MCBS
Medicare Current Beneficiary Survey

2022 | METHODOLOGY REPORT

Centers for Medicare & Medicaid Services (CMS)

Office of Enterprise Data and Analytics (OEDA)

Version Control Log

Date	Version	Revisions
08/21/2024	1.0	Initial version published.

OVERVIEW OF MCBS DOCUMENTATION

The Centers for Medicare & Medicaid Services (CMS) releases a comprehensive suite of documentation products to support researchers in using the Medicare Current Beneficiary Survey (MCBS). This section provides a concise summary of each documentation product.

- **Data User's Guides**: A Data User's Guide is produced for each MCBS Limited Data Set (LDS) and Microdata Public Use File (PUF) data release. There are three broad categories of Data User's Guides.

  - *Survey File Data User's Guide*: Updated annually for each new data year, the *Survey File Data User's Guide* supports researchers in understanding and analyzing Survey File LDS data. This Data User's Guide contains detailed information about the Survey File LDS, including changes between years, important data user considerations, and sample code, as well as basic background information on the MCBS, including sampling, questionnaires, data collection, and data processing. Along with the *New User Tutorial* (see below), this Data User's Guide is the recommended starting point, particularly for researchers new to studying MCBS data.

  - *Cost Supplement File Data User's Guide*: Updated annually for each new data year, the *Cost Supplement File Data User's Guide* functions as a supplement to the corresponding *Survey File Data User's Guide* and supports researchers in understanding and analyzing Cost Supplement File LDS data. This Data User's Guide focuses on providing detailed information about the Cost Supplement File LDS, including changes between years, important data user considerations, and sample code.

  - *Public Use File Data User's Guides*: A Data User's Guide is also produced for each MCBS Microdata PUF release, including the annual Survey File PUF, the annual Cost Supplement File PUF, and the three COVID-19 Supplement PUFs. These Data User's Guides provide detailed, focused information to support researchers in understanding and analyzing PUF data.

- **Methodology Report** (this document): Updated annually for each new data year, the *Methodology Report* provides detailed background information on the methods used to conduct the MCBS and process MCBS data. This includes information on sampling methodology, questionnaire development and programming, interviewer recruitment and training, data collection procedures, data processing and editing, including weighting and imputation, and response rates.

- **Data User Tutorials**:

  - *New User Tutorial*: Aimed at new data users who are unfamiliar with the MCBS, the *New User Tutorial* provides an overview of MCBS history, policy relevance, survey design, data products, and best practices for analysis. Along with the *Survey File Data User's Guide* (see above), the *New User Tutorial* is the recommended starting point for researchers.

  - *Advanced Topic-Based Tutorials*: In addition to the *New User Tutorial*, CMS has released a series of tutorials on more advanced topics, with the goal of supporting researchers in better understanding how to analyze and interpret MCBS data by providing detailed analytic guidance and examples. Topics of these tutorials include the differences between MCBS Community and Facility data, weighting and variance estimation, using data from the MCBS COVID-19 Supplements, conducting longitudinal analysis, and conducting pooled cross-sectional analysis with MCBS data.

- **Glossary**: Formerly included as an appendix in MCBS documentation products, this standalone resource provides the definitions for common key terms used by the MCBS.

TABLE OF CONTENTS

LIST OF EXHIBITS

1. INTRODUCTION TO MCBS

1.1 Purpose and Goals

The Medicare Current Beneficiary Survey (MCBS) consists of a representative national sample of the Medicare population sponsored by the Centers for Medicare & Medicaid Services (CMS).[1] The MCBS is designed to aid CMS in administering, monitoring, and evaluating the Medicare program, and provides important information on beneficiaries that is not otherwise collected through operational or administrative data on the Medicare program.

The MCBS is a continuous, multi-purpose longitudinal survey, representing the population of beneficiaries aged 65 and over and beneficiaries aged below 65 with certain disabling conditions, residing in the United States. The MCBS has conducted continuous data collection since 1991, completing more than 1.2 million interviews provided by thousands of respondents. The MCBS collects this information in three data collection periods, or rounds, per year. Most interviews were traditionally conducted in-person in households and facilities using computer-assisted personal interviewing (CAPI). However, due to the coronavirus disease 2019 (COVID-19) pandemic, data collection switched to phone-only interviews in March 2020 and throughout most of 2021 with a gradual return to some in-person interviewing beginning in November 2021. MCBS data collection will include both in-person and phone interviewing going forward.

This *MCBS Methodology Report* provides an operational perspective on the collection of survey data for the 2022 MCBS data year. The 2022 data year includes both data collected in 2022 and data collected in 2023 from questionnaire sections that have a reference period in 2022. The *Methodology Report* complements other MCBS documentation (i.e., the *Data User's Guides*) with an overview of all activities carried out in support of the 2022 data files, including sampling, instrument design, interviewer training, data collection, data processing, and weighting. Please also see Section 11: Glossary for definitions of key terms used in this Report. Data users can access this *Methodology Report* along with other data documentation at https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.[2]

1.2 Survey Overview

Early in the inception of the MCBS, a design decision was made to limit beneficiary participation in the longitudinal panel to no more than four years. Initial interviews of newly-selected beneficiaries take place once per year in the fall round; these are referred to as the Incoming Panel. Since 2016, the fall round begins in late July or early August to allow more time to conduct outreach and collect information from the new Incoming Panel survey respondents. 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 as one round ends and another begins are acceptable design features of the survey.

Subsequent rounds, which occur every four months, involve re-interviewing of the same beneficiary (or appropriate proxy respondents) until they have completed four years of participation (up to 11 interviews in total); these are referred to as Continuing Panels. Interviews are conducted regardless of whether the

---

[1] The MCBS is authorized by section 1875 (42 USC 139511) of the Social Security Act and is conducted by NORC at the University of Chicago for the U.S. Department of Health and Human Services. The OMB Number for this survey is 0938-0568.

[2] This communication was printed, published, or produced and disseminated at U.S. taxpayer expense.

beneficiary lives 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
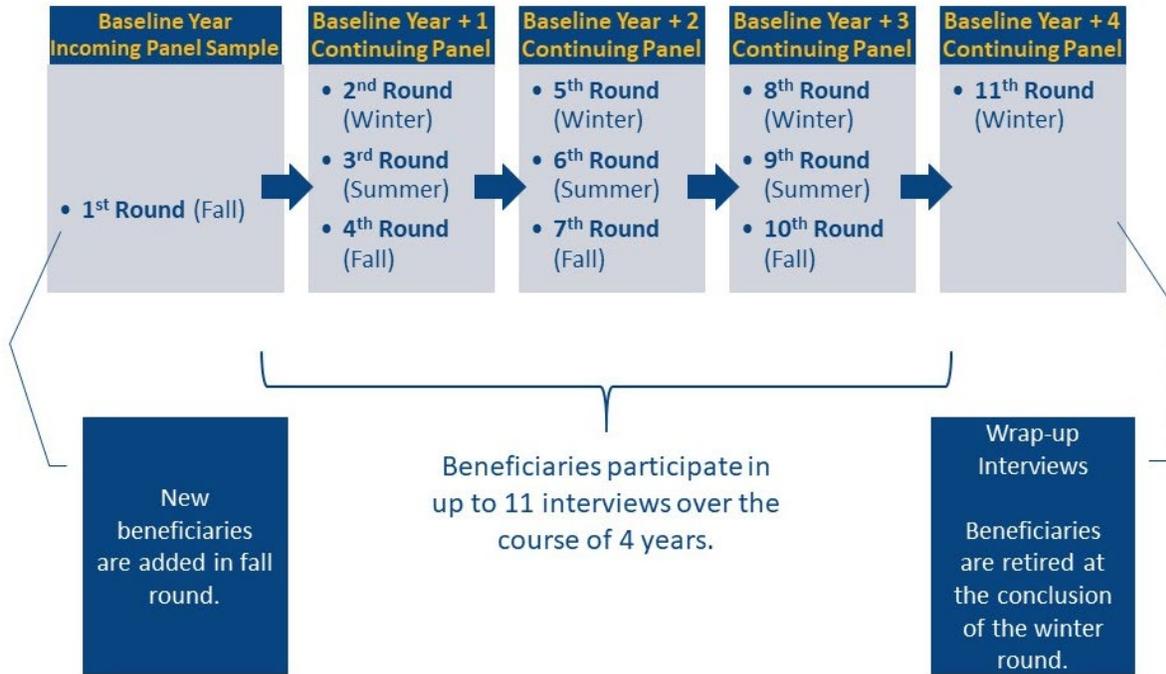
Detailed information on the sampling design can be found in Chapter 3 of this report. Chapter 6 describes the data collection fielding procedures, including eligibility for each round of the interview. Chapter 9 summarizes the results of data collection, including response rates.

1.3 Key Data Products and Analyses

MCBS data are made available via releases of annual files. For 2022, two annual Limited Data Set (LDS) releases, the Survey File and the Cost Supplement File, and two Microdata Public Use Files (PUFs) (based on the Survey File and Cost Supplement File, respectively) were released.[3] For more information on the releases, see the corresponding *MCBS Data User's Guides*.

Chapter 4 of this report provides information on the questionnaire sections associated with each data file. More details on the questionnaires, including item- and section-level changes, can be found in the *MCBS Data User's Guide: Survey File*. Chapter 7 describes the creation of the 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: Survey File* and *MCBS Data User's Guide: Cost Supplement File*.

---

[3] In addition to the annual MCBS Survey File PUF and MCBS Cost Supplement File PUF, CMS has released three special topic Microdata PUFs with data from the three MCBS COVID-19 Community Supplements, which correspond to the 2019 and 2020 data years.

2. CHANGES UNIQUE TO 2022

Several key changes were made to the MCBS during 2022, 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

There were no changes to sampling for the 2022 data year.

2.2 Questionnaires

For more information about Community Questionnaire and Facility Instrument content, including item- and section-level descriptions of 2022 questionnaire changes, see the *2022 MCBS Data User's Guide: Survey File* and *2022 MCBS Data User's Guide: Cost Supplement File*.

2.3 Data Collection

Outreach and interviews were conducted both by phone and in-person throughout 2022. This mixed mode approach began once CMS was safely able to transition from phone-only data collection at the onset of the COVID-19 pandemic. Mode determination involved several factors, including field interviewer location and respondent preference. Field staff considered local COVID-19 policies and rates and used personal protective equipment to ensure the safety of respondents and interviewers alike.

2.4 Documentation

Formerly included as Appendix A of this document, the definitions of common key terms used by the MCBS have been moved into a standalone Glossary document available on the CMS website: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks

2.5 Data Processing

MCBS data files undergo thorough editing and quality control checks prior to release. For more detailed information regarding data editing, imputation, and weighting procedures conducted for the 2022 LDS releases, please consult the *2022 Data User's Guide: Cost Supplement File* and the *2022 Data User's Guide: Survey File* available on the CMS MCBS website.

Imputation:

The 2022 Income and Assets imputation added separate variables for monthly earnings from work for the beneficiary and for their spouse or partner. Previously, these variables were combined into a single variable and this variable was imputed on its own.

Additionally, health insurance premiums for private plans and MA plans have been standardized to reflect monthly premiums and, when missing, premium information has been imputed.

The 2022 imputation process utilized a larger proportion of Medicaid payment amounts that were obtained from Transformed Medicaid Statistical Information System (T-MSIS) for Medicaid-covered services among Medicare beneficiaries who are dually-eligible for Medicaid.

The 2022 Cost Supplement File LDS accounts for modifications to the sequestration payment adjustments that occurred within 2022 due to The Protecting Medicare and American Farmers from Sequester Cuts Act:

- No payment adjustment through March 31, 2022

- 1% payment adjustment April 1 to June 30, 2022

- 2% payment adjustment beginning July 1, 2022

3. SAMPLE DESIGN FOR THE MCBS 2022 PANEL

3.1 Overview of MCBS Sample Design

The MCBS employs a three-stage cluster sample design (see Exhibit 3.1.1). At the first stage, the MCBS used the set of 104 primary sampling units (PSUs) employed for sampling for the MCBS, all of which are in the continental United States.[4] At the second stage, the MCBS used the set of 685 census tract-based secondary sampling units (SSUs) selected within those PSUs. At the third stage, the MCBS selected Medicare beneficiaries, the ultimate sampling units (USUs), from within the selected tract-based SSUs.[5]

Exhibit 3.1.1: MCBS Sample Design Process

The MCBS employs a three-stage cluster sample design:

1 — **Primary sampling units (PSUs)** are made up of major geographic areas consisting of metropolitan areas or groups of rural counties.

2 — **Secondary sampling units (SSUs)** are made up of census tracts or groups of tracts within the selected PSUs.

3 — Medicare beneficiaries, the **ultimate sampling units (USUs)**, are then selected from within the selected SSUs using a systematic sampling scheme with random starts.

In 2022, the MCBS continued to use the sample rotation pattern used historically. In particular, the newly selected 2022 Panel, and the panels selected in 2019, 2020, and 2021 continued into Fall 2022 and beyond according to their established rotation schedules.[6] The 2018 Panel (which was first fielded in Fall 2018) exited at the conclusion of the Winter 2022 round and was replaced with a new sample of beneficiaries in Fall 2022. Exhibit 3.1.2 displays the MCBS rotating panel design from 2018 to the present panel.

---

[4] Note, Puerto Rico was originally included in the MCBS sample and removed in 2017. See prior *MCBS Methodology Reports* for historical sampling information: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey-data-documentation-codebooks.

[5] 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.

[6] 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 four years, participating in a total of 11 rounds.

Exhibit 3.1.2: 2018-2022 MCBS Rotating Panel Design

Data Collection Schedule			Panel				
Calendar Year	Season	Round#	2018	2019	2020	2021	2022
2018	Winter	80					
	Summer	81					
	Fall	82					
2019	Winter	83					
	Summer	84					
	Fall	85					
2020	Winter	86					
	Summer	87					
	Fall	88					
2021	Winter	89					
	Summer	90					
	Fall	91					
2022	Winter	92					
	Summer	93					
	Fall	94					

This section documents the procedures used to select the new sample for Fall 2022 (i.e., the 2022 Panel). The 2022 Panel will be retained in the survey for the four years specified under the MCBS sample rotation scheme and is designed to: (a) replace approximately one-third[7] of the respondents in the existing MCBS sample; and (b) extend survey coverage to persons added to the Medicare rolls during the current year (see Section 3.4 for details).

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 2022 MCBS Panel. A first extract of the enrollment data 5-percent file was provided in March 2022, and the bulk of the 2022 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.[8] The combination of these extracts constitutes the full frame from which the 2022 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

---

[7] 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.

[8] 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.

in the selected SSUs was then selected from each stratum.[9] 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 3.1.3.

Exhibit 3.1.3: 2022 MCBS Sampling Strata

Hispanic	Non-Hispanic
Under 45 Hispanic	Under 45 non-Hispanic
45 - 64 Hispanic	45 - 64 non-Hispanic
65 - 69 Hispanic	65 - 69 non-Hispanic
70 - 74 Hispanic	70 - 74 non-Hispanic
75 - 79 Hispanic	75 - 79 non-Hispanic
80 - 84 Hispanic	80 - 84 non-Hispanic
85 and over Hispanic	85 and over non-Hispanic

Sampling rates varied by stratum, with the strata containing younger beneficiaries with disabilities (under 45), elderly beneficiaries (85 and over), and Hispanics being oversampled to permit more detailed analysis of these subpopulations because of interest in their special health care needs. 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 14 sampling strata. The MCBS sample is designed to yield about 14,500 completed cases annually in the MCBS Survey File and about 9,000 completed cases annually in the MCBS Cost Supplement File.

For 2022, the number of responding beneficiaries across all panels to comprise the 2025 Cost Supplement File was estimated to be 8,449.[10] This is expected to be comprised of approximately 500-800 beneficiaries from each of the under 65 (disability) age groups and approximately 1,300-1,600 beneficiaries from each of the remaining age groups.

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 in a manner that maximized overlap with the original PSU sample.[11] The 28 largest PSUs in the continental U.S. and the largest PSU in Puerto Rico were designated as certainty PSUs. The remaining non-certainty PSUs were grouped by census region and Metropolitan Statistical Area (MSA) status (where Puerto Rico was treated as a separate "region" for sampling purposes). Thirty-eight non-certainty 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

---

[9] Note that the MCBS surveys beneficiaries living in the community (e.g., households) and living in a facility (e.g., nursing home); however, residence status is not known at the time of sampling and is therefore not included among the MCBS sampling strata.

[10] The number corresponds to the 2025, rather than the 2022, Cost Supplement File because 2025 is the final year that the 2022 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.

[11] See prior *MCBS Methodology Reports* for more information on the 2001 reselection of PSUs: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

procedures designed to maximize overlap with the existing MCBS sample. A total of 107 PSUs (including 29 certainty and 78 non-certainty) was selected in 2001.

The PSUs are examined periodically for representativeness of 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. With the removal of Puerto Rico from the sample in 2017, the three Puerto Rico PSUs were removed. Thus, the final 2022 Panel was selected from the remaining 104 MCBS PSUs, all of which are in the continental United States, and include 28 certainty PSUs and 76 non-certainty PSUs.

3.3 Selection of MCBS SSUs

In 2014, the MCBS SSUs were selected using census tracts or clusters of adjacent tracts. Use of census tracts requires minimal maintenance and facilitates merging of MCBS data with U.S. Census Bureau data and other aggregate level geographic or environmental extant data.[12]

A total of 703 core SSUs, comprised of 242 SSUs from the 29 certainty PSUs and 461 SSUs from the 78 non-certainty PSUs, were selected in 2014 within the 107 PSUs. An additional reserve sample of 339 SSUs (122 from the certainty PSUs and 217 from the non-certainty PSUs) was also selected to provide CMS the possibilities to expand the sample or to study special rare populations in future years. With the removal of Puerto Rico from the sample in 2017, the 18 SSUs selected from the three Puerto Rico PSUs were removed from the sample, leaving a set of 685 core SSUs to be used for sample selection. After being phased in over four years, all panels are now selected from the census tract-based SSUs.

3.4 Selection of Beneficiaries for the 2022 MCBS Panel

The third stage of sampling is the selection of Medicare beneficiaries from each SSU.

3.4.1 Current-Year Enrollee Sample

Since 2015, the year t cohort[13] of beneficiaries (i.e., the set of current-year enrollees) was included in the sampling frame of beneficiaries from which the year t panel[14] was selected.[15] The MCBS used multiple enrollment data extracts for sampling and multiple sample draws because not all 2022 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 2022 Panel, were required. The first, and largest, extract, which contained the bulk of the 2022 sampling frame, was created by CMS in March of 2022. The majority of the 2022 Panel was selected from this initial extract. Additional enrollment data extracts of 2022 enrollees were pulled in early August and late September 2022, and additional samples of 2022 enrollees were drawn from these extracts. The sampling frame for the 2022 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 2023 and used to fully enumerate the 2022 population of Medicare enrollees. Because data collection had already ended for Fall 2022, no sample was drawn from the January extract by design; however, the information was used for

---

[12] See prior *MCBS Methodology Reports* for more information on the pre-2014 selection of SSUs using Zip Codes and the creation and selection of SSUs using census tracts: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

[13] An annual cohort is the set of beneficiaries who are enrolled in Medicare and appear in the Medicare enrollment data within a given year.

[14] An annual panel is the set of beneficiaries sampled in a given year and initially interviewed in the fall round of that year.

[15] Historically, to be eligible for sample selection, beneficiaries had to be eligible for Medicare and enrolled by January 1st of the sampling year, instead of at any time during the year. See prior *MCBS Methodology Reports* for historical sampling information: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

weights calibration. Please see the coverage analysis discussion later in this section 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 were all 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 beneficiaries 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 individuals 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).

3.4.2 Hispanic Oversample

Oversampling of Hispanics has been conducted using a variety of methodologies throughout the MCBS and continues as a design feature.[16] For 2022, Hispanics were oversampled relative to their non-Hispanic counterparts within the general MCBS sample. The sampling frame was stratified using a flag provided by CMS based on Census records of Hispanic surnames and other enrollment information, such as language preference, and the Hispanic stratum was oversampled relative to the non-Hispanic stratum. The target of 1,500 annual Hispanic completes was nearly achieved in the 2022 Survey File (total of 1,475 Hispanic completes).

3.4.3 Sample Selection Overview

The sample of MCBS beneficiaries was selected using systematic sampling within each PSU, and specifically only within the 685 core SSUs selected within the 104 PSUs. In May 2022, the majority of the 2022 Panel was selected. In August and October, additional small samples of 2022 enrollees were selected using the same sampling rates as for the initial sample. The sample sizes for the 2022 Panel were determined in early Spring 2022 based on the most up-to-date response rates available at that point in time. As a matter of routine, reserve samples, typically sized at around an additional 10 to 20 percent of the core sample, are also selected as part of each annual sample. In 2022, a reserve sample of an additional 25 percent was selected as part of the 2022 Panel. This allowed for a larger potential sample release in the event of lower response rates in the fall round, when data collection for this panel began. During data collection, only a small portion of the selected reserve sample (a total of 51 cases from the second and third extracts) was released, as detailed further below.

For the 2022 Panel, an initial sample of 20,962 beneficiaries (including the reserve and the Hispanic oversample) was selected in May. In August, an additional 197 current-year enrollees were selected using the sampling rates computed for the first extract and added to the 2022 Panel. In October, 179 additional current-year enrollees were selected, again using the sampling rates computed for the first extract, and added to the 2022 Panel. As of October, the 2022 MCBS Panel was complete, with a total of 21,338 beneficiaries (including the reserve and 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 2022 Panel are given below.

---

[16] Prior to 2017, this was accomplished primarily via sampling in Puerto Rico. Beginning in 2015 through 2018, as an enhancement to this traditional oversample, there was an *additional* oversample of Hispanic beneficiaries living outside of Puerto Rico in new panels to allow for improved precision of estimates of health disparities experienced by these populations. Hispanics were also further oversampled to compensate for the removal of Puerto Rico in 2017. See prior *MCBS Methodology Reports* for historical sampling information: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

3.4.4 Sample Size Determination

The sample size requirements for the 2022 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 sampling year and are determined to be deceased at the first or second interview and whose date of death is confirmed by a proxy to be prior to the sampling year but for whom no death is recorded in CMS administrative updates;

b) Persons in the CMS sampling frame who enrolled prior to the sampling year and are determined to be ineligible for Medicare in the first or second interview and whose loss of entitlement is confirmed by the respondent or a proxy to be prior to the sampling year but for whom there is no record of having lost eligibility in CMS administrative updates; or

c) Persons who enrolled prior to 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 at the time of sampling.[17] 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).[18] 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,[19] response rates, and immortal rates were computed within each age group.[20] The immortal and death rates used were an average of historical rates and actual rates from Fall 2019, Fall 2020, and Fall 2021. 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 2022 Panel. These rates were used in each of the Hispanic and non-Hispanic sampling strata within each age group.

---

[17] Note that members of the 2022 cohort (i.e., 2022 sampled panel members who first became eligible for Medicare during 2022) who died or lost eligibility during the sampling year (i.e., sometime during 2022 after becoming eligible) are not immortals and should still be sampled. These cases contribute to the 2022 Cost Supplement File.

[18] 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.

[19] Included in the calculation of death rates is a small number of persons who lost Medicare eligibility.

[20] 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 2022 Panel, by Age Group

Sampling Rate (in percent)	Age Group (as of 12/31/2022)							
	<45	**45-64**	**65-69**	**70-74**	**75-79**	**80-84**	**85+**	**Total**
Estimated "immortal" rate	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Estimated selection year death rate	1.8	2.3	1.4	2.4	3.8	5.5	12.8	4.6
Selection year response rate	33.9	39.2	38.4	37.6	38.8	38.7	38.4	38.1
Post-fall round death/loss rate	1.3	0.3	0.0	0.0	0.0	0.1	0.1	0.2
First year response rate	55.0	59.7	59.2	59.8	61.6	61.7	61.4	60.1
Estimated first year death rate	1.3	2.9	1.8	3.3	4.7	6.6	12.6	5.0
Second year response rate	69.7	67.9	74.6	75.4	73.8	70.2	69.6	72.0
Estimated second year death rate	3.2	3.0	2.0	2.9	4.2	7.3	12.0	5.2
Third year response rate	79.2	77.5	84.7	85.4	83.1	79.8	74.8	81.3
Year 1 Retention rate[1]	18.1	22.8	22.4	21.9	23.0	22.5	20.6	21.8
Year 2 Retention rate[2]	68.8	65.9	73.3	72.9	70.3	65.6	60.9	68.4
Year 3 Retention rate[3]	76.6	75.2	83.1	82.9	79.6	74.0	65.9	77.0

SOURCE: 2022 MCBS Internal Sample Control File

[1] 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.

[2] 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.

[3] 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.

When initial sample size calculations were made, the response rate for the selection year (i.e., the proportion of sampled beneficiaries, excluding deaths and immortals, who complete the initial fall interview) was assumed to be approximately 38 percent for the 2022 Panel and for future panels. 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), the second year in the survey (i.e., the proportion of living beneficiaries in the first Cost Supplement File who also provide substantially complete data for the second Cost Supplement File), and the third year in the survey (i.e., the proportion of living beneficiaries in the second Cost Supplement File who also provide substantially complete data for the third Cost Supplement File) were based on averages of corresponding rates from 2020 and 2021.

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 (under 45 years), the oldest age category (85 years and over), and the 65 to 69 year-old age category. As the panel ages, the oldest beneficiaries in the under 45 age category will move to the next age category, and there will be no migration into the under 45 age category. On the other hand, there will not be any migration out of the oldest age category (85 years and over), while about 17 to 19 percent of the beneficiaries from the 80 to 84 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 8 percent) than the migration out of this category (about 19 to 25 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 2022 Panel was originally determined to be 17,109 beneficiaries, with the plan to select an additional 25 percent reserve sample.

3.4.5 2022 Sampling Frame

Three extracts of enrollment data were used to create the 2022 MCBS sampling frame and support sampling for the 2022 Panel. The first, or initial, extract of the enrollment data, delivered in March, included:

- Beneficiaries who were first eligible for Medicare before January 1, 2022 and still alive and eligible on January 1, 2022; and

- Beneficiaries who were first eligible for Medicare between January 1, 2022 and March 1, 2022 (inclusive) or who would be automatically enrolled in Medicare during the four months after the first extract (through July), regardless of vital status.

A second extract, delivered in early August, included beneficiaries not included in the first extract and who were first eligible for Medicare between January 1, 2022 and August 1, 2022 (inclusive) or who would be automatically enrolled in Medicare between August 1 and November 30, regardless of vital status.

A third extract, delivered in late September, included beneficiaries not included in the first or second extract and who were first eligible for Medicare between January 1, 2022 and September 1, 2022 (inclusive) or who would be automatically enrolled in Medicare between September 1 and December 31, 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[21] 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 specified for the MCBS for use in sampling beneficiaries for the annual panels. This is referred to as the 2022 enrollment data subsample.

CMS subset each of its enrollment data extracts as described above, keeping only beneficiaries meeting the criteria for the 2022 enrollment data subsample. These enrollment data subsample extracts are further subset to include only beneficiaries falling within the 685 selected MCBS SSUs. Exhibit 3.4.2 shows the number of beneficiaries by sampling stratum (age group by ethnicity) in the three 2022 enrollment data subsample extracts and the resulting 2022 sampling frame. Of the 3,279,455 beneficiaries in the combined 2022 enrollment data subsample extracts, a total of 50,679 beneficiaries fell within the selected MCBS PSUs and SSUs and were eligible for sampling in 2022.

Exhibit 3.4.2: Number of Beneficiaries in 2022 Enrollment Data Subsample Extracts (Combined) and 2022 Sampling Frame, by Stratum

Stratum	Age Group/Ethnicity	Three Extracts Combined	2022 Sampling Frame
1	<45, Hispanic	9,727	124
2	45-64, Hispanic	31,698	397
3	65-69, Hispanic	71,322	988
4	70-74, Hispanic	63,367	885
5	75-79, Hispanic	43,584	598
6	80-84, Hispanic	27,213	401
7	85+, Hispanic	26,046	363
8	<45, non-Hispanic	72,611	1,092
9	45-64, non-Hispanic	275,991	4,289

---

[21] The enrollment data include over 100,000,000 beneficiaries.

Stratum	Age Group/Ethnicity	Three Extracts Combined	2022 Sampling Frame
10	65-69, non-Hispanic	786,149	12,065
11	70-74, non-Hispanic	700,539	11,038
12	75-79, non-Hispanic	521,223	8,249
13	80-84, non-Hispanic	327,716	5,126
14	85+, non-Hispanic	322,269	5,064
Total		3,279,455	50,679

SOURCE: 2022 MCBS Internal Sample Control File and 2022 enrollment data extracts.

Using the initial 2022 enrollment data subsample extract in combination with previous annual enrollment data subsamples, the size of the total 2022 enrollment data subsample (containing all projected 2022 Medicare enrollees, through December 31, 2022, that would be available for sampling) could be forecast at the time of initial sampling (May 2022). 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 2023 and used to fully enumerate the 2022 cohort to (a) inform undercoverage of the 2022 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.

3.4.6 Sample Selection for the 2022 Panel

The goal for the 2022 Panel was to select a sample of beneficiaries, which includes a core sample of 17,109 beneficiaries and a 25 percent reserve sample, with targeted oversamples of the youngest (64 and younger) and oldest (85 and over) age groups and of Hispanic beneficiaries across all age groups.

Sampling Fractions. The sampling fractions for the Hispanic and non-Hispanic strata 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 sampling fractions for the MCBS were completed at the national level within the 14 strata (seven age groups by the Hispanic/non-Hispanic flag).

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, and let f_{-1a} be the national sampling fraction for the non-Hispanic stratum in age group a. The inclusion probability for the i-th PSU is denoted by π_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

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

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 $\rho_{1ak|ij}$ and $\rho_{-1ak|ij}$ 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 3.4.3 displays the final sampling fractions used for calculating probabilities of selection, by stratum, for the 2022 Panel.

Exhibit 3.4.3: 2022 MCBS Panel, Final Sampling Fractions by Stratum

Stratum	Age Group/Ethnicity	Final Sampling Fraction, in Percent
1	<45, Hispanic	1.2798
2	45-64, Hispanic	1.0639
3	65-69, Hispanic	0.6443
4	70-74, Hispanic	0.6127
5	75-79, Hispanic	1.7378
6	80-84, Hispanic	2.8208
7	85+, Hispanic	1.3986
8	<45, non-Hispanic	1.5183
9	45-64, non-Hispanic	0.6742
10	65-69, non-Hispanic	0.4170
11	70-74, non-Hispanic	0.3815
12	75-79, non-Hispanic	0.5487
13	80-84, non-Hispanic	1.0105
14	85+, non-Hispanic	1.1433
Total		0.6749

SOURCE: 2022 MCBS Internal Sample Control File

Selection of the 2022 Panel. The 2022 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. Beneficiaries were ordered within each PSU by age group, SSU (to approximate geographic serpentine sorting), ethnicity flag, and extract.[22] There were 3,414 beneficiaries with a conditional probability of selection equal to 1. These beneficiaries were selected with certainty, given the selection of their PSUs and SSUs.

3.4.7 Sampling Results

As described above, a 25 percent reserve sample was selected as part of the 2022 Panel to allow for flexibility. At the time of sample estimation in March of 2022, we identified two special considerations related to the calculation of the core and reserve sample sizes. First, Continuing Panel sizes had been trending downward throughout the pandemic, impacting the overall size of resulting Survey File and Cost Supplement File data files. Second, continued uncertainty related to conducting in-person interviews remained. Both of these considerations required flexibility in planning. In response, (a) the selection year response rate used in sample size calculations was the low end of expectations, resulting in a "maximum" core sample size expected to be required to achieve our target for the Incoming Panel; and (b) a larger than typical reserve sample (25 percent compared to the typical 10 percent) was selected to provide flexibility if budget and staffing allowed for additional sample releases. The following two exhibits reflect the total 2022 Panel sample, including selected

---

[22] 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.

core sample cases and *released* reserve sample cases. These tables present the total number of beneficiaries in the 2022 Panel, including the Hispanic oversample.

Exhibit 3.4.4: 2022 MCBS Panel, Number of Beneficiaries Selected by Age Group

Age Group (as of December 31, 2022)	Total Selected Beneficiaries
<45	1,216
45-64	1,517
65-69	3,102
70-74	2,538
75-79	2,657
80-84	2,944
85+	3,177
Total	17,151

SOURCE: 2022 MCBS Internal Sample Control File

Exhibit 3.4.5: 2022 MCBS Panel, Number of Beneficiaries Selected by Stratum

Stratum	Age Group/Ethnicity	Total Selected Beneficiaries
1	<45, Hispanic	124
2	45-64, Hispanic	196
3	65-69, Hispanic	355
4	70-74, Hispanic	296
5	75-79, Hispanic	309
6	80-84, Hispanic	353
7	85+, Hispanic	363
8	<45, non-Hispanic	1,092
9	45-64, non-Hispanic	1,321
10	65-69, non-Hispanic	2,747
11	70-74, non-Hispanic	2,242
12	75-79, non-Hispanic	2,348
13	80-84, non-Hispanic	2,591
14	85+, non-Hispanic	2,814
Total		17,151

SOURCE: 2022 MCBS Internal Sample Control File

The number of current-year enrollees (beneficiaries who enrolled in 2022) selected into the 2022 Panel (including the Hispanic oversample) is displayed in Exhibit 3.4.6 below.

Exhibit 3.4.6: 2022 MCBS Panel, Number of Current-Year Enrollees Selected by Age Group

Age Group	Initial Extract	Three Extracts Combined
<45	>30*	>50*
45-64	47	102
65-69	218	513
70-74	0	<11*
75-79	0	<11*
80-84	0	<11*
85+	<11*	<11*
Total	306	682

SOURCE: 2022 MCBS Internal Sample Control File
* Cell sizes of less than 11 are suppressed. Complementary suppression is used so that suppressed cells cannot be derived.

Several quality checks were performed after sample selection. These included the comparison of the weighted 2022 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.

3.4.8 Coverage Analysis of the 2022 Sampling Frame

As discussed above, a final enrollment data 5-percent file extract was provided in mid-January 2023. This extract was used to fully enumerate the 2022 cohort to (a) inform undercoverage of the 2022 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 2022 enrollment data subsample and the 2022 MCBS population. To construct the full 2022 enrollment data subsample, all records of eligible beneficiaries enrolled through December 31, 2022, from the four extracts were combined. From that universe, the 2022 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 2022, 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 2022 MCBS population includes all beneficiaries who were enrolled in Medicare in 2022 and reside in the MCBS PSUs and SSUs.

Exhibit 3.4.7 displays the full 2022 enrollment data subsample and the estimated 2022 eligible U.S. Medicare population, by stratum. This table builds on Exhibit 3.4.3, which displayed the 2022 enrollment data subsample file through the third extract.

The fourth enrollment data subsample extract is smaller in size than the third extract (27,932 beneficiaries overall in the fourth extract, compared to 37,757 in the third extract). The number of cases from the fourth extract falling into the MCBS PSUs and SSUs is also smaller than those in the third extract (371 in the fourth extract versus 552 in the third extract).

Overall, the fourth extract accounts for 0.84 percent of the total 2022 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 2022 Panel. Any imbalance will be accounted for in adjustments made to the weights, discussed in Section 8.

Exhibit 3.4.7: Number of Beneficiaries in 2022 Enrollment Data Subsample and Estimated 2022 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	9,761	195,220
2	45-64, Hispanic	31,800	636,000
3	65-69, Hispanic	73,436	1,468,720
4	70-74, Hispanic	63,470	1,269,400
5	75-79, Hispanic	43,615	872,300
6	80-84, Hispanic	27,230	544,600
7	85+, Hispanic	26,057	521,140
8	<45, non-Hispanic	72,906	1,458,120
9	45-64, non-Hispanic	276,821	5,536,420
10	65-69, non-Hispanic	809,746	16,194,920
11	70-74, non-Hispanic	701,136	14,022,720
12	75-79, non-Hispanic	521,346	10,426,920
13	80-84, non-Hispanic	327,764	6,555,280
14	85+, non-Hispanic	322,299	6,445,980
Total		3,307,387	66,147,740

SOURCE: 2022 MCBS Internal Sample Control File

Exhibit 3.4.8 compares the original forecast of the full sampling[23] 2022 enrollment data subsample, including all cases expected to be in the enrollment data through the end of the 2022 and available for sampling, to the actual count of beneficiaries in the combined three enrollment data subsample extracts. As described above, the forecast was used to develop sampling fractions for use in the selection of the 2022 MCBS Panel sample. The comparisons in Exhibit 3.4.8 are given by stratum and overall. The counts are quite close; the total actual overall count is only slightly lower than the forecast (3,279,455 actual versus 3,279,727 forecast beneficiaries), and the differences by stratum are small.

---

[23] This includes the first three extracts of the 2022 enrollment data. Because we only sample from the first three extracts, our forecast only projects to the three-extract total. This allows for the highest degree of accuracy in the construction of sampling fractions.

Exhibit 3.4.8: Forecast Compared to Actual Beneficiaries in Full Sampling 2022 Enrollment Data Subsample, by Stratum

Stratum	Age Group/Ethnicity	Forecast[1] of Beneficiaries in Full 2022 Enrollment Data Subsample	Actual[2] Beneficiaries in Full 2022 Enrollment Data Subsample
1	<45, Hispanic	9,764	9,727
2	45-64, Hispanic	31,833	31,698
3	65-69, Hispanic	71,413	71,322
4	70-74, Hispanic	63,336	63,367
5	75-79, Hispanic	43,565	43,584
6	80-84, Hispanic	27,193	27,213
7	85+, Hispanic	26,027	26,046
8	<45, non-Hispanic	73,332	72,611
9	45-64, non-Hispanic	277,141	275,991
10	65-69, non-Hispanic	784,365	786,149
11	70-74, non-Hispanic	700,552	700,539
12	75-79, non-Hispanic	521,234	521,223
13	80-84, non-Hispanic	327,712	327,716
14	85+, non-Hispanic	322,260	322,269
Total		3,279,727	3,279,455

SOURCE: 2022 MCBS Internal Sample Control File
NOTE: The Full 2022 Enrollment Data Subsample in this table includes all current-year enrollees through December 31, 2022 that were available for sampling through the end of 2022 (i.e., all beneficiaries included in the first three enrollment data extracts).
[1]Forecast was calculated at the time of sampling (May 2022) and includes projected beneficiaries in the first three enrollment data extracts.
[2]Actual counts based on enrollment data records received for 2022 in the first three enrollment data extracts but excluding counts in the final extract delivered in January 2023.

3.5 Continuing Sample (2018-2021 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 11 rounds, starting in the fall round of the year the panel is selected. In Winter 2022, the 2018 Panel completed its 11th and final round, the 2019 Panel was in its 8th round of participation, the 2020 Panel was in its 5th round, and the 2021 Panel was in its 2nd round.

3.6 Fielded Sample Sizes by Panel and Round

During 2022, sampled beneficiaries were interviewed during three rounds: a winter round, a summer round, and a fall round. As mentioned earlier, during Winter 2022, the 2018 Panel was interviewed for its final time, and in Fall 2022, the 2022 Panel was interviewed for its first time. The fielded sample sizes,[24] by panel, for each round are given in Exhibit 3.6.1.

---

[24] Note that these are not the original sample sizes when the panel was selected (except in the case of the 2022 Panel), but the sample remaining in the round, less attrited beneficiaries and other sample losses, which are fielded in that round.

Exhibit 3.6.1: 2022 Fielded Sample Sizes by Round for Each Panel

Round	Panel	Fielded Sample Sizes
Winter 2022	2018	2,153
	2019	2,480
	2020	3,486
	2021	5,789
	All	13,908
Summer 2022	2019	2,247
	2020	3,001
	2021	4,440
	All	9,688
Fall 2022	2019	2,090
	2020	2,662
	2021	3,598
	2022	17,139
	All	25,489

SOURCE: 2022 MCBS Internal Sample Control File

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. Questionnaire items often ask respondents to recall events or experiences during a certain time period. A reference period is the timeframe to which a questionnaire item refers. See Exhibit 4.1 for a depiction of the MCBS Questionnaire structure.

- Community Component: Survey administered for beneficiaries living in the community (i.e., not in a long-term care facility such as a nursing home) during the reference period covered by the MCBS. An interview may be conducted with the beneficiary or a proxy.

- Facility Component: Survey administered for beneficiaries living in facilities, such as long-term care nursing homes or other institutions, during the reference period covered by the MCBS interview. Interviewers conduct the Facility component with staff members located at the facility (i.e., Facility respondents); beneficiaries are not interviewed if they reside at a facility.

Exhibit 4.1: MCBS Questionnaire Overview
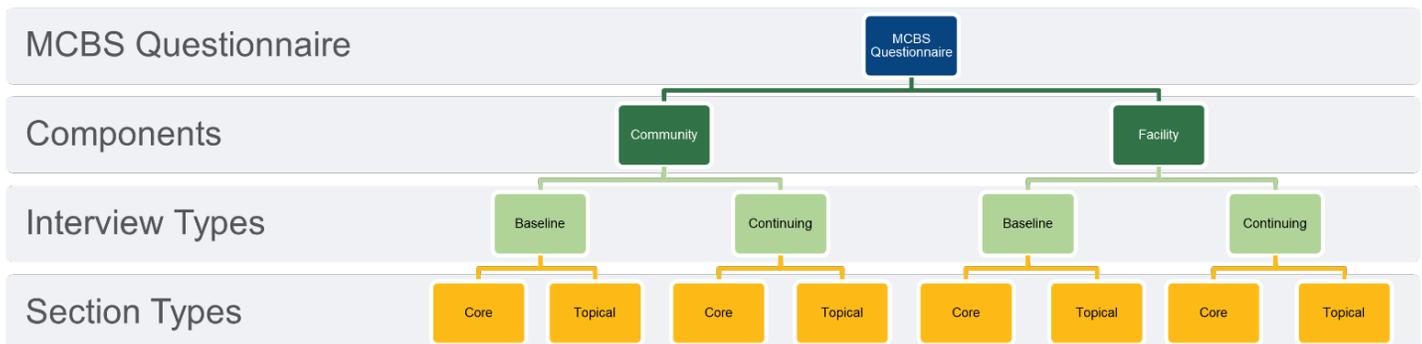
Interviews are conducted in one or both components in a given data collection round, depending on the beneficiary's living situation. Procedures for these "crossover" interviews (where the beneficiary moves from one component to another) are described in Section 6.2.

Within each component, there are two types of interviews – a Baseline interview and a Continuing interview.

- Baseline: The initial questionnaire administered in the fall round of the year the beneficiary is selected into the sample (interview #1).

- Continuing: The questionnaire administered as beneficiaries progress through the study (interviews #2-11).

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 health 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 the *2022 MCBS Data User's Guide: Survey File* for tables of the 2022 Core and Topical sections.

Data collected by the Community and Facility components are released to users via two primary 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 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 beneficiary has been in the MCBS, and the component of the most recent interview.

For more information about Community Questionnaire content, including descriptions of each questionnaire section, see the *2022 MCBS Data User's Guide: Survey File* and *2022 MCBS Data User's Guide: Cost Supplement File*.

4.1.1 Interview Type

As the 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 eight interview types, summarized in Exhibit 4.1.1 below. Three 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. The most recent interview was the fall Baseline interview. The second round interview is the first time utilization and cost data are collected.	Winter
C010	Second round "holdover," meaning the winter interview was skipped. Most recent interview was the fall Baseline interview; therefore, the third round interview is the first time in which utilization and cost data are collected.	Summer

*Interview types for exit panel Community cases in the summer round (INTTYPEs C008 and C009) were removed from the questionnaires in 2019.

4.1.2 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. 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 for the 2022 calendar year, which is synonymous with the 2022 data year for Baseline cases. The Community Questionnaire flow varies based on fielding and operational factors; Exhibit 4.1.3 illustrates the most common flow for standard Community Continuing sample in 2022. This flow reflects the data collection year, rather than the data year, meaning interviews conducted in 2022 used the flow depicted. The 2022 LDS includes data collected in 2022 as well as in other years, which may have slightly different questionnaire flows. This means that some questionnaire sections included in the Survey File will not be reflected in this exhibit if they were first fielded in winter or summer of the following year. To continue to prioritize the collection of seasonal section data and facilitate telephone data collection in response to the COVID-19 pandemic, the questionnaire order was maintained to administer the seasonal sections after Health Insurance (HIQ) and before all utilization and Cost Series sections. This change was made permanent in Summer 2022 due to positive feedback from field interviewers on the revised questionnaire flow.

Starting in Summer 2021, COVID-19 items were fielded within the MCBS Community Questionnaire in the COVID-19 Questionnaire (CVQ). More information on this update can be found in the *2021 MCBS Data User's Guide: Survey File*.

Exhibit 4.1.2: 2022 MCBS Community Questionnaire Flow for Baseline Interview

Introduction

INQ
INTRODUCTION

Enumeration Summary and Housing Characteristics

ENS
ENUMERATION
SUMMARY

→

HAQ[a]
HOUSING
CHARACTERISTICS

Health Insurance

HIQ
HEALTH
INSURANCE

Topical, Experiences with Care, and Health Status Sections

MBQ[a]
MOBILITY OF
BENEFICIARIES

→

PVQ[a]
PREVENTIVE
CARE

→

CVQ[a]
COVID-19

→

HFQ
HEALTH STATUS AND
FUNCTIONING

NAQ[a]
NICOTINE AND
ALCOHOL USE

→

SCQ
SATISFACTION WITH
CARE

→

CMQ[a]
COGNITIVE MEASURES

Demographics and Income

DIQ
DEMOGRAPHICS AND
INCOME

Closing

Interviewer Remarks (IRQ)

a = Topical Section

Exhibit 4.1.3: 2022 MCBS Community Questionnaire Flow for Continuing Interview
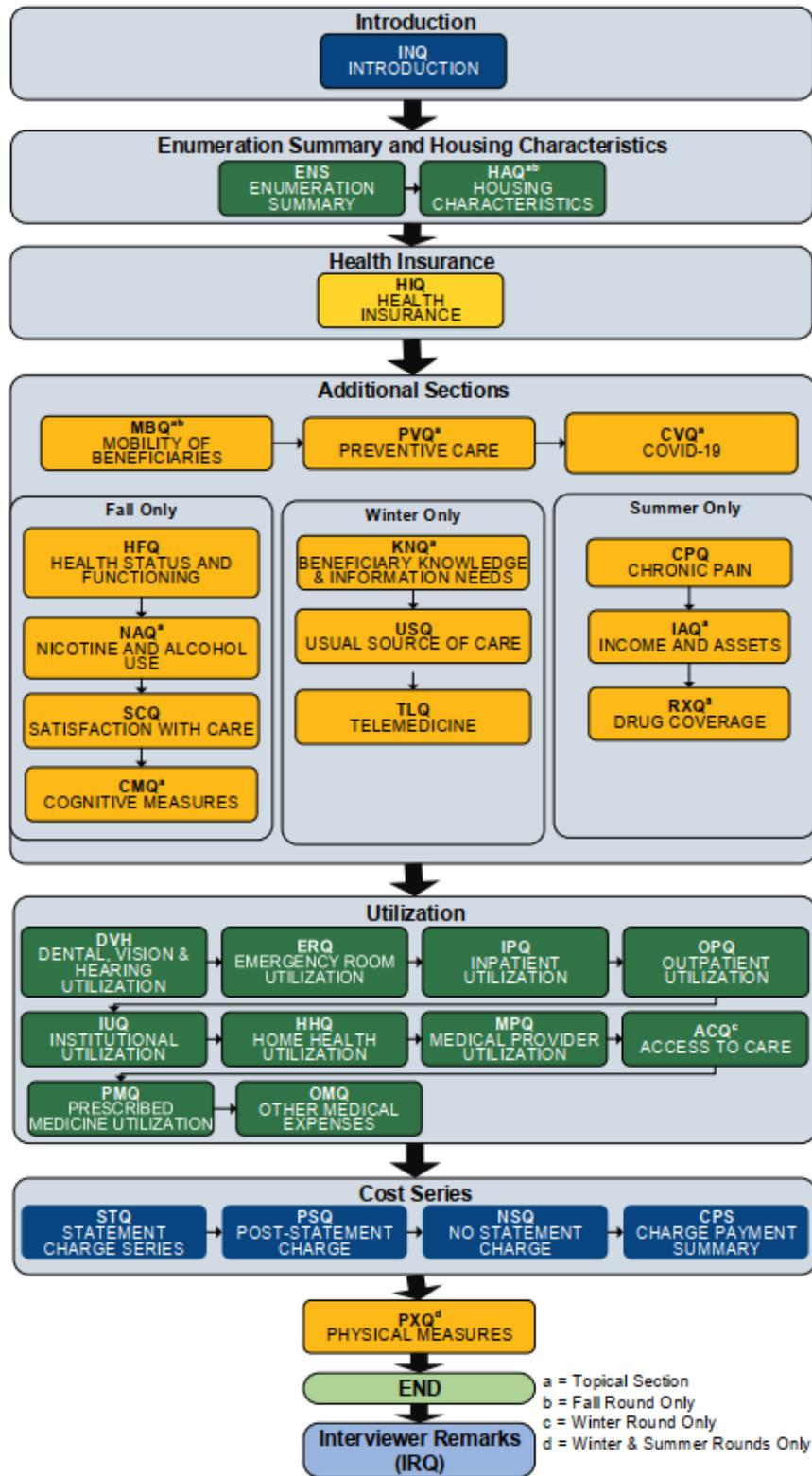
4.2 Facility Instrument Content

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.

For more information about Facility Instrument content, including descriptions of each questionnaire section, see the *2022 MCBS Data User's Guide: Survey File*.

4.2.1 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

INTTYPE	Description	Season
CFR	Continuing Facility Resident. Beneficiary for whom the previous round interview was in a facility and who currently lives at the same facility.	Any
CFC	Community-Facility-Crossover. Beneficiary who was interviewed in the community previously and has now moved to a long-term care facility.	Any
FFC	Facility-Facility-Crossover. Beneficiary for whom an interview 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 an interview in a facility 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 lives in a facility.	Fall

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

4.2.2 Facility Screener

When an interviewer learns that a beneficiary who was previously living in the community has moved into a facility, or a beneficiary who was living 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 component.

As a first step in determining eligibility for the Facility component, 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.

4.2.3 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 component – instead, Facility administrators and staff answer questions on behalf of the beneficiary.

Just like the Community Questionnaire, the sections administered in the Facility Instrument vary by interview type and data collection season (fall, winter, or summer). Similar to the Community Questionnaire, the Facility Instrument flows reflect the data collection year, rather than the data year, meaning interviews conducted in 2022 used the flows depicted. 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: 2022 MCBS Facility Instrument Flow for Baseline Interview
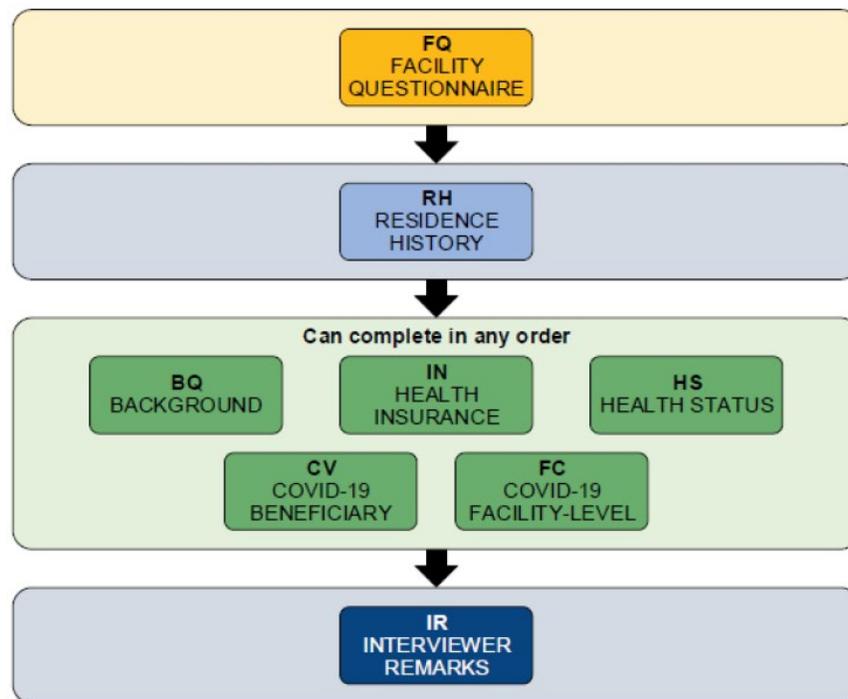
Exhibit 4.2.3 shows the flow for the Continuing and crossover interview types. Because the Facility Instrument is administered to Facility staff and not directly to the beneficiary, the Facility Instrument is designed to have a modular, flexible flow. The interviewer first completes the Facility Questionnaire (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.

Due to the redesign of the MCBS Facility Instrument in Fall 2019, the instrument flow varies for Medicare and/or Medicaid-certified facilities and facilities not certified by Medicare and/or Medicaid. Facilities that report a CMS Certification Number (CCN) and are therefore certified by Medicare and/or Medicaid receive a shortened MCBS Facility Instrument, as the FQ and HS sections skip variables redundant with Minimum Data Set (MDS)

and Certification and Survey Provider Enhanced Reports (CASPER) administrative data. Variables skipped during interview administration are instead populated using MDS and CASPER administrative data sources during data processing; this is described in detail in section 7.1.2. Facilities that do not report a CCN receive the full MCBS Facility Instrument.

Additionally, starting in Fall 2021, COVID-19 items were fielded within the MCBS Facility Instrument in the COVID-19 Beneficiary (CV) and COVID-19 Facility-Level (FC) as Topical sections. More information on these items can be found in the *2022 MCBS Data User's Guide: Survey File*.

Exhibit 4.2.3: 2022 MCBS Facility Instrument Flow for Continuing and Crossover Interviews
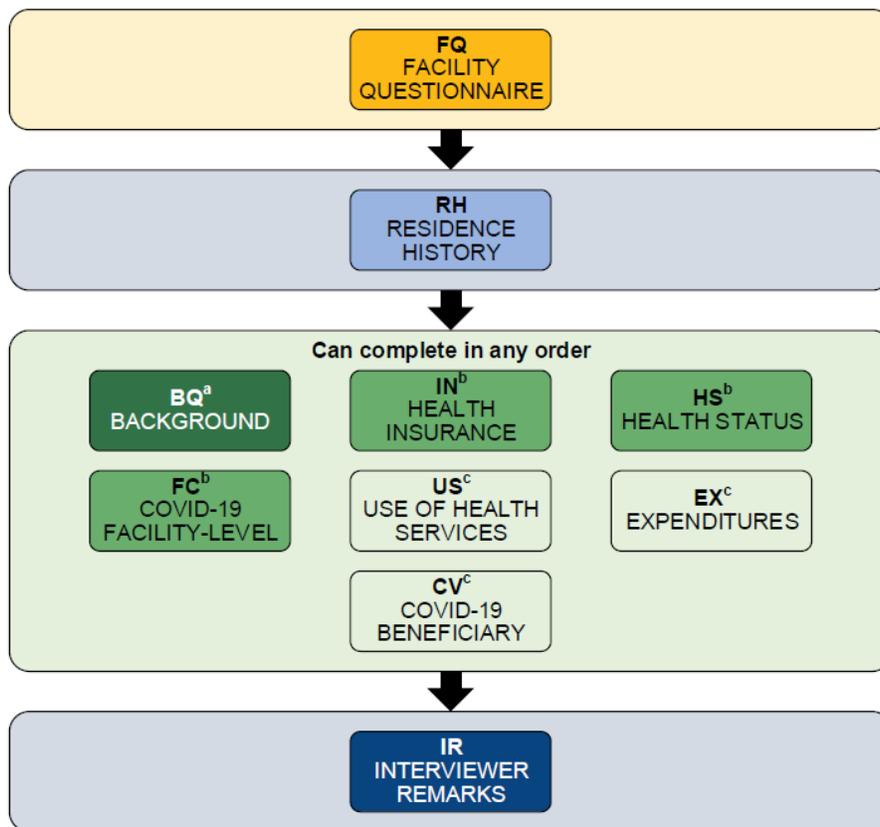
a = Administered only for Community to Facility interviews

b = Administered to all sample types in Fall round. Otherwise, administered only for Community to Facility, Facility to Facility, and for beneficiaries living in a facility whose last interview was a Community interview and who completed a Facility interview in a prior round.

c = Administered for all Facility interviews

4.3 CAPI and Case Management System Programming and Testing

CMS contracts with NORC at the University of Chicago (NORC) to administer the MCBS. A national team of specially trained and certified NORC field interviewers conduct Community interviews with MCBS beneficiaries or their designated proxies or they conduct Facility interviews with Facility staff on behalf of beneficiaries. MCBS interviewers receive project laptops with 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. The CAPI program automatically guides the field interviewer through the questions, records the answers, and contains logic and skip flows that increase the output of timely, clear, and high-quality data. The CAPI also contains follow-up questions where data were missing from the previous interview. When the interview is completed, the CAPI system allows the field interviewer to transmit the data electronically to the NORC central office in a secure manner. This section describes the CAPI and case management systems.

4.3.1 Community Questionnaire

The MCBS Community Questionnaire used in 2022 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 (MA) 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, on-screen interviewer help text is available to assist interviewers with definitions and additional instruction.

4.3.2 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 Instrument includes built-in data quality checks such as range and logic checks, dynamic text fills, and respondent exit and re-entry. The Facility Instrument also features a facility stay history timeline and a lookup for the facility's CCN.

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.[25]

4.3.3 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

---

[25] Not all interviewers may complete Facility interviews – additional training and certification is required beyond the standard Community interview training.

Facility Instrument, and the Facility Screener. 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. The ACP automates transfers of all cases from the Community component to the Facility component and between Facilities. The ACP also automates transfers for most cases from the Facility component to the Community component. Occasionally, transfers of cases from the Facility component to the Community component are completed manually.

4.4 Letters and Other Respondent Materials

A series of materials and other resources provide respondents with information about the MCBS and request their cooperation and participation in the survey. Medicare beneficiaries selected to participate in the MCBS receive letters in the mail, introducing the study and explaining that an interviewer from NORC will contact them to schedule an appointment. Respondents may also receive additional materials from interviewers. In addition, an 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 legitimacy of the survey to entities such as state resources for senior citizens. Materials are tailored to whether respondents live 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.

Interviewers or managers may use various materials provided at their discretion to assist in gaining cooperation.

5. INTERVIEWER RECRUITMENT AND TRAINING

5.1 Interviewer Recruitment and Staffing

A professional interviewer staff is required to complete interviews throughout the year. In 2022, most MCBS interviewers were experienced, having conducted interviews for at least a year or more. New interviewers were recruited to conduct Baseline interviews in Fall 2022, with a subset of these interviewers replacing staff who left the project; annual hiring is targeted based on 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.

5.2 Interviewer Training Programs for 2022

Nationally, the MCBS employs an average of approximately 170[26] field interviewers, who participate in a combination of targeted training initiatives and careful coaching and monitoring activities throughout data collection.

The 2022 MCBS Training Program consisted of remote and in-person trainings which varied based on the level of experience of the interviewer (new to MCBS or MCBS-experienced), the type of interview component (Community or Facility), the sample type (Incoming Panel or Continuing), and season-specific requirements (new or changing questionnaire sections 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. Following the same model used in previous years, new staff were on onboarded remotely in Fall 2022. NORC continued to leverage remote technology to ensure adequate training for new staff, including video conferencing software such as Zoom to hold roundtable discussions with experienced interviewers and field managers. This included question-and-answer sessions, gaining cooperation role playing, protocol demonstrations, and screen-sharing to facilitate real-time feedback as the trainer or interviewer navigated the case management system or questionnaire. NORC also incorporated e-learning technology to develop high-quality, responsive content grounded in adult-learning theory. For example, NORC leveraged the Articulate suite of e-learning software to program and deliver highly interactive trainings with software simulation activities. Later in the Fall round, interviewers were trained in-person on the Continuing interview. This 4.5-day training focused on the essential skills and protocols in the Continuing interview that require in-person instruction, such as correctly prompting for health events and purchases, organizing and abstracting from health and insurance documentation, and balancing complex caseloads.

In addition to the regular round-based trainings, select experienced field interviewers were trained to administer the Physical Measures Questionnaire (PXQ section). In this section, interviewers collect six physical measures from the respondent: height, weight, balance, timed walk, chair stands, and grip strength. In Winter 2022, a small cohort of interviewers were trained on the PXQ section remotely, with a combination of self-directed modules, video conference calls, and practice administering the measures. Live practice sessions were held over video conference while trainers observed the interviewers administering each measure with a partner to provide feedback and ensure proper protocol was followed. In Summer 2022, two additional small cohorts of interviewers were trained on the PXQ section: one remotely and one in-person. The remote cohort

---

[26] The fall round starts with a target of 200 field interviewers which, over the course of the year, is reduced due to staff turnover. A new cohort of interviewers is also hired for the MCBS each fall.

followed the same structure and agenda that was used in the Winter 2022 PXQ training. The in-person training consisted of 1.5 days of lecture and practice administering these measures.

In addition to formal trainings, throughout data collection the MCBS Training Program emphasized proper protocols through continuous quality improvement, featuring skill specialization, reinforcement of key behavior, and targeted messaging to boost interviewer performance. To meet all interviewers' skill-building and training needs, NORC continued to work with field managers to ensure interviewers received additional training via weekly field memos, interviewer group call sessions, and interviewer observations, referred to as "ride-alongs" or "call-alongs." These methods covered important data collection tips, provided answers to interviewer questions, and offered reminders about how to handle complex scenarios.

Field interviewer training stresses the importance of maintaining privacy, and project protocols are documented within the field interviewer manual. Field outreach and contacting procedures also maintain and ensure confidentiality. These procedures include the utilization of standard computer security protocol (dual authentication password protection for each interviewer laptop) and restrictions on submitting personally identifiable information (PII) through electronic mail. All MCBS survey staff directly involved in data collection and/or analysis activities are required to sign a Non-Disclosure Agreement and a confidentiality agreement.

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. 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.

As stated in the MCBS materials submitted for OMB approval, the information collected for MCBS is protected by NORC and by CMS. Respondent data are used only for research and statistical purposes. As required under the Privacy Act of 1974, identifiable information is not disclosed or released without the consent of the individual or the establishment, except to persons involved in research (Public Law 93-579).

6.1 Clearance

6.1.1 OMB Approval

CMS maintains a current OMB clearance for the MCBS. This typically requires annual revisions to the OMB clearance package to obtain approval for changes to the questionnaires or respondent materials.

For the 2022 MCBS, CMS received OMB approval on February 4, 2021 (OMB control number 0938-0568, expiration date 2/29/2024). There were no updates in this revision that impacted the 2022 MCBS questionnaires. CMS received approval for the 2023 MCBS on 8/26/2022 (OMB control number 0938-0568, expiration date 8/31/2025). This revision included updating the Beneficiary Knowledge and Decision-Making Questionnaire (KNQ) to include two items about use of the Medicare Savings Program (MSP) and one item about the beneficiary's right to file a complaint or appeal with Medicare starting in Winter 2023.

6.1.2 IRB Approval

The NORC 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 MCBS data collection, the research protocol and consent procedures were first approved by NORC's IRB in July 2014, with subsequent changes to the protocol approved through amendments and annual renewal.

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 11 rounds over a four-year period.

6.2.1 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 for the Incoming Panel. The first interview conducted for an Incoming Panel beneficiary is somewhat shorter as it does not collect health care utilization or cost data. Continuing interviews are longer as field interviewers collect information about the beneficiary's health care utilization and associated costs.

In 2022, Winter 2022 (Round 92) data collection started January 10, 2022 and concluded April 24, 2022; Summer 2022 (Round 93) data collection started May 9, 2022 and concluded July 31, 2022. Fall 2022 (Round 94) data collection started July 18, 2022 and concluded December 31, 2022.

6.2.2 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 Section 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 11 rounds of data collection to minimize nonresponse bias and the impact of sample attrition over time. 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 the respondent (i.e., the beneficiary, a proxy, or a staff member located at a facility where the beneficiary lives known as the Facility respondent) 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. These contacts may be made by phone or in-person. Phone contacting is more efficient but requires interviewers verify the validity of phone-matches obtained prior and during data collection to first reach the beneficiary before gaining cooperation could begin.

Following CMS guidance, interviewers use contacting strategies that promote efficiency and ensure continuity in contacts across all respondents actively fielded during a given round. The contacting effort required often corresponds to the number of rounds a respondent 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 3rd through 11th interviews typically require a reduced number of attempts necessary to contact respondents and schedule appointments.

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 Community Continuing 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.

6.2.3 Beneficiary Eligibility for MCBS Survey

Eligibility to participate in the survey depends on several 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 2nd and 11th interview is to attempt an interview with a proxy 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 2022, any Incoming Panel beneficiary who enrolled in Medicare prior to 2022) 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 2022, any Incoming Panel beneficiary who became eligible for Medicare in 2022) 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 to collect utilization and costs associated with the period 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. If an Incoming Panel beneficiary permanently moved or relocated more than 30 miles outside of MCBS sampled PSU boundaries before their first interview is completed, the case is finalized as Moved out of Area and not fielded in future rounds. Due to the advent of phone interviewing, we are able to continue to interview beneficiaries after a move, as long as they still reside within the United States.

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. Cases assigned status such as final refusal or final unlocatable are no longer fielded in future rounds.

Holdover Rules for Participation. 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 next round, 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.

6.2.4 MCBS Data Collection Protocols

A primary objective of the MCBS is to collect complete information about medical care, services, and costs for each beneficiary living in a community or a facility setting across all eleven 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 responds on behalf of the beneficiary to complete the interview.

Community Questionnaire Administration. 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 discuss past medical visits with respondents, as well as statements 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 work with respondents to match these documents into charge bundles to streamline entry within the questionnaire whenever possible (see the *2021 Data User's Guide: Survey File* for more information on how these statements are used during the cost series).

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 components 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 component is conducted when the beneficiary lives in a long-term care or other residential facility with three or more long-term 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, 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 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 component.

As a first step in determining eligibility for the Facility component, 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 component. 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.

Due to the emergence of the COVID-19 pandemic, starting in March 2020 the Facility interview was no longer administered in-person and instead administered by phone for the remainder of 2021. Beginning in 2022, Facility interviews could be administered in-person again, however, the majority of Facility interviews were administered by phone in 2022. Much of the content of the Facility component can be found in medical documentation. Therefore, Facility staff may refer to records, such as the beneficiary's medical chart, during the interview. In past years, Facility staff may have allowed the interviewer to abstract responses directly from medical records, but due to phone administration, abstraction was no longer conducted by the interviewer. While administering the Facility interview by phone, interviewers use their knowledge of the instrument and medical charts to help guide the Facility respondent to the appropriate records needed for the Facility interview.

6.2.5 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.[27] 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.

---

[27] Crossovers do not include respondents that have moved but remained within the Community setting.

Survey administration of Incoming Panel cases in Winter and Summer 2022 followed a different protocol that depended on when the beneficiary entered the new component and when s/he gained Medicare entitlement. All other crossover cases in their 3rd-11th interviews follow the crossover procedures outlined below.

Community-to-Facility Crossover. When a contact attempt with a Community Continuing 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 interview allows both the Community and Facility components 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. As a result of an automated crossover process for staging a Community interview, the Community interview in CAPI is then made available within the same round and can be administered then or in the following round. For cases in which a manual crossover is performed, the Community interview will be made available in CAPI for administration 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 new facility. As the result of an automated crossover process for staging a new Facility interview, the Facility interview for the new facility is then made available within the same round and can be administered then or in the following round.

6.2.6 Proxy Interviews and Assistants

Beneficiaries often require assistance in providing the detailed information needed to accurately respond to survey items. During 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 component 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. Interviews completed in 2022 had proxy usage ranging from 12-13 percent.

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 if 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. Interviews completed in 2022 largely by phone had assistant usage ranging from 9-10 percent.

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 components conducted with proxies follow a slightly different path than those administered directly to the beneficiary (see Section 4.1 for the Community Questionnaire flow and the *Data User's Guide: Survey File* for a description of the INQ).

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 whom an appropriate person would be to serve as a proxy or assistant.

6.2.7 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 components were conducted in Spanish in 2022.

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.

6.2.8 Questionnaire Breakoffs

Interviewers can suspend the interview prior to completion while administering both the Community and Facility components. 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 Section 8 for more information on weighting and imputation procedures).

6.3 Data Collection Results

An interview is 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 2022, the change in interview mode required in Winter 2022 impacted the length of interviews and removed the ability to compare to prior rounds.

Exhibit 6.3.1 provides the count of completed interviews by round and component for 2022. Exhibit 6.3.2 provides the ratio of completed interviews by interview mode. Detailed information on response rates can be found in Section 9.

Exhibit 6.3.1: 2022 Completed Interviews by Component

Round	Component	Completed Interviews	Mean Interview Duration (minutes)
Winter 2022	Community	10,904	65.9
	Facility	839	27.7
Summer 2022	Community	7,657	65.6
	Facility	634	30.6
Fall 2022	Community	12,882	75.2
	Facility	554	44.8

Exhibit 6.3.2: 2022 Completed Interviews by Interview Mode

Round	Component	Phone Interviews	In-Person Interviews
Winter 2022	Community	92%	8%
	Facility	99%	<1%
Summer 2022	Community	89%	11%
	Facility	99%	<1%
Fall 2022	Community	77%	23%
	Facility	94%	6%

6.4 Item Non-Response

As in any other survey, some respondents could not, or would not, supply answers to some questions.[28] Item non-response rates are generally low in the MCBS data, but the researcher still needs to be aware of the missing data and be cautious about patterns of non-response.[29] Some of the missing data are attributable to the fact that some of the Community interviews and all of the Facility interviews are conducted through a proxy respondent. In other words, the respondent may not have had knowledge of the information sought on the sample person. In other situations, the respondent may have simply refused to answer.

6.5 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 observations of interviews are used to 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 is recorded

---

[28] This is different from when an individual refuses to participate in the survey altogether, which is called unit non-response. Unit non-response is discussed in detail in Section 9.

[29] In the LDS files, item non-response types are indicated by missing type codes in SAS®, including refusal to answer, don't know the answer, and invalid skip. The code .D represents a "don't know" response, the code .R represents a "refused" response, and the code .N represents an "invalid skip" response.

with respondent consent. By listening to 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 2022, 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 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 to collect additional information for coaching purposes.

Additional interviewer review, such as additional CARI review and thank you calls made to respondents, and on-going large-scale analyses monitor the potential impact phone-only interviewing may have on data quality. Investigations of 2020 and 2021 data found stability in the majority of MCBS questionnaire measures after the initial mode transition but revealed challenges collecting information in the cost sections over the phone, particularly details from physical documentation (e.g., insurance statements). These impacts have been mitigated through tailored interviewer training and respondent coaching, the ability to conduct all or part of the interview in-person if required, as well as statistical techniques including imputation, claims matching, and ratio adjustments (see sections 8.3.3 and 8.3.4), though mode effects cannot be completely eliminated in any multimode data collection effort.

These ongoing analyses suggest interview mode impacts the quality of data collected, but overall, these shifts are relatively small and do not impact sample representativeness enough to be of concern.

7. DATA PROCESSING AND DATA DELIVERY

Longitudinal data collection requires both interim and final post-processing of the data to prepare data for release. These activities include data editing for preloading subsequent round instruments and final file production, data concatenation and reconciliation for the annual LDS files, and the development of other post-processing inputs to the files. This section describes both the data editing process and the annual data concatenation and reconciliation process.

7.1 Data Processing Overview

During the interview, respondents may provide information that is either incomplete or inconsistent with previously reported survey responses or 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 Questionnaire and Facility Instrument preloads, the 2022 Survey File, and the 2022 Cost Supplement File.

The same types of data review and processing protocols are used for each effort, with different source instruments and editing protocols. These data review and processing procedures are described in Section 7.1.1. For Facility data only, additional processing steps (described in Section 7.1.2) are applied to populate certain variables using administrative data. Section 7.1.3 outlines data processing decisions and impacts specific to 2022.

7.1.1 Process Description

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

Exhibit 7.1.1: Data Review and Editing Process
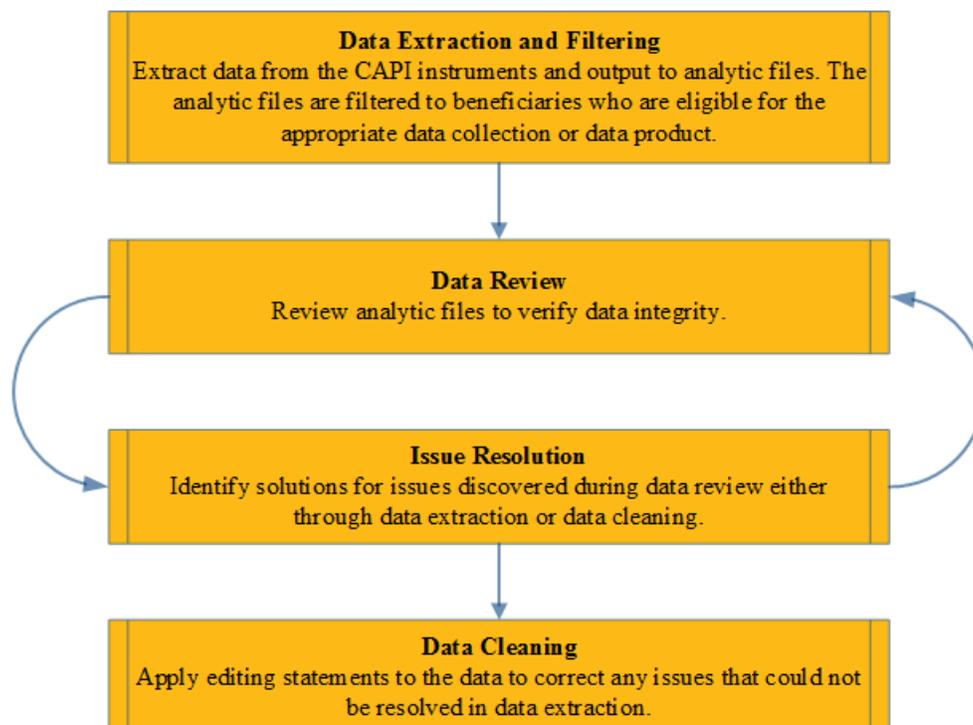
Data Extraction and Filtering. At the conclusion of data collection in each round, data are extracted from the raw Community and Facility CAPI instruments and transformed into SAS® analytic files for further processing. This extraction includes the development of appropriate questionnaire metadata. Prior 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 confirm that all necessary variables are on the file, verify variable attributes, and identify high rates of missingness or out of range values. Row checks confirm the inclusion of expected BASEIDs and check 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 are implemented 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 requires particular attention to the CAPI routing routines 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 via checks for values that are not explicitly disallowed by the questionnaire. For example, in the Community Questionnaire, respondents should not report female-only conditions like cervical cancer for male beneficiaries. Continuous variables are reviewed to identify illogical extreme values. For instance, in the Community Questionnaire, the number of living children reported for 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.

In the MCBS Community Questionnaire data, open-ended responses for other specify variables are reviewed and backcoded into existing codeframes when possible. Prescription medicine data undergo tailored cleaning in a two-step process. First, for the medicines interviewers entered into the questionnaire using the Prescription Medicine Lookup (PMLU) tool, details including prescription medicine name, strength, brand name, generic name, form, and form code are confirmed against values from the First Databank (FDB) list of prescribed medicines and updated as needed. Second, for medicines that interviewers were not able to find in the PMLU tool and entered manually with verbatim fields, a number of cleaning steps are applied to fix common misspellings and typos and standardize spacing, punctuation, abbreviations, and other formatting. These steps simplify the subsequent CMS process of matching the Prescription Medicine (PMED) file list to the FDB list of prescription medicines and administrative claims information.

Many items (referred to as "ever variables") in the MCBS ask respondents whether they have ever had certain experiences, such as ever been told they have a chronic condition, received a treatment, or done a specific activity (such as ever accessed the official Medicare website). Example questions include "Have you ever been diagnosed with diabetes?" and "Have you smoked at least 100 cigarettes in your entire life?" Their responses are coded affirmatively if the respondent has ever reported "yes" to having had that condition or experience. Once respondents report "yes" to these items, the item is not asked again. In the MCBS Community Questionnaire, "yes" responses from prior rounds are pulled forward in data processing to the current round of data. In this manner, the current round of data contains a cross-sectional snapshot of beneficiaries who have ever had certain experiences. See Section 10.1.5 in the *2022 Data User's Guide: Survey File* for further information on these items.

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

include setting inconsistent or improbable values to missing (e.g., a female-only health condition is reported for a male beneficiary) or correcting questionnaire skip logic (e.g., if a respondent erroneously skipped a question, the field is set to a special value indicating invalid skip).

7.1.2 Additional Facility Data Processing

As noted in Section 4.2.3, a subset of items in the FQ and HS sections of the Facility Instrument are skipped during data collection if a CCN is reported. This routing leads to valid missingness for the skipped variables in the analytic files storing data collected in those sections. These variables are populated during an administrative data matching step using two data sources: the CASPER and the MDS. All variables that are populated during data processing using administrative data are indicated in the Codebooks.

CASPER Data Source Details. CASPER data are obtained from a vendor on an annual basis in the form of a cumulative file including records for facilities with both active and terminated CMS certifications. The CASPER database comprises seven unique data sets, two of which are utilized in MCBS data collection and data processing.

CASPER Part 8 is a provider-level file that contains information such as the facility name, address, ownership, and accreditation and is used to populate the CCN questionnaire lookup tool used during data collection. CASPER Part 2 is a certification-level file that contains information such as bed counts, resident censuses, and services provided by the facility. These data are used to populate skipped questionnaire variables during data processing using the most recent survey certification for each facility.

MDS Data Source Details. MDS data are retrieved via the CMS Chronic Conditions Data Warehouse (CCW) and are available on a rolling basis approximately two months after data entry by facilities. MDS data are obtained at least two months after the conclusion of each round of data collection to ensure alignment between the date ranges present in the MCBS survey data and the MDS administrative data.

The MDS data contain one record per MDS assessment conducted. Prior to data processing, the MDS data are restricted to Nursing Home Comprehensive and Nursing Home Quarterly assessments.

FQ-CASPER Data Linkage Process. For beneficiaries living in a facility for which the Facility respondent reported a CCN, the analytic file storing data collected in the FQ section can be linked directly to CASPER data on CCN. The CCN reported in the Facility Instrument is matched to the corresponding record on the CASPER Part 8 file with the same CCN.

HS-MDS Data Linkage Process. For beneficiaries living in a facility for which the Facility respondent reported a CCN, the Facility respondent is asked during the interview to identify the date and type (comprehensive or quarterly) of any MDS assessments conducted for the sampled beneficiary on or around the survey reference date provided in the HS section.[30] During data processing, the analytic file storing data collected in the HS section is linked to MDS data via a match protocol that identifies the "best" administrative data record, that is, the record most likely to be the MDS assessment reported by the Facility respondent during the interview.

Prior to matching, the MDS data source is first restricted to records matching the BASEID and reported Facility CCN for each beneficiary. If the BASEID and CCN combination cannot be found in the administrative data, the data cannot be linked. All other records are run through the match protocol displayed in Exhibit 7.1.2.

---

[30] The HS reference date varies by data collection round and beneficiary circumstances, but in fall rounds, the reference date is set to September 1st of the current year for most beneficiaries.

Exhibit 7.1.2: HS-MDS Data Linkage Process
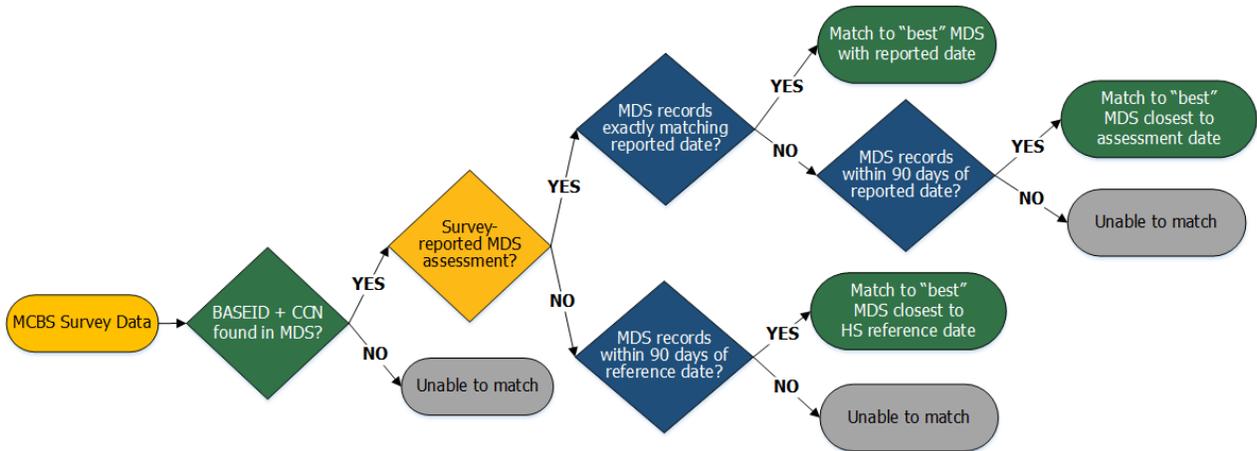
The match algorithm prioritizes date matches, searching for MDS administrative data with the same target date (TRGT_DT) as the survey-reported MDS assessment date.[31] When there is no target date matching to the survey-reported assessment date, MDS administrative records with a target date within ±90 days of the assessment date are considered for matching.

If no MDS assessment is reported by the Facility respondent in the interview, MDS administrative records with a target date within ±90 days of the HS survey reference date are considered for matching since a quarterly assessment is required to be conducted every 90 days.

When multiple MDS administrative records corresponding to the same date are present, the "best" MDS record is chosen by comparing the survey-reported MDS assessment type (Comprehensive or Quarterly) to the MDS administrative assessment type. When multiple MDS administrative records within the ±90 date range are present, the "best" MDS record is chosen by first looking for the target date closest to the assessment or reference date and then comparing the survey-reported MDS assessment type (Comprehensive or Quarterly) to the MDS administrative assessment type.

Match Outcomes. In 2022, 100 percent of FQ records where a CCN was reported were successfully matched to CASPER directly on CCN. As shown in Exhibit 7.1.3, 97 percent (i.e., 497/511) of HS records were successfully matched to MDS via the match algorithm.

Exhibit 7.1.3: 2022 HS-MDS Match Outcomes

Match Type	Record Count
MDS record identified via an exact date match between the survey-reported assessment date and MDS assessment date	432
MDS record identified via a non-exact date match between the survey-reported assessment date and MDS assessment date	48
MDS record identified via the survey reference date	17
No match found	14

---

[31] Centers for Medicare & Medicaid Services. "Target Date (Date of Assessment) (MDS)." ResDAC. Accessed February 26, 2021. https://www.resdac.org/cms-data/variables/target-date-date-assessment-mds.

7.1.3 Data Year Notes

This section outlines data collection changes, questionnaire changes, or data issues that have had an impact in data processing or methodology in the given data year.

"Escape Hatch" Functionality. In response to the shift to phone interviewing during the COVID-19 pandemic, an "escape hatch" option to skip the cost series to reduce respondent burden in the interview was added to the questionnaire beginning in Fall 2020 (Round 88). This option allowed field interviewers to skip out of cost series questionnaire sections if they believed pressing the respondent for this information would cause them to end the interview. This change resulted in unresolved costs and events that would have otherwise been resolved within the round and some potential missing details on costs and events that were routed around. The "escape hatch" functionality impacts both the collection of cost information for new events as well as the follow-up on costs reported for existing events in prior rounds, potentially resulting in imputation of the missing cost information if it is not collected in a subsequent round. Due to the return to some in-person data collection and decreased use of the escape hatch by field interviewers via the phone, this functionality was phased out in Summer 2023 (Round 96).

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 both prevent asking MCBS respondents the same questions in subsequent rounds and 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 their 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 include 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 is 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 Instrument and Community Questionnaire reference dates, and the reason a cost is sent through Charge Payment Summary (CPS reason). For more details on the CPS section of the Community Questionnaire, see the *2022 MCBS Data User's Guide: Survey File*.

7.2.1 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 is the rollover process. After data review and editing, 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. Beginning in Fall 2020 (Round 88), costs linked to events that are ineligible for inclusion in the current year's Event Cost Consolidation process (based on their round or date) are not assigned a CPS reason and further cost information is not collected.

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

Thorough QC 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.

7.2.2 Which Community Data Are Included in the MCBS LDS?

Community data that are incorporated into the Survey File LDS and Cost Supplement LDS for sampled beneficiaries eligible for Medicare in a given benefit year depend on a variety of factors, including beneficiary panel type, the four annual panels of sampled beneficiaries, multiple rounds of data collection, and different types of questionnaire items. The data that are collected in each round depend on the type of panel and the reference periods used by the questionnaires in the interview. Although one data year consists of one calendar year, data included in the LDS are collected over three years. Some data collected in the previous year are pulled forward to fill in data for the current data year. This happens when questionnaire items are administered only once (such as demographics) or when data are missing for the data year, but valid values exist for the previous year. However, most data are collected during and after the data year.

When information for the data year is collected in the following year, it is generally because the reference period for the questionnaire extends back into the data year, and the items are asked of the Medicare population enrolled and eligible in the data year. In the example below (Exhibit 7.2.1), the data year is 2022. There are four panels involved in data collection for 2022: one Incoming Panel (selected in 2022), two Continuing Panels (selected in 2020 and 2021), and one exit panel (selected in 2019). The rounds of data collection that fall within the data year are the Winter, Summer, and Fall rounds of 2022, with additional data for 2022 collected in the Winter and Summer rounds of 2023.

The Survey File LDS consists of questionnaire items collected as part of the Community Core questionnaire sections as well as items collected in the Community Topical questionnaire sections. The Core Survey File data for 2022 were collected in Summer and Fall 2022 and in Winter 2023. The Topical Survey File data were collected in Fall 2022 and Winter and Summer 2023. Each round's interview is based on reference periods, which extend from the time of the previous interview. For example, the Core Survey File HIQ asks about changes to insurance plans during the reference period. In the Fall 2022 interview, this period would cover the time since completion of the Summer 2022 interview, while in Winter 2023 it would cover the time since completion of the Fall 2022 interview, meaning the reference period extends back into 2022. A reference period may also cover the entire data year. For example, the Income and Assets Questionnaire (IAQ) is a Topical questionnaire section collected in the Summer 2023, but it collected beneficiaries' financial information for the 2022 data year. Beneficiaries in the Incoming Panel provided 2022 Survey File LDS data through participation in their first and second interviews in Fall 2022 and Winter 2023 and provided additional Topical data in Summer 2023. Members of the Continuing Panels have some of their data pulled forward from rounds collected in 2021 but provided most of their data through participation in the Summer and Fall of 2022 and Winter and Summer of 2023. Members of the exit panel likewise have data pulled forward from 2021 and were interviewed in Summer and Fall 2022 and Winter 2023 but were not interviewed in Summer 2023.

The Cost Supplement LDS consists of utilization and cost data for the 2022 data year. These data are collected from the four rounds that can have reference periods covering any part of the data year; for the 2022 data year, this includes the rounds from Winter 2022 through Winter 2023. Each interview's reference period covers the time between completion of the previous round and the current round. In the case that a beneficiary skips a round, the reference period for the following round covers the missing period by extending back to the date of the most recently completed interview. The Incoming Panel does not provide Cost Supplement data until their second interview in the winter following the data year. For beneficiaries who are current-year enrollees, meaning beneficiaries who enrolled in Medicare in 2022, the winter round reference period extends back to the date of completion of the Fall 2022 round, collecting utilization and cost data for the latter part of 2022. Members of the Incoming Panel who enrolled prior to 2022 have a Winter 2023 reference period that began on 1/1/2023 and will only provide Cost Supplement data for 2023 and later. The Cost Supplement data for the Continuing and exit panels were collected through participation in the rounds from Winter 2022 through Winter 2023. The exit panel exited the survey in Winter 2023 and had a reference period that ended at 12/31/2022.

Exhibit 7.2.1: Rounds of Data Collection and Reference Periods for Community Data Included in the 2022 Limited Data Sets*

Incoming Panel

Calendar Year 1 (CY1) - 2022												Calendar Year 2 (CY2) - 2023							
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Winter			Summer				Fall					Winter				Summer			
							Topical					Topical				Topical			
							Survey File					Survey File				Survey File			
												Cost Supplement 2022 Incoming Panel Current-Year Enrollees: (data only collected through 12/31/CY1) 2022 Incoming Panel (data collected only from 1/1/CY2)				Cost Supplement			

Continuing Panel

CY1 - 2022												CY2 – 2023							
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Winter			Summer				Fall					Winter				Summer			
							Topical					Topical				Topical			
Survey File			Survey File				Survey File					Survey File				Survey File			
Cost Supplement			Cost Supplement				Cost Supplement					Cost Supplement (data through 12/31/CY1)				Cost Supplement			

Continuing Panel - Exit Year

CY1 - 2022												CY2 – 2023							
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Winter			Summer				Fall					Winter				Summer			
							Topical					Topical							
Survey File			Survey File				Survey File					Survey File							
Cost Supplement			Cost Supplement				Cost Supplement					Cost Supplement (data through 12/31/CY1)							

*Note: Data in this table were collected in calendar years 2022 and/or 2023 and included in the 2022 LDS release. Cost Supplement data reflect the data year of interest (i.e., 1/1/CY1 – 12/31/CY1). In other words, the data included in the 2022 Cost Supplement LDS are based on survey reported information within the year of interest, not rounds of data collection. In contrast, for the 2022 Survey File LDS, data were collected in Summer and Fall 2022 and Winter 2023. Data collected in Winter 2023 are included in the 2022 Survey File LDS if the survey items ask about experiences/coverage since the date of the beneficiary's last fall interview in 2022. For some Topical sections such as RXQ and the food insecurity items in the IAQ, the data are collected in Summer 2023 but included with the 2022 LDS's given the reference period is between 1/1/CY1 – 12/31/CY1.

7.2.3 Which Facility Data Are Included in the MCBS LDS?

Like Community, Facility data that are incorporated into the Survey File LDS and Cost Supplement LDS for sampled beneficiaries eligible for Medicare in a given benefit year depend on a variety of factors, including beneficiary panel type, the round of data collection, and type of questionnaire item. As with Community data, some Facility data collected in a previous year are pulled forward to fill in data for the current data year. This happens when questionnaire items are administered only once or when data are missing for the data year, but valid values exist from a previous year. However, most data are collected during and after the data year.

The Survey File LDS Facility data for 2022 were collected in Winter, Summer, Fall 2022, and Winter 2023 as part of the Facility Core questionnaire sections and Facility Topical questionnaire sections. The Facility Topical questionnaire sections capture data on COVID-19 topics at the facility- and beneficiary-level and are included as part of the FBENCVFL and FFACCVFL Survey File segments.

Like Community, the Cost Supplement LDS consists of Facility utilization and cost data for the 2022 data year. These data are collected from the four rounds that can have reference periods covering any part of the data year; for the 2022 data year, this includes the rounds from Winter 2022 through Winter 2023. Each interview's reference period covers the time between completion of the previous round and the current round. In the case that a beneficiary skips a round, the reference period for the following round covers the missing period by extending back to the date of the most recently completed interview.

The Incoming Panel is not included in the Facility Cost Supplement data. The Cost Supplement data for the Continuing Panel was collected through participation in the rounds from Winter 2022 through Winter 2023. The exit panel exited the survey in Winter 2023 and had a reference period that ended at 12/31/2022.

Exhibit 7.2.2: Rounds of Data Collection and Reference Periods for Facility Data Included in the 2022 Limited Data Sets*

Incoming Panel

Calendar Year 1 (CY1) - 2022												Calendar Year 2 (CY2) – 2023			
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Winter			Summer				Fall					Winter			
							Survey File					Survey File			

Continuing Panel

CY1 - 2022												CY2 - 2023			
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Winter			Summer				Fall					Winter			
Survey File			Survey File				Survey File					Survey File			
Cost Supplement			Cost Supplement				Cost Supplement					Cost Supplement (data through 12/31/CY1)			

Continuing Panel - Exit Year

CY1 - 2022												CY2 - 2023			
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Winter			Summer				Fall					Winter			
Survey File			Survey File				Survey File					Survey File			
Cost Supplement			Cost Supplement				Cost Supplement (data through 12/31 CY1)					Cost Supplement (data through 12/31/CY1)			

*Note: Data in table were collected in calendar year 2022 and calendar year 2023 and included in the LDS released in 2022.

7.3 MCBS 2022 Survey File

The 2022 Survey File release is built from 42 analytic data files encompassing Community and Facility data collection from five rounds of data including Winter 2022, Summer 2022, Fall 2022, Winter 2023, and Summer 2023. These files are input into CMS processes that generate the final data files available to the public. More detail about the 2022 Survey File LDS is provided in the *2022 Data User's Guide: Survey File*. This section describes the eligibility criteria for the analytic files, file preparation, and file contents.

7.3.1 File Eligibility Criteria

The Core Survey File data for 2022 were collected in Winter, Summer, and Fall 2022. The Topical questionnaire data were collected in Fall 2022 and Winter and Summer 2023. The inclusion criteria for these analytic files include beneficiaries continuously living in the community or facility, beneficiaries who move between a facility and the community, proxy respondents for deceased beneficiaries, or individuals who lost entitlement to Medicare. A beneficiary only needs to have completed a Community or Facility component in one of the data collection rounds of interest to be included in these analytic files. That is, if a beneficiary has a completed interview in any eligible round in any component (i.e., Community or Facility), then that beneficiary's data are included in the analytic files. However, specific files have additional criteria that a case needs to meet for inclusion. For example, some segments require that beneficiaries reside in a facility at the time of their fall interview in order to be included in the file.

7.3.2 File Contents

Community. There are two subcategories of Community analytic files included in the MCBS Survey File. First, the Community Continuing questionnaire section 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. Second, the Community Topical questionnaire section analytic files contain data collected in questionnaire sections that cover special interest issues. Topical data may be collected every round or on a seasonal basis. See the *Data User's Guide: Survey File* for a list of the Community Questionnaire sections included in each data file.

Community Continuing Questionnaire Sections. The Community Survey File data contain information about access to medical care, health status and functioning, health insurance plans, medical providers, and income and assets. The Survey File does not include survey-reported cost, health care utilization, or case management data.

Community Topical Questionnaire Sections. 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 2022 Survey File contains data from some sections that were administered in the Winter 2023 (Round 95) and Summer 2023 (Round 96) but have reference periods for 2022. These files are processed in combination with the 2022 Survey File deliveries and as a result, Topical analytic files are considered part of the MCBS 2022 Survey File.

Facility. The Facility analytic files contain Core questionnaire sections critical to the purpose of the MCBS and Topical questionnaire sections on COVID-19.

Facility Continuing Questionnaire Sections. Facility Survey File data contain information about access to medical care, health status and functioning, health insurance plans, facility characteristics, and beneficiary characteristics. See the *Data User's Guide: Survey File* for a list of Facility Instrument sections included in each data file. The Survey File does not include cost, health care utilization, or case management data.

7.3.3 Reference Period

Reference Period is a data editing process that uses case management data to define time periods in 2022 covered by Community and Facility survey data. Along with Insurance Timeline discussed below, it is an interim data product that is not part of the final Survey File or Cost Supplement File LDS's because it feeds into the final segments. Reference Period is run for all beneficiaries who had interviews in 2022 and includes all beneficiaries with a positive Survey File Ever Enrolled weight. Reference Period creates a calendar history of a beneficiary's MCBS interviews as it compares to the beneficiary's residence in the community and/or in the facility during the year. This calendar of residence and interview activity is used to create the Residence Timeline (RESTMLN) segment file and to determine in which files to include Community and Facility data for each beneficiary.

7.3.4 Insurance Timeline

Insurance Timeline is a production process that 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 2022 was produced for the same population as was assigned a Survey File ever enrolled weight. A combination of survey-collected data and administrative data are used to create the timeline of health insurance coverage for the period in which a beneficiary participated in the survey. For beneficiaries who leave the survey prior to completion of their full study tenure, the end date of their insurance coverage is recorded as the date of the last completed

interview. It should be noted that in all likelihood, their insurance coverage extends beyond this date, but no data are available to determine the actual coverage end date.

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 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, which are released as part of the Health Insurance Timeline (HITLINE) segment in the Survey File LDS.

In Winter 2023, several updates were made to improve the Health Insurance Questionnaire (HIQ). The purpose of these changes was threefold: align collection of health insurance information across different plan types, reduce respondent burden by discontinuing collection of detailed information with little analytic utility, and improve the quality of information collected. As a result, plans reported in Winter 2023 with coverage extending back to 2022 were reported under the new questionnaire structure. The redesigned structure also obtains more accurate start and end dates for health insurance plan coverage, which makes it easier to determine whether a plan should be included in the 2022 Survey File or the 2023 Survey File only. The full scope of these changes will be apparent in the 2023 Survey File and detailed in the 2023 LDS documentation.

7.4 MCBS 2022 Cost Supplement File

The Cost Supplement File data include information on beneficiaries' medical events occurring in 2022 and the cost of those events. The Cost Supplement File LDS contains cost and utilization data collected in Winter, Summer, and Fall 2022 about utilization and expenditures occurring in 2022. Cost and utilization collected in Winter 2023 are also included if the events fall within the 2022 reference period. More detail about the 2022 MCBS Cost Supplement File LDS is described in the *2022 Data User's Guide: Cost Supplement File*.

Substantial post-processing is applied to the questionnaire items related to health care events, the costs and payments associated with those events, and the source of payments. Four processes are used to create the inputs to the final data files. The four processes build annualized files, define eligibility for the Cost Supplement File, and create events that are linked to defined payers and the cost of the services provided. The first three 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 Questionnaire and adds the claims match into a single step. The facility stay process then generates data files to produce the Cost Supplement File release.

These processes (Event Cost Consolidation, Prescription Medicine file, and Facility Stay File) are described below. The shared goals 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 data products are considered interim inputs into the final Survey File or Cost Supplement File releases and are therefore not released on their own. Each interim analytic product is described below.

7.4.1 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 living 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 coverage 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 File.

7.4.2 Prescription Medicine File

The PMED file is a list of all prescription medicines that are collected by the MCBS. For 2022, the list included every combination of prescription medicine names, forms, and strengths provided by MCBS respondents during interviews conducted in 2022 (including a total of four rounds between Winter 2022 and Winter 2023). The process of creating the PMED file includes assembling a full list of all beneficiaries' reported prescribed medicines for 2022 from the Community Questionnaire and de-duplicating it to create a unique list of medicines.

The PMED file includes both medicines that were reported by MCBS respondents for the first time during one of these four rounds and refilled medications that were originally reported earlier but updated as being currently prescribed during one of these four 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.

7.4.3 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 living 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 and edit outliers and other anomalies. The Facility Stay File population in 2022 included any beneficiaries in the Continuing Panels completing one or more Facility Instrument interviews covering residence in an MCBS-eligible facility for one or more days in 2022.

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. Weights 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 are also calibrated 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

8.2.1 Overview

Weighting activities for the 2022 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 2022). 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 segments. A listing of all the weights for the MCBS is presented in Exhibit 8.2.1.

Exhibit 8.2.1: 2022 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	CEYRSWGT	CEYRS001-CEYRS100	Continuously enrolled from 1/1/2022 through the fall of 2022
Survey File	Ever Enrolled Cross-Sectional Weights	EVRWGTS	EEYRSWGT	EEYRS001-EEYRS100	Ever enrolled for at least one day at any time during 2022
Survey File	Continuously Enrolled Two-Year Longitudinal Weights	LNG2WGTS	L2YRSWGT	L2YRS001-L2YRS100	Continuously enrolled from 1/1/2021 through the fall of 2022

Limited Data Set	Description	Segment	Full-Sample Weight	Replicate Weights	Population
Survey File	Continuously Enrolled Three-Year Longitudinal Weights	LNG3WGTS	L3YRSWGT	L3YRS001-L3YRS100	Continuously enrolled from 1/1/2020 through the fall of 2022
Survey File	Continuously Enrolled Four-Year Longitudinal Weights	LNG4WGTS	L4YRSWGT	L4YRS001-L4YRS100	Continuously enrolled from 1/1/2019 through the fall of 2022
Cost Supplement File	Ever Enrolled Cross-Sectional Weights	CSEVRWGT	CSEVRWGT	CSEVR001-CSEVR100	Ever enrolled for at least one day at any time during 2022
Cost Supplement File	Two-Year Longitudinal Weights	CSL2WGTS	CSL2YWGT	CSL2Y001-CSL2Y100	Enrolled at any time during both 2021 and 2022
Cost Supplement File	Three-Year Longitudinal Weights	CSL3WGTS	CSL3YWGT	CSL3Y001-CSL3Y100	Enrolled at any time during each of 2020,2021, and 2022
Survey File Topical Section	KNQ Survey File Ever Enrolled	MCREPLNQ	KNSEWT	KNSE1-KNSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	KNQ Survey File Continuously Enrolled	MCREPLNQ	KNSCWT	KNSC1-KNSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	KNQ Cost Supplement Ever Enrolled	MCREPLNQ	KNCEWT	KNCE1-KNCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	ACQ Survey File Ever Enrolled	ACCSSMED	ACSEWT	ACSE1-ACSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	ACQ Survey File Continuously Enrolled	ACCSSMED	ACSCWT	ACSC1-ACSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	ACQ Cost Supplement Ever Enrolled	ACCSSMED	ACCEWT	ACSFCE1-ACSFCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	USQ Survey File Ever Enrolled	USCARE	USSEWT	USSE1-USSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	USQ Survey File Continuously Enrolled	USCARE	USSCWT	USSC1-USSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023

Limited Data Set	Description	Segment	Full-Sample Weight	Replicate Weights	Population
Survey File Topical Section	USQ Cost Supplement Ever Enrolled	USCARE	USCEWT	USCE1-USCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	TLQ Survey File Continuously Enrolled	TELEMED	TMSEWT	TMSE1-TMSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	TLQ Cost Supplement Ever Enrolled	TELEMED	TMSCWT	TMSC1-TMSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	TLQ Survey File Ever Enrolled	TELEMED	TMCEWT	TMCE1-TMCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Winter 2023
Survey File Topical Section	IAQ Survey File Ever Enrolled	INCASSET	INSEWT	INSE1-INSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	IAQ Survey File Continuously Enrolled	INCASSET	INSCWT	INSC1-INSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	IAQ Cost Supplement Ever Enrolled	INCASSET	INCEWT	INCE1-INCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	IAQ Survey File Ever Enrolled	FOODINS	FDSEWT	FDSE1-FDSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	IAQ Survey File Continuously Enrolled	FOODINS	FDSCWT	FDSC1-FDSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	IAQ Cost Supplement Ever Enrolled	FOODINS	FDCEWT	FDCE1-FDCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	PAQ Survey File Enrolled	PNTACT	PASEWT	PASE1-PASE100	Ever enrolled for at least one day at any time during 2022
Survey File Topical Section	PAQ Survey File Continuously Enrolled	PNTACT	PASCWT	PASC1-PASC100	Continuously enrolled from 1/1/2022 through the fall of 2022
Survey File Topical Section	PAQ Cost Supplement Ever Enrolled	PNTACT	PACEWT	PACE1-PACE100	Ever enrolled for at least one day at any time during 2022
Survey File Topical Section	RXQ Survey File Ever Enrolled	RXMED	RXSEWT	RXSE1-RXSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023

Limited Data Set	Description	Segment	Full-Sample Weight	Replicate Weights	Population
Survey File Topical Section	RXQ Survey File Continuously Enrolled	RXMED	RXSCWT	RXSC1-RXSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	RXQ Cost Supplement Ever Enrolled	RXMED	RXCEWT	RXCE1-RXCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CPQ Survey File Ever Enrolled	CHRNPAIN	CPSEWT	CPSE1-CPSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CPQ Survey File Continuously Enrolled	CHRNPAIN	CPSCWT	CPSC1-CPSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CPQ Cost Supplement Ever Enrolled	CHRNPAIN	CPCEWT	CPCE1-CPCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CVQ Survey File Ever Enrolled	COVIDVAC	VSSEWT	VSSE1-VSSE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CVQ Survey File Continuously Enrolled	COVIDVAC	VSSCWT	VSSC1-VSSC100	Continuously enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023
Survey File Topical Section	CVQ Cost Supplement Ever Enrolled	COVIDVAC	VSCEWT	VSCE1-VSCE100	Ever enrolled in 2022 and still alive, entitled, and not living in a facility in Summer 2023

8.2.2 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 94 for the 2022 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 sections 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.

8.2.3 2022 Initial Weighting

In the initial weighting stage, the initial nonresponse adjusted weights for the Incoming Panel of Medicare beneficiaries, which for the 2022 data year is referred to as the "2022 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 2022.

Full-sample and Replicate Raked Base Weights. A full-sample base weight is derived for all beneficiaries in the 2022 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 $\pi k \mid ij$ be the conditional probability of selection for beneficiary k given the PSU i and the SSU j, such that $\pi k \mid ij = \rho Iak \mid ij$ for beneficiaries in the Hispanic sampling stratum and age group a, and similarly equals $\rho\text{-}Iak \mid ij$ for beneficiaries in the non-Hispanic sampling stratum and age group a, as described in Section 3. Then, for all selected beneficiaries, the base weights are defined by

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

where πi is the probability of selection for the PSU, $\pi j \mid i$ is the conditional probability of selection for the $j\text{-}th$ SSU given the PSU, and $\pi k \mid ij$ is the conditional probability of selection for the $k\text{-}th$ 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 $\alpha = 1, \ldots, 100$ 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

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

where τ is a compositing factor between zero and one, δ_{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 2023 contained additional records for beneficiaries who became eligible near the end of 2022. Due to the timing of this file, these newly-added beneficiaries were not subjected to sampling and could not be included in the 2022 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 2022 Panel sum to the correct population total.

The raked full-sample weight is defined by

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

where φ_{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 five dimensions used at this raking step are

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

5. Medicare Advantage (MA) plan enrollment (2-level; MA plan or traditional Medicare)

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

Initial Nonresponse Adjustments. The raked base weights for the 2022 Panel are then adjusted for nonresponse at the first interview in Fall 2022. The response statuses in Fall 2022 are determined, where a respondent is a beneficiary that is alive and entitled and completed the Fall 2022 interview. Nonresponse adjustment cells are constructed prior to performing the adjustment. First, the beneficiaries are divided into three primary adjustment cells: alive community, deceased community, and facility residents.

Separately within each of these main adjustment cells, response propensity models are fit using logistic regression to model the probability of response at Fall 2022 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 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. A total of 111 adjustment cells formed following the response modeling process. 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 2022 Panel.

8.2.4 2022 Survey File Weights

The 2022 Survey File data were collected in Summer and Fall 2022 from beneficiaries sampled in the 2019 through 2022 annual panels. To facilitate estimation from the resulting data, five sets of full-sample and replicate weights are derived. These include 2022 continuously enrolled cross-sectional weights; 2-year longitudinal weights for analysis of 2021-2022 data; 3-year longitudinal weights for analysis of 2020-2022 data; 4-year longitudinal weights for analysis of 2019-2022 data; and finally, the 2022 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 who were sampled in the 2019, 2020, 2021, and 2022 Panels. The 2019, 2020, and 2021 Panels are referred to as Continuing Panels, while the 2022 Panel is referred to as the Incoming Panel as members of this sample were interviewed for the first time in Fall 2022. 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 1st of the data year and completion of the fall interview. The ever enrolled weights represent the population of beneficiaries who were enrolled in Medicare for at least one day at any time during the data year.

The 2022 Survey File continuously enrolled cross-sectional weights are populated for the subset of records with a completed Fall 2022 interview that are alive and entitled at the time of the interview. The resulting cross-sectional weights represent the population of beneficiaries who were continuously enrolled in Medicare from January 1, 2022, through Fall 2022. The continuously enrolled cross-sectional weights are the traditional Survey File weights and have been provided every year.

The two-year longitudinal weights are populated for members of the 2019, and 2020, and 2021 Panels that were continuously enrolled in both 2021 and 2022. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before January 1, 2021, and are still alive and entitled as of Fall 2022. The three-year longitudinal weights are populated only for members of the 2019 and 2020 Panels who were continuously enrolled in each of the years 2020, 2021, and 2022. The population represented by these weights is the population of beneficiaries enrolled on or before January 1, 2020, and surviving and entitled as of Fall 2022. Finally, the four-year longitudinal weights are populated only for members of the 2019 Panel who were continuously enrolled during all of the years 2019-2022. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before January 1, 2019, and are still alive and entitled as of Fall 2022.

The 2022 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 2022 interview. Beneficiaries who first became enrolled in 2022 are also included; these current-year enrollees were sampled and interviewed for the first time in 2022. The resulting weights represent the population of beneficiaries who were enrolled in Medicare on at least one day at any point in 2022.

Fall 2022 Nonresponse Adjustment for Continuously Enrolled Weights

Continuing sample from the 2019, 2020, and 2021 Panels are adjusted for nonresponse through Fall 2022. The process begins with weights for these panels that were previously adjusted through Fall 2021. Response status in Winter 2022, Summer 2022, and Fall 2022 is then identified, where a respondent is a beneficiary that was alive and entitled with a complete Fall 2022 interview, or who died or lost entitlement at some time in prior to Fall 2022 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 2021 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 2022 as a function of covariates derived from the Fall 2021 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 123 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 2022.

Derivation of the Continuously Enrolled Weights

The next step takes the weights for Continuing Panels that are now adjusted through Fall 2022 and combines them with the weights for the 2022 Panel that were separately adjusted for initial nonresponse at the first interview (Fall 2022) as part of the initial weighting process. Next, the process removes cases that either died or lost entitlement prior to the Fall 2022 interview or cases from the 2022 Panel that enrolled after January 1, 2022.

At this stage there is quadruple coverage of beneficiaries who accreted before January 1, 2020, triple coverage of beneficiaries who accreted from January 1, 2020 through December 31, 2020, and double coverage of beneficiaries who accreted from January 1, 2021 through December 31, 2021. 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

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

Where n_{pd}^{eff} is the effective number of Fall 2022 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

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

where n_{pd}^{act} is the actual number of completed interviews, $\overline{w}_{id}$ is the average of the Fall 2022 adjusted weights for the panel, and S_{id} is the standard deviation of these weights.

The resulting weights are the final continuously enrolled cross-sectional weights for the 2022 Survey File. They represent the 2022 continuously enrolled population.

Longitudinal Weights for the 2022 Survey File.[32] The derivation of two-year longitudinal weights begins with the weights adjusted through Fall 2022 from the 2019, 2020, and 2021 Panels, subset to beneficiaries who were alive and entitled at the Fall 2022 interview. A ratio adjustment accounted for cases that did not have complete Survey File data in both 2021 and 2022. The weights were then further adjusted to account for triple coverage of those accreting on or before January 1, 2019, and double coverage of those accreting from January 2, 2019 through December 31, 2020, using compositing factors derived similarly as described in the previous section. The final resulting weights represent the two-year longitudinal population, which is the population of beneficiaries who enrolled on or before January 1, 2021, and were alive and entitled as of Fall 2022.

The derivation of three-year longitudinal weights begins with the weights adjusted through Fall 2022 from the 2019 and 2020 Panels, subset to beneficiaries who were alive and entitled at the Fall 2022 interview. A ratio adjustment accounted for cases that did not have complete Survey File data in both 2020 and 2022. The weights are then further adjusted to account for double coverage of those accreting on or before January 1, 2019, using compositing factors. The final resulting weights represent the three-year longitudinal population, which is the population of beneficiaries who enrolled on or before January 1, 2020, and were alive and entitled as of the Fall 2022 interview.

The four-year longitudinal weights are comprised of members of the 2019 Panel and are equal to the weights adjusted through Fall 2022 for this panel, subset to beneficiaries who were alive and entitled at the Fall 2022 interview. There is no need for further adjustment by compositing factors because there is only one panel

---

[32] 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.

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 who enrolled on or before January 1, 2019, and were alive and entitled as of the Fall 2022 interview.

Ever Enrolled Cross-Sectional Weights. Ever enrolled Survey File weights represent the population of Medicare beneficiaries who were ever enrolled at any time during 2022 (i.e., enrolled on at least one day in 2022). 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 2022 Nonresponse Adjustment for Ever Enrolled Weights

The Continuing sample from the 2019, 2020, and 2021 Panels is adjusted for nonresponse through Fall 2022. As with the continuously enrolled weights, the process begins with weights for these panels that were previously adjusted through Fall 2021. The response status in Winter 2022, Summer 2022, and Fall 2022 is then identified. Under the ever enrolled design, respondents include beneficiaries with a complete Fall 2022 interview, beneficiaries who lost entitlement prior to Fall 2022 and had a final complete interview, beneficiaries who died prior to Fall 2022 whether or not a final proxy interview was obtained, and Fall 2022 nonrespondents who were successfully re-fielded in Winter 2023.

Next, the weights are adjusted for nonresponse through Fall 2022, 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 2022 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 2022 in the previous step and combines them with the weights for the 2022 Panel that are separately adjusted for initial nonresponse at the first interview (Fall 2022). Next, the small number of cases that died or lost entitlement prior to January 1, 2022, and hence were never enrolled in 2022, are removed.

At this stage, beneficiaries from the Continuing Panels who died or lost entitlement during 2022 are included. However, the 2022 Panel cases include only beneficiaries who were respondents to the Fall 2022 initial interview, and as such they do not include any beneficiaries who died or lost entitlement prior to Fall 2022. Beneficiaries who enrolled before January 1, 2022, who died or lost entitlement are accounted for by the Continuing Panels. Enrollees on or after January 1, 2022, 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 2022 are combined separately to account for the fact that these beneficiaries are not represented by the 2022 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 2022 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 2022. They

represent the population of beneficiaries who were enrolled for at least one day at any time in 2022. 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

Age Group	Control Total
< 45 Years	1,651,700
45 - 64 Years	6,167,300
65 - 69 Years	17,651,160
70 - 74 Years	15,289,300
75 - 79 Years	11,296,640
80 - 84 Years	7,097,880
85+ Years	6,963,040
Total	**66,117,020**

SOURCE: Medicare Administrative enrollment data

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

Enrollment Year	Control Total
< 2018	47,923,260
2018	3,590,240
2019	3,694,580
2020	3,724,300
2021	3,628,780
2022	3,555,860
Total	**66,117,020**

SOURCE: Medicare Administrative enrollment data

8.2.5 2022 Cost Supplement Weights

Data for the 2022 Cost Supplement File were collected in Winter 2022 through Winter 2023. The weights include beneficiaries sampled in the 2019 through 2021 Panels, plus members of the 2022 Panel who were enrolled in Medicare during 2022. These Cost Supplement File weights are ever enrolled weights representing the population of beneficiaries who were enrolled for at least one day in 2022. 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 2022 Cost Supplement weights include beneficiaries who were sampled in the 2019, 2020, 2021, and 2022 Panels. The 2019, 2020, and 2021 Panels are referred to as Continuing Panels and provide survey-reported cost and utilization for 2022 through participation in the MCBS during Winter 2022 through Winter 2023. Members of the 2022 Panel who were first enrolled in 2022 are referred to as "current-year enrollees." They were first interviewed in Fall 2022 and did not provide cost and utilization data for the period of time between enrollment and completion of the Fall 2022 interview; cost and utilization data for the period between the Fall 2022 interview and the end of 2022 were collected in Winter 2023. A combination of the survey-collected data for the end of the year and Medicare claims data were used to impute beneficiary-level data for the entire period of enrollment in 2022. The final weights, which include both the Continuing Panels and the recent enrollees, represent the population of beneficiaries who were enrolled in Medicare at any time during 2022.

Adjustment Derivation of Cross-Sectional Weights for the Continuing Panels

The process begins with weights for the 2019, 2020, and 2021 Panels that were previously adjusted through Fall 2022 as part of the 2022 Survey File weights. These weights are further adjusted based on a product of the 2022 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 2022 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 2022 Panel who completed the initial Fall 2022 interview. Beneficiaries who died or lost entitlement prior to January 1, 2022 are ineligible and removed at this stage. Beneficiaries who survived into 2022 but do not meet the above criteria are considered to be nonrespondents for the 2022 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 2022 Survey File weights.

Note that at this stage there is triple coverage of beneficiaries who accreted before January 1, 2020 in the Continuing Panels, and double coverage of beneficiaries who accreted from January 1, 2020 through December 31, 2020. 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 beneficiaries who enrolled between January 1, 2022, and December 31, 2022, inclusive. This step begins with the initial weights for the 2022 Panel, adjusted for nonresponse at the Fall 2022 interview. The subset of all Fall 2022 respondents from the 2022 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.

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 2022 Panel) provides an estimate of the ever enrolled population in 2022, 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 Exhibits 8.2.2 and 8.2.3. The resulting weights are the final weights for the 2022 Cost Supplement File. They represent the population of beneficiaries who were enrolled for at least one day at any time in 2022.

Longitudinal Weights for the 2022 Cost Supplement. The two-year longitudinal weights are populated for members of the 2019, 2020, and 2021 Panels who were enrolled in both 2021 and 2022 and provided utilization and cost data for both years. Members of the 2019 and 2020 Panels provided data for the 2021 and 2022 data years through participation in the MCBS during Fall 2020 through Winter 2023. Members of the 2021 Panel who first enrolled in 2021 provided data for the end of 2021 in the Winter 2022 interview and provided data for the 2022 data year in Winter 2022 through Winter 2023. The final two-year longitudinal weights represent the population of beneficiaries who were ever enrolled in Medicare at any time during both 2021 and 2022.

The three-year longitudinal weights are populated for members of the 2019 and 2020 Panels who were enrolled in 2020, 2021, and 2022, and provided utilization and cost data for all three years. Members of the 2019 Panel provided data for the 2020-2022 data years through participation in the MCBS during Fall 2019 through Winter 2023. Members of the 2020 Panel who first enrolled in 2019 provided data for the end of 2020 in the Winter 2021 interview and provided data for 2021 and 2022 in Winter 2021 through Winter 2023. The

final three-year longitudinal weights represent the population of beneficiaries who were ever enrolled in Medicare at any time during each of 2020, 2021, and 2022 implying continuous enrollment during 2021.

8.2.6 2022 Topical Weights

Topical segment weights pertain only to data collected in certain sections of the Community Questionnaire. The Patient Activation (PA) questions were administered in Fall 2022 (Round 94) as part of the Satisfaction with Care Questionnaire (SCQ) to living beneficiaries not responding by proxy. The Beneficiary Knowledge Questionnaire (KNQ), Access to Care Questionnaire (ACQ), Usual Source of Care Questionnaire (USQ), Telemedicine Questionnaire (TLQ), and forgone care items in the Dental, Vision, and Hearing Utilization Questionnaire (DVH), Medical Provider Utilization Questionnaire (MPQ), and Prescribed Medicine Utilization Questionnaire (PMQ) were administered in Winter 2023 (Round 95) to all living respondents. The Income and Assets Questionnaire (IAQ) and the Drug Coverage Questionnaire (RXQ) were administered to all respondents in Summer 2023 (Round 96). The Chronic Pain Questionnaire (CPQ) was administered in Summer 2023 (Round 96) to beneficiaries responding without a proxy. Additionally, weights are included on the COVIDTOP segment for the COVID-19 Questionnaire (CVQ), which was administered to all respondents in Fall 2022 (Round 94), Winter 2023 (Round 95), and Summer 2023 (Round 96), but only Winter 2023 and Summer 2023 CVQ data were released on the COVIDTOP segment. Note that the COVID-19 Experiences segment (COVIDEXP) includes the Fall 2022 CVQ data and therefore does not have separate seasonal survey weights.

To facilitate estimation from the resulting data, three sets of full-sample and replicate weights were derived for each set of seasonal segments: one based on the 2022 Survey File ever enrolled population, one based on the 2022 Survey File continuously enrolled population, and the last based on the 2022 Cost Supplement ever enrolled population. These weights can be used to conduct joint analyses of Topical segment data, Survey File data, and Cost Supplement data. Exhibit 8.2.4 lists the Topical weights for these rounds.

Note that counts of cases with positive Topical 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. Please see Exhibit 8.2.1 for the segment location and name of Topical weights provided with the 2022 Survey File LDS. Please see Exhibit 8.2.4 for further details regarding the 2022 Topical survey weights including the record and variable counts and descriptions of additional adjustments to the weights.

Exhibit 8.2.4:   2022 Data Year Topical Survey Weights Datasets and Contents

Segment Name	Record Count	Variable Count	Full-Sample Weight	Replicate Weights	Description	Adjustment
PNTACT	10,978	102	P94SFWT	P94SF1-P94SF100	PA R94 Survey File Ever Enrolled	Fall non-proxy adjustment
PNTACT	10,795	102	PA94CWT	PA94C1-PA94C100	PA R94 Survey File Continuously Enrolled	Fall non-proxy adjustment
PNTACT	5,989	102	PA94EWT	PA94E1-PA94E100	PA R94 Cost Supplement	Fall non-proxy adjustment
MCREPLNQ, USCARE, ACCSSMED, TELEMED	10,677	102	W95SFWT	W95SF1-W95SF100	Winter R95 Survey File Ever Enrolled	Winter non-response adjustment

Segment Name	Record Count	Variable Count	Full-Sample Weight	Replicate Weights	Description	Adjustment
MCREPLNQ, USCARE, ACCSSMED, TELEMED	10,257	102	W95CWT	W95C1-W95C100	Winter R95 Survey File Continuously Enrolled	Winter non-response adjustment
MCREPLNQ, USCARE, ACCSSMED, TELEMED	6,383	102	W95EWT	W95E1-W95E100	Winter R95 Cost Supplement	Winter non-response adjustment
INCASSET, RXMED, COVIDTOP	7,663	102	S96SFWT	S96SF1-S96SF100	Summer R96 Survey File Ever Enrolled	Summer non-response adjustment
INCASSET, RXMED, COVIDTOP	7,385	102	S96CWT	S96C1-S96C100	Summer R96 Survey File Continuously Enrolled	Summer non-response adjustment
INCASSET, RXMED, COVIDTOP	4,229	102	S96EWT	S96E1-S96E100	Summer R96 Cost Supplement	Summer non-response adjustment
CHRNPAIN	6,756	102	C96SFWT	C96SF1-C96SF100	CPQ R96 Survey File Ever Enrolled	Summer non-response and non-proxy adjustments
CHRNPAIN	6,514	102	CP96CWT	CP96C1-CP96C100	CPQ R96 Survey File Continuously Enrolled	Summer non-response and non-proxy adjustments
CHRNPAIN	3,750	102	CP96EWT	CP96E1-CP96E100	CPQ R96 Cost Supplement	Summer non-response and non-proxy adjustments

Composition of Sample and Populations of Interest. The PAQ data were collected from beneficiaries selected in the 2019, 2020, 2021, and 2022 Panels who were alive, entitled, living in the community, and completed the Community Questionnaire without use of a proxy in Fall 2022 (Round 94).

The winter Topical segment data were collected from beneficiaries selected in the 2019, 2020, 2021, and 2022 Panels who were alive, entitled, living in the community, and completed the Community Questionnaire in Winter 2023 (Round 95).

The summer Topical segment data were collected from beneficiaries selected in the 2020, 2021, and 2022 Panels who were alive, entitled, living in the community, and completed the Community Questionnaire in Summer 2023 (Round 96). Because the oldest panel does not receive the final summer interview, the summer round sections are limited to only three sample panels of beneficiaries rather than four.

The CPQ data were collected from beneficiaries selected in the 2020, 2021, and 2022 Panels who were alive, entitled, living in the community, and completed the Community Questionnaire without use of a proxy in Summer 2023 (Round 96).

The weights for the Topical segments are all derived to represent a common population: beneficiaries who were alive, entitled, and living in the community during the round of data collection. Some beneficiaries with populated winter Topical weights do not have ACQ data because they did not have any ER, IP, or OP events in the year leading up to interview and were not in an MA plan. For the release of ACQ data, CMS fills in information reflecting no such events for these cases. In addition, the IAQ was administered to proxy respondents for deceased and institutionalized beneficiaries, so some collected IAQ data is forfeited by the population definition. Imputed total income for all respondents, including Community and Facility interviews, will appear on the LDS file containing demographic information.

Derivation of Topical Segment Weights

Each of the Topical segment weights is based on a starting weight, which is a Round 94 nonresponse-adjusted weight derived during the process of creating the final 2022 Survey File ever enrolled, Survey File continuously enrolled, or Cost Supplement ever enrolled weights. The choice of starting weight determines the population that the derived Topical segment weights represent, but the process for each Topical weight 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 2022 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 section to beneficiaries 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 we have no indication of mortality or residential (community or facility) status. As expected, the number of cases with unknown eligibility was smaller in Round 95 because this round immediately followed the Fall 2022 Survey File interviews, whereas in Round 96 there was an intervening round in which some members of the sample became nonrespondents. For the PA question series, which was fielded in Round 94 as part of SCQ, there is no unknown eligibility. In all cases, this first adjustment for unknown eligibility makes the implicit assumption that, if we were able to observe the eligibility 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, we limit the set of beneficiaries to persons who were eligible to receive the respective questionnaire sections. A beneficiary was considered ineligible if they had died, lost entitlement, or were living in a facility during the round. The nonresponse adjustment is then made, in which the weights for the eligible nonrespondents are distributed to the eligible respondents.

Finally, to account for the overlap between panels in accretion year, the weights of the different panels are then adjusted by compositing factors. These compositing factors were derived from the effective number of completes by accretion year and age group across the set of panels that were administered the seasonal section (the 2019-2022 Panels for PAQ/KNQ/ACQ/USQ/TLQ and the 2020-2022 Panels for IAQ/RXQ/CPQ/CVQ).

8.3 MCBS Imputation Processes

8.3.1 Overview

As noted earlier, MCBS imputation falls under three umbrellas that all focus on imputing monetary amounts: IA imputation, insurance premium imputation, and Event, Payer, and Cost imputation, which consists of imputation for PM and Non PM events and costs. IA imputation completes income and asset information for the beneficiary and spouse/partner, insurance premium imputation completes the monthly cost of medical

insurance plans, and PM and Non PM imputation completes medical event and cost data. For all four types, two groups of variables are imputed:

- Probes: Yes/no variables indicating whether the type of income, asset, insurance premium, or payer should have a nonzero amount.

- Amounts: The value of the income, asset, insurance premium, or cost paid for a medical event. For IA imputation, amounts are nonzero if the associated probe indicates the income or asset exists; otherwise, the amounts are missing. For insurance premium imputation, amounts are nonzero if the associated probe indicates an insurance premium has been paid for the insurance plan; otherwise, the monthly premium amount is set to zero. For PM and Non PM imputation, amounts are nonzero if the associated probe indicates that the payer paid; otherwise, the amounts are zero.

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

8.3.2 Income and Asset Imputation

Overview. The 2022 IA imputation imputes detailed information about income and assets of the beneficiary and spouse/partner for Community Questionnaire respondents. For Facility Instrument respondents and Community Questionnaire and Facility Instrument nonrespondents,[33] 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 appropriate. 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

- Correct inconsistent values that appear to be the result of data entry errors (for example, an extra "0" was entered)

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 efficient. This method takes the non-missing IA value directly from another beneficiary with the same socio-economic characteristics 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 do not 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 provided 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 from the prior year to impute the current year missing value. This prior year carry-forward method provides reasonable

---

[33] The Income and Assets questionnaire section (IAQ) is only administered once per year. Nonresponse to this section may be due to nonresponse in the round the questionnaire section is administered, or nonresponse to questions in the IAQ. For more information on IAQ, see Section 4.1.

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 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 age, and metropolitan status.

Monthly earnings from work are reported and imputed separately for the beneficiary and for their spouse or partner. Prior to 2022, these variables were combined into a single variable and this variable was imputed on its own.

8.3.3 Insurance Plan Premium Imputation

Overview. The 2022 insurance plan premium imputation imputes the monthly cost of medical insurance plans under two sets of conditions. The first set requires that the medical plan is offered by a private company and that the main insured person paid any or all the premium for the plan coverage. The second set is for MA plans and requires that there was an additional cost for coverage beyond Medicare Part B premiums.

Process. For private insurance plans, respondents were asked if they paid any or all of the premium for the plan coverage. For MA plans, respondents were asked if there was an additional cost to the plan beyond the cost of their Medicare Part B premium. These probe questions typically have low nonresponse rates, but any nonresponse values are imputed via a hot deck method. This method takes a non-missing response from another beneficiary with the same socio-economic characteristics to fill in a response to the probe question. If the imputed response to the probe is negative, then the monthly premium for the insurance plan is set to zero.

The monthly premium amount is imputed after the probes. Some respondents refuse to provide or do not know the amount of their insurance plan premium. For these cases, the monthly premium is imputed via hot deck imputation. 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. For insurance plan premium imputation, the auxiliary variables that define imputation cells include insurance coverage type (for example, general medical, dental, or vision), beneficiary's age, Medicare eligibility status, and metropolitan residence status.

8.3.4 Event, Payer, and Cost Imputation

Overview. Event, Payer, and Cost imputation fills in missing payer and payment information for beneficiaries' medical events. Event, Payer, and Cost imputation is conducted through two separate processes to account for differing payment scenarios for some event types. Imputation for PM events is done separately from all other events because the rules associated with Medicaid payments for PM events are different. Imputation for all event types other than PM (Non PM) are conducted separately. Also, no PM imputation is conducted for beneficiaries living in a facility as the Medicare Part D administrative claims data for this group are considered complete. The imputation procedures used for PM events versus all other event types (Non PM) are similar but not identical. Note, observed payments from the Veterans Administration (VA) are combined into the "'Other Sources" payer.

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.

For step one, events are imputed 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

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

In step two, events are imputed 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[34], payment amounts are imputed individually from different donors.

---

[34] 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.

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

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)	200

In the third and final step, 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 taken from the same donor when possible or are imputed individually if a donor with the required payer pattern does not exist[35]. 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 (0.07 percent of records in 2022).

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

Non PM Imputation

For beneficiaries living in a facility, all provided event data are claims-only. For these claims-only facility events, the total charge and Medicare payments are known. Medicare pays the full amount of the total charge for 10 to 20 percent of these events and pays a partial amount for the remaining events. For these remaining events, the payers and payment amounts are imputed.

---

[35] 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.

Since 2015, current-year enrollee sample beneficiaries are included in the Non PM imputation.[36] 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 unit-level imputation procedure addresses 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 survey-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 donors with a Medicare and not MA payment, as these would be covered by claims 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.

Since 2019, the MA Encounter Data were utilized to improve estimation of medical events and costs for beneficiaries enrolled in MA. Encounter Data from the prior three calendar years were matched to survey reported events over that same period. The goal was to estimate the ratio of utilization counts from matched event data to the utilization counts reported in the survey. Utilization rates for event types affected by the lack of claims data for MA beneficiaries were calculated for each of the three years under two scenarios. In the first scenario, all matched records were kept, all unmatched MA encounter records were kept, and most survey-only records were kept, but events where the MA payment amount was greater than zero and the Medicare payment amount was equal to zero were dropped. In the second scenario, all survey-reported records were kept. The utilization rates from the first scenario were divided by the utilization rates for the second scenario yielding ratio adjustments for each event type for each year. The ratio adjustments were then averaged over the three years. The resulting multipliers were then applied to current data year payment amounts for the events of MA beneficiaries during their MA enrollment periods. The results of these adjustments were summarized within the service level summary and person level summary files and are not applied at the event level.

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 person level (one record per beneficiary), and at the service level (one record per beneficiary and event type). Event-level records are first summed to the service

---

[36] See Section 3.4, "Current-Year Enrollee Sample", for more information on these beneficiaries.

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 jth variable, y_j is the data for observation y and the jth variable, and wj is the weight for the jth 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 $xj = yj$ and 0 otherwise. For an ordinal, interval, or ratio variable, $d_{x,y}^j = 1 - |x_j - y_j|$.[37,38,39]

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 30th 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.

8.3.5 Details on the MA Encounter Data-Informed Ratio Adjustments

Ratio adjustments were applied to qualifying event-level records. Beginning with 2020, the ratio adjustments were based on the encounter setting as well as age category and reported general health status. Subgroups based on age and health are displayed in Exhibit 8.3.4. The applied ratio adjustments are displayed by setting, event type, and age and health subgroup in Exhibit 8.3.5.

---

[37] SAS Institute Inc. 2017. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.

[38] Podani, János. "Extending Gower's General Coefficient of Similarity to Ordinal Characters." *Taxon* 48, no. 2 (1999). 331-340.

[39] Gower, John C. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27, no. 4 (1971). 857-871.

Exhibit 8.3.4: Ratio Adjustment Age and Health Subgroups

Age	General Health Status[1]	Age and Health Subgroup
<65 years	All responses	1
65+ years	Excellent or Very good	2
65+ years	Good, Fair, Poor, or Unknown	3

[1] Respondents are asked to report the beneficiary's general health status based on a 5-item scale.

Exhibit 8.3.5: Ratio Adjustments by MA Setting, Event Type, and Age and Health Subgroup

Setting (EVNTTYPE)	Age and Health Subgroup 1 (<65) Ratio Adjustment	Age and Health Subgroup 2 (65+, Excellent/Very good) Ratio Adjustment	Age and Health Subgroup 3 (65+, Good/Fair/Poor) Ratio Adjustment
Carrier (MP/SD/SL/OM)	2.29	1.90	2.43
Durable Medical Equipment (OM)	1.48	1.14	1.29
Inpatient (IP)	1.44	1.10	1.46
Outpatient (OP)	1.75	1.31	1.45
Skilled Nursing Facility (IU)	2.63	2.63	2.63

SOURCE: 2019-2021 MA Encounter Data and 2019-2021 survey-reported utilization data.

The MA Encounter ratio adjustment was applied to all payers for event-level records that met the following criteria:

1. Event type (EVNTTYPE) was:
    a. IP: Inpatient
    b. IU: Skilled Nursing Facility (SNF) (excluding hospice)
    c. MP: Medical Provider services
    d. OM: Other Medical services
    e. OP: Outpatient
    f. SD: Separately Billing Doctors
    g. SL: Separately Billing Labs
2. The event occurred within a month covered by MA based on the administrative enrollment data
3. The event was reported in the survey, either matched or unmatched (i.e., not a claims-only event)

Records with an EVNTTYPE of Other Medical Expense (OM) were assigned to either the Carrier or Durable Medical Equipment (DME) setting. There was a large difference in the ratio adjustment for these two settings, so care was taken to assign records to each of these settings. If the claim record was determined to be for durable medical equipment, then the record was assigned to the DME setting, otherwise the record was assigned to the Carrier setting. However, the setting could not be determined for unmatched survey reported OM events. These events were randomly assigned to the Carrier or DME settings at a proportion determined by historical norms. An examination of matched claims data from the most recent three years was conducted and found that 95 percent of OM event type records were categorized as durable medical equipment claims. Therefore, 95 percent of the unmatched survey reported OM events were assigned to the DME setting with the remaining 5 percent assigned to the Carrier setting.

The unweighted impact of the MA Encounter ratio adjustment on the overall sum of total costs was an increase of 12.6 percent.

9. RESPONSE RATES AND NONRESPONSE

This section presents the response rates and describes the derivation of those rates for the 2022 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 (2022) and responded to the survey in 2022. 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 2021 and also responded during 2022. Conditional response rates use the sample who responded during 2021 as the baseline in their calculation. In other words, they are conditioned on response in year 2021.

9.1.1 2022 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 $t = 2022$ 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 1, 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 2022 Cost Supplement File unconditional and conditional response rates by panel.

Exhibit 9.1.1: 2022 MCBS Annual Cost Supplement File Unconditional Response Rates

Panel	Released	Complete	Eligible	Ineligible	Unconditional Response Rate
2019	11,615	2,008	9,910	1,705	20.3%
2020	15,952	2,479	14,349	1,603	17.3%
2021	15,950	3,174	15,101	849	21.0%
2022	>679	223	672	<11*	33.2%
Total	>44,196	7,884	40,032	>4,164	19.7%

SOURCE: 2022 MCBS Internal Sample Control File
* Cell sizes of less than 11 are suppressed. Complementary suppression is used so that suppressed cells cannot be derived.

Exhibit 9.1.2: 2022 MCBS Annual Cost Supplement File Conditional Response Rates

Panel	Complete	Eligible	Subset of Eligibles Not Released	Conditional Response Rate
2019	2,008	9,910	7,454	81.8%
2020	2,479	14,349	10,903	71.9%
2021	3,174	15,101	9,359	55.3%
2022	223	672	-	33.2%
Total	7,884	40,032	27,716	64.0%

SOURCE: 2022 MCBS Internal Sample Control File

9.1.2 2022 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 1st 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 ineligibles 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 nonrespondents 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 ineligibles 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 nonrespondent.

Exhibits 9.1.3 and 9.1.4 display the 2022 annual Survey File unconditional response rates by panel for ever enrolled and continuously enrolled beneficiaries.

Exhibit 9.1.3: 2022 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
2019	11,615	2,021	11,219	396	18.0%
2020	15,952	2,495	15,739	213	15.9%
2021	15,950	3,204	15,914	36	20.1%
2022	17,139	6,259	16,306	833	38.4%
Total	60,656	13,979	59,178	1,478	23.6%

SOURCE: 2022 MCBS Internal Sample Control File

Exhibit 9.1.4: 2022 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
2019	11,615	1,862	10,645	970	17.5%
2020	15,952	2,304	14,878	1,074	15.5%
2021	15,950	2,895	15,013	937	19.3%
2022	17,139	6,054	16,306	833	37.1%
Total	60,656	13,115	56,842	3,814	23.1%

SOURCE: 2022 MCBS Internal Sample Control File

Exhibits 9.1.5 and 9.1.6 display the 2022 Survey File conditional response rates by panel for ever enrolled and continuously enrolled beneficiaries.

Exhibit 9.1.5: 2022 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
2019	2,021	11,219	8,756	82.1%
2020	2,495	15,739	12,287	72.3%
2021	3,204	15,914	10,132	55.4%
2022	6,259	16,306	-	38.4%
Total	13,979	59,178	31,175	49.9%

SOURCE: 2022 MCBS Internal Sample Control File

Exhibit 9.1.6: 2022 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
2019	1,862	10,645	8,262	78.1%
2020	2,304	14,878	11,540	69.0%
2021	2,895	15,013	9,430	51.9%
2022	6,054	16,306	-	37.1%
Total	13,115	56,842	29,232	47.5%

SOURCE: 2022 MCBS Internal Sample Control File

9.2 Nonresponse Bias Analysis

The most recent nonresponse bias analysis report based on the 2021 data year is available in the *2021 MCBS Methodology Report* at https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.[40]

9.3 Change to Multimode Data Collection

Due to the COVID-19 pandemic, MCBS data collection switched from in-person to phone-only interviews in March 2020 and throughout most of 2021, with a return to some in-person interviewing beginning in November 2021. Moving forward, interviewers will use a combination of phone and in-person data collection approaches based on ongoing analyses, interviewer availability, and respondent preference. For the percentage of interviews completed in each mode in 2022, see Exhibit 6.3.2.

CMS and NORC have engaged in ongoing evaluations to understand the impacts of multimode data collection on respondent recruitment, retention, and response rates. Beginning in 2020, NORC re-engineered the recruiting and contacting protocols used for the Incoming Panel to focus on phone data collection, including enhanced matching of beneficiary sample file information to phone numbers and incorporating repeated reminder mailings. In an analysis of multimode data collection efforts in 2022, NORC found differing impacts of in-person outreach and interviewing between the Incoming Panel and Continuing Panel respondents and between beneficiary demographic subgroups. Phone outreach and phone interviews work well for many beneficiaries across panels, meaning that in-person data collection is not needed universally. These findings and continued analysis will allow the MCBS to refine mode choices based on beneficiary characteristics and project priorities, informing multimode strategies on the MCBS going forward.

---

[40] A nonresponse bias analysis is conducted every three years for the MCBS. The next analysis will take place in 2026 and will be based on the 2024 data year.

10. USING MCBS DATA FILES

10.1 MCBS Data User's Guides

The MCBS Data User's Guides offer a publicly available, easily searchable resource for data users. The Guides are 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.

For each MCBS data year, two stand-alone Data User's Guides are released. For 2022, the first is entitled *2022 MCBS Data User's Guide: Survey File*. This Guide documents the key features of the study and MCBS data products overall. It also provides technical information on the Survey File LDS including the derivation of variables and any significant changes in the variables and/or file structure. The second is entitled *2022 MCBS Data User's Guide: Cost Supplement File*. This Guide provides technical information on the Cost Supplement File LDS and also describes the derivation of variables and any significant changes in the variables and/or file structure. In addition, the *Data User's Guide: Cost Supplement* contains detailed information about matching survey and administrative data as well as imputation.

10.2 MCBS Microdata Public Use Files (PUFs)

Beginning with data collected in the 2013 MCBS, a Survey File PUF and accompanying documentation are available free for download under the MCBS PUF link at https://www.cms.gov/data-research/statistics-trends-and-reports/mcbs-public-use-file. The MCBS Survey File PUF is an easy-to-use data file with select data items that allow researchers to conduct analysis on topics such as health disparities, access to and satisfaction with health care, and medical conditions for Medicare beneficiaries living in the community. Additionally, a Cost Supplement File PUF and accompanying documentation are also available free for download at the website linked above. The MCBS Cost Supplement PUF allows researchers to conduct analysis on health care service use and expenditures for Medicare beneficiaries living in the community, as well as sources of payment. The MCBS Microdata PUFs are not intended to replace the more detailed LDS files. Rather, they provide a publicly available alternative for researchers interested in the health, health care cost and use, access to and satisfaction with Medicare of beneficiaries.

In addition to the Survey File and Cost Supplement File PUFs, data collected in the COVID-19 Summer and Fall 2020 and Winter 2021 Community Supplements are available in separate COVID-19 PUFs. Given that the MCBS Microdata PUFs meet all necessary requirements regarding de-identification of the data and mitigation of disclosure risk, they provide the very highest degree of protection to the Medicare beneficiaries' protected health information.

10.3 MCBS Limited Data Set (LDS) Files

There are two MCBS LDS's available to data users for the 2022 data year. Requests for the MCBS LDS files must be made through the CMS Data Use Agreement (DUA) tracking system known as the Enterprise Privacy Policy Engine or EPPE. EPPE can be used to initiate a new LDS DUA request or to amend/update an existing LDS DUA. Instructions for accessing and using EPPE to make a request can be found here: https://www.cms.gov/data-research/files-for-order/data-disclosures-and-data-use-agreements-duas/limited-data-set-lds.