

# DATA USER'S GUIDE



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

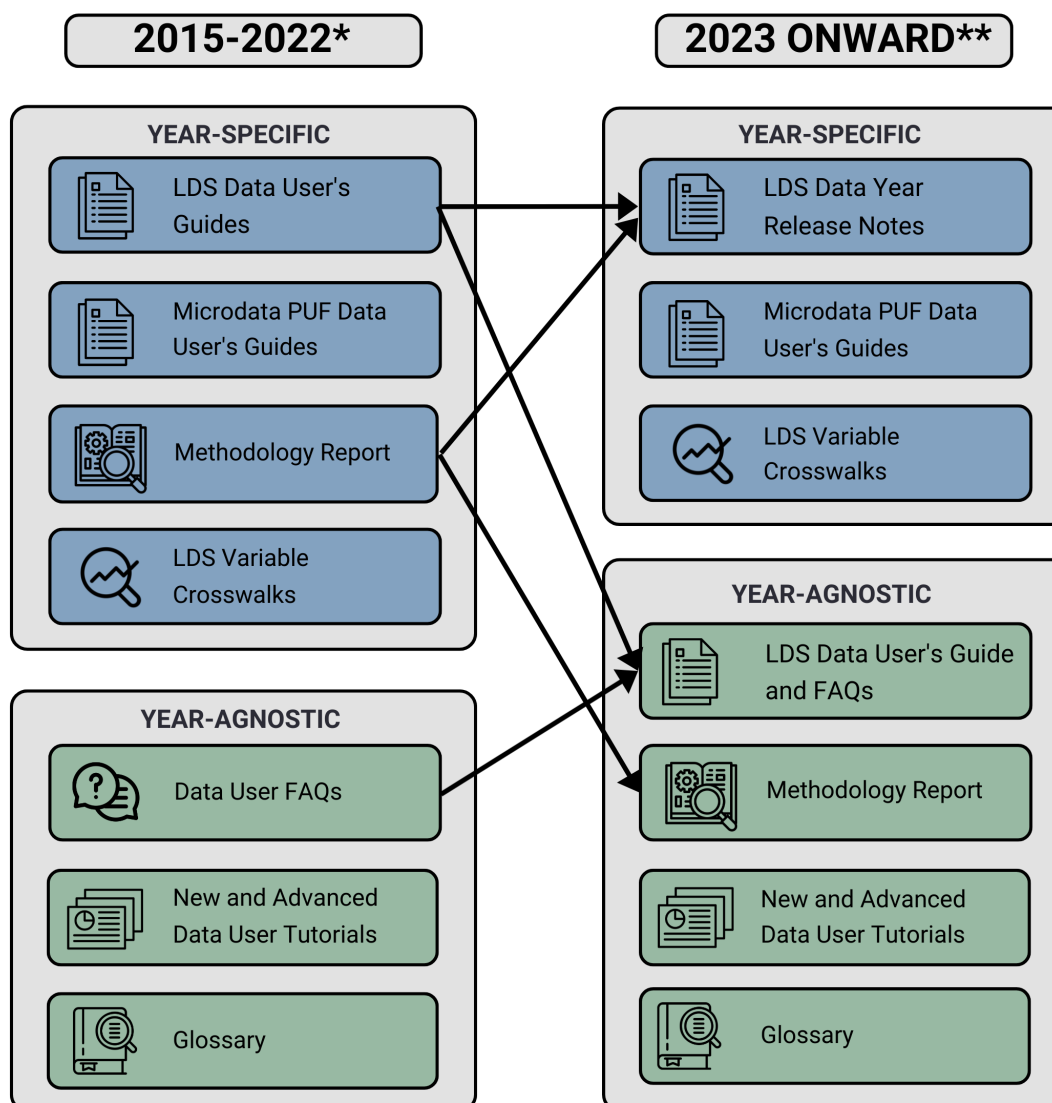
## Version Control Log

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

<b>MCBS DOCUMENTATION CROSSWALK AND OVERVIEW.....</b>	<b>ii</b>
<b>ACRONYM LIST.....</b>	<b>vii</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 Contents of the Data User's Guide.....	2
<b>2. GENERAL GUIDELINES FOR DATA USE.....</b>	<b>4</b>
2.1 Data Access.....	4
2.2 Guidelines for Citation of Data Source .....	4
<b>3. OVERVIEW OF THE MCBS .....</b>	<b>6</b>
3.1 Design of MCBS .....	6
3.2 Sample Design.....	6
3.3 Case Types.....	7
3.4 Questionnaires.....	7
3.5 Medicare Population Covered by MCBS Data .....	7
<b>4. DATA PRODUCTS &amp; DOCUMENTATION.....</b>	<b>9</b>
4.1 Contents of Data Release .....	9
4.1.1 MCBS Survey File - Early Release .....	10
4.1.2 MCBS Survey File .....	10
4.1.3 MCBS Cost Supplement File .....	11
4.1.4 Using the Data.....	11
4.2 Which File Do I Need? .....	12
4.2.1 Using the Survey File - Early Release.....	12
4.2.2 Using the Survey File and Cost Supplement File .....	12
4.2.3 Using Both Community and Facility Data.....	13
<b>5. SURVEY FILE - EARLY RELEASE NOTES .....</b>	<b>15</b>
5.1 LDS Specifications.....	15
5.2 File Structure.....	15
5.3 Analytic Guidance .....	15
<b>6. SURVEY FILE NOTES.....</b>	<b>16</b>

6.1 LDS Specifications.....	16
6.2 File Structure .....	16
6.2.1 Sort Order for Merging the Survey File LDS Segments .....	16
6.3 Claims Files .....	17
6.3.1 Utilization Detail Records .....	17
6.4 Analytic Guidance .....	18
<b>7. COST SUPPLEMENT FILE NOTES.....</b>	<b>19</b>
7.1 LDS Specifications.....	19
7.2 File Structure .....	19
7.3 Analytic Notes.....	21
<b>8. DATA FILE DOCUMENTATION.....</b>	<b>22</b>
8.1 LDS Contents.....	22
8.2 LDS Components .....	22
8.2.1 Codebooks .....	22
8.2.2 Questionnaires.....	23
8.2.3 Data User Resources.....	23
8.3 Data Edits and Imputation .....	23
8.3.1 Data Edits .....	23
8.3.2 Imputation .....	24
8.4 Weighting.....	25
8.4.1 Preparing Statistics (Using the Full Sample Weights) .....	25
8.4.2 Special Topical Segment Weights .....	28
8.5 Using the Data.....	29
8.5.1 Merging Segments .....	29
8.6 Variance Estimation (Using the Replicate Weights) .....	29
8.6.1 Variables Available for Variance Estimation .....	29
8.6.2 Variance Estimation for Analyses of Single Year of MCBS .....	30
8.6.3 Subgroup Analysis.....	30
8.7 Combining Multiple Years of Data.....	31
8.7.1 Longitudinal Analysis.....	31
8.7.2 Repeated Cross-Sectional or Pooled Analysis.....	32
8.8 Data File Notes: Global Information.....	34
8.8.1 Key Variables.....	34
8.8.2 Item Non-Response and Missing Values.....	34
8.8.3 Derived and Administrative Variables.....	34
8.8.4 Initial Interview Variables .....	35

8.8.5 Ever Variables.....	35
8.8.6 Other Specify Questions .....	35
8.8.7 Interview Mode Indicator.....	36
8.8.8 Analytic Notes for Non PM Event Segments.....	36
8.8.9 Do Administrative Data Override Survey-Reported Data? .....	36
<b>9. REFERENCES .....</b>	<b>37</b>
<b>10. APPENDICES.....</b>	<b>39</b>
Appendix A: Sample Code .....	39
Example A.1 Joining Segments within the Survey File and Survey File - Early Release LDS....	39
Example A.2 Merging Segments in the Survey File and Cost Supplement File .....	40
Example A.3 Using Event-Level Data in the Cost Supplement File.....	42
Example A.4 Using Service Summary Data in the Cost Supplement File .....	43
Example A.5 Combining Topical Survey File and Cost Supplement File Data .....	44
Example A.6 Variance Estimation .....	47
Example A.7 Repeated Cross-Sectional or Pooled Analysis .....	55
Example A.8 Longitudinal Analysis .....	57
Appendix B: Table of Links to MCBS Documentation.....	59
Appendix C: Frequently Asked Questions.....	60
C.1 Data Requests.....	60
C.2 File Corrections .....	61
C.3 Content and Methodology .....	61
C.4 Sampling .....	67
C.5 Analysis .....	68

LIST OF EXHIBITS

Exhibit 3.1.1: Typical MCBS Data Collection Year.....6

Exhibit 3.5.1: Medicare Population Covered by the MCBS Data .....8

Exhibit 4.1.1: Contents of MCBS LDS Data Releases .....10

Exhibit 4.2.1: Comparison of Ever Enrolled MCBS Populations in Survey File - Early  
Release and Survey File.....12

Exhibit 4.2.2: Comparison of MCBS Populations in Survey File and Cost Supplement  
File.....13

Exhibit 6.2.1: Sort Order by Segment in the Survey File LDS .....16

Exhibit 8.3.1: Data Review and Editing Codes.....24

Exhibit 8.7.1: Constructing a Longitudinal Analytic File.....31

Exhibit 8.7.2: Constructing a Repeated Cross-Section or Pooled Analytic File.....32

Exhibit A.2.1: Example A.2 Recoding Variables .....41

Exhibit A.5.1: Example A.5 Recoding Variables .....46

Exhibit C.3.1: Timeline of MCBS Structure Changes .....62

Exhibit C.3.2: Overview of MCBS Limited Data Sets Beginning with the 2023 Data Year ...63

Exhibit C.3.3: COVID-19 Data Collection and Corresponding Survey File Releases.....66

## ACRONYM LIST

BRR	Balanced repeated replication (or Fay's method)
CAPI	Computer-Assisted Personal Interviewing
CASPER	Certification and Survey Provider Enhanced Reports
CCN	CMS Certification Number
CCW	Chronic Conditions Warehouse
CENWGTS	Survey File Continuously enrolled weights
CMS	Centers for Medicare & Medicaid Services
CSEVWGTS	Cost Supplement File Ever enrolled weights
CSL2WGTS	Cost Supplement File Longitudinal weights (2-year)
CSL3WGTS	Cost Supplement File Longitudinal weights (3-year)
DME	Durable Medical Equipment
DUA	Data Use Agreement
EOB	Explanation of Benefit Statement
EPPE	Enterprise Privacy Policy Engine
ER_EVRWGTS	Survey File - Early Release Ever enrolled weights
EVRWGTS	Survey File Ever enrolled weights
FFS	Fee-for-Service
HIPAA	Health Insurance Portability and Accountability Act
LDS	Limited Data Set(s)
LNG2WGTS	Survey File Longitudinal weights (2-year)
LNG3WGTS	Survey File Longitudinal weights (3-year)
LNG4WGTS	Survey File Longitudinal weights (4-year)
MA	Medicare Advantage
MCBS	Medicare Current Beneficiary Survey
MDS	Minimum Data Set
NORC	NORC at the University of Chicago
OEDA	Office of Enterprise Data and Analytics
PDE	Prescription Drug Event
PHI	Protected Health Information
PSU	Primary Sampling Units
PUF	Public Use File
RIC	Record Identification Code
SAS	Statistical Analysis System
SNF	Skilled Nursing Facility
SSU	Secondary Sampling Units
VRDC	Virtual Research Data Center



# 1. INTRODUCTION

Medicare is the nation's health insurance program for persons 65 years and over and for persons younger than 65 years who have a qualifying disability. 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).<sup>1</sup> The MCBS is designed to aid CMS in administering, monitoring, and evaluating the Medicare program. A leading source of information on Medicare and its impact on beneficiaries, the MCBS provides important information on beneficiaries that is not otherwise collected through operational or administrative data on the Medicare program and plays an essential role in monitoring and evaluating beneficiary health status and health care policy.

The MCBS is a continuous, multi-purpose longitudinal survey, representing the population of beneficiaries aged 65 and over and beneficiaries aged 64 and below with certain disabling conditions, residing in the United States. Interviews are conducted in-person and over the phone using computer-assisted personal interviewing (CAPI). The MCBS has conducted continuous data collection since 1991, completing more than 1.2 million interviews provided by thousands of respondents.

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

Annually, CMS releases five sets of files – 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. The data within the LDS releases are organized into data segments.

**Data from the MCBS have been used to inform many advancements to the Medicare program, including the creation of new benefits such as Medicare's Part D prescription drug benefit.**

- The **Survey File - Early Release** serves as a standalone research file and is generally released within nine months after the close of the calendar year for that data collection cohort. The Survey File - Early Release contains timely data on a subset of key topics released on the Survey File, including beneficiaries' demographic information, health status and conditions, and access to and satisfaction with care.
- The **Survey File** serves as a standalone research file and is generally released 18 months after the close of the calendar year for that data collection cohort. The content of the Survey File is governed by its central focus of serving as a unique source of information on beneficiaries' health and well-being that cannot be obtained through CMS administrative sources alone. The Survey File includes data related to Medicare beneficiaries' access to care, health status, and other information regarding beneficiaries' knowledge, attitudes towards, and satisfaction with their health care. The data release also contains demographic data and information on all types of health insurance coverage as well as Fee-for-Service (FFS) claims data, which provide information on medical services and payments made by Medicare under this plan type. Some data for the Survey File are collected into the next calendar year to provide a complete picture of beneficiaries' health and well-being for analysis.

<sup>1</sup> The MCBS is authorized by section 1875 (42 USC 1395II) 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.

- The **Cost Supplement File** is usually released approximately three months after the Survey File, when data collection has ended and final administrative and claims data for that calendar year become available. The Cost Supplement File contains a comprehensive accounting of beneficiaries' health care use, expenditures, and sources of payment. Note that for analyses of beneficiaries' health care costs and utilization, data users will need to use the Cost Supplement File in conjunction with the Survey File.

Beginning with the 2023 data year, each LDS data release is documented in the annual *Data Year Release Notes*, which offers a concise, publicly available, and easily searchable resource for data users that is tailored to new information for each data year. Detailed descriptions of each LDS release, including the contents of the files, file structure, and data sources are included in this document.

Information on content and access to the MCBS Microdata PUFs, including codebooks and additional documentation on the data release, can be found at <https://data.cms.gov/medicare-current-beneficiary-survey-mcbs>.

For questions or suggestions on this document or other MCBS data-related questions, please email [MCBS@cms.hhs.gov](mailto:MCBS@cms.hhs.gov).

## 1.1 Contents of the Data User's Guide

This *Data User's Guide* contains detailed information about the annual MCBS LDS releases and specific background information to help data users understand and analyze the data. Data users can access this *Data User's Guide* along with other data documentation at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>. For a detailed description of the questionnaires, data collection protocols, and file creation, weighting, and imputation processes, see the *Methodology Report*.

Here is an overview of the contents of the *Data User's Guide*:

- Section 2: General Guidelines for Data Use – This section describes the main requirements for data use.
- Section 3: Overview of the MCBS – This section provides a high-level overview of the MCBS, including the survey and sample design, questionnaires, and populations covered by the MCBS.
- Section 4: Data Products & Documentation – This section provides an overview of the MCBS data releases.
- Sections 5-7: LDS Notes – These sections provide information specific to each LDS release.
- Section 8: Data File Documentation – This section provides technical information on the LDS data releases, the data editing, imputation, and weighting processes, and use guidance.
- Sections 9-10: References and Appendices – These sections provide references and key supporting documentation, including sample code, links to resources, and frequently asked questions (FAQs) for data users.

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.<sup>2</sup> 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).

<sup>2</sup> <https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html>

- 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.

## 2. GENERAL GUIDELINES FOR DATA USE

The LDS files contain beneficiary-level health information, but exclude specific direct identifiers as outlined in the Health Insurance Portability and Accountability Act of 1996 (HIPAA). LDS files are considered identifiable, even without the inclusion of specific direct identifiers, due to the potential capability to link other sources of data, creating an increased risk of re-identification of individuals. Since the information provided on an LDS is considered identifiable, it also remains subject to the provisions of the Privacy Act of 1974.

### 2.1 Data Access

All requested LDS files require a signed LDS Data Use Agreement (DUA) between CMS and the data requestor to ensure that the data remain protected against unauthorized disclosure. LDS requestors must show that their proposed use of the data meets the disclosure provisions for research. The research purpose must relate to projects that could ultimately improve the care provided to Medicare patients and policies that govern the care. This type of research includes projects related to improving the quality of life for Medicare beneficiaries, improving the administration of the Medicare program, cost and payment related projects, and the creation of analytical reports. In addition, these research projects must contribute to generalizable knowledge.

Data users can submit an LDS request via a CMS DUA tracking system, the Enterprise Privacy Policy Engine or EPPE. EPPE can be used to initiate a new LDS DUA request or to amend/update an existing LDS DUA.

Questions about LDS files or the process for requesting LDS files can be sent to [datauseagreement@cms.hhs.gov](mailto:datauseagreement@cms.hhs.gov). For additional information on data access and the DUA process, including instructions for accessing and using EPPE to make a request, data users can visit the CMS LDS website at <https://www.cms.gov/data-research/files-for-order/data-disclosures-and-data-use-agreements-duas/limited-data-set-lds>.

Information on the administrative processing fees for obtaining the LDS files is available here: <https://www.cms.gov/data-research/files-for-order/limited-data-set-lds-files/medicare-current-beneficiary-survey-mcbs>. The processing of the DUA takes approximately six to eight weeks. Upon approval and payment, CMS releases the data within ten business days, depending on the size of the data request. Data users will receive the data via physical data shipment, or via the CMS Virtual Research Data Center (VRDC) for use with SAS® or other statistical software packages; each data release contains multiple files that are linkable through a key identification variable (BASEID).

Questionnaires, codebooks, and Bibliographies for each survey year are available for download on the CMS MCBS website at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey>. A link to this documentation is also visible when approved data users log in to the VRDC.

### 2.2 Guidelines for Citation of Data Source

This communication was printed, published, or produced and disseminated at U.S. taxpayer expense. All material appearing in this document is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated. Accordingly, CMS requests that data users cite CMS and the MCBS as the data source in any publications or research based upon these data. Suggested citation formats are below.

**Tables and Graphs:** The suggested citation to appear at the bottom of all tables and graphs should read:

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, [Data File (e.g., Survey File Limited Data Set)], [Year].

**Bibliography:** The suggested citation for the *MCBS Data User's Guide* should read:

SOURCE: Centers for Medicare & Medicaid Services. *Medicare Current Beneficiary Survey Data User's Guide*. Retrieved from [URL], [Year Accessed].

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

SOURCE: Centers for Medicare & Medicaid Services. Medicare Current Beneficiary Survey, [Data File (e.g., Cost Supplement File Limited Data Set)] data. Baltimore, MD: U.S. Department of Health and Human Services, [Year].

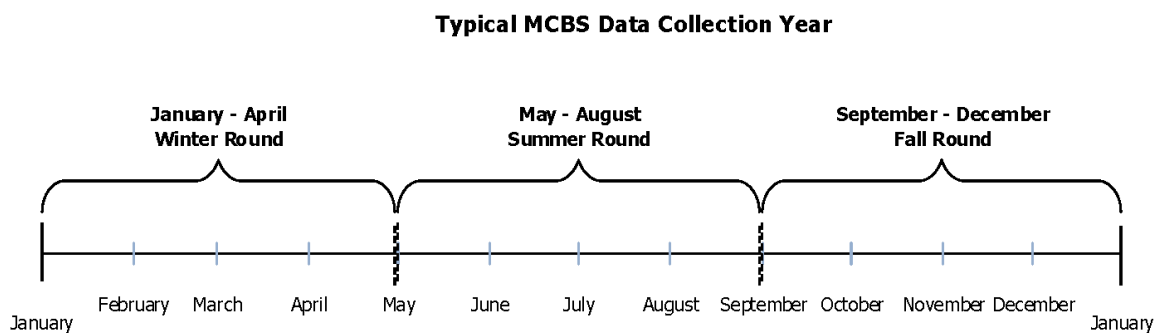
## 3. OVERVIEW OF THE MCBS

### 3.1 Design of MCBS

MCBS data collection is continual throughout the year with three distinct seasons (i.e., rounds) of data collection per year. In general, the three rounds are: winter (January through April); summer (May through August); and fall (September through December). The primary reason for the round by round configuration (rather than interviewing on an annual basis) is to have shorter periods of recall during the year in order to capture more complete health care costs and utilization from beneficiaries.

The MCBS data releases for each data year reflect data collected from January through December of that data collection year, as well as data on income and assets, access to care, usual source of care, preventive care, COVID-19, beneficiary knowledge and information needs, drug coverage, and chronic pain information collected through the winter and summer rounds of the following data collection year.<sup>3</sup> Exhibit 3.1.1 depicts an MCBS data collection year and the typical span of the rounds.

**Exhibit 3.1.1:** Typical MCBS Data Collection Year



Initial interviews of newly selected beneficiaries take place in the fall round. For these initial interviews, the fall round typically begins in July, rather than September, to allow more time to conduct outreach and collect information from the new survey respondents who are selected to participate in the MCBS. Subsequent rounds, which occur every four months, involve re-interviewing of the same beneficiary (or appropriate proxy respondents or Facility staff) until they have completed four years of participation (up to 11 interviews in total). Interviews are conducted regardless of whether the beneficiary resides at home or in a long-term care facility, using a questionnaire version appropriate to the setting.

### 3.2 Sample Design

The MCBS uses a rotating panel sample design, covering the population of Medicare beneficiaries residing in the continental U.S. (48 states and the District of Columbia) for the data collection year.<sup>4</sup> Each MCBS panel, an annual statistical sample of all Medicare enrollees, is interviewed up to three times a year over a four-year

<sup>3</sup> Due to the nature of some survey items, LDS data for each data year may include data pulled forward from a prior data collection year and/or data added from a future data collection year due to the specific reference period.

<sup>4</sup> Alaska and Hawaii are not included among the states from which the sample is selected due to the high cost of data collection in those areas; however, they are included in control totals for weighting purposes. Beginning in 2017, sampling from Puerto Rico was discontinued. Beginning in 2018, all data collection in Puerto Rico was discontinued.

period, creating a continuous profile of selected beneficiaries' health care experiences.<sup>5</sup> One panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round. The size of the new panel is designed to provide a stable number of beneficiaries across all panels participating in the survey annually. For more information about MCBS sampling processes, see the *MCBS Methodology Report*.

### 3.3 Case Types

MCBS respondents are classified by their phase of participation (i.e., Incoming or Continuing) and interview participation (i.e., Community or Facility), which is determined by residence status. These case types are described below. For more information about MCBS data collection procedures and interviewing, see the *MCBS Methodology Report*.

- **Incoming and Continuing Cases:** Every fall, a new panel of sampled beneficiaries is added to the total sample to replace the panel of beneficiaries completing a final interview and exiting the MCBS in the prior winter round. Newly selected beneficiaries who begin in the fall round are referred to as Incoming Panel cases. Their initial interview is referred to as the Baseline interview. After the initial interview, they are referred to as Continuing cases.
- **Community Interviews and Facility Interviews:** Approximately 93 percent of the interviews are conducted with beneficiaries or proxies who are living in their own residence or with family or friends. These interviews are called Community interviews; the remaining 7 percent of the interviews are for beneficiaries living in a facility. Over the course of a four-year period, it is not uncommon for beneficiaries to enter long-term care facilities (e.g., nursing homes) or go back and forth between the community and a facility setting (these cases are called Crossovers). To obtain an accurate representation of the experiences of all Medicare beneficiaries, the MCBS includes beneficiaries wherever they reside, even if they reside in and/or enter a facility for the duration of their four years with the study. The MCBS does not conduct Facility interviews with the beneficiary directly; instead, specially trained Facility interviewers administer the survey to Facility administrative staff.

### 3.4 Questionnaires

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 round (fall, winter, or summer). There are two types of interviews (Baseline and Continuing) containing two types of questionnaire sections (Core and Topical). For more information about the MCBS questionnaire design, see the *MCBS Methodology Report*. Additionally, see the annual *MCBS Questionnaire User Guide* for descriptions of each questionnaire section and details on yearly changes.

### 3.5 Medicare Population Covered by MCBS Data

The MCBS data releases reflect enrolled Medicare beneficiaries residing in the continental United States. The sample for the MCBS is drawn from a subset of the Medicare enrollment data, which is a list of all Medicare beneficiaries. Excluded from both populations are residents of foreign countries and U.S. possessions and territories. The beneficiaries included in the MCBS LDS releases represent a random cross-section of all beneficiaries who were ever enrolled in either Part A or Part B of the Medicare program for any portion of the data collection year. A subset of these beneficiaries represents a random cross-section of all beneficiaries who were continuously enrolled from January 1 of the data collection year up to and including interviews conducted during the fall round of the data collection year. The ever enrolled and continuously enrolled populations are described in further detail below and in Exhibit 3.5.1:

<sup>5</sup> The three rounds per year are referred to seasonally. Respondents are interviewed in the winter round, the summer round, and the fall round each year.

- The **ever enrolled** population represents individuals who were enrolled in Medicare at any time during the calendar year. This population includes beneficiaries who enrolled during the calendar year as well as beneficiaries who dis-enrolled or died prior to their fall interview.<sup>6</sup> The ever enrolled population includes beneficiaries who were enrolled in Medicare for at least one day at any point during the data collection year.
- The **continuously enrolled** population represents only individuals continuously enrolled in Medicare from January 1 of the data collection year up to and including their fall interview; this specifically excludes beneficiaries who enrolled during the calendar year and beneficiaries who dis-enrolled or died prior to their fall interview. The concept of continuously enrolled is consistent with the concept of being exposed or “at risk” for using services up to and including the fall interview.

Beneficiaries who became eligible for Medicare Part A or B and enrolled anytime during the year were eligible to be sampled as part of the annual panel.<sup>7</sup> Thus, each LDS release includes data from four separate MCBS panels, the three Continuing Panels and the Incoming Panel. The inclusion of the current-year enrollees allows data to be released in a timelier manner.

### **Exhibit 3.5.1:** Medicare Population Covered by the MCBS Data

<b>Ever Enrolled Population</b>	<b>Continuously Enrolled Population</b>
<ul style="list-style-type: none"> <li>■ A random cross-section of all beneficiaries who were ever enrolled in either Part A or Part B of the Medicare program for any portion of the calendar year (i.e., for at least one day at any point during the calendar year)</li> </ul>	<ul style="list-style-type: none"> <li>■ A subset of the ever enrolled population who were continuously enrolled from January 1 up to and including interviews conducted during Fall of the data collection year</li> </ul>
<ul style="list-style-type: none"> <li>■ Represents individuals who were enrolled in Medicare at any time during the calendar year</li> </ul>	<ul style="list-style-type: none"> <li>■ Represents only individuals who were continuously enrolled in Medicare from January 1 up to and including their fall interview</li> </ul>
<ul style="list-style-type: none"> <li>■ Includes beneficiaries who enrolled during the calendar year</li> <li>■ Includes beneficiaries who dis-enrolled or died prior to their fall interview</li> </ul>	<ul style="list-style-type: none"> <li>■ Excludes beneficiaries who enrolled after January 1</li> <li>■ Excludes beneficiaries who dis-enrolled or died prior to their fall interview</li> </ul>

<sup>6</sup> Note that data collection for beneficiaries who enrolled during the data collection year and died that same year after enrollment but before their fall interview was still pursued through attempts at conducting proxy interviews.

<sup>7</sup> These beneficiaries are referred to as “current-year enrollees.”



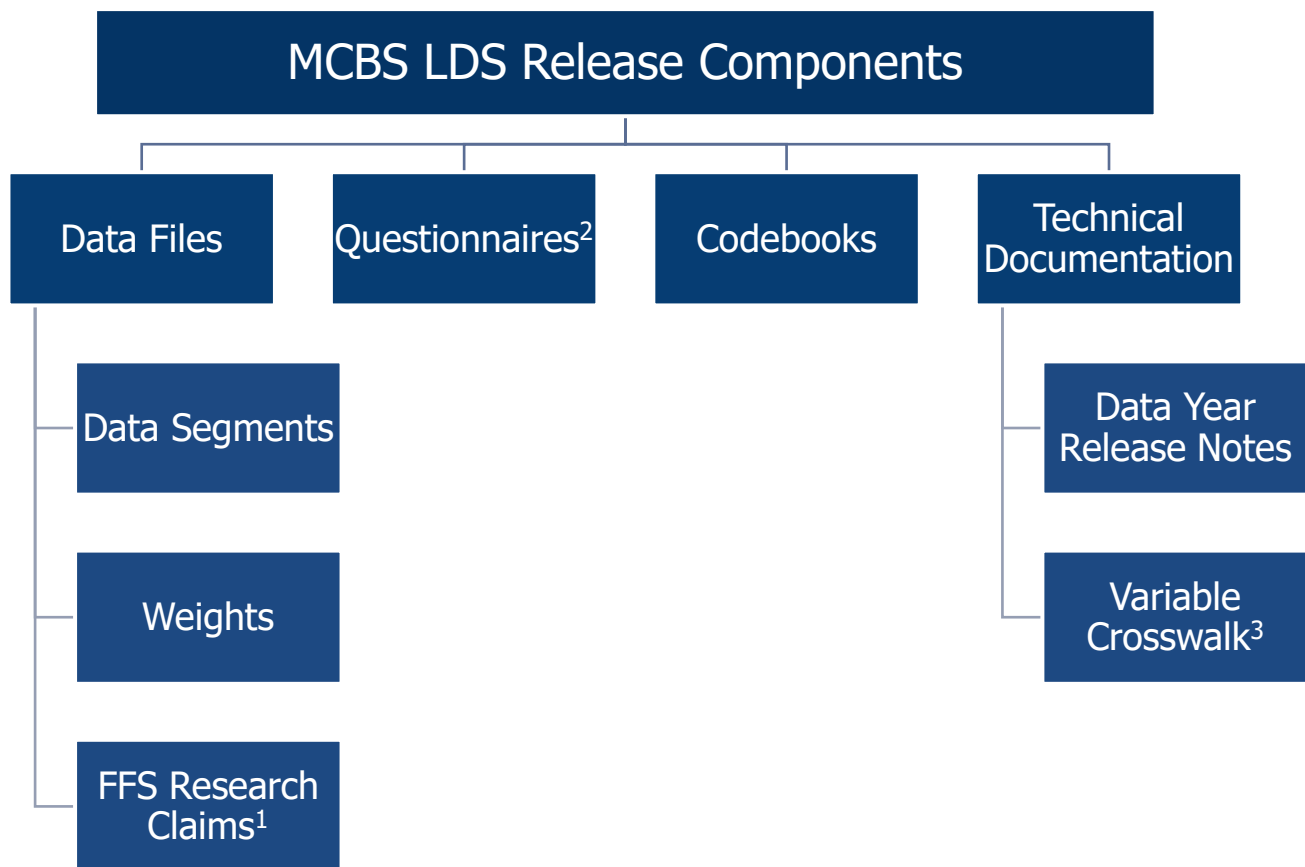
## 4. DATA PRODUCTS & DOCUMENTATION

### 4.1 Contents of Data Release

MCBS data are made available via releases of annual files. For each data year, three annual LDS releases (the Survey File - Early Release, the Survey File, and the Cost Supplement File) and two PUFs (based on the Survey File and Cost Supplement File, respectively) are planned. The LDS releases contain multiple files, called segments, which are easily linkable through a common beneficiary key ID (i.e., the BASEID).

This section provides general information on the LDS releases and LDS data segments. Beginning with the 2023 data year, detailed information on each segment is available in the *Data Year Releases Notes*. Detailed information on the 2015-2022 segments is available in the historical annual *Survey File Data User's Guides* and *Cost Supplement File Data User's Guides*. All MCBS documentation resources are available for download at: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

Exhibit 4.1 displays the components of each LDS release. Each LDS contains data segments, codebooks, questionnaires, and technical documentation. The Survey File release contains the FFS claims data, which provide CMS administrative information on medical services and payments paid by Medicare claims; Prescription Drug Event (PDE) events for Medicare Part D are not included and claims data for Medicare Advantage (MA) beneficiaries are not available. Users can conduct standalone analyses with the Survey File - Early Release and the Survey File. However, users interested in the Cost Supplement File data will also need the Survey File LDS files to link cost and utilization variables with demographic or health insurance coverage variables.

**Exhibit 4.1.1:** Contents of MCBS LDS Data Releases

<sup>1</sup> The FFS Research Claims are only provided with the Survey File.

<sup>2</sup> All MCBS Questionnaires for the data collection year are provided with the Survey File and Survey File - Early Release, even if some of the questionnaire sections do not contribute data to that LDS. Since the Cost Supplement File is only provided with the Survey File, the MCBS Questionnaires are just provided once for both LDS's.

<sup>3</sup> The Variable Crosswalks are only released for the Survey File and Cost Supplement File.

#### 4.1.1 MCBS Survey File - Early Release

First released for the 2023 data year, the Survey File - Early Release contains a subset of the data segments that are made available on the Survey File, including information on beneficiaries' demographics, health status, conditions, and functioning, household characteristics, and access to and satisfaction with care. The Survey File - Early Release contains data collected directly from Community respondents during the fall data collection round supplemented by some administrative data. The Survey File - Early Release does not include information collected in Topical questionnaire sections after the data collection year, facility (non-cost) information or information collected in Facility interviews, or FFS claims. The Survey File - Early Release LDS includes beneficiaries who were alive, enrolled in Medicare, and completed a Community interview in the fall of the data collection year.

#### 4.1.2 MCBS Survey File

The Survey File contains data collected directly from respondents and supplemented by administrative items plus facility (non-cost) information and FFS claims. The Survey File includes multiple topic-related segments, including health status and limitations, access to care, health insurance coverage, and household characteristics. The Survey File also includes information on Facility beneficiaries, including a residence

timeline, facility characteristics, and facility assessment measures (e.g., Minimum Data Set). Finally, data from Topical questionnaire sections administered after the data collection year (e.g., on beneficiary knowledge, drug coverage) are included with this release. To facilitate analysis, the information collected in the survey is augmented with data on enrollment and the use and program cost of Medicare services from Medicare FFS claims data and administrative data. The Survey File includes beneficiaries enrolled for at least one day in the data collection year who completed an interview in the data collection year or winter of the following year, or who died during the data collection year. Beneficiaries who refused to complete a later interview or became nonrespondents during the data collection year are excluded.

### *4.1.3 MCBS Cost Supplement File*

The Cost Supplement File contains both individual event and summary files and can be linked to the Survey File to conduct analyses on health care cost and utilization. The Cost Supplement File links survey-reported events to Medicare FFS claims and provides a comprehensive picture of health services received, amounts paid, and sources of payment, including those not covered by Medicare. Survey-reported data include information on the use and cost of all types of medical services, as well as information on supplementary health insurance costs. Medicare FFS claims data include administrative and billing information on the use and cost of inpatient hospitalizations, outpatient hospital care, physician services, home health care, durable medical equipment, skilled nursing home services, hospice care, and other medical services.<sup>8</sup> The Cost Supplement File supports a broader range of research and policy analyses on the Medicare population than would be possible using either survey data or administrative claims data alone.

The Cost Supplement File contains a subset of the beneficiaries included in the Survey File who have complete cost and utilization data for the year. For each MCBS data year, the Cost Supplement File includes beneficiaries sampled in the previous three panels (i.e., Continuing Panels), plus members of the current panel (i.e., Incoming Panel) who were enrolled in Medicare during the data collection year for at least one day.

For beneficiaries enrolled in MA, cost and utilization information is available. As is done with services not covered by Medicare (e.g., most dental, vision, and hearing care), when a beneficiary reports health care events, the MCBS uses the explanation of benefits (EOB) form from MA providers to report the payments, as well as the capitation information from the administrative data for total MA payments. Actual claims-based information for MA beneficiaries, referred to as Encounter data, is not currently available for these individual events. The Cost Supplement File undergoes a careful reconciliation process to separately identify and flag health care services reported: 1) from the survey alone, 2) from the claims data alone, and 3) from both sources. This process results in a file with a much more complete and accurate picture of health services received, amounts paid, and sources of payment. Due to the added processing time required to reconcile survey reported events with the claims data, this file is generally released 21 months after the close of the calendar year for data collection.

### *4.1.4 Using the Data*

The MCBS data releases are made available in two formats: SAS formatted files and comma delimited files for use with Stata® and R®. Directions and sample SAS code are given below to help users read the dataset into

<sup>8</sup> Only Medicare claims for beneficiaries enrolled in Medicare FFS (often called “traditional” Medicare), are available for linkage; similar claims information for MA beneficiaries is not available. To the extent that health care use and costs may be underreported in the survey or reported differentially between FFS and MA beneficiaries, this will be reflected in the data as MA beneficiaries’ information will not be supplemented by claims data.

SAS (see Appendix A). Files with programming code to create formats and labels are provided for both SAS users and for use with comma delimited files.

## 4.2 Which File Do I Need?

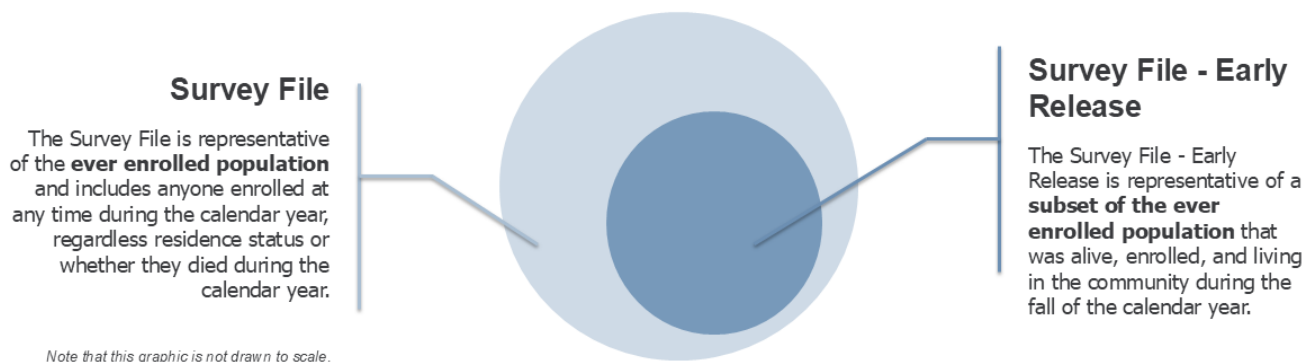
The identification of the target population for a given research question will influence both the selection of weights and the particular segments that a data user will need to conduct analyses.

### 4.2.1 Using the Survey File - Early Release

Users who wish to focus on health-related research questions for beneficiaries living in the community who do not require access to information that is only released on the Survey File, including facility information, complete enrollment data, Topical questionnaire data collected in the winter and summer following the data collection year, and/or Medicare FFS utilization, only need the Survey File - Early Release. The Survey File - Early Release is intended for standalone analysis only.

Exhibit 4.2.1 depicts the relationship between the beneficiaries included in the Survey File - Early Release and Survey File data releases.<sup>9</sup> The ever enrolled population from the Survey File is the largest population available from the MCBS, including anyone enrolled at any time during the calendar year corresponding to the LDS data year. This represents the population of beneficiaries who were ever enrolled in Medicare for at least one day at any time during the year. The ever enrolled population includes beneficiaries who died or lost entitlement prior to completing the fall interview. Beneficiaries who first became enrolled in Medicare during the year are also included. The ever enrolled population from the Survey File - Early Release is a subset of the ever enrolled population from the Survey File, and includes anyone who was alive, enrolled, and living in the community during the fall round of the calendar year.

**Exhibit 4.2.1:** Comparison of Ever Enrolled MCBS Populations in Survey File - Early Release and Survey File



### 4.2.2 Using the Survey File and Cost Supplement File

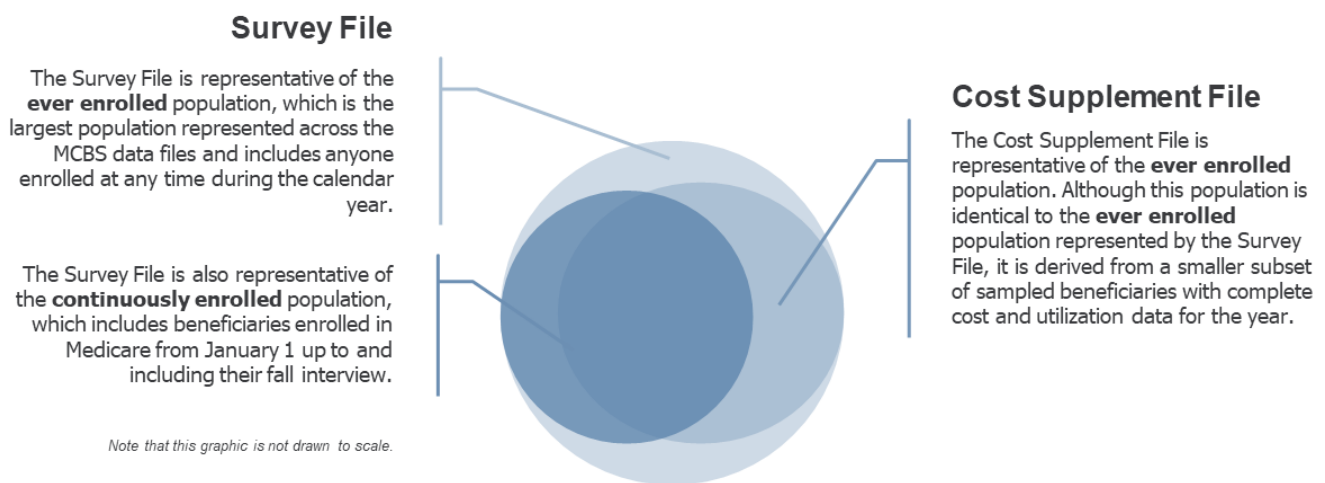
Users who wish to focus on health-related research questions, such as health status and access to care and/or Medicare FFS utilization, only need the Survey File. Users that wish to use demographic and health insurance information to conduct research on the cost and utilization of medical services need both the Survey File and the Cost Supplement File. As such, the Survey File supports both standalone analysis and joint analysis with the Cost Supplement File, while the Cost Supplement File is only intended to be used in combination with the Survey File.

<sup>9</sup> Exhibit 4.2.1 is not drawn to scale, but provided as a visual reference for the relationship of populations between data files.

Exhibit 4.2.2 depicts the relationship between the beneficiaries included in the Survey File and Cost Supplement File data releases.<sup>10</sup> The ever enrolled population from the Survey File is the largest, including anyone enrolled at any time during the calendar year corresponding to the LDS data year. The continuously enrolled population includes beneficiaries who were enrolled continuously between January 1 through the completion of their fall interview. Beneficiaries who died during the year, newly enrolled beneficiaries who enrolled in Medicare during the year that they were sampled, and beneficiaries who lost eligibility are not included in the continuously enrolled group. Thus, the continuously enrolled beneficiaries are a subset of the ever enrolled beneficiaries. The Survey File LDS includes weight segments that allow for subsetting the data by the ever enrolled and continuously enrolled populations.

The Cost Supplement File is representative of the ever enrolled population, but is smaller than the Survey File population because it is derived from a smaller subset of sampled beneficiaries with complete cost and utilization data for the entire year. The Cost Supplement File includes a weight segment that allows for subsetting the data by the ever enrolled population.

#### Exhibit 4.2.2: Comparison of MCBS Populations in Survey File and Cost Supplement File



### 4.2.3 Using Both Community and Facility Data

Analytic decisions about whether to include all beneficiaries regardless of residence status or beneficiaries living only in the community or only in facilities should be driven by both the research question and data limitations. However, there are differences in the data collection protocols and questionnaire instruments for the MCBS Community and Facility components. Thus, caution should be observed when combining data across these populations to address questions requiring analysis of all Medicare beneficiaries.

To determine which population should be included in an analysis, the following steps are recommended:

1. Define the population based on the research question(s) and identify the living in community and living in facility populations. The variable INT\_TYPE on the Demographics (DEMO) segment is the recommended variable for defining the two populations.
2. Identify the LDS segments and variables associated with each of the analysis' domains to determine what data are available for the Community and Facility components.
3. Assess whether the universe, level of measurement, and response categories for the variables of interest are similar for both Community and Facility components.

<sup>10</sup> Exhibit 4.2.2 is not drawn to scale, but provided as a visual reference for the relationship of populations between data files.

4. If needed, recode the LDS variables to align the coding between Community and Facility components and create analytic variables.
5. Merge the Community and Facility segments with the appropriate weights segments. Assess preliminary estimates for variation between community and facility.
6. Review MCBS documentation to determine if there are underlying differences in data collection and processing between community and facility that result in analytic limitations.
7. Conduct analysis and document any potential limitations.

For more information on using community and facility data, including a series of analytic examples with sample SAS code, see the *MCBS Advanced Tutorial on Using Community and Facility Data*. Data users can access this tutorial at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials>. Note that the Survey File and Cost Supplement File include data on all beneficiaries (regardless of residence status), while the Survey File - Early Release only includes data on beneficiaries living in the community during the fall round.

## 5. SURVEY FILE - EARLY RELEASE NOTES

### 5.1 LDS Specifications

The MCBS Survey File - Early Release contains survey-collected data from fall Community interviews augmented with some administrative data to allow for timely analysis regarding the beneficiaries' health status and conditions, access to health care, and satisfaction with health care, among other topics. The specific contents of the LDS release are subject to change year-to-year. For a detailed description of the contents of the LDS release for each data year, see the *Data Year Release Notes*: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

### 5.2 File Structure

Unlike the Survey File, the Survey File - Early Release segments only contain survey data with related Medicare administrative variables, as applicable. The Survey File - Early Release does not include Medicare FFS claims data. The Survey File - Early Release comprises a subset of the segments released on the Survey File (i.e., all segments released on the Survey File - Early Release are also released on the Survey File). To distinguish the Survey File - Early Release segments from the Survey File segments, the Early Release versions begin with the abbreviation "ER\_." Most segments released on both LDS's include the same variables. However, for certain segments, the Survey File - Early Release offers an abbreviated version of the segment that is reprocessed and annualized for release on the Survey File. For more information, see the *Data Year Release Notes*.

All MCBS segments begin with the same three variables: a unique number that identifies the person who was sampled (the BASEID), the survey reference year, and the version of release. These elements serve to identify the type of record and to provide a link to other types of records. To obtain complete survey information for an individual, a researcher must link together records for that individual from the various segments using the variable BASEID. Beneficiaries may not have a record on every segment.

#### 5.2.1 Sort Order for Merging the Survey File - Early Release LDS Segments

Sort order is often important to understand when data users are merging segments within or across LDS releases. Most LDS segments are sorted by BASEID. However, some are sorted on other fields to create appropriate and unique sort keys for matching and merging the data. See Exhibit 5.2.1 below.

**Exhibit 5.2.1:** Sort Order by Segment in the Survey File - Early Release LDS

Segment	Sorted by
Assistance (ER_ASSIST)	BASEID HLPRNUM

### 5.3 Analytic Guidance

The Survey File - Early Release contains a subset of the data released on the Survey File. While analyses using the same variables from the Survey File - Early Release and the Survey File are expected to produce similar results, there may be small differences in these results because of changes between preliminary and final administrative data as well as minor differences in data editing, processing, and weighting for the files. As such, if a user has access to both the Survey File - Early Release and Survey File (once available), it is recommended to use the Survey File.

The Survey File - Early Release is intended for standalone analysis only. It is not intended to be used in combination with the Survey File or Cost Supplement File. The Survey File - Early Release only offers cross-sectional weights. As such, it is also not intended to be used for longitudinal analysis.



## 6. SURVEY FILE NOTES

### 6.1 LDS Specifications

The MCBS Survey File contains survey-collected data from Community and Facility interviews augmented with administrative and claims data to allow for analysis regarding the beneficiaries' health status, access to health care, satisfaction with health care, and usual source of care, among many other topics. The specific contents of the LDS release are subject to change year-to-year. For a detailed description of the contents of the LDS release for each data year, see the *Data Year Release Notes* or the historical 2015-2022 *Survey File Data User's Guides*: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

Several Survey File LDS segments contain information on beneficiaries' Medicare eligibility and enrollment data, including monthly coverage indicators, coverage start and end dates, the type of plan, the source of coverage information for the plan, eligibility codes, and detailed Medicare-Medicaid dual eligibility indicators. See the historical *2015-2022 Survey File Data User's Guides* and the *Survey File Data Year Release Notes* for detailed information.

### 6.2 File Structure

The Survey File segments can be divided into two subject matter groups: files containing survey data with related Medicare administrative variables and files containing Medicare FFS claims data. The claims records represent services provided during each calendar year and processed by CMS. To facilitate analysis, the Administrative Utilization (ADMNUTLS) segment contains a detailed summary of the utilization enumerated by these claims.

All MCBS segments begin with the same three variables: a unique number that identifies the person who was sampled (the BASEID), the survey reference year, and the version of release. These elements serve to identify the type of record and to provide a link to other types of records. To obtain complete survey information for an individual, a researcher must link together records for that individual from the various data files using the variable BASEID. Beneficiaries may not have a record on every data file.

#### 6.2.1 Sort Order for Merging the Survey File LDS Segments

Sort order is often important to understand when data users are merging segments within or across LDS releases. Most LDS segments are sorted by BASEID. However, some are sorted on other fields to create appropriate and unique sort keys for matching and merging the data. See Exhibit 6.2.1 below.

**Exhibit 6.2.1:** Sort Order by Segment in the Survey File LDS

Segment	Sorted by
Assistance (ASSIST)	BASEID HLPRNUM
Facility Characteristics (FACCHAR)	BASEID RECADMN
Health Insurance Timeline (HITLINE)	BASEID PLANTYPE PLANNUM
Interview Characteristics (INTERV)	BASEID SEQNUM
Minimum Data Set (MDS3)	BASEID TRGT_DT A2300
Outcome and Assessment Information (OASIS)	BASEID HHA_ASMT_INT_ID



## 6.3 Claims Files

The MCBS Research Claims are a subset of items from the claims available on the Chronic Conditions Warehouse (CCW). All Research Claims are sorted by BASEID and CLAIMID. Item (variable) names are listed in alphabetical order. The MCBS Research Claims have a unique and de-identified BASEID and CLAIMID so that these cannot be linked back to the original claims. The full descriptions of the items on the MCBS Research Claims can be found on the public facing CCW Claims Data Dictionary, located at:

<https://www2.ccwdata.org/documents/10280/19022436/codebook-ffs-claims.pdf>.

The MCBS Research Claims (also known as the FFS claims or fixed-length claims) are abbreviated versions of the full claim record layout. Each claim type has a subset of variables selected for their relevancy to data analysis of that service. Additionally, institutional claim types have a corresponding revenue center file that links back to the claim-level data file through a unique claim identifier. There is one observation per data record for all MCBS claims files except the Physician/Supplier Claims and Durable Medical Equipment (DME) Claims. Those claim types treat each line item as a separate observation with the claim-level detail repeating for each line item. The MCBS Research Claims are provided as SAS and CSV files.

MCBS data can be linked to Medicare Part A and Part B claims data for beneficiaries who participated in the MCBS. MCBS data cannot be linked to electronic medical records, or to any other records that record lab values or physiologic data.

Beginning with 2021, the Survey File LDS includes five years of Research Claims. This means that depending on their original enrollment date and enrollment type (Medicare FFS vs. Medicare Advantage), a beneficiary could have up to five years of claims included with their MCBS data.

### 6.3.1 Utilization Detail Records

#### Core Content

The following rules are used to select claims records for the claims files.

1. Inpatient claims were included if the discharge or "through" date fell on or after January 1 and on or before December 31 of the calendar year.
2. Skilled nursing facility (SNF) claims were included if the admission or "from" date fell on or after January 1 and on or before December 31 of the calendar year.
3. Home health agency and outpatient facility claims were included if the "through" date fell on or after January 1 and on or before December 31 of the calendar year.
4. Hospice claims were included if the admission or "from" date fell on or after January 1 and on or before December 31 of the calendar year.
5. Physician or supplier claims were included if the latest "service thru" date fell on or after January 1 and on or before December 31 of the calendar year.
6. DME claims were included if the latest "service thru" date fell on or after January 1 and on or before December 31 of the calendar year.

There are no claims records in this file for survey respondents who did not use Medicare reimbursed services in an FFS setting. These individuals may have used no services at all, services only in a managed care plan, or services provided by a payer other than Medicare.<sup>11</sup> For the other respondents, the MCBS captures claims meeting the date criteria that are processed and made available by CMS through June of the following calendar year.<sup>12</sup>

## 6.4 Analytic Guidance

The Survey File is released approximately nine months after the Survey File - Early Release. The Survey File contains all segments released on the Survey File - Early Release, plus additional information not available for that release, including complete enrollment information, data from Facility interviews and assessments, and information collected in Topical questionnaire sections in the winter and summer after the data collection year. The Survey File also offers additional types of weights that are not available on the Survey File - Early Release, including weights that represent the continuously enrolled population and weights that enable longitudinal analysis.

As noted above, while analyses using the same variables from the Survey File - Early Release and the Survey File are expected to produce similar results, there may be small differences in these results because of changes between preliminary and final administrative data as well as minor differences in data editing, processing, and weighting for the files. As such, if a user has access to both the Survey File - Early Release and Survey File, it is recommended to use the Survey File.

Additionally, only the Survey File is intended to be used in combination with the Cost Supplement File, not the Survey File - Early Release. When combining data from the Survey File and Cost Supplement File, users should use the Cost Supplement File weights.

Each year, a subset of the Survey File LDS data are released on the corresponding Survey File Microdata PUF. The MCBS Survey File PUF is not intended to replace the more detailed LDS; rather, it offers a free, publicly available alternative that provides the highest degree of protection to the Medicare beneficiaries' protected health information (PHI). Unlike the Survey File LDS, the Survey File PUF only supports cross-sectional analysis of the ever enrolled population living in the community and cannot be combined with other MCBS LDS files. Download the Survey File PUF documentation for more information on this release: <https://data.cms.gov/medicare-current-beneficiary-survey-mcbs>.

<sup>11</sup> Other MCBS LDS segments provide data on types of insurances, the coverage eligibility timeline, and the source information for the coverage use of services (i.e., Medicare Administrative enrollment data and/or survey data) as well as self-reported data on access and satisfaction with visits. See the *Data Year Release Notes* and codebooks for more information on the contents of these segments.

<sup>12</sup> Note that claims "mature" through the midpoint of the following calendar year. That is, CY1 claims are pulled from CMS' administrative data after June of CY2 to ensure that the CY1 claims have been finalized.

## 7. COST SUPPLEMENT FILE NOTES

### 7.1 LDS Specifications

The following information is represented in the MCBS Cost Supplement File:

- Survey-reported data, including information on the cost and utilization of medical services, which contain all sources of payment and out-of-pocket costs. Certain years also provide data on physical measures.
- Medicare FFS and prescription drug claims data, including administrative and billing information on the cost and utilization of inpatient hospitalizations, outpatient hospital care, physician services, home health care, durable medical equipment, skilled nursing home services, hospice care, and prescription drugs.
- MA cost and utilization information: When a respondent reports health care events, the EOB form from their MA provider is used to report the payments. This is the same approach taken for services that are not covered by Medicare, such as most dental care. Actual claims-based information for MA beneficiaries, referred to as Encounter data, are not currently available for these individual events.

### 7.2 File Structure

The Cost Supplement File data on utilization and expenditures are provided at three different levels of summarization: Event-level, Service Summary (SS) level, and Person Summary (PS) level. The tri-level structure allows researchers to potentially avoid having to process all the detailed event records in the file when summaries may suffice. For example, an analysis of differences in total health spending per person by type of Medicare coverage could use the PS level and thereby avoid having to process the numerous event-level records. Similarly, an analysis of differences in use of Medicare hospital payments by race/ethnicity could use the SS records. Researchers could use event-level records for more detailed analyses, such as looking at average length of long-term facility stays or average reimbursements per prescription drug type. Users should determine whether a segment that contains already summarized costs, utilization, and payment distributions would best serve the analysis.

The Cost Supplement segments are assembled at three levels:

1. The **event-level** reports all payers, costs, and utilization at the most detailed level available. Service types at the event-level are dental (DU), facility (FA), hearing (HU), home health (HH), medical provider (MP), inpatient hospital (IP), outpatient hospital (OP), institutional (IU), prescription medicine (PM), and vision (VU). Note that hospice (HP) services are only included at the service and person level; there are no event-level data provided for these services. For the provided service types, the event-level data has a record for each reported medical event for each person.
2. The **Service Summary level** summarizes all payers, costs, and utilization for a person during the calendar year at the service level. There are 11 service categories: dental, facility, hearing, home health, hospice, inpatient hospital, institutional, medical provider, outpatient hospital, prescribed medicines, and vision. The data include a record for each of the 11 service categories for each person. Note, that there are two service-level records for home health services, one for Home Health Friend events and one for Home Health Provider events. Within each type of service record, separate payer totals for 11 different payers are also shown.

Payer totals are summarized in two ways: one summarizes the event-level records with no adjustment, and the other is summarized adjusting for survey interview gaps that occurred for Medicare covered days within the year. For example, if a beneficiary visited the dentist for an annual checkup and again for a cavity filling, all costs and utilization for these two dental visits would be summarized in one row of data under the dental category for this beneficiary. If the dental costs were covered by different payers, the overall cost will be split by payer in the SS record. The adjusted totals correct for any survey interviewing gaps during the year.

The service summaries exclude unmatched survey event records that are considered duplicative of unmatched Medicare bill record events. The adjusted totals also include an upward ratio adjustment to Non PM, non HH, and non HP utilization and expenditure data for beneficiaries enrolled in MA plans in the data year.

3. The **Person Summary level** summarizes all payers and costs across service categories and summarizes type of service amounts. The data include only one record for each person, which shows the totals for each service and payer for that person. Again, payment amounts are shown two ways: summarized from event records and adjusted to compensate for Medicare covered days that were not covered by interview reference periods.

As with the SS level, the person summaries exclude unmatched survey event records that are considered duplicative of unmatched Medicare bill record events. In addition, the adjusted totals include an upward ratio adjustment to Non PM, non HH, and non HP utilization and expenditure data for beneficiaries enrolled in MA plans in the data year.

All MCBS records begin with the same three variables: the survey reference year, version, and a unique number that identifies the person who was sampled (the BASEID). The BASEID provides the link to the other segments within the Cost Supplement File and Survey File LDS releases. To obtain complete survey information for an individual, a researcher must link together records for that individual from the various data files using the variable BASEID. Beneficiaries may not have a record on every data file.

The MCBS Cost Supplement File segments contain both survey-reported data and administrative claims data on health service utilization, costs, and payers. Exhibit 7.2.1 provides a mapping of the data sources for the service types available on the Cost Supplement File, highlighting that there are differences in data sources by beneficiary residence status and health insurance type.

**Exhibit 7.2.1:** Data Sources for Cost Supplement File Service Types

Beneficiary Residence Status		Community			Facility		
Utilization and Cost Data Sources		Survey Reported	Claims*		Survey Reported	Claims*	
Beneficiary Health Insurance			FFS	Part D		FFS	Part D
Service Type							
Dental utilization events		X	+			+	
Hearing utilization events		X	+			+	
Inpatient hospital events		X	X			X	
Institutional events		X	X			X	
Medical services, equipment, and supplies		X	X	X		X	X
Outpatient hospital events		X	X			X	
Prescribed medicine events		X		X			X
Vision utilization events		X	+			+	
Nursing home or other long-term care facility	■ Cost of Facility <b>Stay</b>				X	X^	
	■ Services provided <b>Within</b> the Facility				X		
	■ Services received <b>Outside</b> of the Facility				X		
Home health events		X	X			X	
Hospice care events		X	X			X	

\*Claims are only available for events paid for by Medicare FFS and Part D. Administrative data pertaining to events paid for by MA, Medicaid, private insurance, or out of pocket are not available to the MCBS. Medicaid T-MSIS claims and historic MA Encounter data are used during imputation for some event types.

+ Medicare FFS does not cover most procedures or supplies for dental, vision, and hearing services. It only covers certain services that are received in a hospital. The MCBS includes claims data for the few services Medicare does cover, but the vast majority of events are Community survey-reported only.

<sup>^</sup> SNF and inpatient Medicare FFS claims are used during data processing of the facility events to impute missing Medicare payment amounts and check the accuracy of the lengths of stays, when available. However, claims-only records are not added to facility events segment whereas they are added for the other event-level segments. See the *MCBS Methodology Report* for more information on claims matching and imputation processes.

## 7.3 Analytic Notes

The Cost Supplement File provides data on a subset of beneficiaries included in the Survey File who provided complete cost and utilization data for the entire year. The Cost Supplement File LDS is not intended for standalone analysis; instead, it is intended to be combined with the Survey File LDS, which provides linkable data on beneficiaries' socio-demographics and health insurance coverage, among other information. As such, the Cost Supplement File is always made available by CMS with the Survey File (i.e., it cannot be purchased alone). When combining data from the Survey File and Cost Supplement File, data users must also use the general or Topical Cost Supplement File weights. For more information on using the weights, please see Section 8.4 Weighting.

Each year, a subset of person-level data from the Cost Supplement File are combined with select demographic indicators from the Survey File LDS and released on the Cost Supplement File Microdata PUF. The MCBS Cost Supplement File PUF is not intended to replace the more detailed LDS's; rather, it offers a free, publicly available alternative that provides the highest degree of protection to the Medicare beneficiaries' PHI. Unlike the LDS's, the Cost Supplement File PUF only supports cross-sectional analysis of the ever enrolled population living only in the community and cannot be combined with other MCBS LDS files. Download the Cost Supplement PUF documentation for more information on this release: <https://data.cms.gov/medicare-current-beneficiary-survey-mcbs>.

## 8. DATA FILE DOCUMENTATION

### 8.1 LDS Contents

CMS provides the following technical documentation for data users with each LDS release:

- Codebooks
- Questionnaires
- Data files (SAS, CSV)
- Research claims (SAS, CSV) (Survey File only)
- Format control files
- Sample SAS code to apply the formats and labels for researchers not using SAS

### 8.2 LDS Components

#### 8.2.1 Codebooks

Codebooks are included with each data release and serve as the key resource for comprehensive information on all variables within a data file. The codebooks list the variables in each of the segments, the possible values, and unweighted frequencies. For variables that are associated with items in the MCBS Questionnaire, the item number and item text are provided. The information provided within each codebook is as follows:

**Variable:** The codebook contains the variable names associated with the final version of the data files. Certain conventions apply to the variable names. All variables that are preceded by the character "D\_", such as D\_ERVIST, are derived variables. Variables preceded by the character "H\_", such as H\_DOB, come from CMS administrative source files.

**Format Name:** This column identifies the format name associated with the variable in the SAS dataset.

**Frequency:** This column shows unweighted frequency counts of values or recodes for each variable.

**Question #:** This column contains a reference to the questionnaire for direct variables, or to the source of derived variables. For example, the entry that accompanies the variable D\_ERVIST in the Access to Care, Medical Appointments segment is "AC1," which refers to the first question in the Access to Care portion of the Community Questionnaire. This column will be blank for variables that do not relate to the questionnaire or relate to CMS administrative source files, which are usually variables created to manage the data and the file.

**Description/Label (variable label and codes):** The variable label provides an explanation of the variable. For coded variables, all the possible values of the variable appear in lines beneath that explanation. For each possible value, the codebook provides a count of the number of times that the variable had that value (in the column labeled "Frequency") and a short format expanding on that value (in the column labeled "Label").

**BASEID:** The BASEID is the unique identifier assigned to each beneficiary. This identifier can be used to link data across the survey files.

**Survey Year:** The Survey Year of interest is included as a variable on the file.

**Version Number:** Files may be re-released due to needed updates, which will be noted by the version number variable.

**Note:** Each variable may be followed by a statement that describes when a question is not asked and results in a missing variable. Questions are not asked when the response to a prior question or other information gathered earlier in the interview would make them inappropriate. For example, respondents who indicated that they never smoked are not asked if they currently smoke. Notes also describe important information about the variable. For variables added to the survey recently, the first year of administration is also listed in the note.

The Cost Supplement File data come from summarized information from MCBS Questionnaire items as well as imputed data. On Cost Supplement File segments, data users will find imputation flag variables where applicable.

Many questions are written to elicit simple “Yes” or “No” answers, or to limit responses to one choice from a list of categories. For other questions, the respondent is given a list of responses and instructed to select all responses that apply. In cases when the question is a “select all that apply” item, each of the responses is coded “Indicated”/“Selected” or “Not Indicated”/“Not Selected.”

If a respondent provides an answer that is not on the list of possible choices, it is recorded verbatim. Verbatim responses are reviewed and categorized, when possible. New codes may be added to the original list of options to accommodate verbatim responses that appear frequently. For this reason, the list of possible values for some variables may not exactly match the questionnaire.

### *8.2.2 Questionnaires*

Data users can view the Questionnaires for each data year along with the questionnaire variable names and question text on the MCBS website at: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/questionnaires>.

### *8.2.3 Data User Resources*

CMS provides technical assistance to researchers interested in using MCBS data and provides free consultation to users interested in obtaining these data products and using these data in research. Users can email [MCBS@cms.hhs.gov](mailto:MCBS@cms.hhs.gov) with questions regarding obtaining or using the data.

## **8.3 Data Edits and Imputation**

### *8.3.1 Data Edits*

A series of checks and edits are conducted to ensure the accuracy, completeness, and reasonableness of data within each data file. Any structural issues are addressed during either data extraction or data cleaning.

Logic checks verify that the questionnaire worked as expected, particularly with respect to questionnaire routing. Errors identified during logic checking are addressed with two categories of data edits: flagging values that were incorrectly skipped and setting incorrectly populated values to null to indicate a valid missing value.

Additional checks identify unreasonable values that are not explicitly disallowed by the questionnaire (e.g., reporting a negative charge for a dental visit). After investigation, such values are then addressed with global edits. The MCBS also conducts consistency checks to identify scenarios where respondents report inconsistent information (e.g., indicating that one is Medicaid eligible due to a certain condition, but not reporting having that same condition when asked about health status). Based on a thorough data review, these types of errors are addressed with edits during data cleaning.



Certain conventions are used in coding all variables to distinguish between questions that beneficiaries would not or could not answer and questions that were not asked. These conventional codes are depicted in Exhibit 8.3.1.

**Exhibit 8.3.1:** Data Review and Editing Codes

Value	Format	Meaning
.	INAPPLICABLE	Valid missing, inapplicable, a valid skip, missing with no expectation that a value should be present. Missing is '.' in numeric variables and blank in character variables.
.R	REFUSED	Valid missing, refused survey response
.D	DON'T KNOW	Valid missing, don't know survey response
.N	INVALID SKIP	Invalid missing, not ascertained, an invalid skip, a response should be present but is not
.E	EDITING CODE	Editing code, extreme value, unreasonable or out of range survey response
.S	SUPPRESSED*	Valid value suppressed due to suppression guidelines applied to Area Deprivation Index (ADI) variables

\*This code is not applied to data collected by the Facility Instrument.

### 8.3.2 Imputation

To compile the most accurate and complete LDS, there are several types of adjustments applied to the MCBS data that compensate for missing information. Although a variety of methods are used in making the adjustments, adjustments of all types are governed by some basic principles. Information reported by the survey respondent is retained, even if it is not complete, unless strong evidence suggests that it is not accurate. When information is not reported during the interview, Medicare claims data and administrative data are the first choice as a source of supplementary, or in some cases, surrogate information.

There are several techniques for handling cases with missing data. One option is to impute the missing data. This can be done in such a way as to improve univariate tabulations, but techniques that retain correlation structure for multivariate analyses are extremely complex. For more discussion of imputation, see Kalton and Kasprzyk.<sup>13</sup> The following sections describe the imputation that occurs for the Survey File and Cost Supplement File data; imputation does not pertain to the Survey File - Early Release data.

### Survey File

For the Survey File, imputation occurs for 1) **income and assets data** and 2) **insurance premiums**.

The MCBS imputes income when income data are missing. Using the hot deck imputation method, the MCBS first imputes whether an income source exists (such as Social Security). If the income source exists, then the amount earned is imputed. A flag is created for each imputed variable indicating whether the corresponding value was imputed. The MCBS imputes different sets of variables for different types of respondents. For respondents who responded to the Income and Assets Questionnaire (IAQ), the MCBS imputes a selection of variables from the questionnaire, including probe variables, which are indicators of whether the beneficiary and/or the spouse/partner had income or asset items, and amount variables, which give the amount of the income or asset items that the beneficiary and/or the spouse/partner had. For respondents who did not

<sup>13</sup> Graham Kalton and Daniel Kasprzyk, "The Treatment of Missing Survey Data," *Survey Methodology* 12, no. 1 (1986): 1-16.



respond to the questionnaire, which includes beneficiaries living in a facility, only the amount of total income is imputed.

Beginning in 2022, the MCBS also imputes medical insurance premiums for beneficiaries who did not provide complete premium information about their private plan or their MA plans.

The MCBS creates one imputation flag for each imputed variable. For the probes, only the hot deck imputation method is used, so the imputation flags indicate whether the probe was imputed or not. For the amounts, the MCBS uses a variety of imputation methods. The imputation flags indicate whether the amount was not imputed, imputed by the hot deck method, imputed by the carry forward method, or imputed by data edits.

## Cost Supplement File

Imputation occurs across three levels for the Cost Supplement File: **imputation of payments; MA Encounter ratio adjustment; part-year ratio adjustment.**

Using information from the Cost Supplement File segments and Medicare claims data, the MCBS imputes missing payer and payment information for reported medical events. For beneficiaries living in a facility, medical event data are provided only from Medicare claims data. The MCBS first imputes whether a payer, such as an insurance plan, paid for a particular event. If the payer paid, then the amount paid was imputed next. Imputation is performed using the hot deck imputation method, and a flag is created for each imputed variable indicating whether the corresponding cost value was imputed. These imputed values are reflected in the event-level data.

Next, an MA Encounter data ratio adjustment accounts for medical events that were not reported by survey respondents who were covered by MA. This adjustment is applied within the outpatient, inpatient, institutional, and medical procedure LDS segments. This adjustment was introduced for the 2019 Cost Supplement File LDS to improve estimation of medical events and costs for beneficiaries enrolled in MA. Beginning with 2020, these MA Encounter data adjustments were improved to better reflect age and general health-related differences.

Finally, a part-year ratio adjustment is applied to account for reference period gaps within the calendar year which are usually caused by Not-In-Round (NIRs).<sup>14</sup> The adjustments are applied to the event-level data and published in the service-level and person-level summaries. Imputation is performed using the hot deck imputation method, and a flag is created for each imputed variable indicating whether the corresponding value was imputed.

Additional information on how the imputation was conducted is available in the *MCBS Methodology Report*.

## 8.4 Weighting

### 8.4.1 Preparing Statistics (Using the Full Sample Weights)

Two types of weights are provided for the MCBS, **cross-sectional weights** and **longitudinal weights**. The data user may choose to conduct analyses of the Survey File - Early Release data alone, or of the final Survey File data alone or in combination with Cost Supplement File data. The Survey File weights are for analysis of Survey File data only; data users cannot use the Survey File weights with Cost Supplement File data. Users who want to analyze Survey File data along with cost and utilization data in the Cost Supplement File should limit analysis to cases with a positive Cost Supplement File weight (i.e., use the provided Cost Supplement File

<sup>14</sup> Not-In-Round refers to cases in which the respondent was not available to be interviewed within the round's time frame.

weights). Survey File - Early Release data should not be combined with the Survey File or Cost Supplement File data but rather is intended to be used on its own for timely analysis.

The following sections describe the full-sample weights. The term “full-sample” distinguishes these weights from the replicate weights used for variance estimation, as discussed in the Section 8.6: Variance Estimation. Each weights segment includes both the full-sample and replicate weights. Users do not need to apply formats to the segments as the weights are real numbers and do not need to be grouped or labeled. Additional information on the specific weights released on each LDS is available in the *MCBS Data Year Release Notes*.

## Cross-Sectional Weights

Cross-sectional weights apply to the entire file of all beneficiaries who completed an interview, either Community or Facility. Cross-sectional weights are available for the Survey File - Early Release, the Survey File, and the Cost Supplement File in each data year.

**Survey File - Early Release.** For analysis of Survey File - Early Release data, there is one population of inference that can be obtained using the weights, the ever enrolled population. There are no weights representing the continuously enrolled population for the Survey File - Early Release.

- *Survey File - Early Release Ever Enrolled Cross-Sectional Weights:* The ever enrolled Survey File - Early Release weight is greater than zero for all beneficiaries in the Survey File - Early Release. This weight segment is ER\_EVRWGTS, and the name of the weight is ER\_EEYRSWGT. The sum of this weight represents the population of beneficiaries who were entitled and enrolled in Medicare for at least one day at any time during the calendar year, and still enrolled and living in the community in the fall.

**Survey File.** For analysis of Survey File data, there are two populations of inference that can be obtained using two distinct cross-sectional weights, the ever enrolled population and the continuously enrolled population.

- *Survey File Ever Enrolled Cross-Sectional Weights:* The ever enrolled Survey File weight is greater than zero for all beneficiaries in the Survey File. This weight segment is EVRWGTS, and the name of the weight is EEYRSWGT. The sum of this weight represents the population of beneficiaries who were entitled and enrolled in Medicare for at least one day at any time during the calendar year.
- *Survey File Continuously Enrolled Cross-Sectional Weights:* The continuously enrolled Survey File weight is greater than zero for the subset of beneficiaries in the Survey File who were continuously enrolled in Medicare from January 1, through completion of their fall interview in the calendar year. This weight segment is CENWGTS, and the weight is named CEYRSWGT. The population represented by the sum of this weight is the continuously enrolled population of Medicare beneficiaries who were enrolled from the first of the year through the fall of the calendar year. Users should use the continuously enrolled Survey File weight for time series analysis of survey data across years.

**Cost Supplement File.** For analysis of Cost Supplement File data, there is one population of inference that can be obtained using the weights, the ever enrolled population. There are no weights representing the continuously enrolled population for the Cost Supplement File.

- *Cost Supplement File Ever Enrolled Cross-Sectional Weights:* The ever enrolled Cost Supplement File weight represents an ever enrolled population of Medicare beneficiaries enrolled in Medicare on at least one day at any time in the calendar year. This weight segment is CSEVWGTS, and the name of the weight is CSEVRWGT. The population represented by the sum of this weight is identical to the population represented by the sum of the ever enrolled Survey File weight, but it is populated for a smaller subset of respondents with complete cost and utilization data.

To define the population, the MCBS creates a calendar history of a beneficiary's MCBS interviews. A number of eligibility checks are run against this calendar history to identify beneficiaries who met eligibility requirements for inclusion in the survey data for the calendar year, either because they were interviewed for a full year or interviewed until death or loss of Medicare entitlement. Beneficiaries who pass these eligibility checks become the population eligible for the Cost Supplement File ever enrolled weight. CSEVWGTS includes records for beneficiaries who were sampled in all four panels (i.e., Continuing and Incoming Panels). The Continuing Panels provide survey-reported cost and utilization for the calendar year through their participation in the MCBS during the winter round of the data collection year through the winter round of the following data collection year. The Incoming Panel is first interviewed in the fall round of the data collection year and do not provide cost and utilization data for the period of time between enrollment and completion of the fall interview; for newly-enrolled beneficiaries from this panel who joined the Medicare program during the calendar year, cost and utilization data for the period between the fall interview and the end of the calendar year are collected in the subsequent winter interview. For these beneficiaries, a combination of the survey-collected data for the end of the year and Medicare claims data are used to impute beneficiary-level data for the entire period of enrollment in the calendar year.

## Longitudinal Weights

Longitudinal weights allow for the study of respondents across data years. The following longitudinal weights are provided with the Survey File and Cost Supplement LDS's.<sup>15</sup> Longitudinal weights are not available for the Survey File - Early Release LDS.

**Survey File.** There are three sets of longitudinal weights released on the Survey File. The descriptions below reflect the longitudinal weights that would be released on the Survey File LDS for Calendar Year (CY) 4.

- *Survey File Two-Year Longitudinal Weights:* Two-year longitudinal weights apply to respondents who completed fall round interviews in the current and the preceding year. This weight segment is LNG2WGTS, and the name of the weight is L2YRSWGT. This set of weights can be used to study data trends over a two-year period and are populated only for members of the CY1, CY2, CY3 panels who had CY3 and CY4 Survey File data and were continuously enrolled for two years. The population represented by these weights is the population of beneficiaries enrolled on or before 1/1/CY3 and surviving and entitled as of completion of the fall interview in CY4. By applying these weights to data in the current and preceding year, users will be able to estimate change among the Medicare population who were alive for the full two-year period.
- *Survey File Three-Year Longitudinal Weights:* Three-year longitudinal weights apply to respondents who completed fall round interviews in the current and the two preceding years. This weight segment is LNG3WGTS, and the name of the weight is L3YRSWGT. This set of weights can be used to study data trends over a three-year period and are populated only for members of the CY1 and CY2 panels who were continuously enrolled during all of the years CY2-CY4 and had Survey File data in CY2 and CY4. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before 1/1/CY2 and were still alive and entitled as of completion of the fall interview in CY4. By applying these weights to data in the current and the three preceding years, users will be able to estimate change among the Medicare population who were alive for the full three-year period.
- *Survey File Four-Year Longitudinal Weights:* Four-year longitudinal weights apply to respondents who completed fall round interviews in the current and the three preceding years. This weight segment is LNG4WGTS, and the name of the weight is L4YRSWGT. This set of weights can be used to study data trends over a four-year period and are populated only for members of the CY1 panel who were

<sup>15</sup> Beginning with the 2016 LDS, the Survey File longitudinal weight names reflect the number of years the beneficiary was enrolled in Medicare (i.e., LNG2WGTS weights are referred to as 'two-year' rather than 'one-year' as they represent the population continuously enrolled for two years). This change was made to align the names of the longitudinal weights in the Survey File LDS with the naming convention used for the Cost Supplement LDS.

continuously enrolled during all of the years CY1-CY4. The resulting weights represent the population of Medicare beneficiaries who enrolled on or before 1/1/CY1 and were still alive and entitled as of completion of the fall interview in CY4. By applying these weights to data in the current and the three preceding years, users will be able to estimate change among the Medicare population who were alive for the full four-year period.

**Cost Supplement File.** There are two sets of longitudinal weights released on the Cost Supplement File. The descriptions below reflect the longitudinal weights that would be released on the Cost Supplement File LDS for CY4.

- *Cost Supplement File Two-Year Longitudinal Weights:* The two-year longitudinal weights are populated for members of the CY1, CY2, and CY3 panels who were ever enrolled in Medicare at any time during both CY3 and CY4 and provided utilization and cost data for both years. This weight segment is CSL2WGTS, and the name of the weight is CSL2YWGT.
- *Cost Supplement File Three-Year Longitudinal Weights:* The three-year longitudinal weights are populated for members of the CY1 and CY2 Panels who were ever enrolled in Medicare at any time during CY2, CY3, and CY4, and provided utilization and cost data for all three years. This weight segment is CSL3WGTS, and the name of the weight is CSL3YWGT.

### 8.4.2 Special Topical Segment Weights

Each data year, a subset of the segments released on the Survey File - Early Release and Survey File should be used with special weights. These segments are referred to as "Topical" segments because most of them were traditionally sourced from the Topical questionnaire sections. To generate estimates using these Topical segment data, on their own or merged with another Survey File - Early Release/Survey File segment, always use the special full-sample and replicate weights. These weights are either included directly in the Topical segment or in standalone Topical weights segment(s). Please refer to the historical 2015-2022 *Data User's Guide* and the *Data Year Release Notes* for more information on the location and use of these weights. Do not use the weights that appear in the non-adjusted weight segments (see above) when using data from a Topical segment.

The questionnaire sections (or specific items within questionnaire sections) that are weighted separately are fielded in the winter and summer rounds following the data year, and/or are not administered to proxy respondents. 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.

**Survey File - Early Release.** For the Topical segments released on the Survey File - Early Release, there is **one set** of full-sample and replicate weights that can be used for each Topical segment based on the Survey File - Early Release ever enrolled population. These weights may be used to conduct joint analyses of Topical segment data and other Survey File - Early Release data.

- The Topical weights that are described as "Survey File - Early Release ever enrolled" correspond to the Survey File - Early Release ever enrolled population and can be used to conduct analyses of the Topical data as representing the ever enrolled population and in conjunction with other Survey File - Early Release data.

**Survey File.** For the Topical segments released on the Survey File, there are **three sets** of full-sample and replicate weights that can be used for each Topical segment, one based on the Survey File ever enrolled population, one based on the Survey File continuously enrolled population, and one based on the Cost Supplement ever enrolled population. These weights may be used to conduct joint analyses of Topical segment data, other Survey File data, and Cost Supplement File data.

- The Topical weights that are described as "Survey File ever enrolled" weights correspond to the Survey File ever enrolled population and can be used to conduct analyses of the Topical data as representing the ever enrolled population and in conjunction with other Survey File data.
- The Topical weights that are described as "Survey File continuously enrolled" weights correspond to the Survey File continuously enrolled population and can be used to conduct analyses of the Topical data as representing the continuously enrolled population and in conjunction with other Survey File data.
- The Topical weights that are described as "Cost Supplement ever enrolled" weights correspond to the Cost Supplement ever enrolled population and can be used to conduct analyses of the Topical data as representing the ever enrolled population and in conjunction with Cost Supplement File data. Weights corresponding to the Cost Supplement File continuously enrolled population are not available for the Topical data. Because the Cost Supplement File data are available for a smaller subset of the Survey File population, for each Topical segment the number of beneficiaries with a continuously enrolled Survey File Topical weight is larger than the number of beneficiaries with an ever enrolled Cost Supplement File Topical weight.

**There are no weights that support joint analysis between two Topical segments.** Each Topical segment has a different set of beneficiaries included. A user could merge data from one Topical segment onto another and then use one of the Topical segment's weights as the Baseline population, but the data will not align and there will be gaps. For some combinations of the different questionnaire sections (e.g., questionnaire sections fielded in the same round to the same population), the amount of missing data may be small enough that users could still conduct analyses.

Prefixes for these weights changed slightly in 2018 to accommodate the Survey File ever enrolled population and make the populations clearer to the data users across weights.

## 8.5 Using the Data

### *8.5.1 Merging Segments*

Data users can merge segments within the Survey File - Early Release. Data users can merge segments within and/or across the Survey File and Cost Supplement File. Appendix A provides sample SAS code for constructing an analytic file and merging data using the MCBS LDS's. Note that although the MCBS data are nationally representative, they are not representative at the regional or state level and cannot be used to produce regional or state-level estimates. However, the data user can use the data to look for national trends across population groups.

## 8.6 Variance Estimation (Using the Replicate Weights)

### *8.6.1 Variables Available for Variance Estimation*

In many statistical packages, the procedures for calculating sampling errors (e.g., variances, standard errors, margins of error) assume that the data were collected in a simple random sample. Procedures of this type are not appropriate for calculating the sampling errors of statistics based upon a stratified, unequal-probability, multi-stage sample such as the MCBS. Unless the complex nature of the MCBS is taken into account, estimates of the variance of a survey statistic may be biased downward.

The MCBS includes variables to obtain weighted estimates and estimated standard errors using either the Taylor-series linearization approach or balanced repeated replication (BRR) method, also known as Fay's method. There is both serial and intra-cluster correlation in the MCBS data, including: sampling second-stage units within primary sampling units; sampling beneficiaries with second-stage units; and repeated observations



of the selected beneficiary across time. Researchers should use the BRR method of variance estimation to account for various correlations. For details on the strengths and weaknesses of the two variance estimation methods, please refer to Wolter.<sup>16</sup>

To estimate variance using the balanced repeated replication method, a series of replicate weights are included in the Survey File release. There are many types of full-sample weights, including those for cross-sectional analyses, longitudinal analyses, and analyses of Topical data. Each of these full-sample weights has a corresponding set of replicate weights. The replicate weights can be used to calculate standard errors of the sample-based estimates as described below.

- For the Survey File - Early Release, the replicate cross-sectional weights are labeled ER\_EEYRS1 through ER\_EEYRS100 corresponding to the ever enrolled weight ER\_EEYRSWGT, which may be found on ER\_EVRWGTS.
- For the Survey File, the replicate cross-sectional weights are labeled CEYRS001 through CEYRS100 corresponding to the continuously enrolled weight CEYRSWGT, and EEYRS001 through EEYRS100 corresponding to the ever enrolled weight EEYRSWGT. These weights may be found on CENWGTS and EVRWGTS, respectively. The Survey File replicate longitudinal weights are found on segments LNG2WGTS, LNG3WGTS, and LNG4WGTS.
- For the Cost Supplement File, the replicate cross-sectional weights are labeled CSEVR001 through CSEVR100 corresponding to the continuously enrolled weight CSEVRWGT, which may be found on CSEVWGTS. The Cost Supplement File replicate longitudinal weights are found on segments CSL2WGTS and CSL3WGTS.

The variables SUDSTRAT (sampling strata) and SUDUNIT (primary sampling unit) are used for variance estimation using the Taylor series linearization method. As with the BRR method, the Taylor series method can be applied for variance estimation using SUDSTRAT and SUDUNIT for any of the Core and Topical segments with full-sample weights. The Taylor series method does not employ the replicate weights that are included on the segments. For examples and guidance on using the Taylor series linearization method of variance estimation or the BRR method, please see Appendix A.

### *8.6.2 Variance Estimation for Analyses of Single Year of MCBS*

Most commercial software packages today include techniques to accommodate the complex design, either through Taylor-expansion type approaches or replicate weight approaches. Among these are R®, STATA®, SUDAAN®, and the complex survey procedures in SAS.

### *8.6.3 Subgroup Analysis*

When analyzing survey data, researchers are often interested in focusing their analyses on specific subgroups of the full population sample (e.g., Medicare beneficiaries aged 65 and over, Hispanics, or females). A common pitfall when performing sub-group analysis of survey data when variance estimation methods such as Taylor-series are used is to delete or exclude observations not relevant to the subgroup of interest. Standard errors for MCBS estimates are most accurate when the analytic file includes all beneficiaries. However, when replicate weights are used for variance estimation, deleting observations not relevant to the subgroup of interest prior to analyzing the subgroup will still produce unbiased standard errors. Almost all statistical packages provide the capability to limit the analysis to a subgroup of the population.

The Taylor Series linearization method of variance estimation is not recommended for subgroup analysis with MCBS data because accidentally excluding any observation in the sample while conducting the subgroup

<sup>16</sup> Kirk Wolter, *Introduction to Variance Estimation* (Springer Science & Business Media, 2007).

analysis using this variance estimation method will result in biased standard error estimates. Variance estimation using the Taylor Series linearization method for subgroup analyses requires a “domain” or “subgroup” statement (available in most statistical packages) to account for estimated domain sizes (i.e., uncertainty in the denominator). The recommended method of variance estimation for subgroup analysis is the BRR method, which does not require any special subgroup considerations. The BRR method allows the researcher to subset data to a subgroup of interest and still produce unbiased standard error estimates.

## 8.7 Combining Multiple Years of Data

The MCBS is based on a rotating panel design, which allows for longitudinal analysis of up to four years when appropriate longitudinal weights are used. Multiple years of MCBS data can also be pooled to perform serial cross-sectional or pooled analysis. The appropriate method to combine data across years will depend on the analytic design of the study. Sample code is presented in Appendix A to demonstrate the steps involved in combining multiple years of data to perform two types of analysis: (1) Longitudinal analysis; (2) Pooled, cross-sectional analysis.

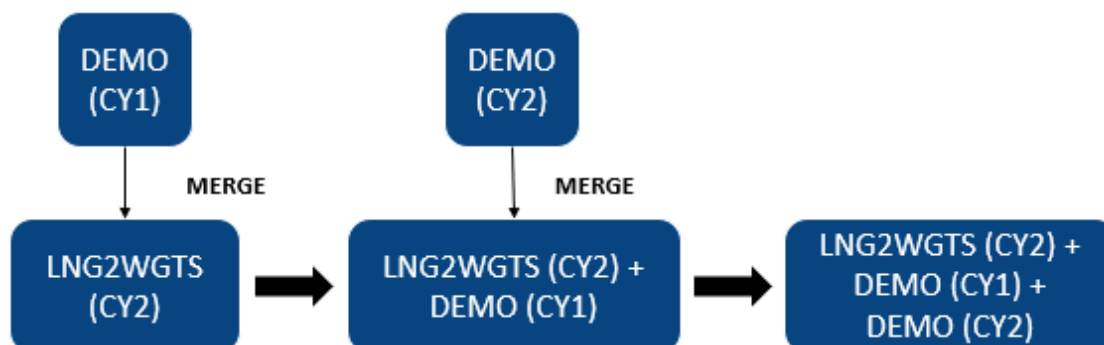
### 8.7.1 Longitudinal Analysis

The study objective in longitudinal analysis is to assess changes over time for each beneficiary. The Survey File cross-sectional and longitudinal population definitions are consistent from year to year, so the data are comparable between years. The Cost Supplement File cross-sectional population definition is also consistent and comparable from year to year.

Most longitudinal analyses require the data to be in long-format (i.e., repeated observations – each representing a calendar year the sample person was surveyed – are stored in a separate row for each sample person). To construct a longitudinal analytic dataset, the first step is to use the appropriate longitudinal weights file. For example, as shown in Exhibit 7.7.1, to assess changes over time beneficiaries who have been in the sample for at least two years – from CY1 to CY2 – the two-year longitudinal weights (i.e., one-year “backward longitudinal weights”) (LNG2WGTS) should be used.

Variables from current year files representing the outcome of interest should then be merged with the current year’s longitudinal weights file. While merging, all observations in the weights file should be preserved. Next, the same variables from the prior year’s files should be merged with the current year’s longitudinal weights file.

**Exhibit 8.7.1:** Constructing a Longitudinal Analytic File



### Variance Estimation for Longitudinal Analysis (Using Replicate Weights)

Just as there are full-sample longitudinal weights, there are corresponding sets of replicate weights. The replicate weights included in the longitudinal weights data files can be used to calculate standard errors of the sample-based estimates. The first set of replicate longitudinal weights is labeled L2YRS001 through L2YRS100

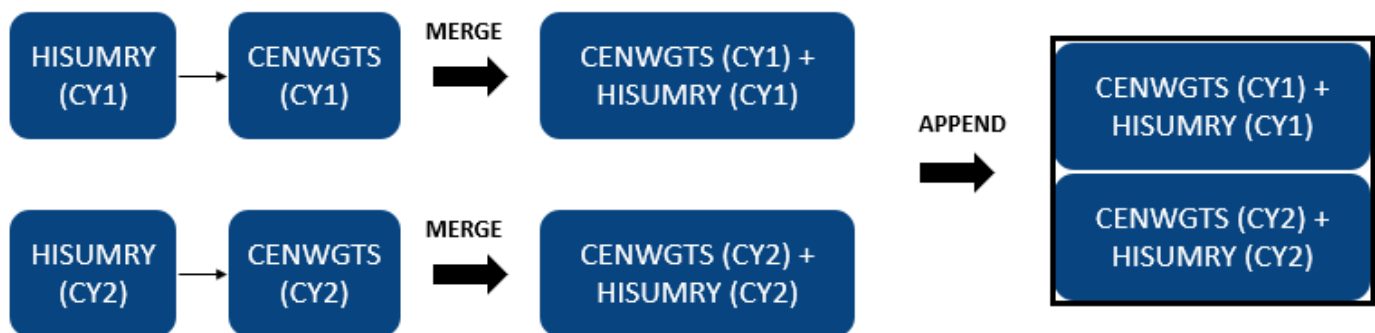
and may be found on the two-year longitudinal weights file (LNG2WGTS). The second set of replicate longitudinal weights is labeled L3YRS001 through L3YRS100 and may be found on the three-year longitudinal weights file (LNG3WGTS). The third set of replicate longitudinal weights in the Survey File LDS is labeled L4YRS001 through L4YRS100 and may be found on the four-year longitudinal weights file (LNG4WGTS).

For additional guidance, see the *MCBS Advanced Tutorial on Longitudinal Analysis*: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials>.

### 8.7.2 Repeated Cross-Sectional or Pooled Analysis

Multiple years of MCBS data can be pooled to perform serial cross-sectional or pooled analysis. Repeated cross-sectional analysis is used for analyzing changes in the Medicare population as a whole over time, while a pooled analysis will produce single cross-sectional estimates that cover a period of multiple years, usually to increase total sample sizes. In contrast, the longitudinal analysis described earlier is used to analyze beneficiary-level changes over time. Pooled data analysis yield estimates that are in effect a moving average of nationally representative year-specific estimates. The pooled estimates can be interpreted as being representative of the midpoint of the calendar year of the pooled period. Exhibit 8.7.2 demonstrates the steps involved in constructing a repeated cross-sectional or pooled analytic dataset using CY1 and CY2 data. For each year in the study, variables representing the outcome of interest should then be merged with the cross-sectional weights file. While merging, all observations in the weights file should be preserved. Next, the year-specific files are appended to produce the analytic dataset.

**Exhibit 8.7.2:** Constructing a Repeated Cross-Section or Pooled Analytic File



### Variance Estimation for Repeated Cross-Sectional or Pooled Analysis (Using Replicate Weights)

Due to the rotating-panel and multistage-sampling design of the MCBS, there is both serial and intra-cluster correlation in the data when pooling multiple years of data. When conducting a pooled analysis, using the balanced half-sample method (also known as the balanced repeated replication, or BRR, method) of variance estimation throughout appropriately accounts for the various correlations due to sampling second-stage units within primary sampling units, sampling beneficiaries within second-stage units, and repeated observations of the selected beneficiary across time. The replicate cross-sectional weights are labeled CEYRS001 through CEYRS100 and can be found in each year's cross-sectional weights file (CENWGTS). A pooled analysis of this type can also be applied to longitudinal weights (LNG2WGTS, LNG3WGTS, LNG4WGTS).

When conducting a repeated cross-sectional analysis to compare between two years, the difference or net change in a population characteristic is often of interest. In this type of analysis, a point estimate of year-to-year difference is straightforward to calculate; simply take difference between the two individual annual cross-sectional estimates. Each cross-sectional estimate included in the comparison can be calculated using the full-sample weights included in that year's data release.



Calculating variance and standard error estimates of net change is more complicated because of correlation between the two annual data sets. Correlation is present because many beneficiaries are retained from one year to the next, and because the same set of Primary Sampling Units (PSUs) and Secondary Sampling Units (SSUs) are used for each year. We refer to these types of correlation as serial and intra-cluster correlation, respectively.

To estimate the variance of net change estimates, the researcher may rely on a program such as SAS or calculate them directly in their own custom program using a closed formula.

For additional guidance, see the *MCBS Advanced Tutorial on Pooled Cross-Sectional Analysis* and the *MCBS Advanced Tutorial on Year-to-Year Estimate Comparisons*: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials>.

### SAS Method

In SAS, point estimates of year-to-year differences, in addition to corresponding estimates of standard errors, can be generated using PROC SURVEYREG. To use this method, first concatenate the two annual datasets by stacking them together vertically, each including its corresponding set of weights, and define a YEAR variable to indicate which data year each of the two files represents. From this concatenated dataset, the example SAS code below will output estimates of the difference in estimates between the two years, using Cost Supplement File weights as an example. The standard errors associated with these estimates are the desired estimated standard errors of the year-to-year net change.

```
PROC SURVEYREG VARMETHOD=BRR(FAY=.30);
  CLASS YEAR;
  MODEL variable = YEAR;
  LSMEANS YEAR / DIFF;
  WEIGHT CS1YRWGT;
  REPWEIGHT CS1YR001-CS1YR00100;
RUN;
```

This process can be repeated for any combination of variable and a complementary set of cross-sectional full-sample and replicate weights (e.g., Survey File continuously enrolled, Survey File ever enrolled, Cost Supplement File ever enrolled weights).

### Direct Method

The variance of a difference can also be calculated directly using the formula below, which a researcher can incorporate into a custom program for producing a variety of estimates of net change. This process does not require concatenating two annual files together, although programmatically it may be useful to do so. Let  $X_0^t$  be the cross-sectional estimate of the mean of population characteristic Y from year t using the full-sample weights from that year, and let  $X_1^t, X_2^t, \dots, X_{100}^t$  be cross-sectional estimates of the same population mean from year t using each of the 100 corresponding replicate weights. Similarly, let  $X_0^{t-1}, X_1^{t-1}, X_2^{t-1}, \dots, X_{100}^{t-1}$  be analogous estimates of the same population characteristic Y from year t-1, using the weights from year t-1. Next, define a set of difference variables as  $D_0 = X_0^t - X_0^{t-1}$ ,  $D_1 = X_1^t - X_1^{t-1}$ , etc.

Then,

$$Var(D_0) = \frac{2.04}{100} \sum_{i=1}^{100} (D_i - D_0)^2$$

is an estimate of the variance of the estimate of net change from year t-1 to year t. The square root of this estimate is the estimated standard error.

## 8.8 Data File Notes: Global Information

This section is a collection of information about various data fields present in the LDS segments. The MCBS does not attempt to present information on every survey data field; rather, it concentrates its efforts on data fields where additional clarity or detail may be useful. The MCBS starts with information that is applicable globally, followed by specific information on individual segments, presented in the same sequence as the segments appear in the codebook.

### 8.8.1 Key Variables

There are several variables that appear on each segment to merge segments and/or identify each segment.

- **BASEID:** This key identifies the person interviewed. It is an 8-digit element, consisting of a unique, randomly assigned 7-digit number concatenated with a single-digit check digit.
- **Version Number:** Often files are re-released due to needed updates, which will be noted by the version number variable.
- **Survey Year:** The Survey Year of interest is included as a variable on each segment.

LDS segments may vary in the number of BASEIDs. This variation may occur for several reasons. First, some segments include data from Community components and others from Facility components with different numbers of beneficiaries providing responses. Second, there are also differences in the number of beneficiaries by the specific round completed. Third, the use of ever enrolled or continuously enrolled weights in constructing the segments may result in differences.

### 8.8.2 Item Non-Response and Missing Values

As in any other survey, some respondents could not, or would not, supply answers to some questions.<sup>17</sup> Item non-response rates are generally low in the MCBS data, but the researcher still needs to be aware of the missing data and be cautious about patterns of non-response. Some of the missing data are attributable to the fact that some of the Community interviews and all the Facility interviews are conducted through a proxy respondent. In other words, the respondent may not have had knowledge of the information sought on the sample person. In other situations, the respondent may have simply refused to answer.

Various special values indicate the reason some data are missing, such as .R for "refused," .D for "don't know," and .N for "invalid skip." See Exhibit 8.3.1 above for additional information.

### 8.8.3 Derived and Administrative Variables

Variables that were derived or created by combining two or more survey variables are preceded with the characters "D\_", such as D\_ERVIST. CMS may create or modify variables to recode data items (e.g., to protect the confidentiality of survey responses) or to globally edit some variables. Variables preceded by the character "H\_", such as H\_DOB, come from CMS administrative source files.

Several segments include variables indicating the length of time the beneficiary spent doing something, such as waiting in the hospital emergency room or waiting for an appointment. In the questionnaire, the length of time is stored in a continuous variable while the corresponding unit is stored in a categorical variable (e.g., hours, minutes, or hours and minutes; days, weeks, or months). These two variables are used to derive a single variable indicating the length of time in the most appropriate unit of time. For example, on the

<sup>17</sup> This is different from when an individual refuses to participate in the survey altogether, which is called unit non-response. Unit non-response is discussed in detail in the *MCBS Methodology Report*.

ACCSSMED segment, D\_ERTIME contains the length of time spent waiting in the hospital emergency room in minutes while D\_MDAPPT contains the length of time spent waiting for a doctor's appointment in days.

#### 8.8.4 Initial Interview Variables

Some questions are asked in only two scenarios: 1) it is the case's Baseline (initial) interview or 2) it is the first time the case has crossed to a new component (e.g., the case crosses from the Community component to the Facility component for the first time). These "initial interview variables" are not asked again during subsequent interviews because the responses are not likely to change. Such questions include "Have you ever served in the armed forces?" and "What is the highest grade of school you ever completed?" To maximize the usefulness of this release as a cross-sectional file, these data are pulled forward from the Baseline interview or the first time the case was interviewed in a given component, as applicable. For each data year, these variables are listed in the *Data Year Release Notes* or the historical 2015-2022 *Data User's Guides*.

#### 8.8.5 Ever Variables

Many items in the MCBS ask respondents whether they have ever had certain experiences, such as ever being told they have a chronic condition, receiving a treatment, or doing a specific activity (such as ever accessing the official Medicare website). Such 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 reports "yes" to having had that condition or experience.

These items are administered to respondents in certain scenarios. For select "ever" variables administered in the Health Status and Functioning Questionnaire (HFQ), there are different versions of each question, depending on whether a respondent is in the Incoming Panel sample or Continuing sample. These versions are combined into recoded variables to provide a complete picture of the response. All Incoming Panel sample respondents are asked if they have ever had certain conditions or experiences. Once a condition or experience is reported, the CAPI questionnaire logic retains that information for subsequent interviews. 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. However, if a negative response is given, respondents are asked annually thereafter if they had that specific condition or experience in the past year. 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. All data from a beneficiary from the current survey year and all previous years are used to determine whether the beneficiary has ever had a condition or experience.

"Ever" variables in the Nicotine and Alcohol Use Questionnaire (NAQ) and Beneficiary Knowledge and Information Needs Questionnaire (KNQ) are collected and processed in a similar manner to the HFQ "ever" variables, except that the NAQ and KNQ "ever" variables use only one version of each question (rather than two separate versions depending on beneficiary sample type).

#### 8.8.6 Other Specify Questions

A subset of MCBS questionnaire items include closed ended responses with "other specify" options. These options allow respondents to provide answers that are not included in the existing code frame and are useful for questions with a wide range of potential responses (e.g., types of problems experienced during attempts to obtain care). In the event that an "other specify" option is selected, interviewers record actual responses verbatim. Verbatim responses are not released.

In Community data processing, the MCBS programmatically identifies "other specify" responses that are sufficiently similar to existing code frame options and back codes responses into existing response option categories as appropriate. This is accomplished by identifying keywords and misspellings corresponding to

each existing response option, programmatically searching the verbatim “other specify” responses for the keywords and misspellings, and then categorizing responses into existing response option categories as appropriate. Codes are then assigned to similar responses to facilitate analysis; there are no verbatim responses provided on the released segments. Often there will be more than one answer to a single question. In these cases, responses are recoded into several variables, all of which contain categorized data. Code lists are updated when necessary to incorporate responses that are frequently provided in “other specify” response options.

### *8.8.7 Interview Mode Indicator*

MCBS data collection is multimode, with both in-person and phone interviewing. The Interview Characteristics (INTERV) segment includes a flag to indicate whether the interview was conducted in-person (INTMODE = 1), by phone (INTMODE = 2), or using a combination of the two modes (“hybrid”, INTMODE = 3). Data collection modes are discussed in further detail in the *MCBS Methodology Report*.

### *8.8.8 Analytic Notes for Non PM Event Segments*

The event-level segments on the Cost Supplement File containing inpatient hospital services, institutional services, and outpatient hospital services include a Survey Only with Medicare Payment (SOWMP) variable. This flag variable is only set to 1 if the FFS administrative claims data are not matched to the survey-reported event data and there is a survey-reported Medicare payment. However, if there is a reported or imputed MA payment for these unmatched records with Medicare payments, the SOWMP flag is not set to 1. The conclusion is that when SOWMP is set to 1, the Medicare payment for these unmatched, survey-reported events is a traditional Medicare FFS payment, which would have already been included in the matched claims or claims-only records. Thus, these payments likely duplicate those claims, so the SOWMP flag is set to 1, and these records are excluded from the summary-level segments.

### *8.8.9 Do Administrative Data Override Survey-Reported Data?*

In linking survey-reported and administrative data, the MCBS keeps records from both sources to provide more complete data. Indicators in the file will usually specify if the information is survey-reported only, administrative data only, or both. Data that are only administrative are indicated as such in the data documentation and codebook.

## 9. REFERENCES

Kalton, Graham and Daniel Kasprzyk. "The Treatment of Missing Survey Data." *Survey Methodology* 12, no. 1 (1986): 1-16.

Wolter, Kirk. *Introduction to variance estimation*. Springer Science & Business Media, 2007.

# APPENDICES

## 10. APPENDICES

### Appendix A: Sample Code<sup>18</sup>

The sample code below provides general examples for using MCBS LDS data. **Please note, the availability and contents of the MCBS LDS segments and variables used in these examples, along with the R syntax, may change year-to-year. Users should refer to the year-specific data documentation (including the annual *Data Year Release Notes* and historical *Data User's Guides*) and *MCBS LDS Variable Crosswalks* for more information on the data available each year.**

In some cases, the Survey File - Early Release LDS can be used in place of the Survey File LDS; however, this LDS is intended to be used for timely, standalone analyses. The Survey File LDS should be used for analyses involving a broader range of topics (including Facility data), finalized enrollment and claims data, data from the Cost Supplement File LDS, and analyses that represent the continuously enrolled or longitudinal populations. For MCBS data or analysis-related questions, please email [MCBS@cms.hhs.gov](mailto:MCBS@cms.hhs.gov).

#### *Example A.1 Joining Segments within the Survey File and Survey File - Early Release LDS*

Data users can join segments within the Survey File, Survey File - Early Release, and Cost Supplement File. This example provides sample SAS code for the construction of an analytic file for a hypothetical research question that involves studying the self-reported general health for Medicare beneficiaries living in the community with diabetes.

First, there are two measures required to identify our study population: residence status and self-reported diabetes. Variables corresponding to these measures can be found in the following Survey File segments, respectively: Demographics (DEMO) and Chronic Conditions (CHRNCOND). General health information is found in the General Health (GENHLTH). To ensure estimates are representative of the continuously enrolled Medicare population, use weights from the CENWGTS file. Please note that if using a Survey File LDS Topical segment (such as CHRNPAIN), users should instead join onto the Topical segment and use the adjusted weights included with that segment. If using the Survey File - Early Release LDS instead of the Survey File, estimates are representative of the ever enrolled Medicare population who completed a fall Community interview (rather than the continuously enrolled population) and use the weights from the ER\_EVRWGTS file.

Below, is an example of how multiple Survey File segments can be joined with the CENWGTS segment in SAS using BASEID as the key variable. Sample code is provided for two alternative methods for joining MCBS data, one using PROC SQL and one using SAS merge; users can use their preferred method. When joining segments, all observations in the CENWGTS segment should be preserved. The same approach applies to joining data on the Survey File - Early Release, just using the ever enrolled weights.

#### **SAS (SAS Merge Method)**

Data merged;

```
merge surveyYY.CENWGTS (in = a)
      surveyYY.DEMO (keep = BASEID H_AGE INT_TYPE)
      surveyYY.CHRNCOND (keep = BASEID D_OCDTYP)
      surveyYY.GENHLTH (keep = BASEID GENHELTH);
```

```
by BASEID;
```

```
if a;
```

```
run;
```

<sup>18</sup> The "YY" in "costYY" and "surveyYY" refers to the data year of the Cost Supplement File and Survey File, respectively. Longitudinal code is represented with the convention of Y1, Y2, etc.



**SAS (PROC SQL Join Method)**

```

PROC SQL;
CREATE TABLE joined AS
  SELECT A.*,
         B.h_age,
         B.int_type,
         C.d_ocdtyp,
         D.genhelth
  FROM   surveyyy.cenwgts AS A
        LEFT JOIN surveyyy.demo AS B
              ON A.baseid = B.baseid
        LEFT JOIN surveyyy.chrncond AS C
              ON A.baseid = C.baseid
        LEFT JOIN surveyyy.genhlth AS D
              ON A.baseid = D.baseid;
QUIT;

```

To segment the file to beneficiaries living in the community only, subset the file on the variable INT\_TYPE.

```

Data joined_surveyfile;
set joined;
where INT_TYPE = 'C'; /* denotes individuals living only in the community */
run;

```

Now there is an analytic file that includes all the Survey File variables and weights required to analyze general health for Medicare beneficiaries living in the community with diabetes. Data users can export the created dataset for use with R and Stata.

*Example A.2 Merging Segments in the Survey File and Cost Supplement File*

Below is the SAS code needed to create an analytic file that includes Survey File and Cost Supplement File variables. This example uses the Cost Supplement File ever enrolled general purpose weights required to analyze out-of-pocket spending for Medicare beneficiaries by age. Because the analysis involves annual out-of-pocket costs, the adjusted total out-of-pocket spending variable (PAMTOOP) in the Person Summary (PS) segment is used. PAMTOOP summarizes out-of-pocket spending for 11 types of events (dental services, hearing services, home health [including home health provider and home health friend events], hospice, inpatient, institutional, medical provider, outpatient, prescribed medication, vision services, and facility) for each beneficiary.

Sample code is provided for two alternative methods for merging/joining MCBS data, one using PROC SQL and one using a SAS merge; users can use their preferred method.

To produce estimates that are representative of the ever enrolled population, data users need to merge the Person Summary segment from the Cost Supplement File with the Demographics (DEMO) segment from the Survey File and limit the observations to respondents listed in the weights segment (CSEVWGTS). Inclusion of "(in=a)" and "if a then output" in the SAS merge code statement ensures that all observations in the weights segment are preserved, regardless of whether or not the other segments include the entire sample, and that the resulting dataset is restricted to the observations included in the weights file.\

**SAS (SAS Merge Method)**

```

data merged_costfile;
  merge      costYY.CSEVWGTS (in = a)
            surveyYY.DEMO (keep = BASEID D_STRAT)
            costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;

```

**SAS (PROC SQL Join Method)**

```

PROC SQL;
CREATE TABLE merged_costfile AS
  SELECT A.*,
         B.d_strat,
         C.pamtoop
  FROM   costYY.csevwgts AS a
        LEFT JOIN surveyYY.demo AS b
          ON a.baseid = b.baseid
        LEFT JOIN costYY.ps AS c
          ON a.baseid = c.baseid
  ORDER BY baseid;
QUIT;

```

This merged file now includes the relevant Survey File and Cost Supplement File variables that will allow for an examination of out-of-pocket spending for Medicare beneficiaries by age.

The code below transforms the age variable (D\_STRAT) from a seven-category variable into a four-category variable. Exhibit A.2.1 presents information regarding recoding this age variable.

```

data mcbs_analyticfile;
  set merged_costfile;

  /*AGE*/
  if D_STRAT in (1,2) then AGECAT = 0;
  else if D_STRAT in (3,4) then AGECAT = 1;
  else if D_STRAT in (5,6) then AGECAT = 2;
  else if D_STRAT = 7 then AGECAT = 3;

run;

```

**Exhibit A.2.1:** Example A.2 Recoding Variables

Measure	Original Variable	Recoded Variable
Age	D_STRAT 1: 0-44 2: 45-64 3: 65-69 4: 70-74 5: 75-79 6: 80-84 7: 85 +	AGECAT 0: 0-64 1: 65-74 2: 75-84 3: 85+

The sample code below demonstrates how to use the constructed analytic dataset to estimate out-of-pocket spending for all Medicare beneficiaries stratified by age. Although the MCBS includes variables to obtain weighted estimates and estimated standard errors using Taylor-series linearization approach, the balanced repeated replication (BRR or Fay's method) method provides more analytic flexibility when analyzing subgroups. Therefore, the examples presented in this section use BRR (Fay's method) variance estimation method.

\* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
    var PAMTOOP;
        weight CSEVRWGT;
        repweights CSEVR001-CSEVR100;
        domain AGECAT;
run;
```

### *Example A.3 Using Event-Level Data in the Cost Supplement File*

A researcher might be interested in using event-level data from the Cost Supplement File, for example to estimate spending on the cost of cardiovascular drugs for Medicare beneficiaries in the given year. In this case, data users need separate records for each prescription drug event a beneficiary had, which allows for identifying beneficiaries who were prescribed cardiovascular drugs and total spending during a given year. This example uses an event-level segment (i.e., the Prescribed Medicine Event segment [PME]), which allows for analysis of each prescription drug event for a given beneficiary.

To identify Medicare beneficiaries who were prescribed cardiovascular drugs during the year, data users need to limit the PME file to prescription drug events related to drugs with a cardiovascular therapeutic classification (i.e., for which variable THERCC=41). Next, data users need to calculate total spending on cardiovascular drugs during the year for each Medicare beneficiary.

#### **SAS (PROC SQL Method)**

```
PROC SQL;
CREATE TABLE drug_spending_bybene AS
    SELECT UNIQUE baseid,
           Sum(amttot) AS drug_spending
    FROM costyy.pme
    WHERE thercc = "41"
    GROUP BY baseid;
QUIT;
```

To produce estimates that are representative of the ever enrolled population, data users need to merge the weights segment (CSEVWGTS) and limit the dataset to the population included in the 'drug\_spending\_bybene' file. Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that observations that are in both the weights segment and the analytic file are preserved.

#### **SAS (SAS Merge Method)**

```
data mcbs_analyticfile;
    merge          costYY.CSEVWGTS
                  drug_spending_bybene (in = a);
    by BASEID;
    if a then output;
run;
```

**SAS (PROC SQL Join Method)**

```
PROC SQL;
CREATE TABLE mcbs_analyticfile AS
  SELECT A.drug_spending,
         B.*
  FROM   drug_spending_bybene AS a
        LEFT JOIN costYY.csevwgts AS b
          ON a.baseid = b.baseid
  ORDER BY baseid;
QUIT;
```

The sample code below demonstrates how to use the constructed analytic dataset to estimate average spending on cardiovascular drugs during the year for Medicare beneficiaries who had at least one prescription for a cardiovascular drug during the year. Balanced repeated replication (BRR or Fay's method) method variance estimation method is used in this example.

\* Estimate Total Spending on Cardiovascular drugs per user for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
  var DRUG_SPENDING;
  weight CSEVRWGT;
  repweights CSEVR001-CSEVR100;
run;
```

***Example A.4 Using Service Summary Data in the Cost Supplement File***

A data user might want to learn about what services beneficiaries use in a given year. This example shows the extent to which the use of outpatient hospital services varies by age. The analysis utilizes the Service Summary (SS) segment, which allows researchers to identify the total number of outpatient hospital events each beneficiary had in a given year. Data users need to subset the SS segment by event type. The sample code below demonstrates this by creating an analytic file that only includes service summary records of outpatient events.

To produce estimates that are representative of the ever enrolled population, data users need to merge the 'op\_summary' analytic file with the Demographics (DEMO) segment, and limit the observations to respondents listed in the weights segment (CSEVWGTS). Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that all observations in the weights segment are preserved, regardless of whether or not the other segments include the entire sample, and that the dataset is restricted to the observations included in the weights file.

```
data op_summary;
  set costYY.SS (keep = BASEID AEVENTS EVNTTYPE);
  if EVNTTYPE = "OP" then output;
run;
```

**SAS (SAS Merge Method)**

```
data merged_costfile;
  merge      costYY.CSEVWGTS (in = a)
            surveyYY.DEMO (keep = BASEID D_STRAT)
            op_summary;
```

```

        by BASEID;
        if a then output;
run;

```

### SAS (PROC SQL Join Method)

```

PROC SQL;
CREATE TABLE merged_costfile AS
  SELECT A.*,
         B.d_strat,
         C.evnttype, C.aevents
  FROM   costYY.csevwgts AS a
        LEFT JOIN surveyYY.demo AS b
          ON a.baseid = b.baseid
        LEFT JOIN op_summary AS c
          ON a.baseid = c.baseid
  ORDER BY baseid;
QUIT;

```

The merged file now includes the relevant Survey File and Cost Supplement File variables that will allow for an examination of the number of outpatient episodes for Medicare beneficiaries by age.

The sample code below transforms the age variable (D\_STRAT) into a four-category variable. Exhibit A.2.1 presents information regarding recoding this age variable.

```

data mcbs_analyticfile;
  set merged_costfile;

  /*AGE*/
  if D_STRAT in (1,2) then AGECAT = 0;
  else if D_STRAT in (3,4) then AGECAT = 1;
  else if D_STRAT in (5,6) then AGECAT = 2;
  else if D_STRAT = 7 then AGECAT = 3;

run;

```

The sample code below demonstrates how to use the constructed analytic dataset to estimate the average number of outpatient episodes for Medicare beneficiaries by age. Balanced repeated replication (BRR or Fay's method) method variance estimation method is used in this example.

\* Estimate Number of Outpatient Episodes for Medicare Beneficiaries by Age (using balanced repeated replication (Fay's method));

```

proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
  var AEVENTS;
  weight CSEVRWGT;
  repweights CSEVR001-CSEVR100;
  domain AGECAT;

run;

```

### *Example A.5 Combining Topical Survey File and Cost Supplement File Data*

The next example uses a hypothetical analysis plan to examine out-of-pocket costs for Medicare beneficiaries with diabetes whose usual source of care is a hospital or emergency room. This analysis requires merging

Topical data on the Survey File to the Cost Supplement File. As in Example A.2 above, the analysis involves annual out-of-pocket costs (the adjusted total out-of-pocket spending variable [PAMTOOP] in the Person Summary [PS] segment) for 11 types of events [dental services, hearing services, home health (including home health provider and home health friend events), hospice, inpatient, institutional, medical provider, outpatient, prescribed medication, and vision services and facility).

To generate estimates using the data from one of the Topical segments (including Usual Source of Care [USCARE]), on their own or merged with another Survey File segment that does not need special non-response adjustment weights, the researcher must always use the special non-response adjustment general and replicate weights designed for the Topical segment **INSTEAD** of using the general and replicate weights that appear in the separate non-adjusted weight segments (CENWGTS, EVRWGTS, CSEVWGTS). In this case, since we are merging the PS segment from the Cost Supplement File, we should use the USCARE Cost Supplement File ever enrolled weights (USCEWT).

### SAS (SAS Merge Method)

```
data merged_surveycostfile;
    merge
        surveyYY.DEMO (keep = BASEID INT_TYPE D_STRAT)
        surveyYY.CHRNCOND (keep = BASEID D_OCDTYP)
        surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-
        USCE100)
        costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;
```

### SAS (PROC SQL Join Method)

```
PROC SQL;
CREATE TABLE merged_surveycostfile AS
SELECT A.*,
       B.int_type, B.d_strat,
       C.d_ocdtyp,
       D.pamtoop
FROM   surveyYY.uscare(keep = baseid placepar placeknd uscewt usce1-usce100) AS a
LEFT JOIN surveyYY.demo AS b
      ON a.baseid = b.baseid
LEFT JOIN surveyYY.chrncond AS c
      ON a.baseid = c.baseid
LEFT JOIN costYY.ps AS d
      ON a.baseid = d.baseid
ORDER BY baseid;
QUIT;
```

This merged file now includes the relevant Survey File and Cost Supplement File variables that will allow for an examination of out-of-pocket costs for Medicare beneficiaries living in the community with diabetes.

There are numerous ways to convert these raw variables into analytic variables. The approach below creates variables for age (AGECAT), residence status (COMMONLY), the presence of Type 1 or Type 2 diabetes (DIABETES), and an indicator for the beneficiary's usual source of care (US\_SOC). Exhibit A.5.1 presents information regarding the recoding and creation of these analytic variables.

```
data mcbs_analyticfile;
```

```

set merged_surveycostfile;
keep baseid commonly diabetes us_soc uscewt usce: pamtoop;
/* AGE */
If D_STRAT in (1,2) then AGECAT = 0;
    else if D_STRAT in (3,4) then AGECAT = 1;
    else if D_STRAT in (5,6) then AGECAT = 2;
    else if D_STRAT = 7 then AGECAT = 3;
/* RESIDENCE STATUS */
if INT_TYPE='C' then commonly=1;
    else commonly=0;
/* DIABETES */
if D_OCDTYP in (1,2) then diabetes=1;
/* indicator variable for Type 1 or Type 2 diabetes */
    else diabetes=0;
/* USUAL SOURCE OF CARE */
US_SOC = 999; /* MISSING */
if PLACEPAR = 2 then US_SOC = 0; /* NONE */
    else if PLACEPAR = 1 then do;
if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
end;
run;

```

**Exhibit A.5.1:** Example A.5 Recoding Variables

Measure	Original Variable	Recoded Variable
Age	D_STRAT 1: 0-44 2: 45-64 3: 65-69 4: 70-74 5: 75-79 6: 80-84 7: 85 +	AGECAT 0: 0-64 1: 65-74 2: 75-84 3: 85+
Residence Status	INT_TYPE B: Both C: Community F: Facility	COMMONLY 1: Community 0: Not community
Diabetes	D_OCDTYP 1: Type 1 2: Type 2 3: Pre-diabetes or borderline 4: Gestational (pregnancy-related) 91: Other type of diabetes	DIABETES 1: Type 1 or Type 2 0: Not Type 1 or Type 2



Measure	Original Variable	Recoded Variable
Usual Source of Care	PLACEKND 1: Doctor's office or group practice 2: Medical clinic 3: Managed care plan center/HMO 4: Neighborhood or family health center 6: Rural Health Clinic 7: Company clinic 8: Other clinic 9: Walk-in urgent care center 10: At home 11: Hospital emergency room 12: Hospital outpatient department 13: Veterans' Administration facility 14: Mental health center 91: Other, specify <i>Note: Applies only if PLACEPAR = 1</i>	US_SOC 1: Doctor's office 2: Medical Clinic 3: Hospital/OPD/ER 4: Other

### Example A.6 Variance Estimation

The sample code below provides examples of how to produce correct estimates for continuous and categorical variables and for the entire sample as well as for a subgroup. The examples below use the Survey File and Cost Supplement File, but this guidance also applies to the Survey File - Early Release. The examples use the file created in Example A.5 above (i.e., `mcbs_analyticfile`) to estimate out-of-pocket spending for all Medicare beneficiaries (Example A.6.1), out-of-pocket spending for Medicare beneficiaries living in the community whose usual source of care is a hospital or emergency room (Example A.6.2), usual source of care for all Medicare beneficiaries (Example A.6.3), and usual source of care for Medicare beneficiaries living in the community with diabetes (Example A.6.4).

Although the MCBS includes variables to obtain weighted estimates and estimated standard errors using Taylor-series linearization approach, the balanced repeated replication (BRR or Fay's method) method provides more analytic flexibility when analyzing subgroups.<sup>19</sup> Therefore, the examples presented in this section pertaining to subgroup analysis (Examples A.6.2 and A.6.4) use BRR (Fay's method) variance estimation method.

CMS generally recommends the BRR method of variance estimation to MCBS users because it requires neither the specification of strata and cluster definitions nor the specification of domain or subgroup definitions in subpopulation analyses, which are required for Taylor-series estimation and are common inadvertent omissions. However, the Taylor series method of variance estimation is also appropriate for experienced users who prefer this method or in instances where the BRR method is not possible in the available software. For these reasons, the MCBS data files include the variables `SUDSTRAT` and `SUDUNIT`, which are needed for Taylor-series estimation. The SAS functions `%surveyglm` and `%surveygenmod` appropriately allow for strata and cluster definitions. When using these functions (and in any other instances where Taylor series estimation is used), specify `SUDSTRAT` as the strata definitions and `SUDUNIT` as the cluster definitions.

<sup>19</sup> For example, performing analysis on small subgroups could lead to instances where one or more strata contain a single observation. In such situations, calculating intra-strata variances using Taylor-series Linearization approach is not possible. Therefore, standard errors cannot be computed.

The examples below provide sample code for SAS, Stata®, and R® to generate estimates, both for the total population and for subgroups.

### Example A.6.1 Total Out-of-Pocket Spending for Medicare Beneficiaries

#### SAS (SAS Merge Method)

```
data mcbs_analyticfile;
  merge      costYY.CSEVWGTS (in = a)
            costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;
```

#### SAS (PROC SQL Join Method)

```
PROC SQL;
CREATE TABLE mcbs_analyticfile AS
  SELECT A.*,
         B.pamtoop
  FROM   costYY.csevwgts AS a
        LEFT JOIN costYY.ps AS b
          ON a.baseid = b.baseid
  ORDER BY baseid;
QUIT;
```

#### SAS

\* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
  var PAMTOOP;
  weight CSEVRWGT;
  repweights CSEVR001-CSEVR100;
run;
```

\* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using Taylor-series Linearization approach);

```
proc surveymeans data= mcbs_analyticfile varmethod = TAYLOR;
  var PAMTOOP;
  weight CSEVRWGT;
  cluster SUDUNIT;
  strata SUDSTRAT;
run;
```

#### Stata<sup>20</sup>

```
* declare survey dataset (using balanced repeated replication (fay's method));
svyset [pweight= CSEVRWGT], brrweight(CSEVR001-CSEVR100) fay(.3) vce(brr) mse
```

<sup>20</sup> Note that the mcbs\_analyticfile for Stata and R code must include the appropriate weights and the stratum variables.

```
* estimate total out-of-pocket spending for Medicare beneficiaries;
svy brr, fay(.3): mean PAMTOOP

* declare survey dataset (using taylor-series linearization approach);
svyset [pweight= CSEVRWGT], strata(SUDSTRAT) psu(SUDUNIT)

* estimate total out-of-pocket spending for Medicare beneficiaries;
svy: mean PAMTOOP
```

R<sup>21</sup>

```
# specify survey design object (using balanced repeated replication (fay's method))
mcbs <- svrepdesign(
  weights = ~CSEVRWGT,
  repweights = "CSEVR[001-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analyticfile,
  combined.weights = TRUE
)

# specify survey design object (using taylor-series linearization approach)
mcbs <- svydesign(
  weights = ~CSEVRWGT,
  id = ~SUDUNIT,
  strata = ~SUDSTRAT,
  nest= TRUE,
  data = mcbs_analyticfile
)

# estimate total out-of-pocket spending for Medicare beneficiaries;
svymean(~PAMTOOP, design=mcbs)
```

### **Example A.6.2 Total Out-of-Pocket Spending for Medicare Beneficiaries Living in the Community Whose Usual Source of Care is the Hospital or Emergency Room**

SAS (SAS Merge Method)

```
data merged_surveycostfile;
  merge
    surveyYY.DEMO (keep = BASEID INT_TYPE)
    surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-
    USCE100)
    costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;
```

SAS (PROC SQL Join Method)

```
PROC SQL;
```

<sup>21</sup> The survey package in R is required to run this R code.

```

CREATE TABLE merged_surveycostfile AS
SELECT A.*,
       B.int_type,
       C.pamtoop
FROM   surveyYY.uscare(keep = baseid placepar placeknd uscewt usce1-usce100) AS a
LEFT JOIN surveyYY.demo AS b
      ON a.baseid = b.baseid
LEFT JOIN costYY.ps AS c
      ON a.baseid = c.baseid
ORDER BY baseid;
QUIT;

```

## SAS

```

data mcbs_analyticfile;
  set merged_surveycostfile;
  keep baseid commonly us_soc uscewt usce: pamtoop;
  /* RESIDENCE STATUS */
  if INT_TYPE='C' then commonly=1;
  else commonly=0;
  /* USUAL SOURCE OF CARE */
  US_SOC = 999; /* MISSING */
  if PLACEPAR = 2 then US_SOC = 0; /* NONE */
  else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
  end;
run;

```

\* Total Out-of-Pocket Spending for Medicare Beneficiaries Living in the Community Whose Usual Source of Care is Hospital or Emergency Room (using balanced repeated replication (Fay's method));

```

proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
  var PAMTOOP;
  weight USCEWT;
  repweights USCE1-USCE100;
  domain COMMONLY * US_SOC;
run;

```

## Stata

```

* declare survey dataset (using balanced repeated replication (fay's method))
svyset [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr) mse

```

```

* total out-of-pocket spending for Medicare beneficiaries living in the community whose usual source of
care is hospital or emergency room
svy brr, fay(.3) subpop(if commonly==1 & US_SOC==3) : mean PAMTOOP

```

R

```
# remove NAs
mcbs_analyticfile <- subset(mcbs, !is.na(USCE1))

# specify survey design object (using balanced repeated replication (fay's method))
mcbs_ussoc <- svrepdesign(
  weights = ~USCEWT,
  repweights = "USCE[1-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analyticfile,
  combined.weights = TRUE
)

#subset survey design object to Medicare beneficiaries living in the community whose usual source of
care is hospital or emergency room
mcbs_ussoc <- subset(mcbs_ussoc, commonly==1 & US_SOC==3)

# total out-of-pocket spending for Medicare beneficiaries living in the community whose usual source
of care is hospital or emergency room
svymean(~PAMTOOP, design=mcbs_ussoc)
```

### Example A.6.3 Number of Medicare Beneficiaries by Usual Source of Care

#### SAS (SAS Merge Method)

```
data merged_surveycostfile;
  merge
    surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-
    USCE100)
    surveyYY.EVRWGTS (keep = BASEID SUDSTRAT SUDUNIT)
    costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;
```

#### SAS (PROC SQL Join Method)

```
PROC SQL;
CREATE TABLE merged_surveycostfile AS
SELECT A.*,
       B.sudstrat, B.sudunit,
       C.pamtoop
FROM   surveyYY.uscare(keep = baseid placepar placeknd uscewt usce1-usce100) AS a
LEFT JOIN surveyYY.evrwgts AS b
      ON a.baseid = b.baseid
LEFT JOIN costYY.ps AS c
      ON a.baseid = c.baseid
ORDER BY baseid;
QUIT;
```

## SAS

```

data mcbs_analyticfile;
  set merged_surveycostfile;
  keep baseid us_soc pamtoop uscewt usce: sudstrat sudunit;
  /* USUAL SOURCE OF CARE */
  US_SOC = 999; /* MISSING */
  if PLACEPAR = 2 then US_SOC = 0; /* NONE */
  else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
  end;
run;

* Number of Medicare Beneficiaries by Usual Source of Care (using balanced repeated replication (Fay's method));

proc surveyfreq data= mcbs_analyticfile varmethod = brr (fay=.30);
  table US_SOC;
  weight USCEWT;
  repweights USCE1-USCE100;
run;

* Number of Medicare Beneficiaries by Usual Source of Care (using Taylor-series Linearization approach);

proc surveyfreq data= mcbs_analyticfile varmethod = TAYLOR;
  table US_SOC;
  weight USCEWT;
  cluster SUDUNIT;
  strata SUDSTRAT;
run;

```

## Stata

```

* declare survey dataset (using balanced repeated replication (fay's method))
svyset [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr) mse

* number of Medicare beneficiaries by usual source of care
svy brr, fay(.3) : tabulate US_SOC, count se

* declare survey dataset (using taylor-series linearization approach)
svyset [pweight= USCEWT], strata(SUDSTRAT) psu(SUDUNIT)

* number of Medicare beneficiaries by usual source of care
svy : tabulate US_SOC, count se

```

## R

```

# specify survey design object
mcbs <- svrepdesign(

```

```

        weights = ~USCEWT,
        repweights = "USCE[1-100]+",
        type = "Fay",
        rho = 0.3,
        data = mcbs_analyticfile,
        combined.weights = TRUE
    )

# specify survey design object (using taylor-series linearization approach)
mcbs <- svydesign(
    weights = ~USCEWT,
    id = ~SUDUNIT,
    strata = ~SUDSTRAT,
    nest = TRUE,
    data = mcbs_analyticfile
)

# number of Medicare beneficiaries by usual source of care
svytable(~US_SOC, design=mcbs)

```

#### **Example A.6.4 Number of Medicare Beneficiaries Living in the Community with Diabetes by Usual Source of Care**

##### **SAS (SAS Merge Method)**

```

data merged_surveycostfile;
    merge
        surveyYY.DEMO (keep = BASEID INT_TYPE)
        surveyYY.CHRNCOND (keep = BASEID D_OCDTYP)
        surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-USCE100)
        costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;

```

##### **SAS (PROC SQL Join Method)**

```

PROC SQL;
CREATE TABLE merged_surveycostfile AS
SELECT A.*,
       B.int_type,
       C.d_ocdtyp,
       D.pamtoop
FROM   surveyYY.uscare(keep = baseid placepar placeknd uscewt usce1-usce100) AS a
LEFT JOIN surveyYY.demo AS b
      ON a.baseid = b.baseid
LEFT JOIN surveyYY.chrncond AS c
      ON a.baseid = c.baseid
LEFT JOIN costYY.ps AS d
      ON a.baseid = d.baseid
ORDER BY baseid;
QUIT;

```



## SAS

```

data mcbs_analyticfile;
  set merged_surveycostfile;
  keep baseid commonly diabetes us_soc uscewt usce: pamtoop;
  /* RESIDENCE STATUS */
  if INT_TYPE='C' then commonly=1;
  else commonly=0;
  /* DIABETES */
  if D_OCDTYP in (1,2) then diabetes=1;
  /* indicator variable for Type 1 or Type 2 diabetes */
  else diabetes=0;
  /* USUAL SOURCE OF CARE */
  US_SOC = 999; /* MISSING */
  if PLACEPAR = 2 then US_SOC = 0; /* NONE */
  else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
  /* OTHER */
  end;
run;

```

\* Number of Medicare Beneficiaries Living in the Community with Diabetes by Usual Source of Care (using balanced repeated replication (Fay's method));

```

proc surveyfreq data= mcbs_analyticfile varmethod = brr (fay=.30);
  table DIABETES * COMMONLY * US_SOC/ row;
  weight USCEWT;
  repweights USCE1-USCE100;
run;

```

## Stata

\* declare survey dataset (using balanced repeated replication (fay's method))  
 svyset \_n [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr) mse

\* number of Medicare beneficiaries living in the community with diabetes by usual source of care  
 svy brr, fay(.3) subpop(if diabetes==1 & commonly==1) : tab US\_SOC, count se

## R

```

# specify survey design object (using balanced repeated replication (fay's method))
mcbs <- svrepdesign(
  weights = ~USCEWT,
  repweights = "USCE[1-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analyticfile,
  combined.weights = TRUE
)

```

```
#subset survey design object to Medicare beneficiaries living in the community with diabetes by usual
source of care
```

```
mcbs_subgrp <- subset(mcbs, diabetes==1 & commonly==1)
```

```
# number of Medicare beneficiaries living in the community with diabetes by usual source of care
svytable(~US_SOC, design=mcbs_subgrp)
```

### *Example A.7 Repeated Cross-Sectional or Pooled Analysis*

The sample code below demonstrates the steps involved in constructing a repeated cross-sectional or pooled analytic dataset and performing analysis. The example below estimates the percentage of Medicare beneficiaries who are dually eligible (i.e., enrolled in both Medicare and Medicaid) during the prior data year and the current data year.

The examples presented in this section involve multiple years of MCBS data and use replicate weights – a form of the BRR technique.

*/\* Join prior data year administrative records (HISUMRY) file with cross-sectional weights (CENWGTS) file \*/*

#### **SAS (SAS Merge Method)**

```
data mcbsY1;
merge surveyY1.CENWGTS (in = a drop = VERSION)
      surveyY1.HISUMRY (keep = BASEID H_OPMDCD);
  by BASEID;
  if a;
run;

data mcbsY2;
merge surveyY2.CENWGTS (in = a drop = VERSION)
      surveyY2.HISUMRY (keep = BASEID H_OPMDCD);
  by BASEID;
  if a;
run;
```

#### **SAS (PROC SQL Join Method)**

```
PROC SQL;
CREATE TABLE mcbsY1 AS
  SELECT A.*,
         B.h_opmdcd
  FROM   surveyY1.cenwgts(DROP=version) AS A
        LEFT JOIN surveyY1.hisumry AS B
          ON A.baseid = B.baseid;
QUIT;
```

*/\* Create Analytic Dataset for Repeated Cross-Sectional or Pooled Analysis \*/*

*/\* Join current data year administrative records (HISUMRY) file with cross-sectional weights (CENWGTS) file \*/*

```
PROC SQL;
CREATE TABLE mcbsY2 AS
  SELECT A.*,
         B.h_opmdcd
```

```
FROM surveyY2.cenwgts(DROP=version) AS A
LEFT JOIN surveyY2.hisumry AS B
      ON A.baseid = B.baseid;
QUIT;
```

```
/* Concatenate prior and current cross-sectional files */
data mcbs_analytic_file;
      set mcbsY1 mcbsY2;
run;
```

## SAS

\* Estimate Percent of Dually Eligible Medicare Beneficiaries (Pooled estimate representing the moving average of nationally representative year-specific estimates) using balanced repeated replication (Fay's method));

```
proc surveyfreq data = mcbs_analytic_file varmethod = brr (fay=.30);
      table H_OPMDCD;
      weight CEYRSWGT;
      repweights CEYRS001-CEYRS100;
run;
```

\* Estimate Percent of Dually Eligible Medicare Beneficiaries by Year (nationally representative, year-specific estimates) using balanced repeated replication (Fay's method);

```
proc surveyfreq data = mcbs_analytic_file varmethod = brr (fay=.30);
      table SURVEYR * H_OPMDCD/ row;
      weight CEYRSWGT;
      repweights CEYRS001-CEYRS100;
run;
```

## Stata

```
* Declare survey dataset
svyset _n [pweight = CEYRSWGT], brrweight(CEYRS001-CEYRS100) fay(.3) vce(brr) mse

* Estimate Percent of Dually Eligible Medicare Beneficiaries (Pooled estimate representing the
* moving average of nationally representative year-specific estimates)
svy brr, fay(.3): tab H_OPMDCD

* Estimate Percent of Dually Eligible Medicare Beneficiaries (nationally representative, year-specific
estimates)
svy brr, fay(.3): tab H_OPMDCD SURVEYR, column
```

## R

Note: Data users will need to install the 'survey' package to use the svrepdesign function below.

```
# Specify survey design object
mcbs <- svrepdesign(
      weights = ~CEYRSWGT,
      repweights = "CEYRS[001-100]+",
      type = "Fay",
      rho = 0.3,
      data = mcbs_analytic_file,
      combined.weights = TRUE
```

```
)

# Estimate Percent of Dually Eligible Medicare Beneficiaries by Year (Pooled estimate representing the
moving average of nationally representative year-specific estimates)
prop.table(svytable(~H_OPMDCD, design=mcbs))

# Estimate Percent of Dually Eligible Medicare Beneficiaries by Year (nationally representative, year-
specific estimates)
prop.table(svytable(~H_OPMDCD + SURVEYR, design=mcbs), 2)
```

### *Example A.8 Longitudinal Analysis*

The sample code below demonstrates the steps involved in constructing an analytic dataset and performing a longitudinal analysis. The example estimates the percent change in out-of-pocket costs between CY1 and CY2 for Medicare beneficiaries enrolled in the Medicare program during both years.

The examples presented in this section involve multiple years of MCBS data and use replicate weights, a form of the BRR technique.

#### **SAS (SAS Merge Method)**

```
/* Create Analytic Dataset for Longitudinal Analysis */
/* Merge Y1 Person Summary (PS) file with Y2 longitudinal weights (CSL2WGTS) file */
data mcbs_analytic_file;
    merge costY2.CSL2WGTS (in = a)
          costY1.PS (keep = BASEID PAMTOOP);
    by BASEID;
    rename PAMTOOP = PAMTOOPY1;
    if a;
run;

/* Merge Y2 Person Summary file with analytic file created above */
data mcbs_analytic_file;
    merge mcbs_analytic_file (in = a)
          costY2.PS (keep = BASEID PAMTOOP rename=(PAMTOOP=PAMTOOPY2));
    by BASEID;

    PAMTOOP_PDIF = PAMTOOPY2 - PAMTOOPY1;

    if a;
run;
```

#### **SAS (PROC SQL Join Method)**

```
/* Create Analytic Dataset for Longitudinal Analysis */
/* Join Y1 Person Summary (PS) file with Y2 longitudinal weights (CSL2WGTS) file */
PROC SQL;
CREATE TABLE mcbs_analytic_file_t1 AS
    SELECT A.*,
           B.pamtoop as pamtoopY1
    FROM   costY2.csl2wgts AS a
           LEFT JOIN costY1.ps AS b
              ON a.baseid = b.baseid
    ORDER BY baseid;

/* Join Y2 Person Summary file with analytic file created above */
```

```

CREATE TABLE mcbs_analytic_file_t2 AS
  SELECT A.*,
         B.pamtoop as pamtoopY2
  FROM   mcbs_analytic_file_t1 AS a
        LEFT JOIN costY2.PS AS b
          ON a.baseid = b.baseid
  ORDER BY baseid;

CREATE TABLE mcbs_analytic_file AS
  SELECT *, pamtoopY2 - pamtoopY1 as pamtoop_pdiff
  FROM   mcbs_analytic_file_t2
  ORDER BY baseid;
QUIT;

```

**SAS**

\* Estimate percent change in out-of-pocket costs between CY1 and CY2 for beneficiaries enrolled in Medicare during both years (using balanced repeated replication (Fay's method));

```

proc surveymeans data=mcbs_analytic_file varmethod = brr (fay = .30);
  var PAMTOOP_PDIF;
  weight CSL2YWGT;
  repweights CSL2Y001-CSL2Y100;
run;

```

**Stata**

\* Declare survey dataset (using balanced repeated replication (Fay's method))  
 svyset \_n [pweight= CSL2YWGT], brrweight(CSL2Y001-CSL2Y100) fay(.3) vce(brr) ms

\* Estimate percent change in out-of-pocket costs between CY1 and CY2 for beneficiaries enrolled in Medicare during both years  
 svy brr, fay(.3) : mean pamtoop\_pdiff

**R**

```

# Specify survey design object (using balanced repeated replication (Fay's method))
mcbs <- svrepdesign(
  weights = ~ CSL2YWGT,
  repweights = "CSL2Y[001-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analytic_file,
  combined.weights = TRUE
)

svymean(~pamtoop_diff, design=mcbs)

```

## Appendix B: Table of Links to MCBS Documentation

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

## Appendix C: Frequently Asked Questions

The Frequently Asked Questions (FAQs) provide answers to common inquiries and points of interest for users of the MCBS data sets. A unique code follows each FAQ (noted in parentheses after each question). The purpose of the code is to provide a reference if one of the FAQs is updated, deleted, or otherwise modified.

**Please note, the availability and contents of the MCBS LDS segments and variables referenced in these FAQs may change year-to-year. Users should refer to the year-specific data documentation (including the annual *Data Year Release Notes* and historical *Data User's Guides*) and *MCBS LDS Variable Crosswalks* for more information on the data available each year.**

### C.1 Data Requests

- **How do I request the MCBS Limited Data Set (LDS) files and how long does it take to receive the data?** (FAQDR1) Requests for the MCBS LDS files must be made through the CMS Data Use Agreement (DUA) tracking system known as the Enterprise Privacy Policy Engine (EPPE). EPPE can be used to initiate a new LDS DUA request or to amend/update an existing LDS DUA. Instructions for accessing and using EPPE to make a request can be found here: <https://www.cms.gov/data-research/files-for-order/data-disclosures-and-data-use-agreements-duas/limited-data-set-lds>.

The processing of DUAs takes approximately 6-8 weeks. If approved, then data processing time varies depending on the number of data years and files requested. Normal data processing time is one week.

- **Does the MCBS produce microdata public use files that are available without a data use agreement (DUA)?** (FAQDR2) Yes, in addition to the Limited Data Sets (LDS), MCBS data are made available to users through three types of Public Use Files (PUF):
  - ▶ The Survey File Microdata PUF, released annually.
  - ▶ The Cost Supplement File Microdata PUF, released annually.
  - ▶ Special Topic Microdata PUFs, such as the COVID-19 PUFs.

Like the LDS files, the Microdata PUFs serve as unique sources of information on beneficiaries' health and well-being that cannot be obtained through CMS administrative sources alone. The MCBS PUFs are not intended to replace the more detailed LDS files; rather, they are publicly available alternatives that do not require a DUA. The MCBS PUFs are available to the public as a free download and can be found through CMS' PUF website at <https://data.cms.gov/medicare-current-beneficiary-survey-mcbs>.

- **What resources are available to data users and how did resources change starting with the 2023 data year?** (FAQDR3) CMS releases a comprehensive suite of documentation products to support MCBS data users. These products were consolidated beginning with the 2023 data year to separate the detailed, background information on the MCBS survey design, methodology, and processes from the focused year-specific content that is most relevant to data users. An overview of the MCBS documentation products beginning with the 2015 data year can be found in the MCBS Documentation Crosswalk and Overview section. All documentation are available for download on the CMS MCBS website: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>. Links to these and other resources are included in Appendix B.

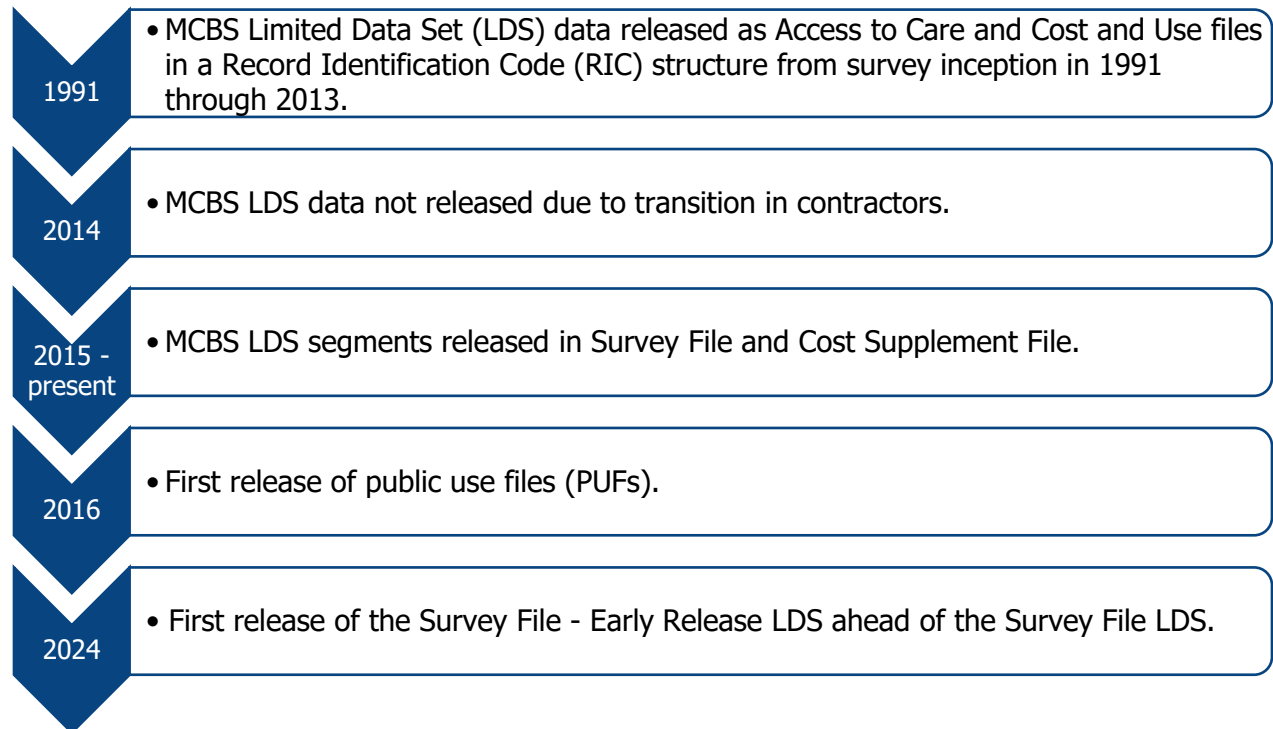


## C.2 File Corrections

- **What happens if there are corrections to data files?** (FAQFC1) Occasionally, there are corrections to the original files that a data user has received. Recipients will be notified via email, and, where possible, a zipped/encrypted correction will be sent via email.
  - ▶ **If you did not receive the updated files.** If you received a password notification but did not receive updated MCBS files, it is likely that your organization has removed the attachments (.zip files). Check your e-mail security settings to determine how attachments are handled by your organization. The updates are also available as executable (.exe) files. If executable files are supported by your organization, these can be sent to you by request.
  - ▶ **If the “password” doesn’t work.** MCBS data products are encrypted using PKWARE and this can sometimes cause problems when opened with other encryption packages. If the error you are receiving is asking for a password, this is the problem. PKWARE uses the term “passphrase” instead of password. PKWARE is freeware. For more information, please see: SecureZIP® by PKWARE, Inc.  
<https://support.pkware.com/>
  - ▶ **If you received multiple emails from the MCBS.** Updates are sent by data use agreement (DUA) #. If you receive password notifications or receive data more than once, it is likely because you are listed as having multiple DUAs that include MCBS data.

## C.3 Content and Methodology

- **Are there changes in the structure of the MCBS over time that may impact analyses?** (FAQCM1) MCBS Limited Data Set (LDS) data have been released for every data year since 1991, except in 2014 due to a contract transition. As described below in Exhibit C.3.1, starting in 2015, the file structure of the MCBS LDS data releases changed. Additionally, new data segments have been added to the MCBS LDS releases over time. To compare variables of interest across two or more LDS data years, please refer to the *MCBS Survey File* and *Cost Supplement File Variable Crosswalks* at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>. Additional structural changes are outlined below:
  - ▶ Beginning with the 2021 Survey File LDS, CMS provides five years of enrollment data and five years of claims data rather than one year.
  - ▶ Starting with the 2023 data year, CMS releases a third LDS, the Survey File - Early Release LDS, providing select data on beneficiaries living in the community from the fall round on an expedited schedule ahead of the Survey File LDS.
  - ▶ Beyond the LDS releases, the annual MCBS Survey File Microdata Public Use File (PUF) was first made available in 2016 with the release of 2013 MCBS data, and CMS has expanded the availability of MCBS data since then with the release of Cost Supplement File and COVID-19 Special Topic Microdata PUFs.

**Exhibit C.3.1:** Timeline of MCBS Structure Changes

- What are the differences and when should I use each of the three MCBS Limited Data Sets (LDS's)?** (FAQCM2) The Survey File - Early Release, Survey File, and Cost Supplement File differ in terms of their contents, data collection timeframe, inclusion of enrollment and claims data, population representation, release timing, and available weights. The Survey File – Early Release is designed to provide expedited access to data, allowing users to conduct timely, cross-sectional analyses. However, this file contains a subset of information on beneficiaries living in the community from the Fall round, so analyses on certain topics cannot be performed. If users need data beyond what is available in the Survey File – Early Release, they should use the Survey File, which contains a more comprehensive dataset. The Survey File can also be merged with the Cost Supplement File and offers additional weights for the continuously enrolled and longitudinal populations. The Cost Supplement File contains survey responses linked to Medicare FFS and Part D claims data, making it the best choice for analyses focused on health care utilization and expenditures. For a detailed comparison of these datasets, refer to Exhibit C.3.2.

**Exhibit C.3.2:** Overview of MCBS Limited Data Sets Beginning with the 2023 Data Year

	Survey File - Early Release	Survey File	Cost Supplement File
File Contents	<ul style="list-style-type: none"> <li>Timely fall data on key topics such as beneficiaries' socio-demographic information; self-reported health status, conditions, and functioning; disability; and access to and satisfaction with care</li> </ul>	<ul style="list-style-type: none"> <li>Fall data released on the Survey File - Early Release as well as annualized data on topics such as health insurance coverage</li> <li>Data on additional topics (e.g., income, debt, knowledge of the Medicare program, etc.) collected in the winter and summer rounds of the next calendar year</li> <li>Facility information, administrative records and assessment data, and Fee-for-Service (FFS) claims data</li> </ul>	<ul style="list-style-type: none"> <li>Comprehensive accounting of beneficiaries' health care use, expenditures, and sources of payment</li> <li>Linked survey-reported and Medicare FFS and Part D claims data</li> </ul>
Data Collection Timeframe	Fall only	Annualized	Annualized
Population Represented	Community only	Community and facility	Community and facility
Weights Available	Cross-sectional ever enrolled and select Topical weights	Cross-sectional ever enrolled, cross-sectional continuously enrolled, longitudinal, and Topical weights	Cross-sectional ever enrolled and longitudinal weights
Inclusion of Enrollment and Claims Data	Limited enrollment information; no FFS claims	Detailed enrollment information; five years of enrollment data; five years of FFS claims	No enrollment information; linked FFS and Medicare Part D claims
Supports Standalone Analysis?	Yes	Yes	No
Intended for Analysis with Other MCBS LDS Files?	No	Yes (with Cost Supplement File)	Yes (with Survey File)
Approximate Release Timeframe	Within nine months after the close of the calendar year	18 months after the close of the calendar year	Three months after the Survey File

- **What is the mode of data collection on the MCBS?** (FAQCM3) Interviews are usually conducted in-person and over the phone using computer-assisted personal interviewing (CAPI).
- **What type of Medicare eligibility/enrollment data are included?** (FAQCM4) The Survey File Limited Data Set (LDS) contains information on Medicare eligibility and enrollment data. Specifically, the Health Insurance Timeline (HITLINE) segment provides monthly coverage indicators, coverage start and end dates, the type of plan, and the source of coverage information for the plan. The Health Insurance Summary (HISUMRY) segment also contains monthly eligibility codes and detailed Medicare-Medicaid dual eligibility indicators. Starting with 2021, the Survey File release includes the Multiple Year Enrollment (MYENROLL) segment, which contains up to five years of enrollment data for each Medicare beneficiary in the Survey File.
- **Are Medicare Advantage (MA)/Medicaid/Part D claims data included?** (FAQCM5) The Survey File release contains the Fee-for-Service (FFS) claims data, which provide CMS administrative information on medical services and payments paid by Medicare FFS claims. Starting in 2021, the FFS Research Claims files include five years of claims data instead of one year. The Cost Supplement File links survey-reported health care utilization and cost data to FFS claims data and Part D Prescription Drug claims data. Claims data for MA beneficiaries are not available, but starting with the 2019 data year, data for beneficiaries with MA coverage have been adjusted to account for unreported utilization that would appear as MA Encounters based on an analysis of previous years' Encounter data files. Prior to 2021, Medicaid claims data were not available and estimated Medicaid amounts, derived from CMS algorithms, were used during imputation for Medicaid payments. Starting with the 2021 data year, Medicaid payment amounts from the Transformed Medicaid Statistical Information System (T-MSIS) administrative claims data were integrated into the imputation. The Medicaid payments continue to be available in event-level segments of the Cost Supplement File release.
- **What cost and utilization information are available for beneficiaries enrolled in Medicare Advantage (MA)?** (FAQCM6) When a respondent reports health care events, the survey uses the explanation of benefits (EOBs) form from their MA provider to report the payments, as well as the capitation information from the administrative data for total MA payments. This is the same approach for services that are not covered by Medicare, such as most dental care. Actual claims-based information for MA beneficiaries, referred to as Encounter data, are not currently available for these individual events. However, starting in data year 2019, utilization and cost data for MA beneficiaries have been adjusted to account for unreported utilization that would appear as MA Encounter data files if they were available.
- **How often do respondents receive each questionnaire?** (FAQCM7) Different combinations of MCBS Questionnaire sections are used depending on a number of criteria, including interview type (Baseline vs. Continuing); the season of the round 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. For more information about the specific questionnaires administered during each round of data collection, please see the historical *MCBS Data User's Guide: Survey File* for the 2015-2022 data years or the *Data Year Release Notes* for 2023 onward, available on the CMS MCBS website. The *Questionnaires* for each data collection year are also available on the CMS MCBS website at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/questionnaires>.
- **How do I find out what proportion of Medicare beneficiaries received a flu shot in a given calendar year?** (FAQCM8) Flu shot data are available for both Community and Facility components,

but data collection and processing methods are different, and the variables are located on different segments in the Survey File Limited Data Set (LDS). To estimate prevalence of flu shots in a given flu season, data users need the prior data year Survey File for beneficiaries living in the community and the current data year Survey File for beneficiaries living in a facility. Note that the response categories of the FLUSHOT variables are similar across the two components (Yes/No), but the coding values associated with the Yes/No categories in the LDS files are different. For Community, Yes = 1 and No = 2, but for Facility Yes = 1 and No = 0. In addition, the reference periods differ between the Community and Facility components. Therefore, users need both the prior and current data year Survey File LDS's to estimate the flu shot prevalence for all Medicare beneficiaries for a given flu season.

- **Why did CMS redesign the Facility Instrument?** (FAQCM9) The MCBS Facility Instrument is comprised of core questionnaire sections that collect information about beneficiaries living in a long-term care facility at the time of interview. In Fall 2019, the MCBS Facility Instrument was redesigned to skip items redundant with Certification and Survey Provider Enhanced Reports (CASPER) and Long-Term Care Minimum Data Set (MDS) administrative data. These items are regularly reported to CMS and skipping them allows the MCBS to shorten the Facility Instrument for interviews conducted at Medicare- or Medicaid-certified facilities. The redesigned instrument provides a link between Facility interview data and the administrative data by collecting the CMS Certification Number (CCN), a unique number that is assigned to all facilities certified by Medicare and/or Medicaid, via a questionnaire lookup tool. For facilities with a CCN, the Facility Instrument then skips more than 100 questionnaire items that are redundant with CASPER and MDS administrative data in the Facility Questionnaire (FQ) and Health Status (HS) sections. For interviews conducted at facilities not certified by Medicare or Medicaid, the full Facility Instrument is administered.

During data processing, survey-reported data from the HS section are merged with MDS data and survey-reported data from the FQ section are merged with CASPER for beneficiaries living in a facility for which the Facility respondent reported a valid CCN. This processing step results in a blended data product. MDS data are released for data users in the MDS segment in the Survey File Limited Data Set (LDS).

Due to the new skip patterns in the Facility Instrument and inclusion of administrative MDS variables when a CCN is reported, values for some variables may be coded as 'Missing' rather than 'Not indicated.' Data users should take this into account when calculating estimates or comparing distributions with prior years.

For additional details on the redesigned Facility Instrument and ensuing changes to data processing, refer to the *2019 MCBS Methodology Report*, Special Section: Facility Redesign.

- **How did the MCBS modify data collection activities due to the COVID-19 pandemic?** (FAQCM10) Due to the COVID-19 pandemic, on 3/13/20 for Facility and 3/22/20 for Community, data collection was paused for all in-person data collection. Testing data collection by phone using the computer-assisted personal interviewing (CAPI) questionnaires ensued, and on 4/9/20 the MCBS determined it was feasible to continue collecting interviews via phone only. Beginning in November 2021, a gradual return to in-person data collection began alongside phone interviewing. MCBS data collection now includes both in-person and phone interviewing.
- **How were data from the MCBS COVID-19 Community and Facility Supplements released and why are the data available in different Limited Data Set (LDS) releases?** (FAQCM11)

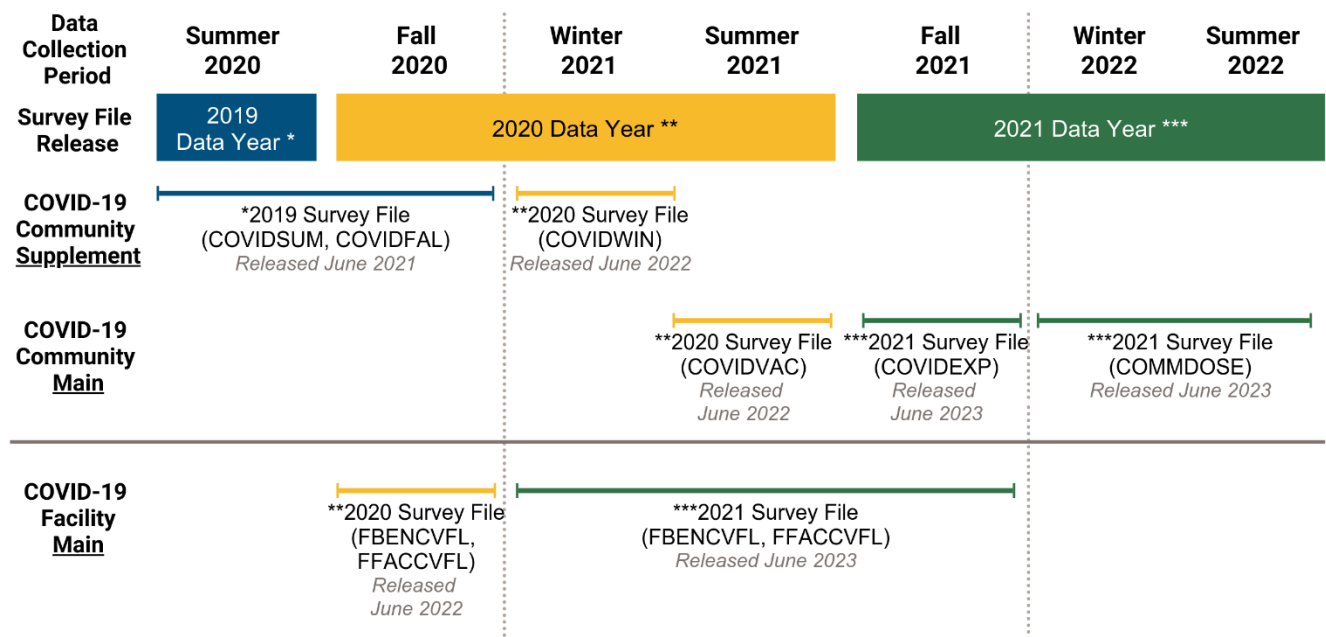
Data collected for MCBS sampled beneficiaries living in the community using the COVID-19 Summer and Fall 2020 Community Supplements were first released as standalone COVID-19 Public Use Files (PUFs) and were also made available as part of the 2019 Survey File LDS. These data were released with the 2019 Survey File LDS because the population administered the COVID-19 Summer and Fall 2020 Community Supplements aligns with the 2019 Survey File population.

Likewise, data collected for MCBS sampled beneficiaries living in the community using the COVID-19 Winter 2021 Community Supplement were made available as a standalone COVID-19 PUF and as part of the 2020 Survey File LDS. These data were released with the 2020 Survey File LDS because the population administered the COVID-19 Winter 2021 Community Supplement aligns with the 2020 Survey File population. Data collected for MCBS sampled beneficiaries living in a facility in Fall 2020 were also released as part of the 2020 Survey File LDS. These data were released with the 2020 Survey File LDS because the population administered the COVID-19 Fall 2020 Facility Supplement aligns with the 2020 Survey File population.

Beginning in Summer 2021, COVID-19 Community Supplement content was incorporated into the main MCBS Community Questionnaire for subsequent interviews as appropriate, including in the COVID-19 Questionnaire (CVQ). In Winter 2021, the COVID-19 Facility Supplement content continued to be incorporated into the main MCBS Facility Instrument. In Fall 2021, COVID-19 items were fielded within the main MCBS Facility Instrument in the COVID-19 Beneficiary Supplement (CV) and COVID-19 Facility-Level Supplement (FC) as Topical sections.

Exhibit C.3.3 shows the schedule of the MCBS COVID-19 Community and Facility Supplements and corresponding LDS and PUF data releases through data year 2021.

**Exhibit C.3.3:** COVID-19 Data Collection and Corresponding Survey File Releases



After the 2021 data year, COVID-19 data continue to be collected through the main questionnaire and included in the annual MCBS LDS releases. For additional details on these data, users can refer to the year-specific documentation and the *MCBS Advanced Tutorial on the COVID-19 Supplement Data*.



## C.4 Sampling

- **What types of beneficiaries are in the continuously enrolled vs. ever enrolled populations?** (FAQS1) The continuously enrolled represent a population of beneficiaries who were enrolled continuously between January 1st of the data collection year and the completion of their fall interview. Beneficiaries who died during the calendar year, current-year enrollees who enrolled in Medicare during the year that they were sampled, and beneficiaries who have lost eligibility are not included in the continuously enrolled group. The ever enrolled represent the population of beneficiaries who were ever enrolled in Medicare for at least one day at any time during the data collection year. The ever enrolled population includes beneficiaries who died or lost entitlement prior to completing the fall interview. Beneficiaries who first became enrolled in Medicare during the data collection year are also included. Thus, the continuously enrolled beneficiaries are a subset of the ever enrolled beneficiaries.  
  
The ever enrolled population from the Survey File is the largest, including anyone enrolled at any time during the calendar year. The Survey File continuously enrolled population is limited to beneficiaries who were enrolled from January 1 of the calendar year through the fall interview date. The Cost Supplement File represents the ever enrolled population derived from a smaller subset of sampled beneficiaries with complete cost and utilization data for the year. The Survey File - Early Release represents a different subset of the ever enrolled population who were alive, still enrolled, and completed a Community interview in the fall.
- **How is the Cost Supplement File ever enrolled population derived?** (FAQS2) The Cost Supplement File ever enrolled population is a subset of the Survey File ever enrolled population with complete cost and utilization data for the year. To be included in the ever enrolled population, 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 the calendar year 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 current Panel who completed the initial fall interview. Beneficiaries who died or lost entitlement prior to January 1 of the calendar year are ineligible and removed. Beneficiaries who survived into the calendar year but do not meet the above criteria are considered to be nonrespondents for the current Cost Supplement File data year and are adjusted for in the resulting weights.
- **Does the survey use a household sample or a list sample?** (FAQS3) The survey uses a list sample. The sample for the MCBS is drawn from a subset of the Medicare enrollment data, which is a list of all Medicare beneficiaries.
- **Do Primary Sampling Units (PSUs) and Secondary Sampling Units (SSUs) align with other federal health surveys, such as the National Health Interview Survey (NHIS)?** (FAQS4) The MCBS selects its own PSUs and SSUs. In late 2000, the current set of PSUs was selected. In 2014, SSUs were reconstructed using Census tracts, and a new sample was drawn. While the MCBS PSUs and SSUs do not align directly with other surveys, they may overlap in some areas with PSUs and/or SSUs used for other federal health surveys.
- **Are populations (given changes to the sample design, e.g., the addition of current-year enrollees to the sample) comparable with past years?** (FAQS5) The Survey File cross-sectional and longitudinal population definitions are consistent from year to year, so the data are comparable between years. The Cost Supplement File cross-sectional population definition is also consistent and

comparable from year to year. The Cost Supplement File two-year longitudinal population changed slightly in 2016 from what was defined the last time the two-year longitudinal weights were supplied (i.e., in 2013). In 2013, the two-year longitudinal (i.e., one-year backward longitudinal weight) Cost Supplement weights represented the population that enrolled on or before 1/1/2011 and was still enrolled in 2013 (i.e., enrollees after 1/1/2011 were not included). Beginning in 2016, the two-year longitudinal weights represent a true two-year ever enrolled population (i.e., the population of beneficiaries who were ever enrolled in both 2015 and 2016).

### C.5 Analysis

- **How do I analyze prescription medicine (PM) data across time?** (FAQA1) Prior to the 2018 Limited Data Set (LDS), Facility PM utilization data were only available in the Prescription Medicine Events (PME) event-level segment but were not included in the summary-level Cost Supplement File LDS segments, the Person Summary (PS) and Service Summary (SS) segments. Researchers interested in analyzing Facility PM utilization with data released prior to the 2018 LDS should use the PME segment instead of PS or SS. The events from the PME segment can be summed to create a dataset that will be comparable to the summary records in the PS or SS segments from data years starting in 2018.
- **Can the MCBS be used to produce subnational estimates? / Are MCBS estimates representative at the state or local level?** (FAQA2) The MCBS is designed to produce nationally representative estimates of the population of all Medicare beneficiaries. Estimates from MCBS data are not representative at the state or local level.
- **Why do I see differences between Medicare published statistics and estimates using MCBS data?** (FAQA3) In general, MCBS estimates may differ from Medicare program statistics using 100 percent administrative enrollment data. There are several reasons for the differences. The most important reason for the difference is that the administrative enrollment data may include people who are no longer alive. This may occur where people have entitlement, such as for Part A only, and receive no Social Security check. When field interviewers try to locate these beneficiaries for interviews, they establish the fact of these deaths. Unrecorded deaths may still be present on the Medicare Administrative enrollment data. The MCBS makes every effort to reconcile the survey information against the administrative data when possible. Other reasons, such as sampling error, may also contribute to differences between MCBS estimates and Medicare program statistics. Lastly, estimates may differ because Medicare program statistics adjust for partial enrollment. Medicare program estimates use a 'person year' calculation where partial enrollment is counted as a fraction for the year. In contrast, the MCBS gives each beneficiary the same weight regardless of full or partial enrollment during the year, thus leading to differences in estimates using Medicare published statistics and MCBS data.
- **Where can I find documentation for longitudinal, pooled cross-sectional, and year-to-year analyses?** (FAQA4) Section 8.7: Combining Multiple Years of Data provides guidance on conducting longitudinal analyses and Appendix A contains analytic examples. The *Variable Crosswalks* for both the Survey File LDS and the Cost Supplement File LDS provide additional guidance to assist data users with determining the feasibility of conducting these analyses on variables of interest. For more information on conducting longitudinal, pooled cross-sectional, and year-to-year analyses using MCBS data, including analytic examples, please see the *MCBS Advanced Tutorial on Longitudinal Analysis Using MCBS Data*, the *MCBS Advanced Tutorial on Pooled Cross-Sectional Analysis*, and the *MCBS Advanced*



*Tutorial on Year-to-Year Estimate Comparisons* available at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials>.

- **Why does the number of BASEIDs differ across segments?** (FAQA5) There are multiple reasons why the number of BASEIDs may differ across segments. First, some segments include data from the Community Questionnaire and others from the Facility Instrument with different numbers of beneficiaries providing responses. Second, there are also differences in the number of beneficiaries by the specific round completed. Third, the use of ever enrolled or continuously enrolled weights in constructing the segments may result in differences.

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.

- **How can I conduct subgroup analyses and maintain the appropriate variance estimation?** (FAQA6)

► *Using the Balanced Repeated Replication (BRR) method of variance estimation*

Variance estimation can be impacted by selecting individuals prior to analysis. If the BRR variance estimation method is used, subgroup analyses can be conducted by limiting the dataset to the desired sub-sample. There are multiple ways to conduct subgroup analyses using BRR.

For indicator variables in three-way tables, you can create flags to help you identify the population of interest. For instance, if you are interested in the prevalence of diabetes by language spoke at home, but only in the over-65 population in Medicare Advantage (MA), you could use the following SAS® code:

```
proc surveyfreq data=mcbsdata VARMETHOD = brr (fay=.30);
table LANGUAGE * DIABETES * FLAG / col notot;
weight CEYRSWGT;
repweights CEYRS001 - CEYRS100;
run;
```

This sample code assumes an analytic data set, including replicate weights, in which the data user has created binary analytic variables for LANGUAGE and DIABETES, as well as a FLAG variable to identify the population of interest for this analysis. In this case, the flag is equal to 1 if the beneficiary is over 65 and in MA, and equal to 0 otherwise.

Since variance estimation using the BRR approach permits limiting the dataset to the desired sub-sample of interest, the following SAS code can also be used to achieve the same result through subgroup analysis:

```
data mcbsdata_subset;
set mcbsdata;
if FLAG = 1 then output;
run;

proc surveyfreq data=mcbsdata_subset VARMETHOD = brr (fay=.30);
table LANGUAGE * DIABETES / col notot;
weight CEYRSWGT;
```

```
repweight CEYRS001 - CEYRS100;
run;
```

► *Using the Taylor Series linearization method of variance estimation*

If other variance estimation methods, such as Taylor Series linearization are used, the correct way to analyze MCBS data is to employ domain statements (in SAS: proc surveymeans, surveylogistic, and surveyreg) or indicator variables in three-way tables (in SAS: proc surveyfreq). The Taylor Series linearization method of variance estimation is not recommended for subgroup analysis with MCBS data because accidentally excluding any observation in the sample while conducting the subgroup analysis using this variance estimation method will result in biased standard error estimates.

For indicator variables in three-way tables, data users can create flags to identify the population of interest. The variables SUDSTRAT (sampling strata) and SUDUNIT (primary sampling unit) are included for variance estimation using the Taylor Series linearization method. This method does not require replicate weights. For instance, if a data user is interested in the prevalence of diabetes by language spoken at home, but only in the over-65 population in MA, they could use the following SAS code:

```
proc surveyfreq data=mcbsdata;
table LANGUAGE * DIABETES * FLAG / col notot;
strata SUDSTRAT;
cluster SUDUNIT;
weight CEYRSWGT;
run;
```

Additional information on variance estimation can be found in Section 8.6: Variance Estimation.

- **Which weights are available in the Survey File - Early Release LDS?** (FAQA7) The Survey File - Early Release LDS includes cross-sectional weights for the ever enrolled population based on preliminary enrollment data. This ever enrolled population is a subset of the Survey File ever enrolled population. Longitudinal weights are not available for this LDS, nor are cross-sectional weights for the continuously enrolled population. For more information, see Section 8.4 Weighting.
- **Can I use the Survey File longitudinal weights with the Cost Supplement File data?** (FAQA8) The Survey File longitudinal weights are for analysis of Survey File data. Data users cannot use the Survey File longitudinal weights with Cost Supplement File data. There are no longitudinal weights for the 2015 Cost Supplement File, because 2014 data were not released. As of 2018, a three-year longitudinal Cost Supplement File weight is included. Users who want to analyze Survey File data along with utilization and cost data in the Cost Supplement File should limit analysis to cases with a positive Cost Supplement File weight.
- **Which cross-sectional weights are available in which data years?** (FAQA9) Cross sectional weights are available for the Survey File - Early Release, Survey File, and the Cost Supplement File in each data year. The Survey File - Early Release LDS contains weights for the ever enrolled (EVRWGTS) population and the Survey File LDS contains weights for both continuously enrolled (CENWGTS) and ever enrolled (EVRWGTS) populations. Given that the Cost Supplement File population represents an ever enrolled population enrolled in Medicare on at least one day at any time in the calendar year, the Cost Supplement File LDS contains cross-sectional weights for the ever enrolled population only (CSEVRWGT). The population represented by the sum of CSEVRWGT is identical to the population

represented by the sum of the ever enrolled Survey File weight, but it is populated for a smaller subset of respondents with complete cost and utilization data.

- **Can I link MCBS to electronic medical records?** (FAQA10) MCBS data cannot be linked to electronic medical records, or to any other records that record lab values or physiologic data. MCBS data can be linked to Medicare Part A and Part B claims data for beneficiaries who participated in the MCBS.
- **Are physical exams performed for the MCBS?** (FAQA11) Beginning in Winter 2022, the MCBS piloted the collection of physical measures data using the Physical Measures Questionnaire (PXQ), which collected data on six physical measures: gait speed, chair stand, balance test, measured height, measured weight, and measured grip strength. Pilot PXQ data collected in Summer 2023 were first released for the 2022 data year for a subset of Continuing Community beneficiaries. Information about the Physical Measures Pilot segment (PILOT\_PXQ) can be found in the *2022 Data User's Guide: Cost Supplement File*. Summer 2024 PXQ data are also available for the 2023 data year.
- **Can I use the MCBS data to estimate deductibles for Part A or Part B?** (FAQA12) MCBS data alone cannot be used to estimate total paid deductibles for Part A or Part B. CMS administrative data have no available compilation of total paid deductible for a beneficiary. However, the fee-for-service (FFS) claims data indicate which portion of the amount was applied to the deductible for FFS claims, so the total out of pocket amount in the Cost Supplement File Limited Data Set (LDS) does include amounts paid for the deductible. If a data user were to link MCBS data to Medicare Part A or Part B claims, summing the deductible fields on the FFS claims would provide an estimate of total paid deductible for the FFS enrollees. Please note that estimating total paid deductibles is only feasible for FFS enrollees. Medicare Advantage (MA) plan enrollees do not usually have a deductible, but instead have a copay. If MA enrollees have a deductible, it would vary by the plan.
- **How can I identify private health insurance plans that offer comprehensive coverage?** (FAQA13) MCBS collects information on a number of plan attributes. In particular, the survey asks beneficiaries whether their plan covers visits to a doctor, prescription medicines, hospital stays, etc. and makes coverage flags available on the Health Insurance Timeline (HITLINE) file. Data users can use these coverage flag to decide whether coverage offered by a particular plan should be considered comprehensive.
- **Why do certain segments have Topical segment weights?** (FAQA14) Some segments in the Survey File Limited Data Set (LDS) and Survey File - Early Release LDS have special non-response adjusted weights because the corresponding questionnaire sections, or specific items within those questionnaire sections, were fielded in the winter and summer rounds following the data year and/or are not administered to proxy respondents. For example, for the 2023 Survey File LDS, the Beneficiary Knowledge and Information Needs Questionnaire (KNQ) was only administered to respondents in Winter 2024, so the corresponding LDS segment with KNQ data (MCREPLNQ) contains Topical weights that are adjusted for the Winter 2024 population. These segments are referred to as Topical segments because most were traditionally sourced from the Topical questionnaire sections in the MCBS Community Questionnaire. Depending on the year, these adjusted weights may be included directly on each Topical segment or in standalone Topical weights segment(s). For more information, including a full list of Topical segments, weights, and weight names for each data year, see the *Data Year Releases Notes* or the historical 2015-2022 *MCBS Data User's Guide: Survey File*.

- **When should I use Topical weights and which Topical weights are available? (FAQ15)** Data users should use Topical segment weights when analyzing data from segments with special non-response adjusted weights. To generate estimates using these Topical segment data, on their own or merged with another Survey File - Early Release or Survey File segment, always use the special full-sample and replicate weights included directly in the Topical segment or in the corresponding standalone Topical weights segment (depending on the year). For example, an analysis that merges DEMO and FOODINS should use the Topical weights included in the FOODINS segment (or the standalone Topical weights segment with the summer non-response adjustment weights, as applicable for the year). Each Topical segment on the Survey File LDS corresponds to three sets of full-sample and replicate special weights, which can be used to conduct joint analysis of Topical segment data, Survey File data, and Cost Supplement File data. For the Topical segments released on the Survey File - Early Release, there is one set of full-sample and replicate special weights that corresponds to each Topical segment based on the Survey File - Early Release ever enrolled population. These weights may be used to conduct joint analyses of Topical segment data and other Survey File - Early Release data. Topical segment data collected in different rounds (e.g., winter and summer) should not be merged with each other, as these Topical segments include a different set of beneficiaries and there are no corresponding weights for this analysis. For additional guidance and analytic examples, see Section 8.4 Weighting and the *MCBS Advanced Tutorial on Weighting and Variance Estimation* at: <https://www.cms.gov/files/document/mcbs-advanced-tutorial-weighting-and-variance-estimation.pdf>
- **Where can I find data briefs and data products released by CMS analyzing MCBS data? (FAQ16)** CMS releases data briefs and other data products on various topics of interest, detailing methodology, results of analysis, and implications for the Medicare population. For example, CMS released a data brief on *Access to Care among Medicare beneficiaries Aged 65 and Over Living with High-Impact Chronic Pain* using MCBS data. This and other data briefs can be accessed through the CMS website: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-briefs-tutorials>. Additional analytic products can be found at <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-tables>.
- **What information is available in the Multiple Year Enrollment segment (MYENROLL)? (FAQ17)** The Multiple Year Enrollment segment (MYENROLL) first became available in 2021 and allows data users to view five years of enrollment information for the current year MCBS beneficiary population in one file. Depending on the original enrollment date of the beneficiary, the file will contain up to five records with monthly flags related to enrollment, dual eligibility status, and types of Medicare coverage. Data users can use this information alongside claims data or analyze the file independently. For example, data users can identify beneficiaries who are newly dually eligible with this file.