary). Event-level records are first summed to the service level, and then adjustments are performed to annualize these amounts and adjust for days the beneficiary was eligible for Medicare but not covered by survey-reported data. This process is described in further detail in the MCBS Data User's Guide: Cost Supplement File. Then, unadjusted and adjusted service-level amounts are summed to the person level.

### **Hot Deck Imputation Procedure**

All PM and Non PM imputation is performed using a hot deck imputation procedure.

While hot deck has been used as a donor selection method for several years on the MCBS, the method to identify a compatible donor was updated, beginning with 2015.

Each imputation step has a unique set of qualification rules and key variables used to identify a similar donor record for a given recipient record. The donor pool for each set of recipients is first

restricted to the group of potential donor records that meets the donor qualification rules, such as requiring that donors have complete data on the item to be imputed. Next, the similarity between a given recipient and each possible donor is measured via the Gower function using SAS/STAT® software's PROC DISTANCE:

$$s_1(x, y) = \frac{\sum_{j=1}^v w_j \delta_{x,y}^j d_{x,y}^j}{\sum_{j=1}^v w_j \delta_{x,y}^j}$$

Where  $v$  is the number of variables,  $x_j$  is the data for observation  $x$  and the  $j^{\text{th}}$  variable,  $y_j$  is the data for observation  $y$  and the  $j^{\text{th}}$  variable, and  $w_j$  is the weight for the  $j^{\text{th}}$  variable. For ordinal, interval, and symmetric nominal variables,  $\delta_{x,y}^j = 1$ . For asymmetric nominal variables,  $\delta_{x,y}^j = 1$  if either  $x_j$  or  $y_j$  is present and 0 if both are absent. For a nominal variable,  $d_{x,y}^j = 1$  if  $x_j = y_j$  and 0 otherwise. For an ordinal, interval, or ratio variable,  $d_{x,y}^j = 1 - |x_j - y_j|$ .<sup>38,39,40</sup>

The Gower function was selected because it can compute a similarity measure across several variable types (nominal, ordinal, and interval). For each recipient, we select donors whose similarity score is less than or equal to the 30<sup>th</sup> largest distance (with a score of 0 representing identical records and 1 representing divergent records). This may result in 30 potential donors, or more if there are ties. Frequently, PM and Non PM donor pools are small, and this method allows us to relax some of the boundaries defining a suitable donor while continuing to find donors that are highly similar to a recipient. After computing donor pools by finding donor records that are similar to recipients, the new imputation procedure goes on to identify the donor record using the hot deck method in SAS/STAT® software's PROC SURVEYIMPUTE.

---

<sup>38</sup> SAS Institute Inc. 2017. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.

<sup>39</sup> Podani, János. "Extending Gower's General Coefficient of Similarity to Ordinal Characters." *Taxon* 48, no. 2 (1999). 331-340.

<sup>40</sup> Gower, John C. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27, no. 4 (1971). 857-871.

## 9. RESPONSE RATES AND NONRESPONSE

This section presents the response rates and describes the derivation of those rates for the 2016 Cost Supplement and Survey File data releases.

### 9.1 Response Rates

This section details the definitions and calculations of Cost Supplement File response rates and Survey File response rates. Response rates presented in this report are unweighted.

In the sections that follow, both unconditional and conditional response rates are presented. The unconditional response rate is the percentage of sample that were released during the fall round of the selection year and responded to the survey in 2016. The unconditional response rates, also called cumulative response rates, use the original selected sample size as the baseline in their calculation. Conditional response rates are the percentage of sample that responded during 2015 and also responded during 2016. Conditional response rates use the sample who responded during 2015 as the baseline in their calculation. In other words, they are conditioned on response in year 2015.

#### *2016 Cost Supplement File Response Rates*

##### **Unconditional Response Rates for the Annual Cost Supplement File**

The response rate for a given data year,  $t$ , in canonical form is simply

$$r_t = \frac{C_t}{E_t},$$

where  $C_t$  is the number of beneficiaries for whom the Cost Supplement File data are taken to be *complete*, and  $E_t$  is the number of beneficiaries who are considered *eligible* for the annual Cost Supplement File data release.

$C_t$  is calculated as the number of beneficiaries with a non-missing, positive Cost Supplement File weight for the given year.

The number of eligible beneficiaries is calculated as

$$E_t = T_t - I_t,$$

where  $T_t$  is the *total sample size* for the given year, and  $I_t$  is the number of beneficiaries who are considered *ineligible* for the given annual Cost Supplement File data release.

For the  $t = 2016$  data year,  $T_t$  includes the following:

- All of the panel selected in year  $t - 3$ , called  $S_{t-3}$ .
- All of the panel selected in year  $t - 2$ , called  $S_{t-2}$ .

- All of the panel selected in year  $t - 1$ , called  $S_{t-1}$ .
- The subset of the panel selected in year  $t$ , called  $S_t$ , consisting of members of both the year  $t - 1$  and the year  $t$  cohorts of beneficiaries.
- The subset of the panel selected in year  $t + 1$ , called  $S_{t+1}$ , consisting of members of the year  $t$  cohort of beneficiaries.

### Conditional Response Rates for the Annual Cost Supplement File

The conditional response rate for the year  $t - 3$  to  $t - 1$  panels in Cost Supplement File year  $t$  is:

$$\frac{C_t}{E_t - N_t},$$

where

$C_t$  =  $S_{t-3}$  to  $S_{t-1}$  panel beneficiaries with positive weights on the year  $t$  Cost Supplement File;

$E_t$  =  $S_{t-3}$  to  $S_{t-1}$  panel beneficiaries still entitled on January 1, year  $t$ ;

$N_t$  = subset of  $E_t$  that were not released in the first round of year  $t$ .

The conditional response rate for the year  $t$  panel in Cost Supplement File year  $t$  is:

$$\frac{C_t}{E_t},$$

where

$C_t$  =  $S_t$  panel beneficiaries with positive weights on the Cost Supplement File;

$E_t$  =  $S_t$  panel beneficiaries enrolled between January 2, year  $t - 1$  to December 31, year  $t - 1$  and still entitled on January 1, year  $t$ .

The conditional response rate for the year  $t + 1$  panel in Cost Supplement File year  $t$  is:

$$\frac{C_t}{E_t},$$

where

$C_t$  =  $S_{t+1}$  panel beneficiaries with positive weights on the Cost Supplement File;

$E_t$  =  $S_{t+1}$  panel beneficiaries enrolled between January 1, year  $t$  and December 31, year  $t$ .

Exhibits 9.1.1 and 9.1.2 display the 2016 Cost Supplement File unconditional and conditional response rates by panel.

**Exhibit 9.1.1:** 2016 MCBS Annual Cost Supplement File Unconditional Response Rates

Panel	Released	Complete	Eligible	Ineligible	Unconditional Response Rate
<b>2013</b>	7,400	2,131	6,332	1,068	33.7%
<b>2014</b>	11,398	3,309	10,253	1,145	32.3%
<b>2015</b>	8,621	2,939	8,145	476	36.1%
<b>2016</b>	518	245	511	--*	47.9%
<b>Total</b>	27,937	8,624	25,241	2,696	34.2%

\*Cell sizes suppressed that are less than 11.

**Exhibit 9.1.2:** 2016 MCBS Annual Cost Supplement File Conditional Response Rates

Panel	Complete	Eligible	Subset of Eligibles Not Released	Conditional Response Rate
<b>2013</b>	2,131	6,332	3,740	82.2%
<b>2014</b>	3,309	10,253	5,798	74.3%
<b>2015</b>	2,939	8,145	3,826	68.0%
<b>2016</b>	245	511	0	47.9%
<b>Total</b>	8,624	25,241	13,364	72.6%

### 2016 Survey File Response Rates

#### Unconditional Response Rates for the Annual Survey File: Ever Enrolled Beneficiaries.

The response rate for a given data year,  $t$ , in canonical form is simply

$$r_t = \frac{C_t}{E_t},$$

where  $C_t$  is the number of beneficiaries for whom the Survey File data are taken to be *complete*, and  $E_t$  is the number of beneficiaries who are considered *eligible* for the annual Survey File data release.

$C_t$  is calculated as the number of beneficiaries with a non-missing, positive Survey File ever enrolled weight for the given year.

The number of eligible beneficiaries is calculated as

$$E_t = T_t - I_t,$$

where  $T_t$  is the *total sample size* for the given year and  $I_t$  is the number of beneficiaries who are considered *ineligible* for the given annual Survey File data release.

For year  $t$ ,  $T_t$  includes the following:

- All of the panel selected in year  $t - 3$ , called  $S_{t-3}$ .
- All of the panel selected in year  $t - 2$ , called  $S_{t-2}$ .
- All of the panel selected in year  $t - 1$ , called  $S_{t-1}$ .
- All of the panel selected in year  $t$ , called  $s_t$ .

$I_t$  is calculated as the number of beneficiaries from panels  $t - 3$  to  $t - 1$  who died or lost entitlement prior to January 1<sup>st</sup> of year  $t$ , plus the number of ineligible or deceased beneficiaries from the year  $t$  panel in the fall round.

### Conditional Response Rates for the Annual Survey File: Ever Enrolled Beneficiaries.

The conditional response rate for the year  $t - 3$  to  $t - 1$  panels in Survey File year  $t$  is:

$$\frac{C_t}{E_t - N_t},$$

where

$C_t = S_{t-3}$  to  $S_{t-1}$  panel beneficiaries with positive weights on the year  $t$  Survey File;

$E_t = S_{t-3}$  to  $S_{t-1}$  panel beneficiaries still entitled and alive prior to fall round, year  $t$  and are not  $I_t$ .

$N_t =$  subset of  $E_t$  that were not released in the first round of year  $t$ .

The conditional response rate for the year  $t$  panel in Survey File year  $t$  is:

$$\frac{C_t}{E_t},$$

where

$C_t = s_t$  panel beneficiaries with positive weights on the Survey File;

$E_t = s_t$  panel beneficiaries still entitled and alive prior to fall round, year  $t$  and are not  $I_t$ .

### Response Rates for the Annual Survey File: Continuously Enrolled Beneficiaries

The formulas for calculating the unconditional and conditional response rates for the continuously enrolled beneficiaries are identical to the corresponding formulas detailed above for the ever enrolled population. The only differences are in the definitions of  $C_t$  and  $I_t$ .

For the continuously enrolled response rate calculations,  $C_t$  is calculated as the number of beneficiaries completing an interview in the fall round of year  $t$  with a non-missing, positive Survey File continuously enrolled weight for the given year  $t$ .

Two subsets of ineligible contribute to  $I_t$  for the continuously enrolled response rate calculations:

- The first subset includes beneficiaries who are found to be ineligible or deceased in any round up to and including the fall round of year  $t$ .
- The second subset includes beneficiaries who finished the fall round year  $t$  interview but are not Survey File completes, or beneficiaries who were non-respondents prior to the fall round of year  $t$  and thus were not fielded in the fall round, and had a final status with no further attempts to field in any previous round. (These are beneficiaries not included in the first subset of ineligible described above.) For these cases, the date of death or lost entitlement date, if any, is compared to the average interview date in the fall round year  $t$ . If date of death or lost entitlement date is prior to the average interview date, the case is determined to be ineligible. Otherwise, it is determined to be an eligible non-respondent.

Exhibits 9.1.3 and 9.1.4 display the 2016 annual Survey File unconditional response rates by panel for ever enrolled and continuously enrolled beneficiaries.

**Exhibit 9.1.3:** 2016 MCBS Annual Survey File Unconditional Response Rates for Ever Enrolled Beneficiaries

Panel	Released	Ever Enrolled Complete	Ever Enrolled Eligible	Ever Enrolled Ineligible	Unconditional Response Rate of Ever Enrolled Beneficiaries
<b>2013</b>	7,400	2,141	7,120	280	30.1%
<b>2014</b>	11,398	3,343	10,987	411	30.4%
<b>2015</b>	8,621	2,968	8,571	50	34.6%
<b>2016</b>	12,145	6,326	11,596	549	54.6%
<b>Total</b>	39,564	14,778	38,274	1,290	38.6%

**Exhibit 9.1.4:** 2016 MCBS Annual Survey File Unconditional Response Rates for Continuously Enrolled Beneficiaries

Panel	Released	Continuously Enrolled Complete	Continuously Enrolled Eligible	Continuously Enrolled Ineligible	Unconditional Response Rate for Continuously Enrolled Beneficiaries
<b>2013</b>	7,400	1,944	6,488	912	30.0%
<b>2014</b>	11,398	2,957	10,243	1,155	28.9%
<b>2015</b>	8,621	2,604	7,993	628	32.6%
<b>2016</b>	12,145	6,107	11,596	549	52.7%
<b>Total</b>	39,564	13,612	36,320	3,244	37.5%

Exhibits 9.1.5 and 9.1.6 display the 2016 Survey File conditional response rates by panel for ever enrolled and continuously enrolled beneficiaries.

**Exhibit 9.1.5:** 2016 MCBS Annual Survey File Conditional Response Rates for Ever Enrolled Beneficiaries

Panel	Ever Enrolled Complete	Ever Enrolled Eligible	Subset of Ever Enrolled Eligibles That Were Not Released	Conditional Response Rate for Ever Enrolled Beneficiaries
<b>2013</b>	2,141	7,120	4,527	82.6%
<b>2014</b>	3,343	10,987	6,528	75.0%
<b>2015</b>	2,968	8,571	4,231	68.4%
<b>2016</b>	6,326	11,596	0	54.6%
<b>Total</b>	14,778	38,274	15,286	64.3%

**Exhibit 9.1.6:** 2016 MCBS Annual Survey File Conditional Response Rates for Continuously Enrolled Beneficiaries

Panel	Continuously Enrolled Complete	Continuously Enrolled Eligible	Subset of Continuously Enrolled Eligibles That Were Not Released	Conditional Response Rate for Continuously Enrolled Beneficiaries
<b>2013</b>	1,944	6,488	3,982	77.6%
<b>2014</b>	2,957	10,243	5,988	69.5%
<b>2015</b>	2,604	7,993	3,828	62.5%
<b>2016</b>	6,107	11,596	0	52.7%
<b>Total</b>	13,612	36,320	13,789	60.4%

## 9.2 Nonresponse Bias Analysis

Please see the 2015 MCBS Methodology Report to obtain the results of the most recent non-response bias analysis.

## 10. USING MCBS DATA FILES

### 10.1 MCBS Data User's Guide

The MCBS Data User's Guide offers a publicly available, easily searchable resource for data users. It is updated for each new data year to ensure that users have current documentation on the survey design, questionnaires, and estimation as well as detailed notes on the structure and contents of the MCBS data releases.

The Data User's Guide features three stand-alone chapters. The General Data User's Guide documents the key features of the study and data products. The Survey File and Cost Supplement Guides provide technical information on each file including the derivation of variables and any significant changes in the variables and/or file structure.

### 10.2 MCBS Public Use Data File

Beginning with data collected in the 2013 MCBS, a public use file (PUF) and accompanying documentation are available free for download under the MCBS PUF link at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index.html>. The MCBS PUF is an easy to use data file with select data items that allow researchers to conduct analysis on health disparities, access to and satisfaction with healthcare, and medical conditions for community dwelling Medicare beneficiaries. The MCBS PUF is not intended to replace the more detailed limited data set (LDS) files. Rather, it provides a publicly available alternative for those researchers interested in the health, health care use, access to and satisfaction with Medicare of beneficiaries. Given that the MCBS PUF meets all necessary requirements regarding de-identification of the data and mitigation of disclosure risk, it provides the very highest degree of protection to the Medicare beneficiaries' protected health information.

### 10.3 MCBS Limited Data Sets

There are two MCBS Limited Data Sets (LDS) available to data users. In order to access these data files, data users must submit a Data Use Agreement (DUA) and complete an LDS Worksheet, which provides CMS with information about the research project, the particular files needed, and payment information for administrative fees associated with the data request. Note that new data users and repeat data users complete distinct forms. Data users should visit CMS' LDS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA - NewLDS.html> for a full description of available LDS files, further details about the request process, and downloadable forms.

## 11. GLOSSARY

**Activities of daily living (ADLs):** Activities of daily living are activities related to personal care. They include bathing or showering, dressing, getting in and out of bed or a chair, walking, using the toilet, and eating.

**Baseline interview:** The initial questionnaire administered to new respondents to the study; administered in the fall of the year they are selected into the sample (interview #1).

**Beneficiary:** An individual selected from MCBS' sample about whom the MCBS collects information. Beneficiary may also refer to a person receiving Medicare services who may or not be participating in the MCBS.

**Claim-only event:** A claim-only event is a medical service or event known only through the presence of a Medicare claim. The event did not originate from an event or service reported by a respondent during an interview.

**Community component:** Survey of beneficiaries residing in the community at the time of the interview (i.e., not in a long-term care facility such as a nursing home).

**Continuing interview:** The questionnaire administered to repeat respondents as they progress through the study (interviews #2-12).

**Continuously enrolled (aka always enrolled):** A Medicare beneficiary who was enrolled in Medicare from the first day of the calendar year until the fall interview and did not die prior to the fall round. This population excludes beneficiaries who enrolled during the calendar year 2016, those who dis-enrolled or died prior to their fall interview, residents of foreign countries, and residents of U.S. possessions and territories other than Puerto Rico.

**Core sections:** These sections of the MCBS Questionnaire are of critical purpose and policy relevancy to the MCBS, regardless of season of administration.

**Crossover:** A respondents who enters a long-term care facility setting (e.g., nursing homes) or who alternates between a community and a facility setting.

**Current-year Enrollee:** Beneficiaries who were eligible and enrolled in Medicare (Parts A or B) anytime from January 1 to December 31 of the year the sample was selected.

**Ever enrolled:** A Medicare beneficiary who was enrolled at any time during the calendar year including those who dis-enrolled or died prior to their fall interview. Excluded from this population are residents of foreign countries and of U.S. possessions and territories other than Puerto Rico.

**Exit interview:** Conducted in the summer round, this interview completes the respondent's participation in the MCBS (interview #12). The exit interview is a special case of the Continuing interview.

**Facility component:** Survey of beneficiaries residing in facilities, such as long-term care nursing homes or other institutions, at the time of the interview. Interviewers do not conduct the Facility component with the beneficiary, but rather, with a staff member located at the facility.

**Fee-for-Service (FFS) payment:** Fee-for-Service is a method of paying for medical services in which each service delivered by a provider bears a charge. This charge is paid by the patient receiving the service or by an insurer on behalf of the patient.

**Field interviewer:** The principal contact for collecting and securing respondent data.

**Field manager:** A supervisor who motivates and manages a group of field interviewers to meet the goals of high quality data collection on time and within budget limits.

**Incoming Panel Sample (formerly known as Supplemental Panel):** A scientifically selected group of sampled beneficiaries that enter the MCBS in the fall of a data collection year. One panel is retired during each summer round, and a new panel is selected to replace it each fall round. Panels are identified by the data collection year (e.g., 2015 panel) in which they were selected.

**Instrumental activities of daily living (IADLs):** Instrumental activities of daily living are activities related to independent living. They include preparing meals, managing money, shopping for groceries or personal items, performing light or heavy housework, and using a telephone. If a beneficiary had any difficulty performing an activity by himself/herself, or did not perform the activity at all, because of health problems, the person was deemed to have a limitation in that activity. The limitation may have been temporary or chronic at the time of the survey. Facility interviewers did not ask about the beneficiary's ability to prepare meals or perform light or heavy housework, since they are not applicable to the beneficiary's situation; however, interviewers did question proxies about the beneficiary's ability to manage money, shop for groceries or personal items, or use a telephone.

**Internal Sample Control File:** A data file that contains every beneficiary sampled back through the beginning of MCBS. The file contains sampling information, year of selection, primary sampling unit, secondary sampling unit, contact information, and other sampling demographic information as well as final disposition codes to indicate completion status per round, component fielded per round, dates of death, and lost entitlement information.

**Long-term care facility:** A facility that provides rehabilitative, restorative, and/or ongoing skilled nursing care to patients or residents in need of assistance with activities of daily living.

**Medicare:** Medicare is the federal health insurance program for people who are 65 or older, certain younger people with disabilities, and people with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant, sometimes called ESRD). The different parts of Medicare help cover specific services:

- Hospital Insurance (Part A): covers inpatient hospital stays, care in a skilled nursing facility, hospice care, and some home health care.

- Medical Insurance (Part B): covers certain doctors' services, outpatient care, medical supplies, and preventive services.
- Medicare Advantage (Part C): an alternative to coverage under traditional Medicare (Parts A and B), a health plan option similar to a Health Maintenance Organization (HMO) or Preferred Provider Organization (PPO) administered by private companies.
- Prescription Drug Coverage (Part D): additional, optional coverage for prescription drugs administered by private companies.

For more information, please visit the Medicare.gov website at <https://www.medicare.gov/sign-up-change-plans/decide-how-to-get-medicare/whats-medicare/what-is-medicare.html>

**Medicare Advantage (MA):** Medicare Advantage Plans, sometimes called “Part C” or “MA Plans,” are offered by private companies approved by Medicare. An MA provides, or arranges for the provision of, a comprehensive package of health care services to enrolled persons for a fixed capitation payment. The term “Medicare Advantage” includes all types of MAs that contract with Medicare, encompassing risk MAs, cost MAs, and health care prepayment plans (HCPPs).

**Medicare beneficiary (aka, beneficiary):** An individual who meets at least one of three criteria (is aged 65 years or older, is under age 65 with certain disabilities, or is of any age with End-Stage Renal Disease) and is entitled to health insurance benefits. (Source: <https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html>).

**Minimum Data Set (MDS):** The Minimum Data Set (MDS) is part of the federally mandated process for clinical assessment of all residents in Medicare and Medicaid certified nursing homes. For more information, please visit <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/Minimum-Data-Set-3-0-Public-Reports/index.html>.

**Panel:** see Incoming Panel Sample

**Personal health care expenditures:** Personal health care expenditures consist of health care goods and services purchased directly by individuals. They exclude public program administration costs, the net cost of private health insurance, research by nonprofit groups and government entities, and the value of new construction put in place for hospitals and nursing homes.

**Prescription medicines:** The basic unit measuring use of prescription medicines is a single purchase of a single drug in a single container. Prescription drug use is collected only for beneficiaries living in the community or in a facility, and does not include prescription medicines administered during an inpatient hospital stay.

**Primary Sampling Unit (PSU):** Primary sampling unit refers to sampling units that are selected in the first (primary) stage of a multi-stage sample ultimately aimed at selecting individual elements (Medicare beneficiaries in the case of MCBS). PSUs are made up of major geographic areas consisting of metropolitan areas or groups of rural counties.

**Race/ethnicity:** Responses to race and ethnicity questions are self-reported by the respondent. Respondents who reported they were white and not of Hispanic origin were coded as white non-Hispanic; those who reported they were black/African-American and not of Hispanic origin were coded as black non-Hispanic; persons who reported they were Hispanic, Latino/Latina, or of Spanish origin, regardless of their race, were coded as Hispanic; persons who reported they were American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, two or more races, or other race and not of Hispanic origin were coded as other race/ethnicity.

**Reference Period:** The timeframe to which a questionnaire item refers.

**Residence status:** Full-year community residents are Medicare beneficiaries who lived solely in household units during the data collection year and who received community interviews only. Full-year facility residents are Medicare beneficiaries who lived solely in a long-term care facility during the data collection year and who received Facility interviews only. Part-year community/part-year facility residents are Medicare beneficiaries who lived part of the year in the community and part of the year in a long-term care facility, and who received both Community and Facility interviews. Skilled nursing facility users are Medicare beneficiaries who lived in either the community or a facility, and who used skilled nursing facility services during the data collection year.

**Respondent:** The person who answers questions about the beneficiary for the MCBS; this person can be the beneficiary themselves, a proxy, or a staff member located at a facility where the beneficiary resides.

**Round:** The MCBS data collection period. There are three distinct rounds each year; winter (January through April); summer (May through August); and fall (September through December).

**Sample person:** An individual beneficiary selected from MCBS' Incoming Panel sample to participate in the MCBS survey.

**Survey-reported event:** A survey-reported event is a medical service or event reported by a respondent during an interview. The event may have been matched to a Medicare claim, or it may be a survey-only event, in which case it was not matched to a Medicare claim and is only known through the survey.

**Secondary Sampling Unit (SSU):** SSUs are made up of census tracts or groups of tracts within the selected PSUs.

**Topical sections:** Sections of the MCBS Questionnaire that collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

**Ultimate Sampling Unit (USU):** USUs are Medicare beneficiaries selected from within the selected SSUs.



## APPENDIX A: TECHNICAL NOTES

**Naming conventions and reorganization:** Starting with the 2015 MCBS, the two annual releases are now referred to as follows:

- **Survey File (formerly Access to Care):** as before, this file contains survey data augmented with administrative and claims data to allow for analysis regarding beneficiaries' self-reported health status, health conditions, access to health care and satisfaction with health care.
- **Cost Supplement File (formerly Cost and Use):** as before, provides cost and utilization data and can be linked to the Survey File to conduct analysis on healthcare cost and utilization. Of note, previously demographic and survey data were included in this file; but now only the MCBS Survey File exclusively contains these data. Users will require both files to conduct analyses.

Additionally, data file segments within the release are no longer referred to as RICs,<sup>41</sup> but are now titled according to the topic of data included in the file. More complete descriptions of the data file contents can be accessed in the MCBS Data User's Guide.

**Data file structure and weights:** For analysis of Survey File data, there are two available populations of inference that can be obtained through the use of two distinct weights. First, the ever enrolled Survey File weights are populated for the largest group of respondents and represent the population of beneficiaries who were entitled and enrolled in Medicare for at least one day at any time during the 2016 calendar year. Second, the continuously enrolled Survey File weights are populated for, and represent the population of, beneficiaries who were enrolled from the first of the year (January 1, 2016) through completion of their fall interview. These weights are identical in terms of estimation to the historical Access to Care (ATC) weights that were available in previous years, and can be used in the same way as the prior weights.

Analyses of the Cost Supplement File data are conducted with the Cost Supplement weights, which represent an ever enrolled population during 2016 identical to the population represented by the ever enrolled Survey File weights but are populated for a smaller subset of respondents with complete cost and utilization data. Joint analysis of both Survey File and Cost Supplement File data should be conducted using these Cost Supplement File weights. Detailed information about the data file structure and weights can be found in Chapter 8.

**Data editing and cleaning procedures:** MCBS data files undergo thorough editing and cleaning prior to release. Quality control checks are conducted to confirm each analytic file is structurally sound. These checks include confirming that all necessary variables are present, checking variable attributes, and identifying high rates of missing data.

---

<sup>41</sup> RIC stands for Record Identification Code

Logic and reasonableness checks are also implemented for each analytic file. Logic checks verify that the questionnaire worked as expected, particularly with respect to questionnaire routing, and reasonableness checks identify values that are not explicitly disallowed by the questionnaire. Based on the results of this data review, new edits are developed to correct the errors during data cleaning. Additional information about data editing can be found in Chapter 7.

## Copyright Information and Guidelines for Data Use

This document was produced, published, and disseminated at U.S. taxpayer expense. All material appearing in this document is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Accordingly, CMS requests that data users cite CMS and the Medicare Current Beneficiary Survey as the data source in any publications or research based upon these data. Suggested citation formats are below.

**The suggested citation for the MCBS survey data files and other documentation should read:**

SOURCE: Centers for Medicare & Medicaid Services. Medicare Current Beneficiary Survey, Survey File data. Baltimore, MD: U.S. Department of Health and Human Services, 2016.