MCBS
Medicare Current Beneficiary Survey

METHODOLOGY REPORT

Centers for Medicare & Medicaid Services (CMS)

Office of Enterprise Data and Analytics (OEDA)

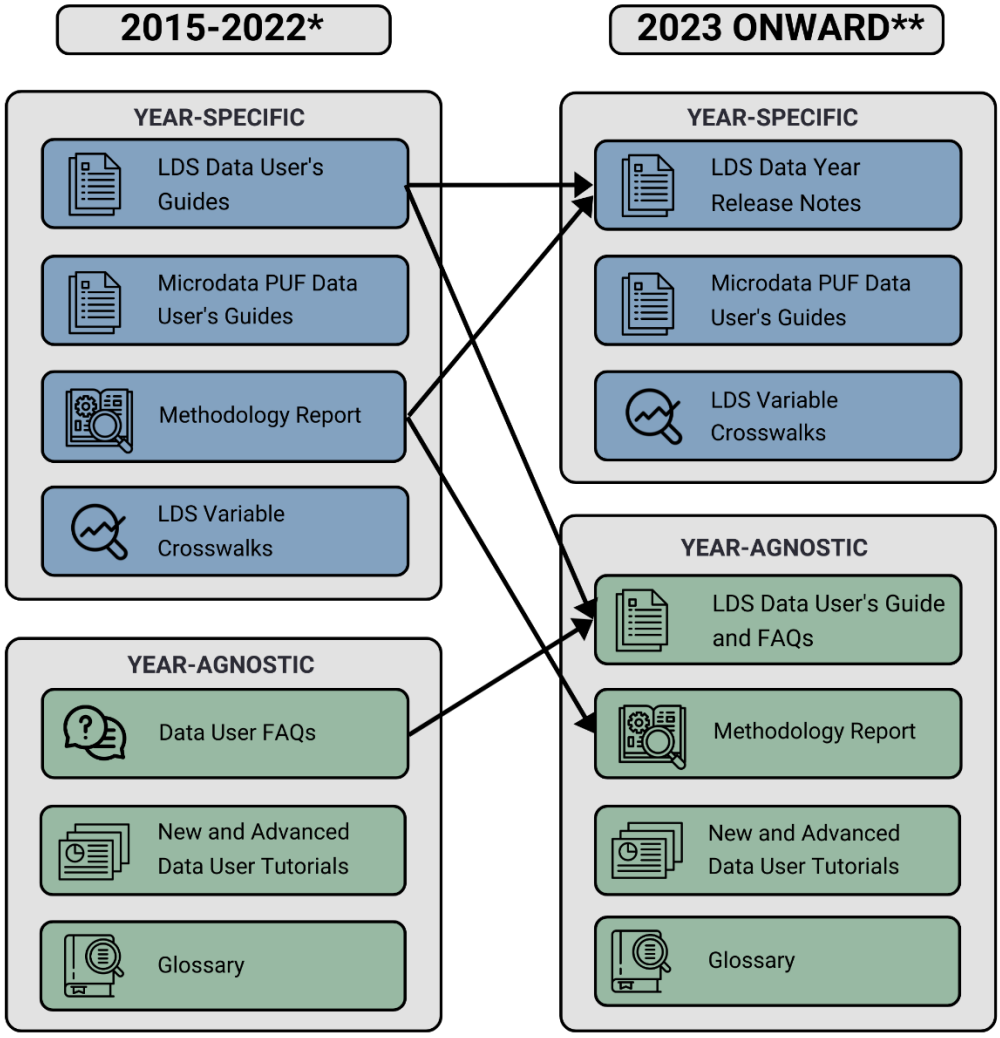
Version Control Log

Date	Version	Revisions
7/2/2025	1.0	Initial version published.

MCBS DOCUMENTATION CROSSWALK AND OVERVIEW

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). These products were consolidated beginning with the 2023 data year to separate the detailed, background information on the MCBS from focused year-specific content that is most relevant to researchers. This section provides a concise overview of MCBS documentation products beginning with the 2015 data year, all available for download on the CMS MCBS website: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

MCBS DOCUMENTATION OVERVIEW



NOTES: The year-specific products are updated annually for each data year. The year-agnostic products are reviewed annually, but only updated as needed.

* For new researchers using the 2015-2022 MCBS LDS, the *Survey File LDS Data User's Guide* and *New User Tutorial* are the recommended starting points. See the CMS MCBS website for information on the pre-2015 MCBS documentation.

** Beginning with the 2023 MCBS LDS, the *LDS Data Year Release Notes* and *New User Tutorial* are the recommended starting points for new researchers.

TABLE OF CONTENTS

LIST OF EXHIBITS

ACRONYM LIST

ACS	American Community Survey
ACP	Automated Crossover Process
BRR	Balanced repeated replication (or Fay's method)
CAPI	Computer-Assisted Personal Interviewing
CASPER	Certification and Survey Provider Enhanced Reports
CCN	CMS Certification Number
CFC	Community-Facility-Crossover
CFR	Continuing Facility Resident
CMS	Centers for Medicare & Medicaid Services
EOB	Explanation of Benefit Statement
FCF	Facility-Community-Facility Crossover
FDB	First Databank
FFC	Facility-Facility-Crossover
FFS	Fee-for-Service
GAD	Generalized Anxiety Disorder screening tool (GAD-2)
HICN	Heath Insurance Claim Number
HMO	Health Maintenance Organization
IA	Income and Assets
IPR	Incoming Panel Respondent
IRB	Institutional Review Board
LDS	Limited Data Set(s)
MA	Medicare Advantage
MAPD	Medicare Advantage Prescription Drug Plan
MBI	Medicare Beneficiary Identifier
MCBS	Medicare Current Beneficiary Survey
MDS	Minimum Data Set
MSA	Metropolitan Statistical Area
MSN	Medical Summary Notice
NCH	Medicare National Claims History
NDC	National Drug Code
Non PM	Non Prescription Medicine
NORC	NORC at the University of Chicago
OEDA	Office of Enterprise Data and Analytics
OMB	Office of Management and Budget
PDP	Prescription Drug Plan
PHQ	Patient Health Questionnaire depression screening tool (PHQ-9)
PII	Personally Identifiable Information
PM	Prescription Medicine
PSU	Primary Sampling Units
PUF	Public Use File
QC	Quality Control
SAS	Statistical Analysis System

SNF	Skilled Nursing Facility
SSU	Secondary Sampling Unit
USU	Ultimate Sampling Unit

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. Interviews are conducted in-person and over the phone_using computer-assisted personal interviewing (CAPI).

This *MCBS Methodology Report* provides an operational perspective on the collection of survey data. The *Methodology Report* complements the *MCBS Data User's Guide* with an overview of all activities carried out in support of the data files, including sampling, instrument design, interviewer training, data collection, data processing, and weighting. 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

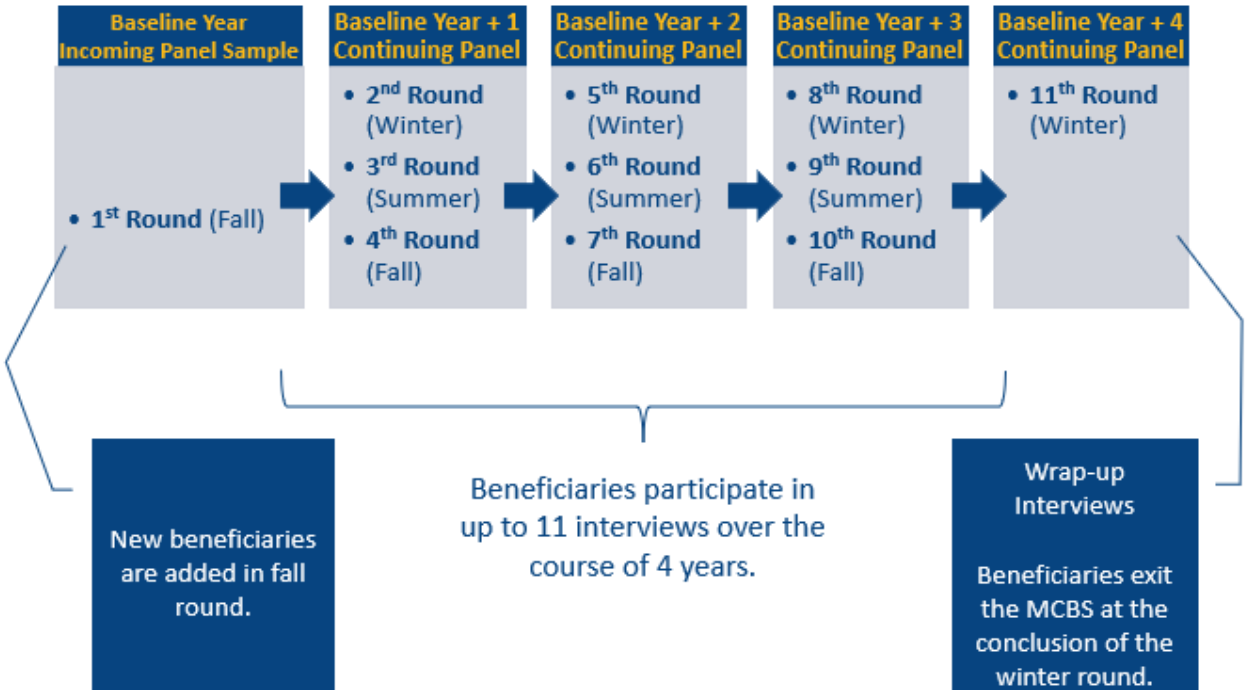
Early in the inception of the MCBS, a design decision was made to limit beneficiary participation in the longitudinal panel to no more than four years. Initial interviews of newly-selected beneficiaries take place once per year in the fall round; these are referred to as the Incoming Panel. For these initial interviews, the fall round begins in July 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 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.

[1] The MCBS is authorized by section 1875 (42 USC 1395ll) 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.

Exhibit 1.2.1: MCBS Participation Timeline



1.3 Contents of the Methodology Report and Key Data Products

Here is an overview of the contents of the *Methodology Report:*

- Section 2: Sample Design for the MCBS Panel – This section provides detailed information on the sampling design and selection of beneficiaries.
- Section 3: Instrument and Materials Design – This section provides information on the questionnaire design and content. More details on the questionnaires, including item- and section-level descriptions and changes, can be found in the annual *MCBS Questionnaire User Guide*.
- Sections 4 & 5: Interviewer Recruitment, Training, and Data Collection – These sections describe interviewing and data collection fielding procedures, including eligibility for each round.
- Sections 6 & 7: Data Processing, Data Delivery, Weighting, and Imputation – These sections describe the creation of the MCBS data files and provide a detailed overview of weighting and imputation procedures.
- Section 8: Response Rates – This section describes the methodology for computing responses rates for the Survey File and Cost Supplement File.
- Sections 9-10: References and Appendices – These sections provide references and key supporting documentation.

Annually, CMS releases five sets of files for the MCBS – two Microdata Public Use Files (PUFs) and three Limited Data Sets (LDS). The LDS releases are referred to as the Survey File - Early Release, the Survey File, and the Cost Supplement File. Detailed descriptions of the MCBS LDS releases, including the contents of the files and file structure, as well as analytic use guidance and sample code can be found in the *MCBS Data User's Guide*. Special notes on each MCBS LDS release, including information on new and changed variables, can be found in the annual *Data Year Release Notes* or the *historical 2015-2022 Data User's Guides*. Special notes on each MCBS Microdata PUF release can be found in the annual PUF documentation at: https://data.cms.gov/medicare-current-beneficiary-survey-mcbs.

Please note the following terminology preferences for the MCBS used throughout this document:

- *Beneficiary* refers to a person receiving Medicare services who may or may not be participating in the MCBS.[3] Beneficiary may also refer to an individual selected from the MCBS sample about whom the MCBS collects information.
- *Respondent* is the person who answers questions for the MCBS; this person can be the beneficiary, a proxy, or a staff member located at a facility where the beneficiary resides (i.e., the Facility respondent).
- The *data collection year* refers to the three rounds of data collection (winter, summer, and fall) that occur within the calendar year. This is sometimes labeled as the *calendar year*.
- The *data year* refers to the data collected over the three years that are included in the LDS release. This includes data collected in the prior data collection year, the current data collection year, and the following data collection year.

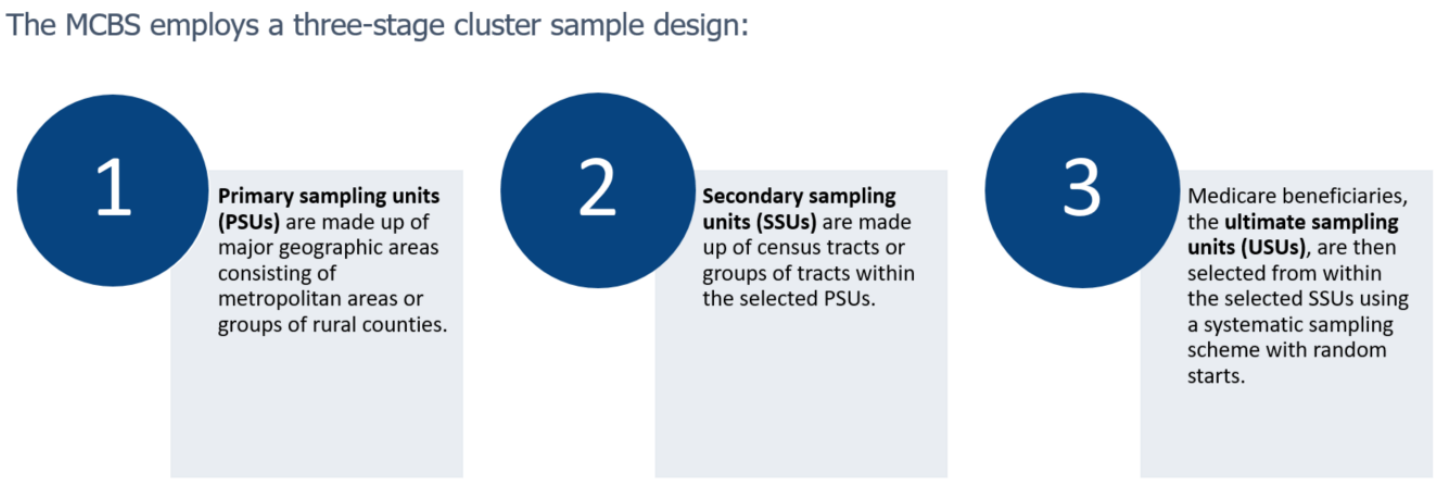
[3] https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html

2. SAMPLE DESIGN FOR THE MCBS PANEL

2.1 Overview of MCBS Sample Design

The MCBS employs a three-stage cluster sample design (see Exhibit 2.1.1). 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 2.1.1: MCBS Sample Design Process

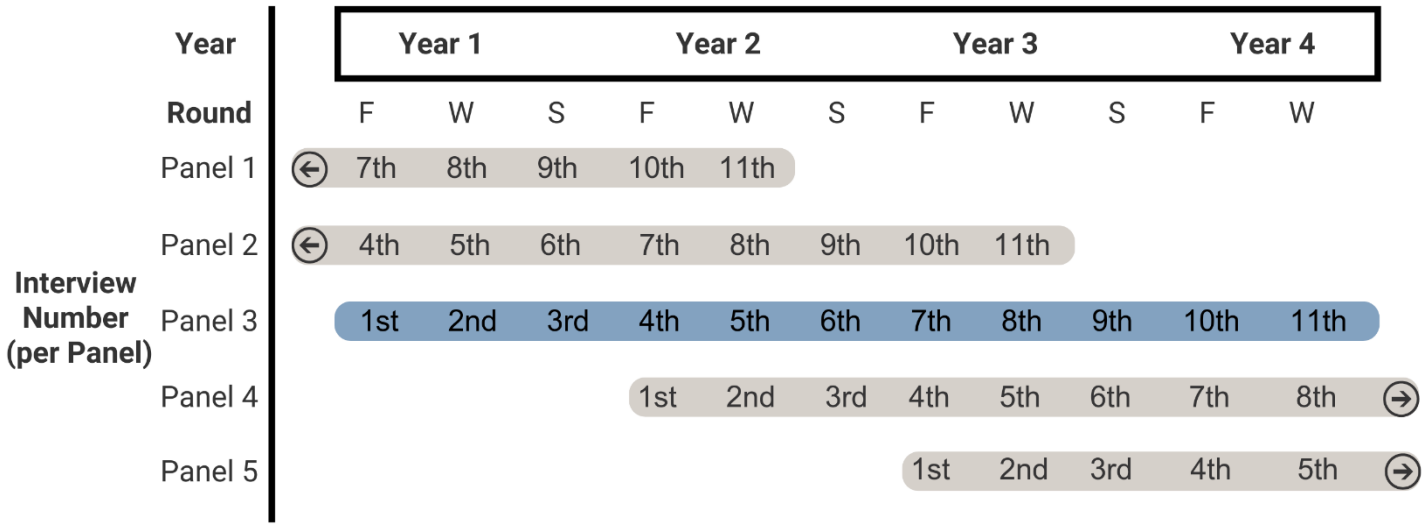


The MCBS continues to use the sample rotation pattern used historically. A new panel is added each fall and retains the year of its entry as its sampling designation for projections and response rate analysis. Once a panel is selected, it remains in the MCBS for four years, participating in a total of 11 rounds.

[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.

Exhibit 2.1.2: MCBS Rotating Panel Design



NOTES: F stands for fall. W stands for winter. S stands for summer. Each panel participates in up to 11 interviews over four years.

This section documents the procedures used to select the new yearly sample in the fall round (i.e., the Incoming Panel). Each panel is retained in the survey for the four years specified under the MCBS sample rotation scheme and is designed to: (a) replace approximately one-third[6] 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 2.4 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, which is used as a basis for selecting the sample for the panel. The targeted population for the MCBS consists of persons enrolled in one or both parts of the Medicare program, that is, Part A or Part B, as of December 31 of the applicable calendar year, and whose address on the Medicare files is in one of the 48 contiguous states (excludes Alaska and Hawaii) or the District of Columbia. Details about the sampling frame construction can be found in Section 2.4.

The MCBS enrollment data 5-percent file extracts are 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 geocode to the selected SSUs constitute the MCBS sampling frame of beneficiaries. The sampling frame is then divided into fourteen sampling strata based on age and Hispanic ethnicity.[7] Age is determined as of December 31 of the calendar year in order to include all beneficiaries enrolling during the calendar year. The age categories are under 45, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 or older. An ethnicity flag (see Section 2.4 for a full description) is 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. Strata are displayed in Exhibit 2.1.3. A random sample of beneficiaries residing in the selected SSUs is then selected from each stratum.

[6] 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.

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

Exhibit 2.1.3: 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 vary 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.[8]

2.2 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.[9] 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 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, each panel is now 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.

2.3 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.[10]

[8] The Survey File - Early Release reflects a subset of the completed cases from the MCBS Survey File.

[9] 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.

[10] 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.

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.

2.4 Selection of Beneficiaries for MCBS Panels

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

2.4.1 Current-Year Enrollee Sample

Since 2015, the year t cohort[11] of beneficiaries (i.e., the set of current-year enrollees) is included in the sampling frame of beneficiaries from which the year t panel[12] is selected.[13] The MCBS uses multiple enrollment data extracts for sampling and multiple sample draws because not all current-year enrollees are included in the enrollment data at the time the initial sampling occurs. Additional extracts, or "updates" to the original enrollment data extract are completed at multiple points during the year. A final extract is delivered in January of the following year and used to fully enumerate the population of Medicare enrollees for year t. Because data collection has already ended, no sample is drawn from the January extract by design; however, the information is used for weights calibration.

Timing of the Interview. Members of the year t cohort of beneficiaries sampled are all enrolled in Medicare sometime during sampling calendar 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 they enroll in Medicare (i.e., on or after their enrollment date in the enrollment data).

2.4.2 Hispanic Oversample

Oversampling of Hispanics has been conducted using a variety of methodologies throughout the MCBS and continues as a design feature.[14] Hispanics are oversampled relative to their non-Hispanic counterparts within the general MCBS sample. The sampling frame is stratified using a flag provided by CMS based on Census records of Hispanic surnames and other enrollment information, such as language preference, and the Hispanic stratum is oversampled relative to the non-Hispanic stratum. The MCBS targets 1,500 annual Hispanic completes in each Survey File.

---

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

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

[13] Historically, to be eligible for sample selection, beneficiaries had to be eligible for Medicare and enrolled by January 1st of the calendar 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.

[14] 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 within the continental US in new panels to allow for improved precision of estimates of health disparities experienced by these populations. Hispanics are 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.

2.4.3 Sample Size Determination

The sample size requirements for each Incoming Panel are 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 calendar 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 calendar year but for whom no death is recorded in CMS administrative updates;

b) Persons in the CMS sampling frame who enrolled prior to the calendar 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 calendar year but for whom there is no record of having lost eligibility in CMS administrative updates; or

c) Persons who enrolled prior to the calendar year and died or lost Medicare eligibility prior to the calendar 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.[15] Sampled beneficiaries who were deceased at the first or second interview and for whom a date of death **after** January 1 of the calendar 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 calendar year (or as of their enrollment date, for current-year enrollees).[16] 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,[17] response rates, and immortal rates were computed within each age group.[18] The immortal and death rates used were an average of historical rates and actual rates from the previous three fall rounds. The immortal rates apply to losses in the first fall interview round only. Similarly, the initial losses due to deaths in the calendar 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.

The response rates 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

---

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

[16] Data for beneficiaries in this group who were newly enrolled (i.e., enrolled during the calendar 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.

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

[18] 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.

complete data for the third Cost Supplement File) are based on averages of corresponding rates from previous years.

The sample size projections also include 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 than the migration out of this category 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.

2.4.4 Sampling Frame

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

- Beneficiaries who were first eligible for Medicare before January 1 of the calendar year and still alive and eligible on January 1 of the calendar year; and
- Beneficiaries who were first eligible for Medicare between January 1 and March 1 of the calendar year (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, includes beneficiaries not included in the first extract and who were first eligible for Medicare between January 1 and August 1 of the calendar year (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, includes beneficiaries not included in the first or second extract and who were first eligible for Medicare between January 1 and September 1 of the calendar year (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[19] 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.

CMS subset each of its enrollment data extracts as described above, keeping only beneficiaries meeting the criteria for the enrollment data subsample for the calendar year. These enrollment data subsample extracts are further subset to include only beneficiaries falling within the 685 selected MCBS SSUs. Using the initial enrollment data subsample extract for the calendar year in combination with previous annual enrollment data subsamples, the size of the total enrollment data subsample for the calendar year (containing all projected calendar year Medicare enrollees, through December 31 of the calendar year, that would be available for sampling) could be forecast at the time of initial sampling (May of each calendar year). This forecast is used to determine how much of the current-year enrollee sample is expected to be selected from the first extract and

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

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 is provided in mid-January of the next year and used to fully enumerate the calendar year cohort to (a) inform undercoverage of the calendar year 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.

2.4.5 Sample Selection

Sampling Fractions. The sampling fractions for the Hispanic and non-Hispanic strata are 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 are computed 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 are 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

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

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

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

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 are adjusted and the conditional probabilities $\rho_{1ak|ij}$ and $\rho_{-1ak|ij}$ are recomputed. Within each stratum, the cumulative sums of the probabilities of selection are formed. In an iterative process, the initial national sampling fractions are repeatedly adjusted until the cumulative sums are as close as possible to the final targeted sample sizes

Selection of the Incoming Panel. Each Panel is drawn by systematic random sampling with probability proportional to the conditional probabilities of selection with an independently selected random start within each PSU. Beneficiaries are ordered within each PSU by age group, SSU (to approximate geographic serpentine sorting), Hispanic flag, and extract.[20] Beneficiaries with a conditional probability of selection equal to 1, which can happen when there are insufficient numbers of beneficiaries within the frame in certain strata,[21] are selected with certainty, given the selection of their PSUs and SSUs.

[20] The second extract is added to the end of the first extract, in the same sort order, and the systematic selection is continued into the range of newly enrolled beneficiaries. The same process is used for the third extract.

[21] This has historically been observed in smaller strata such as the Hispanic and Non-Hispanic Under 65 and the Hispanic 85 and Older strata.

2.4.6 Coverage Analysis of the Sampling Frame

As discussed above, a final enrollment data 5-percent file extract is provided in mid-January. This extract is used to fully enumerate each cohort to (a) inform undercoverage of the sampling frame, and (b) contribute to weighting adjustments to account for this undercoverage.

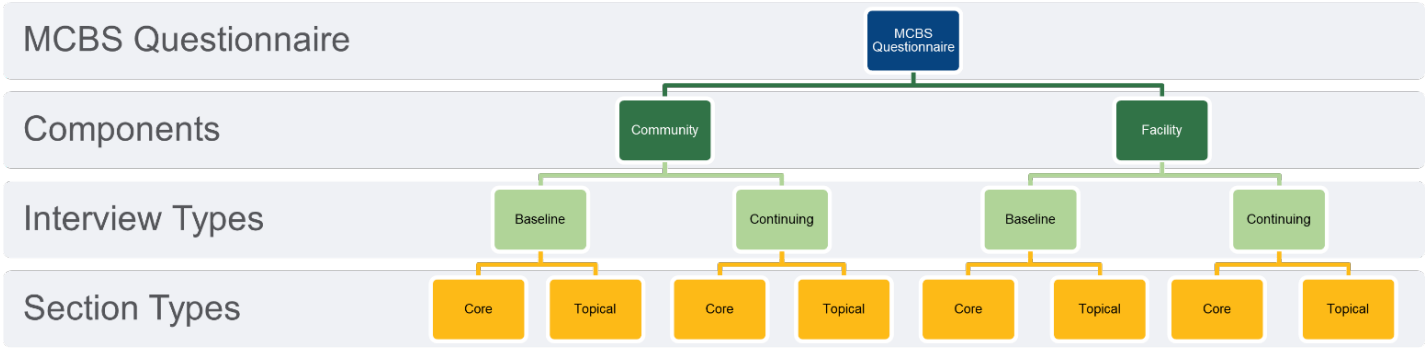
Coverage Analysis. The fourth enrollment data subsample extract, along with the first three extracts, are used to fully enumerate both the enrollment data subsample and the MCBS population. To construct the full enrollment data subsample, all records of eligible beneficiaries enrolled through December 31 of the calendar year from the four extracts are combined. From that universe, the MCBS population is 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 the selection year, 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 MCBS population includes all beneficiaries who were enrolled in Medicare in the calendar year and reside in the MCBS PSUs and SSUs.

3. 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 3.1 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 3.1: 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 5.2.

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 age, 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 6.2 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 *MCBS Questionnaire User Guide* for tables of the Core and Topical sections for each data collection year.

Data collected by the Community and Facility components are released to users via three primary LDS – the Survey File - Early Release, 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 Survey File - Early Release includes a subset of data released on the Survey File that were collected via Core and Topical sections in the fall round. 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.

The MCBS questionnaire contains content from a variety of sources that are adapted for inclusion in the MCBS. Some questionnaire items on the MCBS come from validated scales that were developed by external researchers and tested for reliability and validity. Two examples of such scales are the Generalized Anxiety Disorder Scale (GAD-2), which is a screening tool for generalized anxiety disorder used by the MCBS Community Questionnaire and the Patient Health Questionnaire (PHQ-9), which is a screening tool for depression used by the MCBS Community Questionnaire and MCBS Facility Instrument. See the *MCBS Questionnaires* for more information.

3.1 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.

The content of the MCBS Community Questionnaire consists of Core and Topical sections. Core survey content is grouped into questionnaire sections that collect data central to the policy goals of CMS, such as information related to socio-demographics, health insurance coverage, health care utilization and costs, beneficiary health status, and experiences with care, as well as operational and procedural data. The questionnaire sections in each of these categories may be asked each round or seasonally (fall, winter, summer). Data from these questionnaire sections are found the Survey File - Early Release, Survey File, and Cost Supplement File. In addition to the Core content, there are several Topical questionnaire sections that capture data on a variety of key topics, such as beneficiary's housing characteristics, health behaviors, knowledge about Medicare, and health-related decision making. Data from the Topical sections are included in the Survey File - Early Release and Survey File data release. Annually, special non-response adjustment weights are included within certain Survey File - Early Release and Survey File segments for use in analysis when data are not collected within the same calendar year or are only collected from non-proxy respondents. This applies to the data collected in many of the Topical questionnaire sections. For information on the specific Topical segments released each year, see the *Data Year Release Notes*.

Different combinations of Core and Topical sections are used depending on a number of criteria, including interview type (Baseline vs. Continuing); the season of data collection (fall, winter, summer); whether the beneficiary is alive, deceased, or in a facility; and whether the interview is being completed with the beneficiary or a proxy.

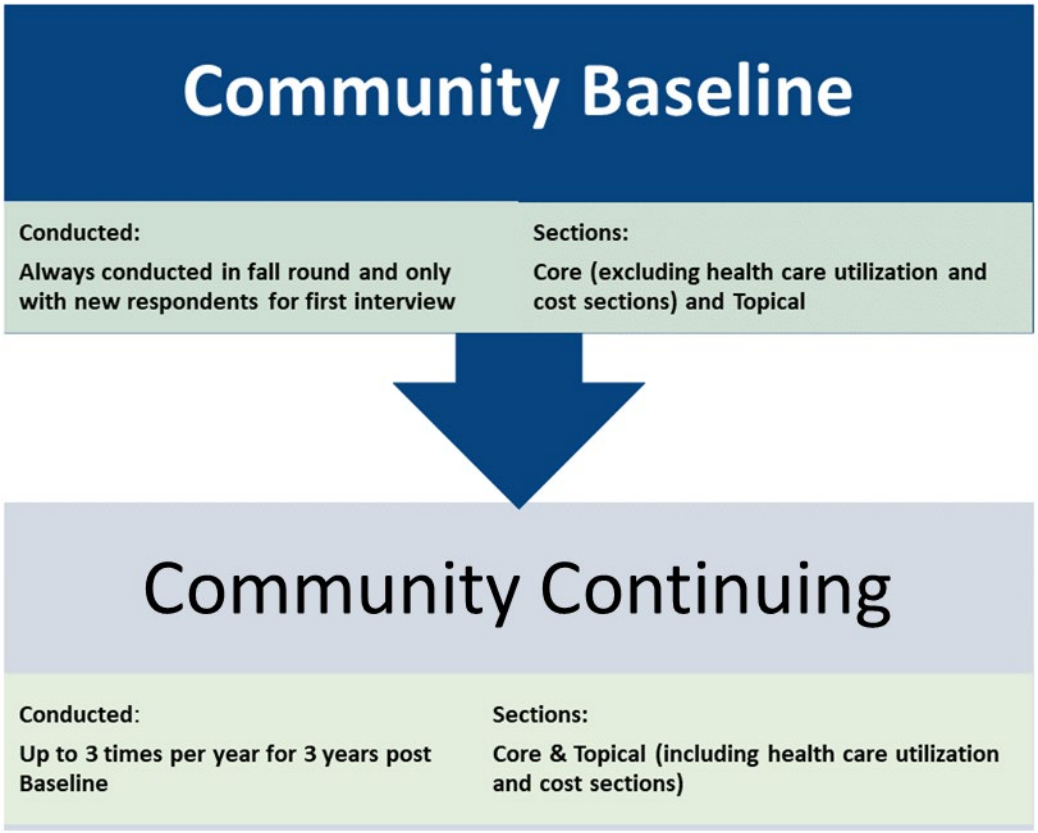
The first Community interview conducted with Incoming Panel respondents is referred to as the Baseline interview. This interview is always conducted in the fall round and consists of a combination of Core and Topical sections. It is important to note that this first interview does not include Core sections that collect health care utilization and cost data. The respondent's 2nd through 11th interviews, also known as the Continuing interviews, consist of Core and Topical sections, including those that collect health care utilization and cost data; these interviews provide three calendar years of reported health care utilization and cost data for each beneficiary.

The Community Questionnaire consists of the following components (see Exhibit 3.1.1):

- **Community Baseline interview**
- **Community Continuing interview**

In addition to including data collected in the three rounds (winter, summer, and fall) administered during the calendar year (i.e., January through December of the data collection year), some data collected in the previous and following calendar years are also included in the LDS. Specifically, some data collected in the previous data collection year are carried forward to fill in data when questionnaire items are administered only once or when data are missing for the data year but valid values exist for the previous year. Some data are also collected in winter and summer of the following data collection year and are "pulled back" for inclusion in the LDS because the section's reference period extends back to the current data collection year.

Exhibit 3.1.1: Overview of the MCBS Community Questionnaire Components



For more information about Community Questionnaire content, including descriptions of each questionnaire section, see the annual *MCBS Questionnaire User Guide.*

3.1.1 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 3.1.2 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 3.1.2: 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.

3.1.2 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. Other 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.

As the first interview conducted, the Baseline interview provides an opportunity for the field interviewer to develop a strong rapport and connection with the respondent, acquaint the respondent with the intent of the survey, and emphasize the importance of keeping accurate records of medical care and expenses. Whenever possible, field interviewers are assigned to the same beneficiary over the course of their participation in the survey, so establishing a positive relationship is critical during the Baseline interview. The Baseline interview contains an abbreviated flow which does not include the utilization or cost sections of the questionnaire. As such, in the data collection year, only persons in the Continuing Panels received the Core sections about health care utilization and health care costs. All panels received the health insurance section.

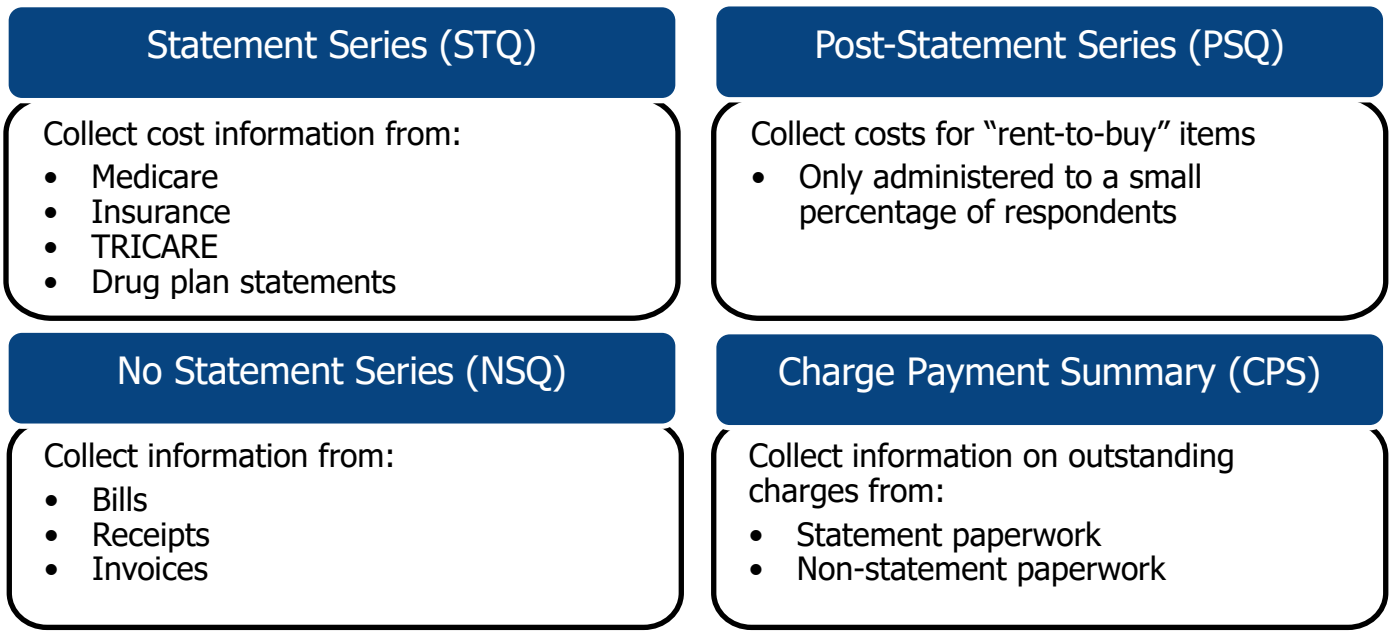
The Continuing interview consists of Core sections that focus on the use of medical services and the resulting costs; these sections are asked in essentially the same way each and every time they are administered. The respondent is asked about new health events and to complete any partial information that was collected in the last interview. For example, the respondent may mention a doctor visit during the health care "utilization" part of the interview. In the "cost" section, the field interviewer will ask if there are any receipts or statements from the visit. If the answer is "yes", the field interviewer will record information about the costs from those statements, but if the answer is "no," the question will be stored until the next interview. The Continuing interview also includes sections about health insurance. During each interview, the respondent is asked to verify ongoing health insurance coverage and to report any new health insurance plans. Continuing interviews also include Topical sections which cover subjects such as mobility and drug coverage.

The Community Questionnaire flow varies based on fielding and operational factors. Annually, the MCBS produces the *Questionnaire User Guide*, which depicts the questionnaire flow for the data collection year and provides detailed descriptions of each questionnaire section. Note that the data year includes surveys administered in other years which may have slightly different questionnaire flows but are included in the data year LDS given the reference period. As such, users may need to look across several years of *Questionnaire User Guide* to understand the questionnaire flow that corresponds to a given data year.

3.1.3 Cost Series

Once all utilization sections are completed, the Community Questionnaire flows to the cost series, wherein the costs of all reported visits and purchases are recorded, along with the amount paid by various sources. Importantly, additional visits and purchases not reported in the utilization sections of the questionnaire could be recorded within the cost series, and all corresponding data for those events are collected within the cost series. The cost series consists of four sections: Statement, Post-Statement, No Statement, and Charge Payment Summary. Each is summarized in Exhibit 3.1.3 and described below.

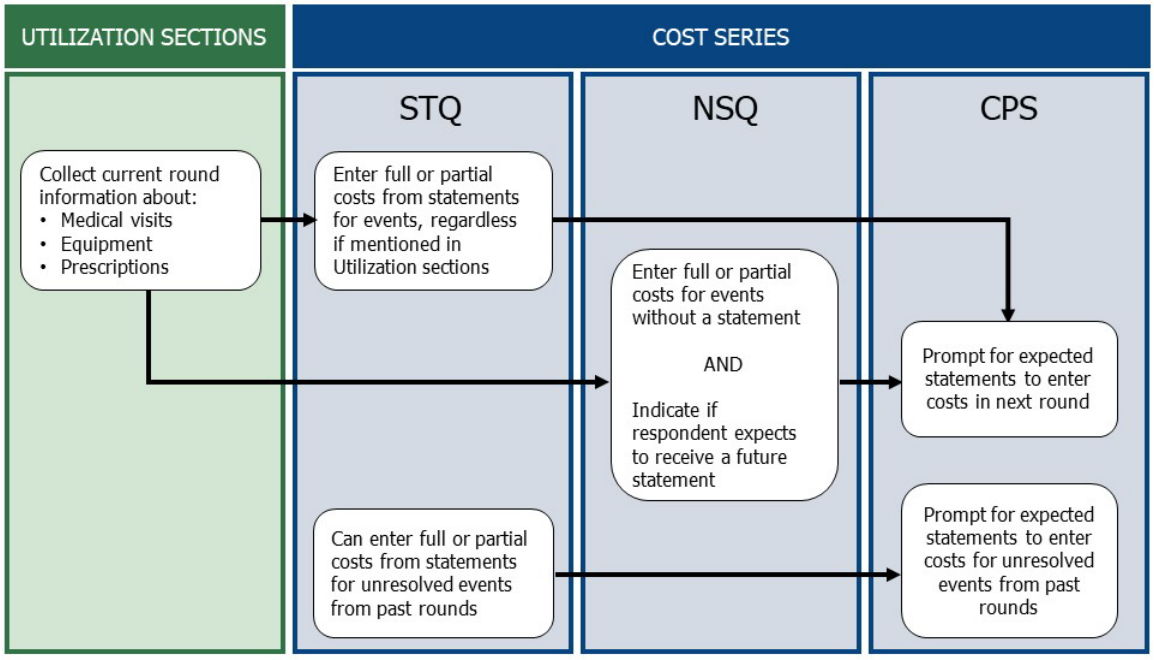
Exhibit 3.1.3: Cost Series Section Overview



The flow of sections and questions within the cost series varies depending on data collected in the current round (e.g., whether the beneficiary had a health insurance statement for a visit reported in the current round) and data collected in prior rounds (i.e., whether there was outstanding cost information reported from

a prior round). Exhibit 3.1.4 illustrates how paths through these sections may vary depending on health care utilization and cost information collected in the current and previous rounds.

Exhibit 3.1.4: Utilization and Cost Section Flow



Costs are considered unresolved when full cost information is not collected due to events being reported 1) without any cost or payment information, 2) with an indication that a statement is expected, so follow-up questions about costs and payments are deferred until the next interview, or 3) with partial information about costs or payments, but there is a remaining dollar amount with pending payment information.

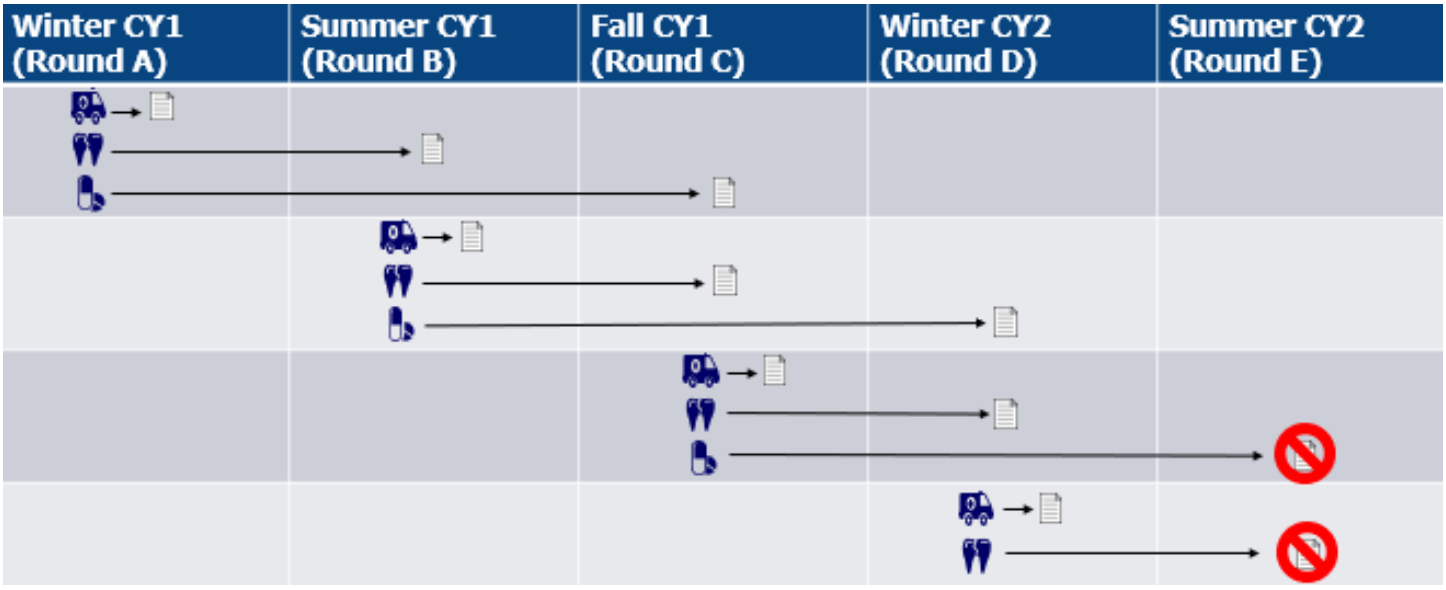
The current MCBS protocol allows for cost resolution attempts up to two rounds later than the events were reported. Exhibit 3.1.5 displays sample paths to resolving cost information. The first row displays a hospital event reported with costs and statement. This cost is resolved within the round. The second row displays a resolved dental event reported in the summer round with the statement provided in the fall round. The third row displays a prescription medicine event reported with a statement in the summer and resolved in the winter after the statement was provided. The final row displays an unresolved event that was reported in the summer round but did not receive cost or statement verification.

Exhibit 3.1.5: Example Paths Toward Cost Resolution

Scenario	Summer Round	Fall Round	Winter Round	Cost Status
Event reported with costs, statement available				**Resolved**
Event reported without costs, awaiting statement				**Resolved**
Event reported with receipt, awaiting statement				**Resolved**
Event reported without costs, statement not received				**Unresolved**

Each data year typically includes events collected from winter of the data collection year through summer of the following data collection year (see Exhibit 3.1.6). The unresolved costs are indicated with a red circle-backslash symbol and are unresolved given that the statement was not received.

Exhibit 3.1.6: Events Collected in the Data Year



3.2 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.

In addition to collecting information from respondents living in the community, the MCBS collects information at the institutional level if the beneficiary is living in a facility at the time of the interview. Information is obtained only by interviewing Facility staff; the beneficiary is never interviewed directly.

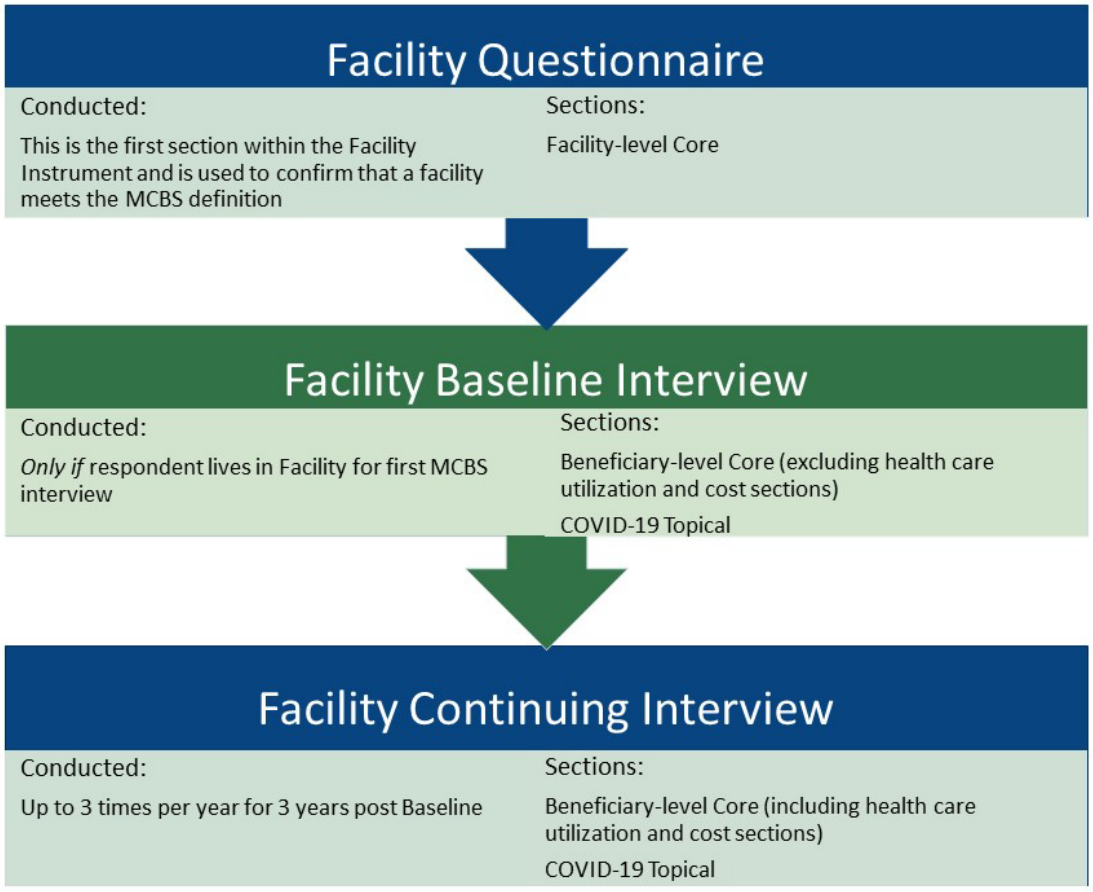
The Facility Instrument consists of the following components (see Exhibit 3.2.1):

- **Facility Questionnaire**
- **Facility Baseline interview**
- **Facility Continuing interview**

Similar to the Community Questionnaire, if a beneficiary is living in a facility when first selected to participate in the MCBS, a Facility Baseline interview is administered. For cases in the 2nd through 11th round, a Facility Continuing interview is conducted. Administration of the Facility Instrument sections varies by season and interview type. The Facility Instrument is comprised of Core sections that collect information that is considered of critical importance to the MCBS and parallel the Core sections for the Community component, as well as a Topical section that captures data on COVID-19 vaccination. Data from the Facility Instrument are found in the Survey File and Cost Supplement File; Facility data are not released on the Survey File - Early Release.

If a person living in a facility returns to the community, that person would receive the Community Questionnaire. If the beneficiary spent part of the reference period in the community and part in a facility, then a separate interview is conducted to collect information pertaining to the beneficiary's experiences covering each distinct period of time. In this way, a beneficiary is followed in and out of facilities and a continuous record is maintained regardless of the location of the beneficiary.

Exhibit 3.2.1: Overview of the MCBS Facility Instrument



For more information about Facility Instrument content, including descriptions of each questionnaire section, see the annual *MCBS Questionnaire User Guide.*

3.2.1 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 3.2.2.

Exhibit 3.2.2: 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.

3.2.2 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.

3.2.3 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). The Baseline interview, administered to Incoming Panel respondents, contains an abbreviated flow, which does not include the utilization or cost sections of the questionnaire. The Facility Baseline interview serves as a reference interview and gathers information on the facility itself as well as the health status, insurance coverage, residence history, and demographic information for the beneficiary.

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 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 section at the end of the interview. Note that beneficiaries who move to a facility from the community (Community to Facility cases), move to a new facility (Facility to Facility cases), or move to the community from the facility (Facility to Community cases) receive a different combination of Facility Continuing sections than beneficiaries who have lived continuously in the same facility.

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 Health Status 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 6.1.2. Facilities that do not report a CCN receive the full MCBS Facility Instrument.

3.3 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.

3.3.1 Community Questionnaire

The MCBS Community Questionnaire is 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.

3.3.2 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 5.2 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.[22]

3.3.3 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 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.

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

3.4 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.

4. INTERVIEWER RECRUITMENT AND TRAINING

4.1 Interviewer Recruitment and Staffing

Each professional interviewer staff is required to complete interviews throughout the year. Annual hiring is targeted based on staffing needs and MCBS-specific skill requirements. The set of preferred skills includes experience with financial data and complex surveys; language skills; working with individuals who have hearing, visual, or cognitive challenges; and interviewing people with disabilities and the elderly.

4.2 Interviewer Training Programs

Nationally, the MCBS employs an average of approximately 170[23] field interviewers, who participate in a combination of targeted training initiatives and careful coaching and monitoring activities throughout data collection.

The MCBS Training Program consists of remote and in-person trainings which vary based on the experience of the interviewer (new to MCBS or MCBS-experienced), the 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 is structured to provide all field staff the same training content, ensuring that the performance of data collection responsibilities is standardized, methodical, and measurable.

Remote trainings target MCBS-experienced interviewers in advance of each round of data collection. New staff are onboarded remotely each fall round. NORC leverages 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 includes question-and-answer sessions, gaining cooperation role playing, and protocol demonstrations, and screen-sharing to facilitate real-time feedback as the trainer or interviewer navigates the case management system or questionnaire. NORC also incorporates e-learning technology to develop high-quality, responsive content grounded in adult-learning theory. For example, NORC leverages the Articulate suite of e-learning software to program and deliver highly interactive trainings with software simulation activities. Later in the fall round, interviewers are trained in-person on the Continuing interview. This multi-day training focuses on the essential skills and protocols for 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 formal trainings, throughout data collection the MCBS Training Program emphasizes 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 works with field managers to ensure interviewers receive additional training via weekly field memos, interviewer group call sessions, and interviewer observations, referred to as "ride-alongs" or "call-alongs." These methods cover important data collection tips, provide answers to interviewer questions, and offer 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.

[23] 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.

5. 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).

5.1 Clearance

5.1.1 OMB Approval

CMS maintains a current OMB clearance for the MCBS (OMB control number 0938-0568). This typically requires annual revisions to the OMB clearance package to obtain approval for changes to the questionnaires or respondent materials.

5.1.2 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.

5.2 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.

5.2.1 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.

5.2.2 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 6: 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 3.3 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.

5.2.3 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 is finalized as deceased without pursuing a proxy interview. Any Incoming Panel beneficiary reported as deceased who enrolled in Medicare during the same year 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 who have moved after their first interview, 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 a 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 who are unavailable in the current 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 without giving a hard refusal. Cases meeting similar criteria are finalized as Unavailable this Round and are staged for fielding in the following round.

5.2.4 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 11 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, conduct the interview in the correct setting, and 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 Medicare 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 Section 3.1.3 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 Individuals with Intellectual Disabilities; 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 COVID-19 pandemic, the Facility interview was no longer administered in-person and instead administered by phone starting in March 2020 through 2021. Beginning in 2022, Facility interviews could be administered in-person again; however, the majority of Facility interviews continue to be administered by phone. 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 is 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.

5.2.5 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.[24] 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.

Survey administration of Incoming Panel cases in the winter and summer rounds follows a different protocol that depends on when the beneficiary entered the new component and when they gained Medicare entitlement. All other crossover cases in their 3rd through 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

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

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. Starting in Winter 2025, interviewers still follow this same process, but the Community interview is made available 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. Starting in Winter 2025, interviewers still follow this same process, but a new Facility interview can only be administered within the same round if the Facility interview at the original facility is not administered within the current round. Otherwise, the new Facility interview is made available in the following round.

5.2.6 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. Less than 15 percent of interviews have proxy usage from year to year.

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 they 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. Less than 15 percent of interviews have assistant usage from year to year.

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 3.1 for information about the Community Questionnaire and the *Questionnaire User Guide* for a description of Introduction Questionnaire).

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.

5.2.7 Interviewing Languages

The Community Questionnaire is programmed for administration in English or Spanish. The Facility Instrument is available for administration in English.

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.

5.2.8 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, the interview is finalized as a Final Breakoff at the end of the round (see Section 7 for more information on weighting and imputation procedures).

5.3 Data Collection and Quality Control

An interview is complete once administration of all questionnaire sections to the respondent has concluded, the Interviewer Remarks section is completed, and data are fully transmitted.

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 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. The data review procedures consist 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

6.1). 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 are used to monitor data quality.

The relationship between data collection mode and data quality is also examined regularly. Following the MCBS transition from in-person interviewing to phone interviewing in response to the COVID-19 pandemic in 2020, early investigations found stability in the majority of questionnaire measures but revealed challenges collecting information in the cost sections by phone, particularly details from physical documentation (e.g., insurance statements). These impacts have since been mitigated through tailored interviewer training and respondent coaching, using model-based strategies to appropriately prioritize in-person interviews for beneficiaries most likely to benefit from them (e.g., beneficiaries who have more cost information to report or are more likely to struggle to do so over the phone), as well as statistical techniques including imputation, claims matching, and ratio adjustments (see Section 7.3). Mode effects cannot be completely eliminated in any multimode data collection effort, but ongoing analyses suggest that overall, these shifts are relatively small and do not impact sample representativeness enough to be of concern.

6. 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.

6.1 Data Processing Overview

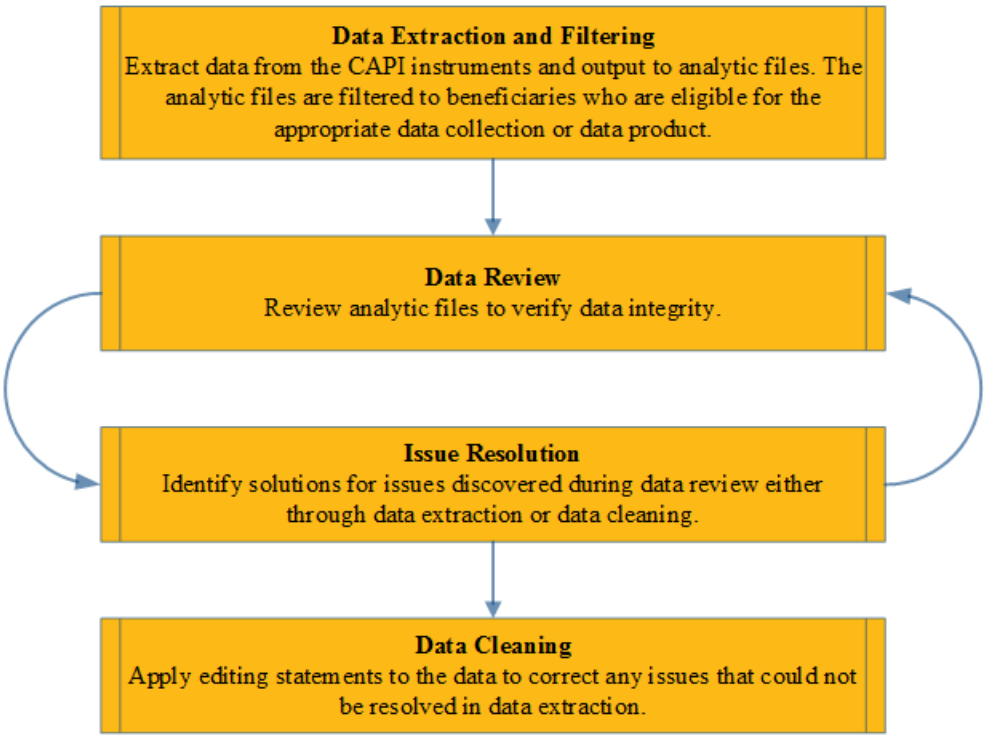
During the interview, respondents may provide information that is either incomplete or inconsistent with previously reported survey responses or administrative records. These data require further processing to ensure the highest quality of estimates produced from the MCBS. The processing may involve 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 Survey File and Survey File - Early Release, and the 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 6.1.1. For Facility data only, additional processing steps (described in Section 6.1.2) are applied to populate certain variables using administrative data.

6.1.1 Process Description

Exhibit 6.1.1 illustrates the steps and iterative nature of the data review and editing process.

Exhibit 6.1.1: 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 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. For variables that ask about conditions that cannot change after diagnosis, such as Alzheimer's, once an affirmative response is given, respondents are not asked again. For conditions that can change after diagnosis or can be reoccurring, such as high blood pressure, respondents are asked annually thereafter if they had that specific condition or experience in the past year. 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.

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 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).

6.1.2 Additional Facility Data Processing

As noted in Section 3.2.3, a subset of items in the FQ and Health Status 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.

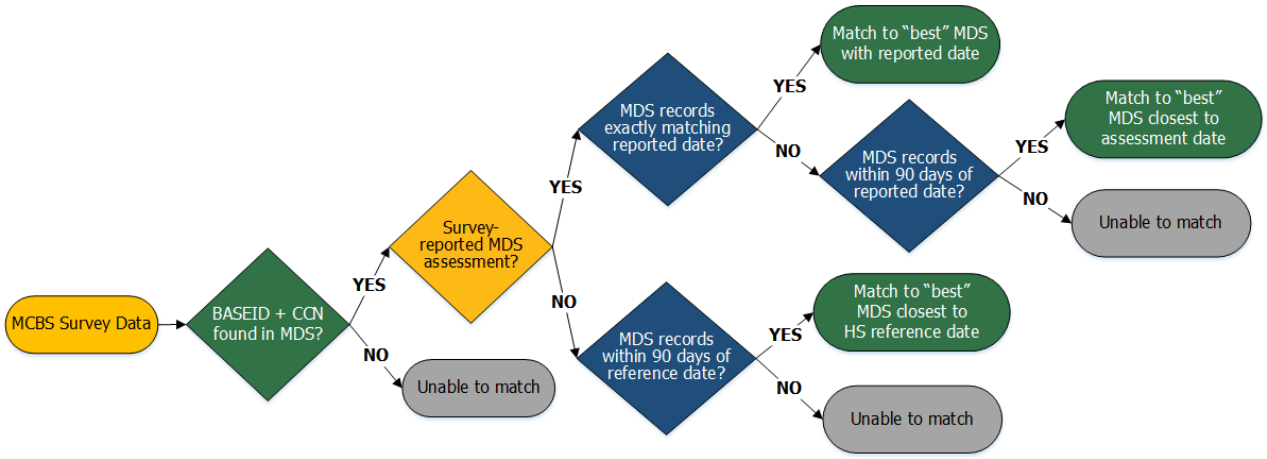
FQ-CASPER Data Linkage Process. For beneficiaries living in a facility for which the Facility respondent reported a CCN, the analytic file storing data collected in the FQ section can be linked directly to CASPER data on CCN. The CCN reported in the Facility Instrument is matched to the corresponding 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.[25] 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 6.1.2.

[25] 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 6.1.2: 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.[26] 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.

The majority of MDS matches are identified through an exact date match between the survey-reported and MDS assessment dates, representing approximately 79-85 percent of all matches. Non-exact date matches account for 9-12 percent, while 3-5 percent are matched using the HS survey reference date. Unmatched cases make up a minimal portion, ranging from 2-4 percent.

6.2 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

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

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 Section 3.1.3 and the *MCBS Questionnaire User Guide*.

6.2.1 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 6.1. 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 need to be in the specified format acceptable to the case management system, which then makes the preload data available to be loaded into the Community Questionnaire and/or Facility Instrument for the upcoming round.

6.2.2 Which Community Data Are Included in the MCBS LDS?

Community data that are incorporated into the Survey File - Early Release LDS, Survey File LDS, and Cost Supplement LDS for sampled beneficiaries eligible for Medicare in a given calendar 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 each data year corresponds to one calendar year, data included in the LDS are collected over three data collection years (i.e., the prior data collection year, the current data collection year (which matches the data year), and the following data collection year). Some data collected in the previous data collection year are pulled forward to fill in data for the current data collection year. This happens when questionnaire items are administered only once (such as demographics) or when data are missing for the data collection year, but valid values exist for the previous data collection year. However, most data are collected during and after the data collection 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. Exhibit 6.2.1 provides an example of the data that contribute to the three LDS's for a given data year (in this case, Calendar Year 1). There are four panels involved in data collection each year: one Incoming Panel (selected in the current calendar year), two Continuing Panels (selected in the two previous calendar years), and one exit panel (selected three calendar years prior). The rounds of data collection that fall within the data year are the winter, summer, and fall rounds of the current data collection year, with additional data collected in the winter and summer rounds of the following data collection year.

The Survey File - Early Release LDS consists of questionnaire items collected in the Community Core and Topical questionnaire sections during the fall round of the current data collection year, covering a subset of the data year. Members of the Continuing and exit panels have some of their data, such as demographics, pulled forward from previous rounds, but the majority of the data released on the Survey File - Early Release LDS were collected in the single fall round. Data collected during winter and summer of the following data collection year are not included in the Survey File - Early Release.

The Survey File LDS consists of questionnaire items collected in the Community Core questionnaire sections as well as items collected in the Community Topical questionnaire sections. The Core Survey File data are collected in summer and fall of the data collection year and in winter of the following year. The Topical Survey File data are collected in fall of the data collection year and winter and summer of the following year. Each round's interview is based on reference periods, which extend from the time of the previous interview.

- For example, the Health Insurance Questionnaire (HIQ) is a Core questionnaire section that asks about changes to insurance plans during the reference period. In the fall interview, this period would cover the time since completion of the summer interview, while in winter of the following year, it would cover the time since completion of the fall interview, meaning the reference period extends back into the prior data collection year (i.e., the current data year).
- A reference period may also cover the entire data collection year. For example, the Drug Coverage Questionnaire (RXQ) is a Topical questionnaire section administered in the summer of the following data collection year, but it collects beneficiaries' prescription drug coverage information for the prior data collection year (i.e., the current data year). Beneficiaries in the Incoming Panel provide current data collection year Survey File LDS data through participation in their first and second interviews in fall of the data collection year and winter of the following data collection year and provided additional Topical data in the summer of the following data collection year.

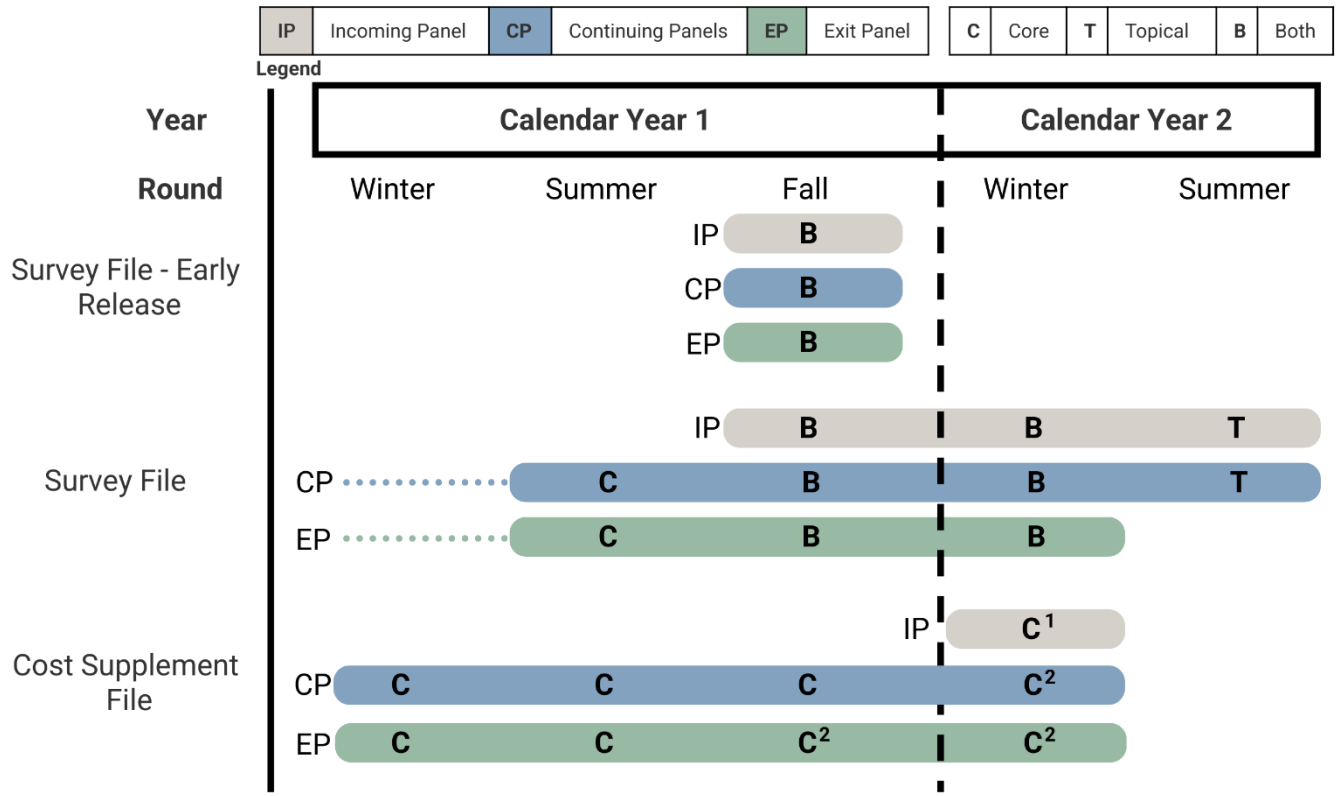
Members of the Continuing Panels have some of their data pulled forward from previous rounds but provide most of their data through participation in summer and fall of the current data collection year and winter and

summer of the following data collection year. Members of the exit panel likewise have data pulled forward from previous rounds and are interviewed in summer and fall of the current data year and winter of the following data collection year but are not interviewed in summer of the following data year.

The Cost Supplement File LDS consists of utilization and cost data for each data year. These data are collected from the four rounds that can have reference periods covering any part of the data year; for each data year, this includes the rounds from winter of the current data collection year through winter of the following data collection year. 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 File data until their second interview in the winter following the data collection year. For beneficiaries who are current-year enrollees, meaning beneficiaries who enrolled in Medicare during the calendar year in which they are sampled, the winter round reference period extends back to the date of completion of the fall round, thus collecting utilization and cost data for the latter part of the data collection year. Members of the Incoming Panel who enrolled prior to the calendar year in which they are sampled have a winter reference period that begins on January 1 of the following data collection year and only provides Cost Supplement File data for that data year and later. The Cost Supplement File data for the Continuing and exit panels are collected through participation in the rounds from winter of the current data collection year through winter of the following data collection year. The exit panel exits the survey in winter of the following data collection year and has a reference period that ended December 31 of the data year.

Exhibit 6.2.1: Rounds of Data Collection and Reference Periods for Community Data Included in the Limited Data Sets*



NOTES: IP stands for Incoming Panel. CP stands for Continuing Panels. EP stands for Exit Panel. C stands for Core questionnaire sections. T stands for Topical questionnaire sections. B stands for both Core and Topical questionnaire sections. Data in this table were collected in CY1 and/or CY2 and included in the CY1 LDS release. Cost Supplement File data reflect the data year of interest (i.e.,

1/1/CY1 – 12/31/CY1). In other words, the data included in the CY1 Cost Supplement LDS are based on survey-reported information within the year of interest, not rounds of data collection. In contrast, for the CY1 Survey File LDS, data were primarily collected in summer and fall of CY1 and winter of CY2, although data from previous rounds could be pulled forward (represented by the dotted line). Data collected in winter of CY2 are included in the CY1 Survey File LDS if the survey items ask about information since the date of the beneficiary's last fall interview in CY1 or the reference period is between 1/1/CY1 – 12/31/CY1. Some Topical data collected in summer of CY2 are included with the CY1 LDS given the reference period is between 1/1/CY1 – 12/31/CY1. The CY1 Survey File - Early Release LDS contains data collected in fall of CY1 only. Note, additional data not shown in this exhibit may be collected during these rounds that contribute to the prior or following data years' LDS's.

[1] Includes Cost Supplement File data through 12/31/CY1 for current-year enrollees on the Incoming Panel. For other Incoming Panel members, data are collected only from 1/1/CY2.

[2] Includes Cost Supplement File data through 12/31/CY1.

6.2.3 Which Facility Data Are Included in the MCBS LDS?

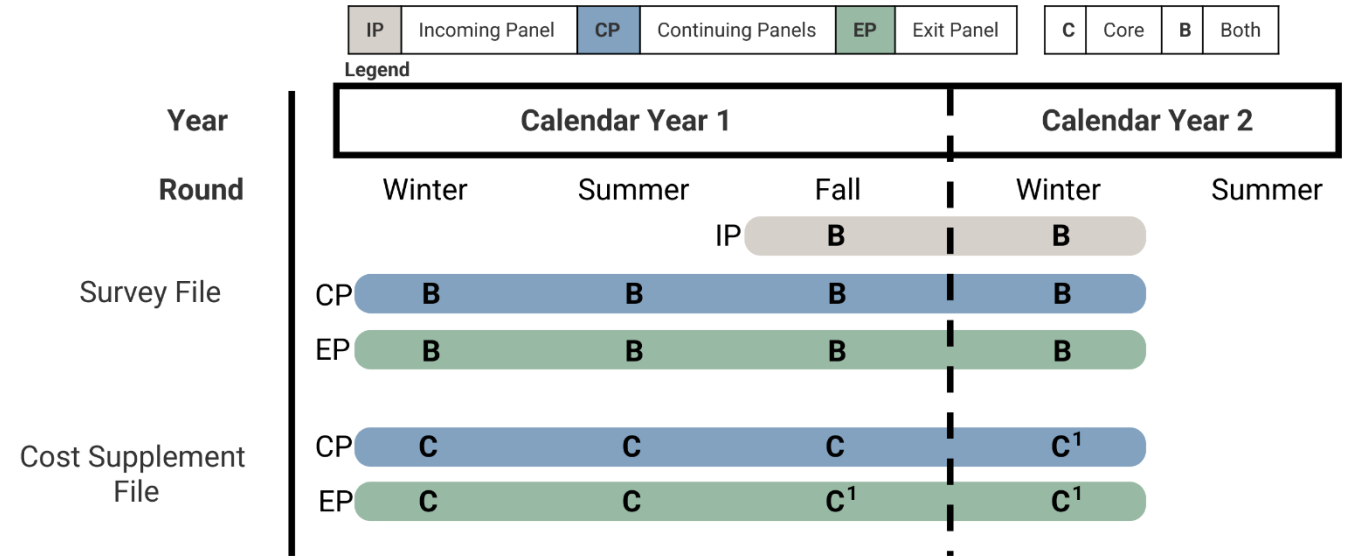
Facility data that are incorporated into the Survey File LDS and Cost Supplement File LDS for sampled beneficiaries eligible for Medicare in a given calendar year depend on a variety of factors, including beneficiary panel type, the round of data collection, and type of questionnaire item. Facility data are not included in the Survey File - Early Release LDS. As with Community data, some Facility data collected in a previous year are pulled forward to fill in data for the current data collection year. This happens when questionnaire items are administered only once or when data are missing for the data collection year, but valid values exist from a previous year. However, most data are collected during and after the data collection year.

The Survey File LDS Facility data are collected in winter, summer, and fall of the data collection year, and winter of the following data collection year as part of the Facility Core questionnaire sections and Facility Topical questionnaire sections.

The Cost Supplement File LDS consists of Facility utilization and cost data for the data year. These data are collected from the four rounds that can have reference periods covering any part of the data year; for each data year, this includes the rounds from winter of the current data collection year and winter of the following data collection year. 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 File data. The Cost Supplement File data for the Continuing Panel is collected through participation in the rounds from winter of the current data collection year through winter of the following data collection year. The exit panel exits the survey in winter of the following data collection year and have a reference period that ends December 31 of the data year.

Exhibit 6.2.2: Rounds of Data Collection and Reference Periods for Facility Data Included in the Limited Data Sets*



NOTES: IP stands for Incoming Panel. CP stands for Continuing Panels. EP stands for Exit Panel. C stands for Core questionnaire sections. B stands for both Core and Topical questionnaire sections. Data in this table were collected in CY1 and/or CY2 and included in the CY1 LDS. Note, additional data not shown in this exhibit may be collected during these rounds that contribute to the prior or following data years' LDS's.
[1] Includes Cost Supplement File data through 12/31/CY1.

6.3 MCBS Survey File - Early Release

6.3.1 File Eligibility Criteria

The Survey File - Early Release contains a subset of Core and Topical questionnaire data collected in the fall of the data year. The inclusion criteria for these analytic files include beneficiaries living in the Community who are alive, enrolled in Medicare, and completed a fall round interview in the Community component. Specific files have additional criteria that a case needs to meet for inclusion. For example, certain segments do not contain interviews conducted with proxy respondents.

6.3.2 File Contents

The Survey File - Early Release data contain core information such as health status and functioning, household composition, access to care, and demographics, as well as special interest topics such as patient activation. The Survey File - Early Release does not include survey-reported cost, health care utilization, or case management data.

6.4 MCBS Survey File

6.4.1 File Eligibility Criteria

The Core Survey File data are collected in winter, summer, and fall of the data year. The Topical questionnaire data are collected in fall of the current data year and winter and summer of the following data year. 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 to be included in the file.

6.4.2 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 Year Release Notes* or the historical *Data User's Guide* 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. Each year, the Survey File contains data from some sections that were administered in the winter and summer of the following data year but have reference periods for the current data year. These files are processed in combination with the current data year Survey File deliveries and as a result, Topical analytic files are considered part of the current year MCBS 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 Year Release Notes* or the historical *Data User's Guides* 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.

6.4.3 Reference Period

Reference Period is a data editing process that uses case management data to define time periods in the data year 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 the data year 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.

6.4.4 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 for each data year is 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 Health Maintenance Organization (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 Survey File LDS.

6.5 MCBS Cost Supplement File

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

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.

6.5.1 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.

6.5.2 Prescription Medicine File

The PMED file is a list of all prescription medicines that are collected by the MCBS. The list includes every combination of prescription medicine names, forms, and strengths provided by MCBS respondents during interviews conducted in the data year (including a total of four rounds between winter of the current data year and winter of the following data year). The process of creating the PMED file includes assembling a full list of all beneficiaries' reported prescribed medicines for the data year 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.

6.5.3 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. 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 includes 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 the data year.

7. WEIGHTING AND IMPUTATION

7.1 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.

7.2 MCBS Weighting Procedures

7.2.1 Overview

Weighting activities for each 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 of the current data year). 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. Throughout this section, the current data year is denoted year t. The specific weights released on each MCBS LDS are documented in the corresponding *Data Year Release Notes* and historical *Data User's Guides*.

Weights for the Survey File - Early Release follow an abridged process that uses data available at the time of their production. These data may change for a small number of beneficiaries after the production of these weights, and not all data that are used in the final weighting processes are available for the Survey File - Early Release weights. However, production of these weights follows the same steps as the Initial and Survey File weighting processes described below; notes related to differences in the Survey File - Early Release weighting process are found in each of these sections.

7.2.2 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 year t. 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 are weighted separately as they are fielded in the winter and summer rounds following the data year or are only fielded to non-proxy respondents.

7.2.3 Initial Weighting

In the initial weighting stage, the initial nonresponse adjusted weights for the Incoming Panel of Medicare beneficiaries 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 of year t.

Full-sample and Replicate Raked Base Weights. A full-sample base weight is derived for all beneficiaries in the Incoming 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 $\pi k \mid i\, j$ be the conditional probability of selection for beneficiary k given the PSU i and the SSU j, such that $\pi k \mid i\, j = \rho 1 ak \mid i\, j$ for beneficiaries in the Hispanic sampling stratum and age group a, and similarly equals ρ-$1ak \mid i\, j$ 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

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

where ni is the probability of selection for the PSU, $\pi j \mid i$ is the conditional probability of selection for the j-th SSU given the PSU, and $\pi k \mid i\, j$ is the conditional probability of selection for the k-th beneficiary in the 5-percent enrollment data extract given the PSU and SSU.

Then, one hundred replicate base weights are derived from the full sample base weights, using the variance stratum and the variance unit of the beneficiary. The variance strata and variance units are derived from the PSUs and SSUs used for sampling. For sampled beneficiary ijk as described above, the α = 1, …, 100 replicate weights for balanced repeated replication (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

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

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 τ = 0.85 is used, continuing the practice used in prior MCBS years.

Base weights among the Incoming Panel for the Survey File - Early Release weights are calculated in exactly the same manner as described as above.

The full-sample and replicate base weights are then adjusted in such a way that the sum of the weights for various demographic domains is 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 of year t-1 contains additional records for beneficiaries who became eligible near the end of the year t. Due to the timing of this file, these newly-added beneficiaries are not subjected to sampling and could not be included in the Incoming 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 Incoming 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.

Base weights for the Survey File - Early Release are not raked to population control totals; this step is skipped for the Early Release weights.

Initial Nonresponse Adjustments. The raked base weights for the Incoming Panel are then adjusted for nonresponse at the first interview in fall of year t. The response statuses in fall are determined, where a respondent is a beneficiary that is alive and entitled and completed the fall interview in year t. 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 of year t 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 Incoming Panel.

For the Survey File - Early Release weights, the nonresponse adjustment follows the same process as described here, but the data that are available for building the response propensity models are more limited and include only demographic information from the enrollment database as well as administrative and claims data. These variables are the most impactful in the final weighting process and perform well at identifying differences between respondents and nonrespondents.

7.2.4 Survey File Weights

Data for the Survey File in year *t* are collected in summer and fall of year *t* from beneficiaries sampled in the last four annual panels (including the Incoming Panel). To facilitate estimation from the resulting data, five sets of full-sample and replicate weights are derived. These include continuously enrolled cross-sectional weights for year *t*; 2-year longitudinal weights; 3-year longitudinal weights; 4-year longitudinal weights; and finally, the ever enrolled weights for year *t*. In addition to the weights, the dataset includes the panel (calendar 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 last four panels (i.e., the three Continuing Panels and the Incoming Panel). 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.

If the Survey File continuously enrolled cross-sectional weights are populated for the subset of records with a completed interview in the fall round of year *t* 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 through the fall round of year *t*. 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 year *t-3*, year *t-2*, and year *t-1* Panels that were continuously enrolled in both year *t-1* and year *t*. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before January 1 of year *t-1* and are still alive and entitled as of the fall round of year *t*. The three-year longitudinal weights are populated only for members of the year *t-3* and year *t-2* Panels who were continuously enrolled in each of the years *t-2*, *t-1*, and *t*. The population represented by these weights is the population of beneficiaries enrolled on or before January 1 of year *t-2*, and surviving and entitled as of the fall round of year *t*. Finally, the four-year longitudinal weights are populated only for members of the year *t-3* Panel who were continuously enrolled during the years *t-3*, *t-2*, *t-1*, and *t*. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before January 1 of year *t-3*, and are still alive and entitled as of the fall round of year *t*.

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

The Survey File - Early Release weights are ever-enrolled weights that represent the population of beneficiaries who were ever enrolled for at least one data in year *t* and were still enrolled and living in the community in the fall of the same year.

Fall Nonresponse Adjustment for Continuously Enrolled Weights

Sample from the three Continuing Panels are adjusted for nonresponse through fall of year *t*. The process begins with weights for these panels that were previously adjusted through fall of year *t-1*. Response status in winter, summer, and fall of year *t* is then identified, where a respondent is a beneficiary that was alive and

entitled with a complete fall interview in year t, or who died or lost entitlement at some time prior to the fall of year t 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 nonrespondents in the fall of year t-1.

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 of year t as a function of covariates derived from the Survey File data from the fall of year t-1. 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 the fall of year t.

Derivation of the Continuously Enrolled Weights

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

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

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

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

Where n_{pd}^{eff} is the effective number of fall completes in year t 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

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

where n_{pd}^{act} is the actual number of completed interviews, $\bar{w}_{id}$ is the average of the fall adjusted weights for the panel in year t, and S_{id} is the standard deviation of these weights.

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

Longitudinal Weights for the Survey File.[27] The derivation of two-year longitudinal weights begins with the weights adjusted through fall of year *t* from the three Continuing Panels, subset to beneficiaries who were alive and entitled at the fall interview in year *t*. A ratio adjustment accounts for cases that do not have complete Survey File data in both year *t-1* and year *t* . The weights are then further adjusted to account for triple coverage of those accreting on or before January 1 of year *t-3*, and double coverage of those accreting from January 2 of year *t-3* through December 31 of year *t-2*, 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 of year *t-1*, and were alive and entitled as of the fall interview of year *t*.

The derivation of three-year longitudinal weights begins with the weights adjusted through fall of year *t* from the year *t-3* and year *t-2* Panels, subset to beneficiaries who were alive and entitled at the fall interview of year *t*. A ratio adjustment accounts for cases that did not have complete Survey File data in both year *t-2* and year *t*. The weights are then further adjusted to account for double coverage of those accreting on or before January 1, year *t-1*, 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, year *t-2*, and were alive and entitled as of the fall interview of year *t*.

The four-year longitudinal weights are comprised of members of the year *t-1* Panel and are equal to the weights adjusted through fall of year *t* for this panel, subset to beneficiaries who were alive and entitled at the fall interview of year *t*. There is no need for further adjustment by compositing factors because there is only one panel providing four-year data, so the weights are equal to the final cross-sectional weights for these beneficiaries. The final weights represent the four-year longitudinal population, which is the population of beneficiaries who enrolled on or before January 1 of year *t-3* and were alive and entitled as of the fall interview of year *t*.

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

The sample from the three Continuing Panels is adjusted for nonresponse through fall of year *t*. As with the continuously enrolled weights, the process begins with weights for these panels that were previously adjusted through fall of year *t-1*. The response status in winter, summer, and fall of year *t* is then identified. Under the ever enrolled design, respondents include beneficiaries with a complete fall interview in year *t*, beneficiaries who lost entitlement prior to fall of year *t* and had a final complete interview, beneficiaries who died prior to fall of year *t* whether or not a final proxy interview was obtained, and nonrespondents in fall of year *t* who were successfully re-fielded in winter of year *t-1*.

Next, the weights are adjusted for nonresponse through fall of year *t*, 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 of year *t* for purposes of the ever enrolled weights.

---

[27] 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 File LDS.

Continuing Panel weights in the Survey File - Early Release are calculated for the ever enrolled population following the same process, but are based on a more limited set of covariates sourced from demographic, enrollment, administrative, and claims data.

Derivation of the Ever Enrolled Weights

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

At this stage, beneficiaries from the Continuing Panels who died or lost entitlement during year t are included. However, the Incoming Panel cases include only beneficiaries who were respondents to the initial interview in fall of year t, and as such they do not include any beneficiaries who died or lost entitlement prior to fall of year t. Beneficiaries who enrolled before January 1 of year t, who died or lost entitlement are accounted for by the Continuing Panels. Enrollees on or after January 1 of year t, 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 year t are combined separately to account for the fact that these beneficiaries are not represented by the Incoming 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 Incoming 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 year t. They represent the population of beneficiaries who were enrolled for at least one day at any time in year t.

Although there are differences in the data available for response propensity modeling for the Survey File – Early Release weights, as described above, the finalization of final early release ever enrolled weights follows the same process as described in this section.

7.2.5 Cost Supplement File Weights

Data for the Cost Supplement File for year t are collected in winter of year t through Winter of year $t+1$. The weights include beneficiaries sampled in the Continuing Panels, plus members of the Incoming Panel who were enrolled in Medicare during year t. These Cost Supplement File weights are ever enrolled weights representing the population of beneficiaries who were enrolled for at least one day in year t. In addition to the weights, the dataset includes panel (calendar year) identifier, and variance strata and unit variables for variance estimation.

Composition of Sample and Populations of Interest. The Cost Supplement File weights include beneficiaries who were sampled in the last four panels (i.e., the three Continuing Panels and the Incoming Panel). The Continuing Panels provide survey-reported cost and utilization for year t through participation in the MCBS during winter of year t through Winter of year $t+1$. Members of the Incoming Panel who first enrolled in year t are referred to as "current-year enrollees." They are first interviewed in fall of year t and do not provide cost and utilization data for the period of time between enrollment and completion of the fall interview of year t; cost and utilization data for the period between the fall interview of year t and the end of

year t are collected in winter of year $t+1$. 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 year t. 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 year t.

Adjustment Derivation of Cross-Sectional Weights for the Continuing Panels

The process begins with weights for the Continuing Panels that were previously adjusted through fall of year t as part of the Survey File weights for year t. These weights are further adjusted based on a product of the Reference Period process for year t 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 year t 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 Incoming Panel who completed the initial fall interview in year t. Beneficiaries who died or lost entitlement prior to January 1 of year t are ineligible and removed at this stage. Beneficiaries who survived into year t but do not meet the above criteria are considered to be nonrespondents for the Cost Supplement File for year t 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 Survey File weights for year t.

Note that at this stage there is triple coverage of beneficiaries who accreted before January 1 year $t-2$ in the Continuing Panels, and double coverage of beneficiaries who accreted from January 1 of year $t-2$ through December 31 of year $t-2$. 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 of year t, and December 31 of year t, inclusive. This step begins with the initial weights for the Incoming Panel, adjusted for nonresponse at the fall interview of year t. The subset of all fall respondents from the Incoming 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 File. The sum of the combined weights across all four panels (the three Continuing Panels plus the recent enrollees from the Incoming Panel) provides an estimate of the ever enrolled population in year t, 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. The resulting weights are the final weights for the Cost Supplement File for year t. They represent the population of beneficiaries who were enrolled for at least one day at any time in year t.

Longitudinal Weights for the Cost Supplement File. The two-year longitudinal weights are populated for members of the three Continuing Panels who were enrolled in both year $t-1$ and year t and provided utilization and cost data for both years. Members of the year $t-3$ and year $t-2$ Panels provided data for the year $t-1$ and year t through participation in the MCBS during fall of year $t-2$ through winter of year $t+1$. Members of the year $t-1$ Panel who first enrolled in year $t-1$ provided data for the end of year $t-1$ in the winter interview of year t and provided data for year t in winter of year t through winter of year $t+1$. The final two-year longitudinal weights represent the population of beneficiaries who were ever enrolled in Medicare at any time during both year $t-1$ and year t.

The three-year longitudinal weights are populated for members of the year $t-3$ and year $t-2$ Panels who were enrolled in year $t-2$, year $t-1$, and year t, and provided utilization and cost data for all three years. Members of

the year *t-3* Panel provided data for the year *t-2*, year *t-1*, and year *t* through participation in the MCBS during fall of year *t-3* through winter of year *t+1*. Members of the year *t-2* Panel who first enrolled in year *t-2* provided data for the end of year *t-2* in the winter interview of year *t-1* and provided data for year *t-1* and year *t* in winter of year *t-1* through winter of year *t+1*. The final three-year longitudinal weights represent the population of beneficiaries who were ever enrolled in Medicare at any time during each of year *t-2*, year *t-1*, and year *t* implying continuous enrollment during year *t-1*.

7.2.6 Topical Weights

Topical segment weights pertain only to data collected in certain sections of the Community Questionnaire. These include sections that are only administered to living beneficiaries not responding by proxy as well as sections that are administered to all living respondents in the winter or summer round after year *t*. A full list of the Topical segments released on each MCBS LDS is available in the *Data Year Release Notes*.

To facilitate estimation from the resulting data, one set of full-sample and replicate weights is derived for each Topical segment on the Survey File - Early Release based on the Survey File - Early Release ever enrolled population. These weights can be used to conduct joint analyses of Topical segment data and other Survey File - Early Release data. Three sets of full-sample and replicate weights are derived for each Topical segment on the Survey File: one based on the Survey File ever enrolled population, one based on the Survey File continuously enrolled population, and the last based on the Cost Supplement File ever enrolled population. These weights can be used to conduct joint analyses of Topical segment data, other Survey File data, and Cost Supplement File data.

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.

Composition of Sample and Populations of Interest.

The winter Topical segment data are collected from beneficiaries selected in all four panels (i.e., the year *t-3*, year *t-2*, year *t-1*, and year *t* Panels) who were alive, entitled, living in the community, and completed the Community Questionnaire in winter of year *t+1*.

The summer Topical segment data are collected from beneficiaries selected in the year *t-2*, year *t-1*, and year *t* Panels who were alive, entitled, living in the community, and completed the Community Questionnaire in summer of year *t+1*. 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.

For questionnaire sections that are only administered to non-proxy beneficiaries, data are collected from beneficiaries selected in the panels that were administered the questionnaire section (i.e., all panels for fall and winter sections; the last three panels for summer sections) who were alive, entitled, living in the community, and completed the Community Questionnaire without use of a proxy in the interview.

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 Access to Care Questionnaire (ACQ) data because they did not have any emergency room, inpatient, or outpatient events in the year leading up to interview and were not in a Medicare Advantage plan. For the release of ACQ data, CMS fills in information reflecting no such events for these cases. In addition, the Income and Assets Questionnaire (IAQ) is administered to proxy respondents for deceased and institutionalized beneficiaries, so some collected IAQ data are forfeited by the population definition. Imputed total income for all respondents, including Community and Facility interviews, appears 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 nonresponse-adjusted weight for the fall of year t derived during the process of creating the final Survey File ever enrolled, Survey File continuously enrolled, or Cost Supplement File ever enrolled weights for year t. 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 Survey File and Cost Supplement File weights are 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 is smaller in the winter of year $t+1$ because this round immediately followed the fall of year t Survey File interviews, whereas in the summer of year $t+1$ there was an intervening round in which some members of the sample became nonrespondents. For the Patient Activation (PA) question series, which is fielded in the fall round as part of the Satisfaction with Care Questionnaire (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 are derived from the effective number of completes by accretion year and age group across the set of panels that were administered the seasonal section.

7.3 MCBS Imputation Processes

7.3.1 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.

7.3.2 Income and Asset Imputation

Overview. The 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,[28] 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 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 have 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 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 residence status.

[28] 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 the annual *MCBS Questionnaire User Guide*.

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.

7.3.3 Insurance Plan Premium Imputation

Overview. The 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 are asked if they paid any or all of the premium for the plan coverage. For MA plans, respondents are 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.

7.3.4 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 Section 7.5. 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 7.3.1 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 7.3.1, 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 7.3.1: 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 7.3.2. 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,[29] payment amounts are imputed individually from different donors.

Exhibit 7.3.2: 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 7.3.3). 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.[30] Total charge is set to the sum of the payment amounts.

[29] 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.

[30] 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.

Exhibit 7.3.3: 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 near 0.10 percent .

The PM imputation produces one file, an event-level data set 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.

Since 2015, current-year enrollee sample beneficiaries are included in the Non PM imputation.[31] 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* 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.

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

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*, 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 (HP) event data are appended to the Non PM events. These data come directly from CMS, are complete, and therefore not imputed. More information on HP data is provided in Section 7.5.

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 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*. Then, unadjusted and adjusted service level amounts are summed to the person level.

Hot Deck Imputation Procedure

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

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

Each imputation step has a unique set of qualification rules and key variables used to identify a similar donor record for a given recipient record. The donor pool for each set of recipients is first restricted to the group of potential donor records that meets the donor qualification rules, such as requiring that donors have complete data on the item to be imputed. Next, the similarity between a given recipient and each possible donor is measured via the Gower function using SAS/STAT® software's PROC DISTANCE:

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

where v is the number of variables, x_j is the data for observation x and the j^{th} 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, $d_{x,y}^{j} = 1 - |x_j - y_j|$.[32,33,34]

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.

7.3.5 Details on the MA Encounter Data-Informed Ratio Adjustments

Ratio adjustments are applied to qualifying event-level records. Beginning with 2020, the ratio adjustments are 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 7.3.4. Exhibit 7.3.5 shows how the ratio adjustments are applied across settings and subgroups.

Exhibit 7.3.4: 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.

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

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

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

Exhibit 7.3.5: Applicability of Ratio Adjustments by MA Setting/Event Type and Age and Health Subgroup[1]

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)	X	X	X
Durable Medical Equipment (OM)	X	X	X
Inpatient (IP)	X	X	X
Outpatient (OP)	X	X	X
Skilled Nursing Facility (IU)	X	X	X

[1] The ratio adjustments are calculated using MA Encounter Data and survey-reported utilization data and are updated year-to-year for each setting and subgroup combination.

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

1. Event type (EVNTTYPE) is:
    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 occurs within a month covered by MA based on the administrative enrollment data
3. The event is reported in the survey, either matched or unmatched (i.e., not a claims-only event)

Records with an EVNTTYPE of Other Medical Expense (OM) are assigned to either the Carrier or Durable Medical Equipment (DME) setting. There is a large difference in the ratio adjustment for these two settings, so care is taken to assign records to each of these settings. If the claim record is determined to be for durable medical equipment, then the record is assigned to the DME setting, otherwise the record is assigned to the Carrier setting. However, the setting is not determined for unmatched survey reported OM events. These events are randomly assigned to the Carrier or DME settings at a proportion determined by historical norms. Typically 90 to 95 percent of the unmatched survey reported OM events are assigned to the DME setting with the remaining 5 to 10 percent assigned to the Carrier setting.

The unweighted impact of the MA Encounter ratio adjustment on the overall sum of total costs is typically an increase of 10 to 15 percent.

More detailed information on these imputation procedures is available in Section 7.5.

7.4 Matching Survey and Administrative Data

This section describes the processes for bringing together survey information, which can only be obtained directly from a beneficiary or proxy respondent with reliable information on services used, and Medicare payments made from administrative bill files for the Cost Supplement File. Survey-reported cost data include information on the use and costs of health care services, including the amount paid by private health insurance and other payers/payer types (if applicable). The survey also collects information on health services not covered by Medicare, most notably, long-term facility care and dental, vision, and hearing services.

Medicare bill data include use and cost information on prescription drugs, inpatient hospitalizations, outpatient hospital care, physician services, home health services, durable medical equipment, skilled nursing home services, and hospice services. This combination Cost Supplement File can support a much broader range of research and policy analyses on the Medicare population than would be possible using either survey-collected cost data or administrative bill data alone.

Use and costs of Medicare-covered services are reported on both the MCBS survey and in the CMS administrative bill files. This overlap in reporting from the two sources is used to verify the accuracy of survey reports of health service use. Survey reports are matched with administrative bill data to adjust for survey under-reporting using more complete administrative bill data, and to fill in and correct survey reported payment amounts with more accurate information from bills submitted to and paid by Medicare. Note that this can only be done for Fee-for-Service (FFS, or Original Medicare) beneficiaries accessing services covered by Medicare such as inpatient hospital services, outpatient hospital services, physician services, home health services, acute skilled nursing facility services, durable medical equipment, and other covered services as well as Medicare Part D beneficiaries accessing prescription drug services. For health services not covered by Medicare FFS or Part D such as long-term facility care, there is no independent source to which survey reports could be matched.

Under-reporting of medical services is an enduring problem in personal interview surveys. While respondents can usually recall significant events like hospitalizations for several months, they often fail to recall more routine care like physician visits after a few weeks. In general, as the time interval between the interview date and the medical event increases, the probability decreases that the event will be recalled and reported accurately in the interview. The MCBS interviews sampled beneficiaries up to three times per year over a four year period, and the average interview recall period is about four months. (More frequent interviews would reduce the recall problem, but it would greatly increase both survey administration costs and the reporting burden on beneficiaries). Given normal rates of memory decay and the frequency with which aged and disabled persons use medical care, it is reasonable to assume that matching survey events to administrative bills would be helpful in identifying medical events that the respondent could not recall during the interviews.

7.4.1 File Building

All matched service records should be added to all unmatched Medicare claim-only records. In addition, unmatched survey reports, excluding the records with a Medicare payment amount and no MA payment amount, should be added to the matched and Medicare claim-only records. The records which are to be excluded are identified with the SOWMP flag on the institutional, outpatient hospital, medical provider, and inpatient hospital event-level segments. This file will be the most complete and accurate file possible, and this combination minimizes the risk of double counting unmatched records.

7.4.2 Filling the Gaps

The Cost Supplement File is designed to provide person-level data for estimating total use of, and total payments for, all health care services, covered and non-covered, received by Medicare beneficiaries during the calendar year.

This section describes the adjustments that are made to the MCBS data to attain the goal of estimating total use of, and total payments for all health care services (covered and non-covered), received by Medicare beneficiaries during the calendar year. The adjustments made are as follows:

- EVENT-LEVEL MATCHING: These operations identified services paid for by Medicare that were not reported on the survey and corrected Medicare payment data reported inaccurately on the survey. A discussion of match results and instructions for building a complete file and avoiding duplication is provided below.

- MISSING PAYMENTS AND PAYERS: These adjustments compensate for missing payment data and for when the respondent did not know how much an event cost and/or how the event was paid for (by whom, and how much by each payer).
- PRESCRIPTION DRUGS: Describes the particular problems encountered in creating the prescription drug event file. Also includes adjustments to compensate for missing payment data and for when the respondent did not know how much a prescription cost or how the prescription was paid for (by whom, and how much by each payer).
- ADJUSTMENTS FOR MISSING DAYS AND UNDATED USE: These adjustments compensate for data that are missing because some periods of time were not covered by interviews and because some types of health services use (particularly prescription drugs) are undated.
- ADJUSTMENTS FOR UNREPORTED HEALTH CLAIMS OF MA BENEFICIARIES: These adjustments compensate for known under-reporting of health events for beneficiaries with MA coverage. For beneficiaries with FFS coverage, similar under-reported events were obtained from the administrative claims data records that were not matched to survey reported events.

Adjustments made to records in the Cost Supplement File LDS are constrained in two ways. First, because CMS administrative data are used to fill in much of the missing information, all adjustments to MCBS utilization, cost, and source of payment data are consistent with CMS administrative data. For example, if CMS records indicate that the beneficiary is dually entitled to both Medicare and Medicaid, then Medicaid must be considered a possible source of payment when source of payment is missing, even if the respondent did not volunteer that information. Second, adjusted data must be consistent with other information for the same person. For example, the source of payment for individual events must be consistent with the beneficiary's health insurance information.

7.4.3 Event-Level Matching

There are two primary objectives in matching survey reports to Medicare administrative bill records: to correct for under-reporting of events on the survey, and to correct errors in payment information collected in the survey.

The first step in matching survey-reported medical events to Medicare bill records is gathering all events for a person together. Because the MCBS sample is drawn from the CMS Medicare Administrative enrollment data, matching the Medicare paid claims and bills with the correct beneficiary is a reasonably straightforward process. The beneficiary's Medicare number (i.e., health insurance claim number (HICN) or Medicare Beneficiary Identifier (MBI)) is part of the information collected from the enrollment data when the sample is drawn. The beneficiary's HICN/MBI is verified in the first MCBS interview. Prior to the match, Medicare paid claims are retrieved from the Medicare National Claims History (NCH) repository, by HICN/MBI. The search file includes all cross-reference numbers and different beneficiary identification codes associated with each beneficiary, ensuring that all bill records are recovered.

Linking and reconciling the retrieved Medicare claims with individual events reported in the survey is a much more complicated process than matching Medicare paid bills with the correct beneficiary. There is no data element, or combination of elements, that provide a consistent basis for matching survey data to Medicare claims across all types of services. There are significant differences in the ways in which medical goods and services are characterized in the survey and in the Medicare claims records.

Neither the MCBS nor CMS claims records capture a consistent set of data elements for all service types. For example, the MCBS does not capture total reimbursement for inpatient hospital services because the respondent is not likely to know that information; it is not typically included on the notice of utilization, and thus, this information cannot be used in matching. In other categories, especially Part B services, the total charge of the service is known because it appears on the explanation of benefits, and it is a key match field.

Similarly, CMS claims data do not always have the same data elements for different claim types. The carrier control number for each claim is included in the CMS claims history files and the MCBS attempts to collect the carrier control number from the beneficiary's explanation of benefits in the interviews. As a result, this item is extremely useful in matching survey reported utilization to Part B claims. On the other hand, the intermediary control number (Intermediaries process claims for Part A of Medicare) is not available in the CMS files, so even though it is collected in the survey, this data element is not helpful in matching the survey data to Part A bill records. Additional details regarding matching are provided below.

Survey-Reported Utilization

In the utilization sections of the MCBS Community Questionnaire, respondents are asked about all medical events, including visits to practitioners of all types, prescriptions, and any medical equipment or supplies used. (Please find copies of the survey instruments and exact wording of the questions at https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/questionnaires).

Types of utilization collected in the MCBS:

- Dental (DU): Dentist visits, including cleaning, x-rays and repair, purchase or repair of dentures, and orthodontic procedures
- Vision (VU): Vision care including vision exams, contact lens fittings or purchases, eye glass frame fittings or purchases, and different kinds of surgeries (e.g., cataract, corneal, etc.).
- Hearing (HU): Hearing care including hearing exams, hearing aid fittings, repairs, or purchases, or hearing rehabilitative services
- Emergency room (ER): Hospital emergency room visits
- Inpatient hospital (IP): Inpatient hospital stays
- Institutional (IU): Other short-term institutional stays, such as skilled nursing home stays or rehabilitation hospital stays
- Medical provider (MP): Doctor visits, including visits with medical doctors (MD); practitioners, such as chiropractors, podiatrists, audiologists, and optometrists; mental health professionals such as psychiatrists, psychologists, and clinical social workers; therapists, such as physical therapists, speech therapists, occupational therapists, and intravenous and respiratory therapists; other medical practitioners, such as nurses and paramedics; and other places offering medical care, such as clinics, neighborhood health centers, infirmaries, and urgent care centers.
- Outpatient hospital (OP): Outpatient visits, including visits to the outpatient department or outpatient clinic of a hospital.
- Other medical expenses (OM): Other medical expenses, including purchase or rental of a variety of items: eyeglasses or contact lenses and hearing aids; orthopedic items such as canes, walkers, wheelchairs, and corrective shoes; diabetic supplies; oxygen supplies and equipment; kidney dialysis equipment; hospital beds, commodes, and disposable supplies such as disposable diapers and bandages.
- Prescription medicine (PM): All prescription medications except those provided by the doctor or practitioner as samples and those provided in an inpatient setting.
- Home health (HH): Home health visits, collected in the survey as visits by professionals (HHP) or friends (HHF). Health professionals include nurses, doctors, social workers, therapists, and hospice workers. Friends include persons who do not live with the beneficiary but help the beneficiary at home with personal care or other daily needs. These persons may be home health aides, homemakers, friends, neighbors, or relatives.

In addition to these categories, the Community Questionnaire is also designed to collect some types of utilization that the respondent may unintentionally omit. This utilization is captured when the beneficiary's Medicare and private health insurance statements are reviewed and is classified as separately billing doctor (SD) and separately billing lab (SL). The SD and SL categories typically include such things as anesthesiology

administered while the beneficiary was an inpatient, lab tests not done at the doctor's office, and the radiologist's interpretation of an x-ray.

The MCBS Facility Instrument captures similar information about people living in long-term care facilities, with the exception of prescribed medicines, which is not collected in the Facility Instrument. CMS administrative data for inpatient hospital stays and short-term skilled nursing home stays are integrated with survey data to ensure date alignment. These administrative data are also used to properly account for the Medicare payments related to the short-term skilled nursing home stays that occur within the encompassing stay at the long-term care facilities.

CMS-Reported Utilization

Medicare claims are organized by type of provider. The categories of Medicare claims records are as follows:

- Inpatient hospital, psychiatric hospital, Tuberculosis hospital, Christian Science facility, and skilled nursing facility bills: Although these records all share the same format, they contain codes that allow them to be separated into these subcategories. For purposes of the match, bills from skilled nursing facilities were separated from the other types of bills, but no further subdivisions were made.
- Home health bills
- Hospice bills
- Outpatient hospital bills
- Part B physician/supplier claims for physician services, diagnostic laboratory and radiology, durable medical equipment, and some prescription medicines.
- Part D bills

Match Categories

In matching the survey-reported utilization to the Medicare claims data, MCBS staff frequently must match a Medicare claim category to multiple MCBS categories, and vice versa. Although there are some clear relationships between the categories of utilization collected in the survey and CMS claims categories, not all categories match neatly.

Event-level matching is actually a series of matches between different categories of Medicare claims and MCBS service types. In conducting these matches MCBS staff employ different match algorithms, depending on the data elements available for the particular categories being matched. Matches are arranged in sequence, so that the most similar survey-reported and Medicare claims categories are compared first. Exhibit 7.3.6 presents an overview of the categorical matches.

Exhibit 7.3.6: Overview of Event Category Matches Conducted During Event-level Matching

Matches between similar service types	Matches between less similar service types
IP to Inpatient hospital bills	ER to Inpatient hospital bills
MP, OM, SD, SL to Part B physician/supplier	OP to Inpatient hospital bills
OP to Outpatient hospital bills	IU to Inpatient hospital bills
IU to SNF bills	IP to SNF bills
DU to Part B physician/supplier claims	IP to Outpatient hospital bills
ER to Outpatient hospital bills	OP to Part B physician/supplier claims
HHF & HHP to Home health agency bills	MP, OM, SD, SL to Outpatient hospital bills

Each match algorithm employs a hierarchy of match criteria which are progressively less restrictive. For example, reported doctor visits are initially compared to claims records by physician's name, date of service, and total charge. If there is not an exact match, the algorithm checks for a match on physician's name and date of service, or total charge and date of service. If there is still no match, the program looks for an exact match on physician's name and total charge, with the date of service match relaxed to dates within one week of each other. The match algorithms not only link survey-reported utilization and Medicare claims records, but also code the records to indicate the strength of the link.

MCBS staff designed the match algorithms to allow survey-reported utilization to be linked to multiple Medicare claims, and vice versa, for two reasons. First, multiple links are often valid. For example, a survey-reported doctor visit may be linked to both a Medicare claim for the physician's services and a Medicare claim for laboratory services connected with the visit. Second, a stronger match may occur later in the series of matches. A survey-reported doctor visit may have a weak link to a Medicare Part B physician/supplier claim and a strong link to a Medicare outpatient claim. MCBS' staff uses the link-strength indicator to resolve situations where the multiple matches are logically inconsistent.

Hospice bills are excluded from the match because there is no clean "hospice" category in the survey data. Facility utilization is matched in only a summary fashion to improve the accuracy of Medicare payment data.

Three outcomes are possible from the attempted match of survey data to Medicare claims data for events where claims are available (i.e., paid for by Medicare FFS or Part D): the information from the two sources agrees (a match); or, information reported in the survey is not present in the Medicare claims data; or, information is present in the Medicare claims data which was not reported in the survey.

- Cost Supplement File "events"
  - The matching programs produce a set of records which reflect the best combination of survey and Medicare claims categories, and present records from both sources (matched and unmatched) in a uniform format. Since the categories of utilization in the Medicare claims do not match the survey categories, utilization groups in the Cost Supplement File are a combination of the two sources.
  - The most disaggregate level of utilization records in the Cost Supplement File is the "event" level record. Event records combine survey-reported information and Medicare claims data in the ten categories presented: IPE, OPE, IUE, DUE, home health, HUE, MPE, PME, VUE, and FAE. Event records

contain a variable to indicate the source of information—Medicare claims data, survey data, or both. An additional variable provides the link from the event record to the bill data if both sources provided the information.

- Emergency room
  - The emergency room (ER) survey category is split between IPE and OPE. Under the prospective payment system, emergency room services that result directly in a hospital admission are included in the diagnosis-related group (DRG) payment for the inpatient stay, and thus are not associated with any separate charges or claims (see DRG variable on IPE segment). Emergency room visits that are not immediately followed by an inpatient admission are classified as outpatient services. For this reason, survey-reported emergency room (ER) utilization is matched to outpatient, then inpatient bill records, and is reflected in the Cost Supplement File as either OPE or IPE records. Several other survey categories (MP, SD, SL, and OM) have been combined to make up the single MPE category. Hospice services do not exist as a separate category of utilization in the survey data, so this category derives from the Medicare claims data.
- Post-match edits
  - For most types of services, the MCBS collects a date of service to assist in matching survey-reported data to claims records. Respondents may not always recall exact dates, so dates are collected in three independent parts (i.e., month, day, and year).
  - Since the year portion of a survey date may be missing or incorrect, records for services in the prior calendar year and following calendar year are not eliminated from the Survey File until the match is concluded. Similarly, respondents may "telescope" events, believing them to have taken place recently when in reality they occurred a year or more in the past. As matching Medicare claims might help to identify and eliminate these responses, the Medicare records were also not edited on date until after the match; for the match records included services rendered in the prior, current, and following calendar years. After matching, the event file was edited to exclude all services that were rendered outside of the current calendar year.
  - If the survey-reported data matched Medicare claims data, the dates of service on the Medicare record were carried into the event record. Dates of service were used as a match criterion in most of the matches, so in many cases, the dates of service in the event record did not change from those reported.

Summary of Match Results

Medicare bill events for beneficiaries living in the community were matched against survey reports. The percentage of dollars matched was considerably higher than the percentage of events matched.

Some small part of the unmatched Medicare records is undoubtedly represented in the survey-reported events that could not be matched under the matching criteria used. However, the unmatched survey events would be expected to include a substantial share of events that are not covered by Medicare, and therefore would not be expected to match a Medicare paid claim. In addition, only a small subset of the unmatched survey-reported events have a Medicare payment amount. The unmatched Medicare billing records strongly suggest that the survey reports seriously understate the number of Medicare services when compared to CMS billing records.

The under-reporting problem is more serious for event counts than for Medicare payments. The unmatched Medicare bill events represent a larger percent of events than percent of total payments. Unmatched Medicare events were a little more than half as expensive on average as matched events. This is consistent with general household survey experience, which has shown that more expensive medical events are more likely to be remembered and reported at the interview. This trend is consistent with the hypothesis that survey

respondents tend to remember major health events/more expensive events better than minor health treatments.

In addition to correcting for unreported events, the match also helped to fill in missing Medicare payment amounts and correct Medicare payment amounts that had been reported incorrectly. The match and reconciliations generated corrections that made Medicare a payer of record on the cases where this information was originally omitted in the survey reports; and made it possible to determine the correct Medicare payment amount in the survey records where this information was omitted.

Not all services could be cleanly and easily matched from the two sources. The match employed "strength of evidence" criteria and "hierarchical algorithms" in order to identify matches, survey reports only, bill file reports only, and a small number of similar events for which it was not clear whether there was duplicate survey and bill reports or not. The methods and criteria used in the match are discussed in more detail in the Event-Level Matching discussion in section of this manual. In addition, Eppig and Edwards' paper, "Computer Matching of MCBS Data with Medicare Claims," presents a full discussion of methods and criteria.

Evidence Supporting Improved Accuracy

On the matched events, Medicare should have been reported as a payer on 100 percent of the survey-reported events. However, Medicare was only reported as a payer for about two-thirds of the events. Consequently, the match corrected the remaining records to make Medicare a payer of record.

On the matched events, the Medicare payment amount was only reported on about half of the survey reports. The match filled in the correct Medicare payment for the remaining survey reports.

Evidence of Survey Under-reporting

The unmatched Medicare paid bill events strongly suggest a high level of under-reporting on the survey. While there are more unmatched survey reports on the other side, many of these events could not be reasonably expected to be undiscovered matches. For example:

Unmatched survey events unlikely to match an unmatched Medicare bill include:

- Survey events for dental services that are rarely covered by Medicare.
- Survey events that had total payments equal to zero. (These were very likely parts of bundles of services that were covered in one global payment on the Medicare side, for example, post-operative services that were covered by a global surgery fee.)
- Survey events were for MA enrollees. Virtually all of the Medicare services for these persons are paid through a capitated payment amount and the likelihood is very small that their events ever match a fee for service Medicare paid bill record.
- Survey events where the beneficiary was only entitled to Part A or Part B of Medicare, but not both. Therefore, a survey-reported service could reasonably not be expected to match a Medicare paid bill record.
- Survey events were provided by the Veteran's Administration (VA) or in a military installation where no Medicare bill would be expected.
- Survey events were for other medical services. While Medicare covers durable medical equipment such as wheelchairs, and supplies such as oxygen, it does not cover many items in the broad other medical services category such as eyeglasses, hearing aids, heating pads, incontinence supplies, etc. Average payments for unmatched survey reports of other medical events were less than the average survey

reported payments for matched. This suggests that most unmatched survey events for other medical services are probably not undiscovered matches.

- Taken together, a majority of the unmatched survey events either definitely could not, or very likely would not, match a Medicare bill event record.

Unmatched survey events likely to be undiscovered matches include:

- Survey-reported events that reported a dollar amount that Medicare paid for the event. These unmatched survey events are very likely to be undiscovered matches.

Ambiguous events:

- A subset of unmatched survey events that are not explained by the above. There are many medical services and supplies that Medicare does not cover such as most alternative medicine services, over the counter remedies, etc.

Building a Complete File

Medicare covered services:

- A complete file would include all matched events. These events, which were reported on both the survey and in Medicare bill event records, will combine the most accurate and complete information possible from both sources.
- All Medicare bill record unmatched events should also be included. These event records are official records of Medicare program payments and will partially correct for survey under-reporting.
- It is more debatable which of the unmatched survey records to include. The data include type of service, adjusted file summaries, and all unmatched survey reports except the records with a Medicare payment. For the reasons discussed above, these records are likely undiscovered matches that would duplicate some of the unmatched Medicare bill event records if they were included. For official MCBS reporting purposes, CMS uses all unmatched survey reports except the records with a Medicare payment.
- Hospice records, which were not entered in the event-level match, should be added into the file.

Total medical services including Medicare covered and non-covered services:

- In addition to the Medicare covered services listed above, Prescription Drug and Long-Term Facility records should be added to the file.

7.5 Detailed Imputation Procedures

There are several adjustments made to fill in payment amounts that are missing because the respondent did not know how much an event cost or did not know how the event was paid for (by whom, and how much for each payer).

First, a target reimbursement or total payment for the event was established, all possible sources of payment were identified, and then the total payment was distributed across all payers. Missing amounts and payers were filled in using either analytic editing or statistical imputation. This process relied heavily on Medicare administrative records. The guiding principle of retaining as much survey data as possible, and filling in around it, was maintained throughout the process. Where feasible, information about the payers for a specific event, known payment amounts, and target reimbursement were used to determine unknown payment amounts by analytic edits. When insufficient information was available and analytic editing was impossible, unknown payment amounts were completed by statistical imputation.

Different approaches were used with different categories of utilization to define payers and determine payment amounts. Records submitted to the survey/administrative match (which was discussed in Section 7.4.3 Event Level Matching) were handled differently from those not matched. Survey-reported records for dental, vision, hearing, medical practitioner, home health from a medical practitioner, inpatient, outpatient, institutional (other than long-term care), and medical equipment and supplies (survey utilization categories DU, VU, HU, MP, SD, SL, HP, IP, OP, IU, and OM) were entered into the match with Medicare claims data. After the match, these events were individually assigned target reimbursement amounts, and then source of payment variables and separate payment amounts were calculated for each payer. Other procedures, usually some adaptation of the procedures sketched above, were used to determine payers, target reimbursements, and payments for other categories of utilization. The next section discusses how target reimbursements were established, explains the procedures used for matched utilization (the largest category of utilization), and then discusses the smaller and more specific non-matched categories.

7.5.1 Determining Target Reimbursement

A primary rule was to establish the target reimbursement for an event with a missing total payment prior to determining or imputing the payment distribution. This was done to establish a target reimbursement that was consistent with payments shown for other similar services in the file. In this way, a credible target reimbursement can be used to inform and control the payment distribution. For Medicare covered services, target reimbursements were developed from Medicare claims; this method is more accurate than determining the amounts paid by individual sources of payment and summing them.

Another primary rule was to retain survey-reported payment data, even when it was only partial data, wherever possible. There are situations where retaining the reported payment amounts and establishing the target reimbursement amount without regard to individual payment amounts are mutually exclusive. On a few occasions, the target reimbursement had to be adjusted in order to retain reported payment data.

The rules for establishing target reimbursements depend first on whether or not Medicare claims data are available. If the survey-reported data match a Medicare claim record, or if the Medicare claim record was the only source of information about the service (i.e., nothing about the service was reported in the survey), the Medicare claims data were used to establish a target reimbursement. The target reimbursements for about 75 percent of the events in this file were established using Medicare administrative bill payment data.

If the utilization was only reported in the survey (i.e., matching to Medicare claims was not successful in identifying a corresponding claims record), the survey data were used to create the target reimbursement. This occurred for about 25 percent of events in this file.

For a small subset of the survey reported events without a matching Medicare claim, but where Medicare was reported as a payer, a different approach was used to create a target reimbursement. A set of regression models, one for each type of event, was developed to predict the target reimbursement from the total charges reported in the survey.[35]

When the respondent did not report a total charge for the event but indicated that Medicaid was a payer, an imputed target reimbursement was created which was consistent with the generally lower payments made by Medicaid.

[35] Note that these records with a target reimbursement amount estimated by the regression models are the same records described in Analytic Notes for Non PM Event Segments of the *Data User's Guide*. Most of these records were not included in the summary files (SOWMP = 1) because they were largely events with FFS Medicare payments. Some were retained in the summary files if the beneficiary had MA coverage (i.e., they were considered MA payments and not expected to have matched FFS administrative claims).

7.5.2 Filling in Missing Payments and Payers for Matched and Not Matched Utilization Records

The following procedures were used to determine who paid for each event, how much an event cost in total, and how much each payer paid. These procedures were applied to events in the Cost Supplement File designated: IPE (inpatient), OPE (outpatient), IUE (institutional), DUE (dental), VUE (vision), HUE (hearing), PME (prescribed medicine), MPE (medical and surgical services, equipment and supplies), and HHE (home health).

See below for information on how missing payments and payers are filled in for the FAE (long-term care facility) events.

Determining Potential Payers

Regardless of the method used for imputation, payment amounts were only imputed for potential payers. The total reimbursement for an event was distributed among 10 sources of payment (SOP):

- Medicare FFS
- Medicaid
- Medicare managed care
- Private insurance managed care
- Employment-based private health insurance
- Individually purchased private health insurance
- Private insurance, source unknown
- Out-of-pocket
- Uncollected liability
- Other public insurance

It was determined that payments made by the VA could not be estimated with sufficient accuracy. Therefore, in 2016 and beyond, payments from the VA are combined into the "other public insurance" source of payment.

Out-of-pocket payments are those payments made by the beneficiary or their family, either as cash or through Social Security or SSI checks to a nursing home. Medicare MCOs coverage (i.e., Medicare Part C/MA) is different enough from FFS coverage to merit its being reported separately. Non-MCO private insurance is characterized as individually purchased or employment-based because there are differences in cost and coverage depending on type. As this information is not known for residents of nursing homes (the nursing home staff are not likely to know, and thus are not asked, how the insurance was purchased), a third category of private, non-MCO insurance was created for private insurers when the source is not known. Uncollected liability refers to unpaid amounts where there is a legal obligation to pay. If there is an agreement between the provider and a payment source, which reduces the amount that the provider can collect for a service, there is no uncollected liability. On the other hand, if the respondent reports a total amount payable and specific payment amounts for all known sources of payment, and the sum of those payments is less than the total amount payable, the difference is considered an uncollected liability. Other public insurance includes payments made by the VA, as well as federal or state programs not included in the other categories, such as state drug programs like PACE in Pennsylvania.

An individual's insurance coverage can change during the course of a year. A health insurance timeline,[36] created for each person in the Survey File, provided the basis for determining the potential payers for each event. The timeline contains complete insurance information, including Medicare entitlement, Medicaid

[36] The Survey File LDS includes the types of insurances, the coverage eligibility timeline, and the source information for the coverage.

eligibility, and enrollment in Medicare MCOs (Medicare Part C/MA), for every day of the beneficiary's Medicare eligibility during the year. Medicare entitlement, Medicaid eligibility, and enrollment in Medicare MCOs are captured from CMS administrative data, while information about private insurance is collected in the insurance portion of the interview and then supplemented by information obtained from statements and Medicare claims. Medicaid eligibility from the CMS administrative data provides the ability to distinguish whether a beneficiary has full Medicaid benefits or only Qualified Medicare Beneficiary (QMB) or Special Low-income Medicare Beneficiary (SLMB) status. If an individual has full Medicaid benefits, then Medicaid is a potential payer for all medical events. If an individual has QMB-only status, then Medicaid is a potential payer for cost-sharing amounts on Medicare-covered services, but not on medical events not covered by Medicare. If an individual has SLMB-only status, then Medicaid is not a potential payer for any medical events.

Payer Indicators

A payer indicator code is used to identify definite and potential payers of the total charge for an event. Source of Payment (SOP) flags are used to initialize the payer indicator. Each SOP flag corresponds to one component of the payer indicator and could have a value ranging from 0 to 4 as shown below.

Source of payment (SOP) flag values:

- 0 - Source definitely did not pay
- 1 - Source definitely did pay, known amount
- 2 - Source definitely did pay, unknown amount
- 3 - Source possibly paid, beneficiary was covered at time of event
- 4 - Source possibly paid, beneficiary may have been covered at time of event

SOP values are set by using survey information about reported events, about the type of provider for the event (that is, whether the service was delivered by a managed care provider or a VA facility), and about the type of insurance coverage and/or program participation. SOP values also depend on Medicare claims data when a survey-reported event corresponds to a Medicare claim (a "matched" event). Based on all of this information, each source is determined to be a payer, a potential payer, or not a payer of charges for the event.

- **Payers**: A source is a definite payer if the SOP for that source has a value of 1 or 2. An SOP value of 1 indicates that the respondent reported that the payer had paid a portion of the charges and also reported a payment amount, or that Medicare claims data provided that information. An SOP value of 2 means that the respondent reported that a payer paid a portion of the charges, but did not know the exact amount, and no matching Medicare claim was found to provide this information.
- **Potential payers**: A source is a potential payer if the corresponding SOP has a value of 3 or 4. An SOP value of 3 means that either the beneficiary definitely had that type of insurance coverage at the time of the event and the payer may have paid some amount, or the beneficiary received the service from that type of payer (i.e., a managed care provider or a VA facility), but did not report it as a payment source. An SOP value of 4 is used when there is doubt about the beneficiary's insurance coverage during the event or about the event date itself.
- **Non-payers**: If neither the respondent nor the Medicare claims data indicate that a payer has been a source of payment for an event, the SOP is set to 0.

A more comprehensive discussion of the rules used for setting the SOP flags is included below.

Translating Payer Indicators into Sources of Payment for Matched Utilization Records Only

A value of 1 for a particular payer indicator means that the payer paid a portion of the total charge for the event, and a value of 0 means that the payer did not contribute. Final payer indicator values are determined in one of three ways: 1) directly from the corresponding SOP values; 2) through analytic edits; or 3) through statistical imputation.

Different rules apply when payer indicator values are set directly from the corresponding SOP values, depending on whether the SOP is determined to be a definite payer, a potential payer, or a non-payer. The payer indicators are initialized as follows:

- If the source is a definite payer and the payment amount is known (SOP=1), the corresponding payer indicator is set to 1.
- If the source is a definite payer but the payment amount is not known (SOP=2), the payer indicator value is set to 1 with one exception.
  - If the event is for dental, vision, or hearing care or for durable or nondurable medical equipment not usually covered by Medicare, the Medicare payer component is set to 0. The rationale is that if the respondent is not able to report the Medicare payment, then it is more likely that Medicare has not actually paid for the ordinarily non-covered services.
- When the SOP is a potential payer (SOP=3 or 4), the corresponding payer indicator is set to missing and imputed (as 0 or 1) in a later step. However, the general rule for setting a payer indicator value based on a corresponding SOP value of 3 or 4 is sometimes modified by analytic edits.
  - For example, the Medicare payer indicator value is never set to missing. It is always equal to 0 or 1. When the SOP for Medicare is listed as a potential payer (SOP=3 or 4), which is rare, the Medicare payer indicator is set to 0 (Medicare did not contribute).
- When the SOP is not a payer (SOP=0), the corresponding payer indicator is set to 0, with exceptions for out-of-pocket payments and uncollected liability.
  - If the SOP is out-of-pocket or uncollected liability and the SOP value is equal to zero, the payer indicator is set to missing, to be imputed as 0 or 1 in a later step.

Analytic Edits

Analytic editing of charge and source of payment data at the event-level also determines some payer indicator values. The general goal of the analytic edits is to resolve as many events as possible (i.e., to fully allocate total charges to payers) and to set as many payer indicator values as possible based on logic and knowledge of payer policies. The edits resolve some events without using a hot deck procedure to impute payment sources or amounts.

The analytic edits rely on having both unambiguous SOP values and external information about interaction among the insurance or payment sources. Edits for two of the payment sources (Medicaid and MCOs) depend on information specific to those payers, but payer indicators for other payment sources are also affected. The analytic edits are discussed fully below, as they apply to each source of payment.

Medicaid. Analytic edits were used extensively when Medicaid was a potential or actual source of payment for an event. One set of edits, designed to reflect the role of Medicaid as the payer of last resort, ensures that Medicaid could not be a payer if payments are reported or imputed for another third-party insurer (except

Medicare), or if the provider is an MCO or VA facility. Another set of edits was developed for dually eligible beneficiaries whose cost-sharing liability is covered by Medicaid. For additional information, see below.

Private and Medicare MCOs (Medicare Part C/MA). MCOs (especially Medicare-contracting MCOs) often operate differently from other third-party payers and tend to have unique payment patterns. For instance, risk and (to a lesser extent) cost Medicare MCOs are paid a set fee per enrolled Medicare beneficiary (called a capitated amount) designed to compensate the MCO for the expected costs of delivering Medicare's package of benefits. There are no Medicare claims or Medicare or insurance statements indicating the total charge for events covered by the capitated amount. Often the respondent only knows the copay amount if there was one. Also, MCOs often provide "Medigap"-type coverage by paying for most of the member's deductibles and copays for Medicare-covered benefits. A beneficiary who belongs to an MCO does not need private Medigap insurance or Medicaid coverage for these amounts. Thus, payment patterns for MCO beneficiaries tend to be simpler than those for FFS beneficiaries. The set of analytic edits for MCOs attempt to account for these simplified patterns and for the respondent's usual inability to report charges and payments for events. The MCO edits also attempt to avoid creating "illogical" payment patterns.

General Edits. At the beginning of the analytic editing, and after each main section of edits, an attempt is made to resolve events through addition or subtraction. Events without a known total charge but with a complete payer indicator vector (i.e., each payer is identified as either having paid or not paid for an event and each payer's amount was known) are completed by summing across all payment sources to derive the total charge. Events with a known total charge and complete except for one missing payment amount or payment source are completed by subtraction. The excess of charges over known payment amounts is attributed to the known payer, or the one missing payer indicator is set to 1 and the excess allocated to that payer.

If a service was provided free of charge, all payer indicators and payment amounts are set to 0. However, if the respondent reported an event as free, but also reported that a source other than Medicare or Medicaid paid something for the event, the total charge is reset to "missing" and imputed.

If a source is a potential payer for an event, or if the respondent reported that the payer had contributed to an event but did not know the amount, it is assumed that the payer is not actually a source if the current sum of reported payments equals the reported total charge.

If Medicaid is a payer, a Medicaid payment amount is calculated as a percentage of coinsurance and deductible for the Medicare service.

Other Utilization (Not Matched)

The procedures described in the next three sections below, are used to determine who paid for each event, how much an event cost altogether, and how much each payer paid, for events that were not matched to Medicare claims data on a service-by-service basis. These procedures are applied to home health and hospice services. (The procedures used for missing payments or payers for prescription drugs and facility utilization are described separately below. For information on the editing and creation of these types of utilization, refer to the Prescription Drugs and Long-Term Facility segments). Long-term facility and prescription drug utilization are presented in the Cost Supplement File as event-level records designated facility and prescription medicines.

Hospice Services

Hospice utilization is unusual in terms of Medicare administrative records because it is the only utilization that is recorded in two different ways, in two different files. The beginning and ending dates of the hospice benefit periods are recorded in Medicare Administrative enrollment data, while the bill records are part of the NCH repository. This dual reporting serves as an internal check on the dates of service on the billing records.

Determining and Imputing Payment Amounts for Hospice Services

With a target reimbursement amount (representing the "total cost" of the event) and payer indicator values indicating which payers contributed some payment toward the total, the share "amounts" paid by the individual payers could be determined.

If Medicare payments are known to be incomplete, then utilization for the missing periods is completed by editing from the existing billing records. The hospice benefit is paid on a per-diem basis, and the missing data are completed with average per diem rates calculated from existing bills. Virtually all services provided to the hospice beneficiary are fully covered by Medicare, and as there are no copayments or deductibles, there is no cost sharing (Prescribed medicines are an exception, as there may be a small copayment for drugs, which are reported separately, and also inpatient respite care for which the patient pays 5 percent of the Medicare allowed rate – typically under $5). Hence, the Medicare reimbursement is the target reimbursement, and Medicare is the sole payer of hospice bills. Hospice bills are not matched; as a result, there is some overlap between hospice utilization and events reported in the survey. The overlapping survey events are usually, but not always, home health events. When an unmatched survey event occurs at the same time as a hospice event (i.e., is an overlapping event), the event is retained but the payer indicators and amounts are zeroed out to prevent double counting.

Home Health

The home health use and payment records in the Cost Supplement File are designed to represent events where medical care, as opposed to personal care and support, is furnished to the beneficiary. The decision to include only medical services in the user file in no way derogates the importance of unpaid assistance in maintaining the health and well-being of Medicare beneficiaries. It simply reflects the primary emphasis of the MCBS Cost Supplement File, which focuses on use of, and payment for, formal medical care services.

For home health events with event periods that spanned two years, the first step is to allocate services proportionately into the calendar year. The rules used to do this were identical to the procedures detailed in section: Adjustments for Missing Days and Undated Use. At this stage, a home health "event" could have represented one or more home health visits. Bundled events with multiple visits were unbundled for the allocation of home health services across years.

Survey event records are originally classified in the interview according to whether a professional or a friend provided the home health services. This distinction is used in separating out home health services that are not medical in nature. In winnowing down the file to medical services only, the following decision rules are used to exclude non-medical home health services:

1. Exclude services provided by a friend where the out-of-pocket payment, if any, is equal to the total charge for the service. (The reasoning is that even if the friend was paid for delivering a service, it is very likely non-medical in nature if there was no other payer).
2. Exclude services provided by a professional where the out-of-pocket payment is equal to the total charge for the service AND the person answered NO to the question asking whether the professional gives nursing/medical treatment.
3. Exclude all housekeeping/cleaning services unless Medicaid is listed as a payer.
4. Exclude all "meals-on-wheels" types of services.

After these allocation and exclusion operations, the remaining survey reported medical home health services are matched (not at the event level but at the person level only) to Medicare bills for home health services. The survey reports and Medicare bills are combined to provide the most accurate and complete summary

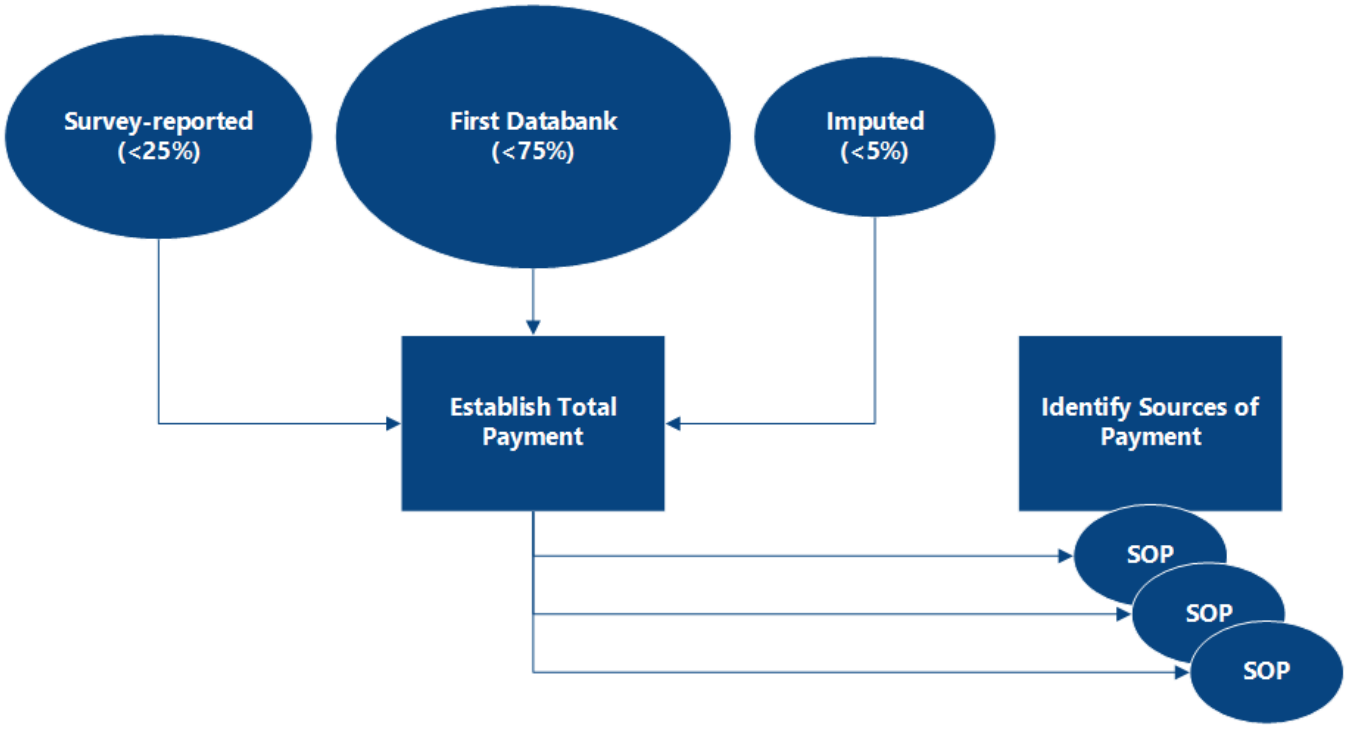
possible of the number of visits and payments (broken down by source of payment such as Medicare, out-of-pocket, etc.).

Prescription Drug Data

Prior to 2006, all prescription drug data were based on information collected in the survey. Due to the advent of Part D, beginning in 2006 Part D Events (PDEs) were available for Medicare beneficiaries enrolled in MA Prescription Drug Plans (MAPDs) or stand-alone Prescription Drug Plans (PDPs). For survey-reported events that matched to PDEs, information on the drug "claim" is used to calculate the total payment field as well as the other payer fields. The approach used to fill in missing drug payment data for unmatched events is similar to that used for other missing payment amounts described above. The first step is to establish a total payment amount for each drug event. First preference is given to using survey reports of the total payment for the drug. For approximately one-quarter of survey-only drug events on the file, the total payment is reported. For about three-quarters of the survey-only drug events, an administrative drug pricing source (FDB MedKnowledge (formerly NDDF Plus)) is used to impute prices. The administrative source is used only when no total payment is reported, and it is rarely used to supersede the survey reported payment. Finally, a small proportion (<5%) of survey-only drug events have total payments established using statistical imputation techniques.

After the total payment is established for each drug event, potential sources of payment are identified using a similar approach to that outlined earlier. In the last step, the total payment amount is distributed across the sources of payment (see Exhibit 7.3.7). In cases where a total payment is available from either a survey report or the FDB, unknown payment amounts for a specific payer are handled by accounting techniques and analytic edits before employing statistical imputation. In a small percentage of cases where the total payment is derived by statistical imputation, the payer amounts are also derived through statistical imputation.

Exhibit 7.3.7: Establishing Total Payment Amount and Allocating to Sources of Payers



Preparation of Survey-Reported Data

Prior to matching or imputation, the prescription drug data collected in the survey are edited for consistent spelling. MCBS staff edit the records to ensure that the same drug is always reported in the same way. All unique drug name spellings supplied in the survey for medications reported for all four survey panels through the end of the calendar year,[37] are gathered together in a single list. Using the FDB, MCBS staff manually assign corrected spellings to each unique supplied spelling.

Preparation of FDB (First Databank) Data

The FDB served as a pricing reference and as a source of therapeutic class for prescription medicines. However, survey reports of total payments were given preference over a FDB price because MCBS records and FDB records could not be matched exactly on all fields. The FDB generally identifies the name, form, strength, and packaging size of the drug in a single entry. The MCBS collected prescription size in the survey but could not collect the packaging size of the drug prescribed. In the survey, form and strength are also collected, but as separate items, not as part of the name. In the initial match, therefore, a FDB name "Septra DS Tab 800 mg" had to be changed to "Septra DS," to increase the likelihood of a match between the two sources on name.

Assignment of Wholesale Prices

In the FDB, a wholesale price is assigned to each National Drug Code (NDC) entry. The NDC is an 11-digit code; the first five digits define the manufacturer or labeler, the next four digits identify the drug product, and the last two digits identify the packaging size. As noted above, the MCBS does not collect packaging size, but instead collects prescription size, and unit average wholesale prices can differ substantially by packaging size. Using a relative frequency distribution of packaging sizes within each drug type, weighted by utilization rates from Part D prescription drug event data, MCBS staff developed a composite price for drugs that come in multiple packaging sizes.

After both survey data and FDB data are cleaned as described, survey prescription data are matched to the modified FDB information by drug name, form, strength, and packaging size, in that order, to develop a wholesale acquisition cost. Often, it is not possible to match on all four variables. If the survey drug name is not known or could not be matched, a wholesale acquisition cost is imputed. If the drug name is known but form or strength is not known, the missing characteristic is imputed and the wholesale acquisition cost is then obtained through a match to the FDB. For example, if the respondent reported a prescription of Diamox but did not know the strength, a wholesale acquisition cost is imputed using the weighted average price of all Diamox prescriptions. This weighted average price is developed using Part D prescription drug event data.

A small number of survey entries could not be translated to any drug listed in the FDB. In general, these entries were either misspellings that made it impossible to determine the drug name or not a specific drug (e.g., "little green pills"). These entries are classed as "untranslatable," and an average price is imputed based on frequency distributions of drug claims from Part D prescription drug event data. In some cases, the size of the prescription is known but the price is not. Average unit costs (per pill, per milliliter, etc.) are then multiplied by the prescription size, to derive a whole prescription cost. In other cases, prescription size is estimated through the respondent's answers to a series of probe questions, which are asked during the interview when the respondent did not know the size of the prescription.

Converting Wholesale Acquisition Cost into Event Price

Establishing a price for prescription drug records with no survey reported price begins with the assignment of wholesale acquisition cost. Event prices that are less than $0.50 are reset to missing and imputed statistically.

[37] The prescription drug data are cleaned for all panels included in the current LDS release.

Non-missing wholesale prices are multiplied by a pricing factor that varies depending on the likely payer(s) of the event. Six pricing factors are developed: retail, MCO, VA, Medicaid, employer sponsored and other public insurer. The retail pricing factor is actually a series of factors which reflect empirical evidence of the relationship between the wholesale acquisition cost and what the respondents reported paying. The managed care pricing factor estimates that MCOs pay approximately 150 percent of the wholesale acquisition cost of prescription medicines. The VA factor is developed using VA drug cost data that are provided by the Department of Veterans Affairs. The Medicaid pricing factor is developed using composite data from the CMS Medicaid Drug Rebate System, and included a dispensing fee, a discount off the wholesale acquisition cost, and a rebate percentage. The employer-sponsored insurer is slightly higher than the wholesale acquisition cost.

Determining Target Reimbursement

Target reimbursements are developed differently for prescription medicines than for other services (Target reimbursements for other types of services are described above). If Medicare claims data are unavailable, adjusted "event prices" are used to develop target reimbursements.

The target reimbursement is defined as the price that the beneficiary paid for a single purchase of a single drug. For single purchases (one unique medicine, purchased only once and not refilled), the price reported by the respondent is the target reimbursement.

If the respondent could not give a price, the event price, adjusted by the appropriate pricing factor (discussed below) is the target reimbursement. For multiple purchases (a single prescription, filled multiple times, or multiple prescriptions), the target reimbursement is developed as for single purchases and then divided by the number of purchases to yield a target reimbursement for each purchase.

If several drugs are reported together ("bundled"), but the total cost is not known, a target reimbursement is developed for each drug in the bundle, based upon the event price adjusted by the appropriate pricing factor. If several drugs are bundled together and a total cost is reported, that total cost is used to control the imputation of the individual drug prices. A relative percentage of the total cost is developed for each drug, using the event price adjusted by the appropriate pricing factor; those percentages are applied to the reported total cost, and the result becomes the target reimbursement for each drug. If the event price for one or more of the drugs in the bundle is missing, an average price for all strengths and forms of the drug is used in the computation, unless the drug name is not known; in those cases, an average event price is used. These averages are then used to calculate relative percentages, which are then applied to the amount reported in the survey for the bundle.

Determining Potential Payers

Potential payers for prescription medicines are determined in essentially the same way that potential payers are identified for matched utilization.

Adjustments for Missing Days and Undated Use

This section describes the adjustments made (at the person and service level, but not at the event-level) to:

1. Compensate for data that are missing because some periods of the beneficiary's Medicare entitlement were not covered by interviews. CMS administrative records are used to establish the exact period of Medicare entitlement during the calendar year and calculating the number of Medicare days;
2. Allocate undated survey events, primarily prescription drugs and some other medical equipment, between years where interview reference periods spanned two years.

Calculating Medicare Covered Days and Residence History

The periods of Medicare entitlement and living situations are established in order to validate and supplement utilization reported in the survey with information reported on claims and bills from the CMS national claims history database. This is accomplished by matching survey-reported utilization to the CMS records that was described earlier in the section on Filling the Gaps.

For most beneficiaries, the period covered by the survey and the period of the beneficiary's Medicare entitlement are identical: they both cover all days of the calendar year. However, the Medicare entitlement period may be longer than the period covered by the survey for beneficiaries who left the survey before the end of the calendar year, or died, and no proxy information is available. The most common reason for incomplete data is the beneficiary's refusal to participate further in the survey. Only beneficiaries who participated in the survey for at least 60 percent of the period they were eligible for Medicare during the year were retained for the Cost Supplement File.[38]

To identify Medicare entitlement and where the beneficiary was residing during the survey period (in the community or in a facility), three variables are provided in the Residence Timeline segment: total number of days entitled to Medicare (D_TDAYS); number of days where the beneficiary was living in the community (D_CDAYS); and number of days where the beneficiary was living in a facility (D_FDAYS).

Allocating Services Between Years

The cost and utilization data collected during the three interviews that occurred during the calendar year cover more than just that calendar year. Each interview serves as a boundary to the next interview – the respondent describes medical care that took place "since the last interview" – and those boundaries are generally not the beginning or ending of the calendar year. As a result, the first or last interviews generally include utilization that covers part of two calendar years. To adjust the utilization in these cases, dated event records are edited to remove those that took place outside of the calendar year, and undated events (prescription medicines) are pro-rated according to the number of calendar days in the interview reference period to total days in the reference period.

Simply pro-rating use between the two calendar years was considered but rejected. By assuming that use occurred in both years, this procedure could overstate the number and rate of persons using services in a year. In place of this, a random number generator is used to assign services to calendar years. The probability of an event being placed in the calendar year is based upon the ratio of calendar days in the reference period to total days in the reference period. For example, assume a reference period had 120 days, and 90 of these days were in the calendar year. For each event, a random number between 1 and 120 was generated. For all events where the random number was 90 or less, the service was allocated to the calendar year. For all events with random numbers between 91 and 120, the service was allocated to the other year.

Filling in Medicare Covered Days Not Surveyed: PM Data

When there is a gap in survey data, that is, a period for which a beneficiary was enrolled in Medicare but was not covered by a survey interview, it is necessary to estimate the medical service usage during that gap period. For beneficiaries with gaps who were interviewed in the calendar year, reported services are simply prorated upward to cover the gap. For example, for prescription drugs, the number of prescriptions per day is calculated for the interview period and multiplied by the number of gap days. This assumes, in effect, that the beneficiary used prescriptions at the same rate in the interview and gap periods. Likewise, to get adjusted

[38] This population aligns with the Cost Supplement File ever enrolled weights population. Please see Section 7.2 for more information on the ever enrolled weights.

sums for all payers, the cost per prescription per payer per day is calculated and multiplied by the adjusted number of prescriptions for each payer.

If the beneficiary is not surveyed due to a missed interview, a different approach is used. To cover these non-interview gap periods, a donor is selected who is similar to the person in terms of personal and economic characteristics. The donor's use of prescription drugs (measured in prescriptions per day and cost per prescription per payer per day) is used to impute use and payment data.

Filling in Medicare Covered Days Not Surveyed: Non PM Data

For non-prescription events with a gap in survey data, reported services are prorated upward to cover the gap. This proration is provided for the Non PM summary data. There is no adjustment made to the event-level data. For these gaps in survey data, no additional events are added, nor are the individual events for these beneficiaries adjusted; however, when the data are summarized to the person level or summarized to the service level, the proration is applied. The ratio applied is simply the ratio of the number of days in the year divided by the number of days for which survey data are available for the beneficiary. For example, if a beneficiary had a 100 day gap in a 365 day year, then survey data were available for 265 days, and the ratio would be (365 / 265) = 1.377.

For Incoming Panel beneficiaries, a different approach is taken to estimate data for gap periods. For most of these beneficiaries, the first survey interview does not cover medical expenses going back to the date of enrollment. Therefore, there is a need to estimate medical events from the beneficiary's date of enrollment through the earliest of either: the date of their next interview, their date of death if they have passed away, their date of lost entitlement if applicable, or December 31st. Since medical claims data that have been processed through FFS Medicare are available, these claims do not need to be estimated. However, there are many medical events that are either covered by MA or not covered by traditional FFS Medicare. Examples of these uncovered events may include dental visits or regular physician visits if the beneficiary does not have Medicare Part B. For these cases, new records are created to fill in Medicare covered days that have not been surveyed in a process known as unit-level imputation.

In the unit-level imputation process, a donor beneficiary is selected from a pool of beneficiaries with similar demographic attributes and insurance coverage to the Incoming Panel beneficiary. Selected data records are copied from the donor beneficiary and assigned to the Incoming Panel beneficiary. Only medical events that occurred within the Incoming Panel beneficiary's gap period are retained. If the donor beneficiary did not have any medical events that were not covered by FFS Medicare during the Incoming Panel beneficiary's gap period, then no additional data records are created. The new data records have payment information and amounts provided by the donor beneficiary's data record, but demographic and other variables associated with the Incoming Panel beneficiary are retained.

Determining and Imputing Payment Amounts for Long-Term Care Facility Events

The payments for time periods when a beneficiary lived in a long-term care facility contained on the FAE segment are determined from the billing period and payment information collected via the Facility Instrument. This information is not always available for collection, so imputation is needed to fill in payments for these missing periods.

To impute payments for the periods of missing data, information is gathered from time periods where payment information was reported. The reported payments for basic and ancillary services from the long-term care facilities are summed over the period of provided data and then divided by the number of days within that period. Medicare payments are excluded from this calculation because these payments are unlikely to be missing. The resulting daily rate is then applied to estimate payments for the periods of missing data by multiplying the daily rate by the number of days within the period of missing data. For some beneficiaries,

either no payment information is available or the only known payments are from Medicare. For these beneficiaries, standard daily or monthly rates are gathered from the given facility, and these are used to estimate a reasonable daily rate, which is then used to estimate payments for the periods of missing payment information. A small number of facilities do not have the standard daily or monthly rates available. For beneficiaries living at these facilities, a median daily rate is calculated based on the rates provided by the facilities that had standard rates available.

Once the total payments for each stay at a facility with missing payment information are estimated, then amounts are allocated to the appropriate source of payment (private insurance, Medicaid, out of pocket, supplemental security, VA, or other) based on the allocations observed for periods where payment information was available. If there is no observed payment information available for a given beneficiary, averages are calculated based on data from beneficiaries that did have payment information available. Finally, the imputed payments are reviewed for extreme values and edited if necessary.

7.5.3 Analytic Edits of Sources of Payment (SOP) Values for Non PM Events

The general goal of the analytic edits is to resolve as many events as possible (i.e., to fully allocate total charges to payers) and to set as many payer indicator values as possible based on logic. The edits resolve some events without using a hot deck procedure to impute payment sources or amounts.

Medicaid

Analytic edits are used extensively when Medicaid is a potential or actual source of payment for an event. One set of edits—designed to reflect the role of Medicaid as the payer of last resort—ensures that Medicaid could not be a payer if payments are reported or imputed for another third-party insurer (except Medicare), or if the provider was an MCO or VA facility. Another set of edits is developed for dually eligible beneficiaries whose cost-sharing liability is covered by Medicaid.

Out-of-pocket payments are allowed when Medicaid was a payer only if the respondent is able to report the out-of-pocket amount(s). Medicaid may cover copays and deductibles for dually eligible beneficiaries and Qualified Medicare Beneficiaries such that the respondent has no out-of-pocket costs for Medicare-covered services.

Private and Medicare MCOs

MCOs (especially Medicare-contracting MCOs) often operate differently from other third-party payers and tend to have unique payment patterns. For instance, risk and (to a lesser extent) cost Medicare MCOs are paid a set fee per enrolled Medicare beneficiary (called a capitated amount) designed to compensate the MCO for the expected costs of delivering Medicare's package of benefits. There are no Medicare claims or insurance statements indicating the total charge for events covered by the capitated amount.

Often the respondent only knows the copay amount if there was one. Also, MCOs often provide "Medigap"-type coverage by paying for most of the deductibles and copays for Medicare-covered benefits. A beneficiary who belongs to an MCO does not need private Medigap insurance or Medicaid coverage for these amounts. Thus, payment patterns for MCO beneficiaries tend to be simpler than those for FFS beneficiaries. The set of analytic edits for MCOs attempt to account for these simplified patterns and for the respondent's usual inability to report charges and payments for events. The MCO edits also attempt to avoid creating "illogical" payment patterns.

General Edits

At the beginning of the analytic editing, and after each main section of edits, an attempt is made to resolve events through addition or subtraction. Events without a known total charge but with a complete payment vector (i.e., each payer is identified as either having paid or not paid for an event and each payer's amount was known) are completed by summing across all payment sources to derive the total charge. Events with a

known total charge and complete except for one missing payment amount or payment source are completed by subtraction. The excess of charges over known payment amounts is attributed to the known payer, or the one missing payer indicator is set to 1 and the excess allocated to that payer. If a service is provided free of charge, all payer indicators and payment amounts are set to 0.[39]

If a source is a potential payer for an event, or if the respondent reported that the payer had contributed to an event but did not know the amount, it is assumed that the payer is not actually a source if the current sum of reported payments equals the reported total charge.

7.5.4 Setting SOP Flags

Each beneficiary's health insurance timeline, survey-reported events and Medicare claims are used to establish an indicator variable (SOP flag) for each of the source of payment (SOP) categories. Information in the SOP flags is, in turn, used to determine the corresponding payer indicator variables, which are used in imputation to determine whether or not a possible source of payment actually paid something toward the cost of an event.

This section outlines the rules that apply to the process of setting the values of the SOP flags. SOP flags can have one of five possible values:

- 0 - Source definitely did not pay
- 1 - Source definitely did pay, known amount
- 2 - Source definitely did pay, unknown amount
- 3 - Source possibly paid, beneficiary was covered at time of event
- 4 - Source possibly paid, beneficiary may have been covered at time of event

SOP Medicare

Medicare Part A and Part B entitlement dates established the period of Medicare coverage.

1. If the sample beneficiary is entitled to Medicare Part A benefits, Medicare is a potential source of payment for: Inpatient hospital – IP events, SNF – IU events, and Home Health – HH events (HHP and HHF events). The initial value of the Medicare SOP flag is 3 (possible payer) for these event types.
2. If the sample beneficiary is entitled to Medicare Part B benefits, Medicare is a potential source of payment for: Outpatient hospital – OP events and Part B Physician/Supplier services – DU, VU, HU, ER, HP, HF, MP, SD, SL, and OM events. The initial value of the Medicare SOP flag is 3 (possible payer) for these event types.

SOP Medicaid

If either the respondent or CMS administrative data indicates that the sample beneficiary had Medicaid coverage, the Medicaid SOP flag is initially set to 3 for all events which occurred during the period of Medicaid coverage.

SOP Managed Care

The managed care flag is set based on information in the beneficiary's health insurance timeline and the CMS administrative records of managed care enrollments.

[39] If the event was reported as free, but the respondent had also reported that a source other than Medicare or Medicaid had paid something for the event, the total charge was set to missing and imputed.

1. If CMS administrative records indicate that the beneficiary is enrolled in a Medicare managed care plan, but the beneficiary did not report the enrollment, the Managed care SOP flag is initialized to a value of 4 for all events that occurred during the beneficiary's enrollment.
2. The MCO SOP flag is set to 4, for all events except DU, VU, HU, and PM, if the health insurance section shows that the beneficiary is in an MCO, whether or not it is a Medicare MCO.
3. For DU, HU, VU, and PM events, the MCO SOP flag is initialized to 3 if the respondent indicated that the MCO covers the associated services (dental, hearing, vision, and prescription medicine, respectively), otherwise the MCO SOP flag is initialized to 4.

SOP Veterans Administration

Beginning with the 2016 data year, the VA SOP flag is combined with the "other public insurance" source of payment flag. Please see section Other Public Insurance for a description of that source of payment.

SOP Private Health Insurance

Employer-based information about private health insurance (PHI) as a payment source is provided in the insurance section of the interview, by the respondent, and through insurance statements. Information about the source of the policy (used to differentiate between employer-sponsored and individually purchased private health insurance) is also provided by the respondent in the insurance section of the interview.

1. The employer-sponsored PHI SOP flag is set to 3 for all types of services, except prescribed medicines, which occur while the sample beneficiary is covered by employer-sponsored health insurance, based on the health insurance timeline and the date of the event.
2. For prescribed medicines, employer-sponsored health insurance is considered a possible source of payment (initial value SOP=3) if the respondent said that the plan covered drugs. If the respondent said that the plan did not cover drugs but reported a specific amount the plan paid for another PM event, the employer-sponsored PHI SOP flag for all PM events during the same round is set to 4.
3. If the event date is missing or ambiguous and the sample beneficiary's insurance coverage changed during the round, the employer-sponsored PHI SOP flag is set to 4 instead of 3 where applicable.

Information about individually purchased private health insurance as a payment source is provided in the insurance section of the interview, by the respondent, and through insurance statements. Information about the source of the policy (used to differentiate between employer-sponsored and individually purchased private health insurance) is also provided by the respondent in the insurance section of the interview.

1. The Individually Purchased PHI SOP flag is set to 3 for all event types, except prescription medicines, which occur while the sample beneficiary is covered by individually purchased private health insurance, based on the beneficiary's health insurance timeline and the date of the event.
2. For prescription medicines, the Individually Purchased PHI SOP flag is set to 3 if the respondent reported that the individually purchased PHI plan covered drugs. If the respondent said the plan did not cover drugs but reported a specific amount the plan paid for another prescription medicine, the Individually Purchased PHI SOP flag is set to 4 for all prescription medicines reported in the same round.
3. If the event date is missing or ambiguous, and the sample beneficiary's insurance coverage changed during the round, the Individually Purchased PHI SOP flag is set to 4 instead of 3 where applicable.

SOP Out-of-Pocket and SOP Uncollected Liability

The out-of-pocket and uncollected liability flags are not set based on health insurance timelines. In many cases, these two categories could not be ruled out as payers based on the health insurance timeline, or even after the claims match.

SOP Other Public Insurance

Beginning with the 2016 data year, the VA SOP flag was combined with the "other public insurance" source of payment. The following describes how the other public insurance source of payment is determined based on VA and other public insurance information provided in the interview by the respondent.

1. For all event types except prescription medicines, if the respondent indicated that the service is provided by a VA hospital or clinic, or if the respondent reported coverage by "other public insurance", then the other public insurance SOP flag is set to 3. If the respondent was not certain that the service was provided by the VA, or if the respondent was uncertain of having "other public insurance" coverage, then the other public insurance SOP flag is set to 4.
2. For prescription medicines, the other public insurance SOP flag is set to 4 if the VA or "other public insurance" paid a known amount for some other drug in the same round.

Updating SOP Flags Using Survey-Collected Cost Data

The initial values of the SOP flags are updated when survey-collected cost data provide more definitive information. If the respondent reported the amount the payer paid, the appropriate SOP flag is set to 1. If the respondent did not know how much the payer paid, the SOP flag is set to 2.

Updating SOP Flags Based Upon Matching Medicare Claims Data

The initial values of the SOP flags are also updated when the utilization could be linked to Medicare claims records.

The Medicare payment amount and the Medicare SOP flag are updated if the survey-reported utilization matches Medicare claims data, or if the Medicare claims data provide the only record of the utilization. If the Medicare claims record shows a positive, non-zero Medicare payment, the Medicare SOP flag is set to 1, to show that the payment amount is known and would not have to be imputed. If the claims record shows that the sample beneficiary's Medicare benefits were exhausted, the Medicare SOP flag is set to 1, and the Medicare payment amount is set to $0.00. If the claims record indicates that the service was not a Medicare covered service or that the beneficiary did not have Medicare coverage for the service, both the Medicare payment amount and the Medicare SOP flag are set to zero.

If the claims record showed that Medicare is a secondary payer, the appropriate SOP flag for the primary payer is set to 1 (identifying the insurer as the primary payer), and the Medicare claim is used to develop the amount paid by the private insurer.

Updating SOP Flags Based on Unmatched "Survey Only" Utilization

The Medicare SOP flag is set to zero for all unmatched survey events unless the Medicare SOP flag already has a value of 1 or 2.

Resolving Conflict Between Person-Level Survey Reported Health Insurance Information and Event-Specific Survey Reported Source of Payment Data

For a very limited number of events (less than 0.5 percent) the reported source of payment data conflict with the individuals' reported health insurance information. In these situations, the payment data are manipulated to conform to the health insurance data for the following payers: Medicaid, Employer Sponsored Health Insurance, managed care coverage, and Medigap insurance. Since Out-of-Pocket and Uncollected Liability are always potential payers, there are no inconsistencies for these payer categories. Inconsistencies in Medicare SOP data are not resolved, but unmatched survey reported events with Medicare payment and no MA payment are excluded from the payment summaries' adjusted totals .

8. RESPONSE RATES AND NONRESPONSE

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

8.1 Response Rates

The unconditional response rate is the percentage of sample that was released during the fall round of the calendar year and responded to the survey in in the calendar year. 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 the previous calendar year and also responded during the current calendar year. Conditional response rates use the sample who responded during the previous calendar year as the baseline in their calculation. In other words, they are conditioned on response in the previous calendar year.

8.1.1 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 $t =$ current 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 the year t cohorts 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.

8.1.2 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.

9. REFERENCES

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

Gower, John C. "A General Coefficient of Similarity and Some of Its Properties." Biotmetrics 27, no. 4 (1997). 857-871

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

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

10. APPENDIX

Appendix A. Table of Links to MCBS Documentation

MCBS Resources	Links
CMS MCBS Website	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey
MCBS LDS File Information	https://www.cms.gov/data-research/files-for-order/data-disclosures-and-data-use-agreements-duas/limited-data-set-lds
MCBS Microdata PUFs	https://data.cms.gov/medicare-current-beneficiary-survey-mcbs
CMS Chronic Conditions Warehouse (CCW)	https://www.ccwdata.org/web/guest/home/
Data User's Guides, Data Year Release Notes, Methodology Reports, Codebooks, and LDS Variable Crosswalks	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks
PUF Table Packages and Chartbook PDFs	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-tables
Early Look, Data Briefs, Infographics, Posters, and Tutorials	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials
Bibliography	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/bibliography
Questionnaires and Questionnaire User Documentation	https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/questionnaires
Glossary	https://www.cms.gov/files/document/mcbs-glossary.pdf