

DATA USER'S GUIDE: MICRODATA PUBLIC USE FILES



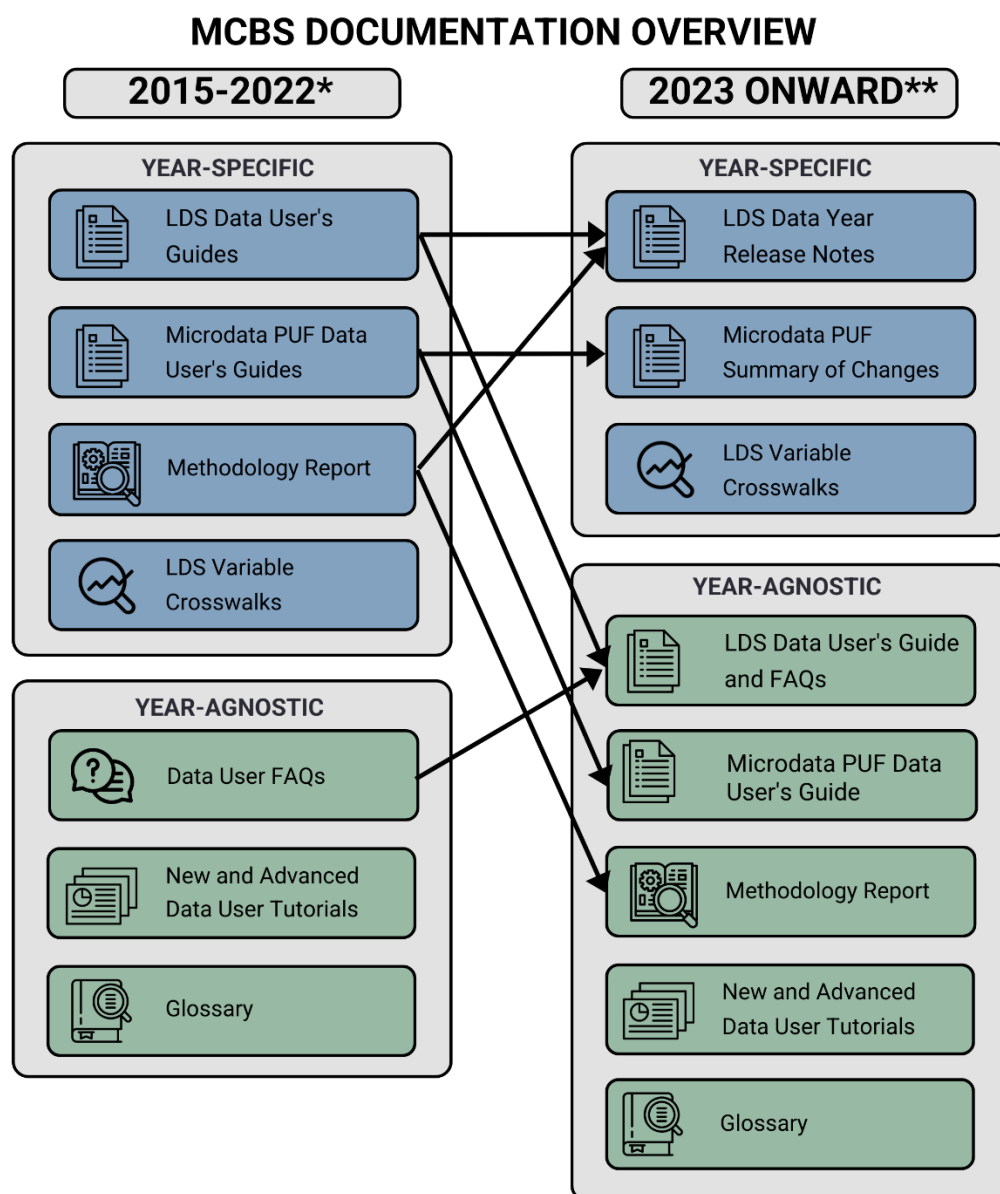
Centers for Medicare & Medicaid Services (CMS)
Office of Enterprise Data and Analytics (OEDA)

Version Control Log

Date	Version	Revisions
9/22/2025	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>.



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

VERSION CONTROL LOG	i
MCBS DOCUMENTATION CROSSWALK AND OVERVIEW	ii
ACRONYM LIST	vi
1. INTRODUCTION	1
2. OVERVIEW OF THE MCBS	2
3. TECHNICAL AND PROGRAMMING INFORMATION	3
3.1 General Information	3
3.2 Data File Information	3
3.3 Comparison to the LDS	4
3.4 Top- and Bottom-coding in the MCBS Cost Supplement File Microdata PUF	6
4. SURVEY OVERVIEW	7
4.1 Design of MCBS	7
4.2 Sample Design	8
4.3 Eligibility	9
4.3.1 Medicare Population Covered by the LDS and Microdata PUFs	9
4.4 Case Types	10
4.5 Interviewing and Training Procedures	10
4.5.1 Overview of Data Collection	10
4.5.2 Interviewer Training	11
4.5.3 Privacy and Data Security	11
5. QUESTIONNAIRES	12
5.1 Overview	12
6. SAMPLING	13
6.1 Medicare Population Covered by the MCBS Microdata PUFs	13
6.2 Targeted Population and Sampling Strata	13
6.3 Primary and Secondary Sampling Units	14
6.4 Sample Selection	14
7. TECHNICAL NOTES ON USING THE DATA	15

7.1 Weights and Variance Estimation 15

7.2 Item Non-Response 16

7.3 Subgroup Analysis..... 16

7.4 Example Research Questions 17

8. APPENDICES..... 19

 Appendix A: Technical Appendix – Survey File Microdata PUF Sample Code and Output..... 19

 SAS Analysis Statements 19

 Stata Analysis Statements 19

 R Analysis Statements 19

 Appendix B: Technical Appendix – Cost Supplement File Microdata PUF Sample Code and Output . 21

 SAS Analysis Statements 21

 Stata Analysis Statements 21

 R Analysis Statements 22

LIST OF EXHIBITS

Exhibit 3.3.1.a: Comparison between the MCBS Survey File Microdata PUF and MCBS Survey File LDS 5

Exhibit 3.3.1.b: Comparison between the MCBS Cost Supplement File Microdata PUF and MCBS Cost Supplement File LDS 5

Exhibit 4.1.1: Typical MCBS Data Collection Year 7

Exhibit 4.1.2: MCBS Beneficiary Participation Timeline 8

Exhibit 4.2.1: MCBS Rotating Panel Design 9

Exhibit 7.4.1: Example Research Questions That Can be Answered Using the MCBS Survey File Microdata PUF or MCBS Survey File LDS..... 17

ACRONYM LIST

AAPOR	American Association for Public Opinion Research
ACCESSCR	LDS Survey File Access to Care segment
ACCSSMED	LDS Survey File Access to Care, Medical Appointments segment
ADMNUTLS	LDS Survey File Administrative Utilization Summary segment
ASSIST	LDS Survey File Assistance segment
CAPI	Computer-Assisted Personal Interviewing
CBSA	Core-based statistical area
CHRNCOND	LDS Survey File Chronic Conditions segment
CMS	Centers for Medicare & Medicaid Services
COVIDTOP	COVID-19 Topical segment
CSV	Comma-separated values file
CSEVWGTS	LDS Cost Supplement File Ever Enrolled weights
DEMO	LDS Survey File Demographics segment
DUA	Data Use Agreement
EVRWGTS	LDS Survey File Ever Enrolled Population Weight segment
FALLS	LDS Survey File Falls segment
FOODINS	LDS Survey File Food Insecurity segment
GENHLTH	LDS Survey File General Health segment
HHCHAR	LDS Survey File Household Characteristics segment
HHS	U.S. Department of Health and Human Services
HIC	Health Insurance Claim
HISUMRY	LDS Survey File Health Insurance Summary segment
HITLINE	LDS Survey File Health Insurance Timeline segment
INTERV	LDS Survey File Interview Characteristics segment
IRB	Institutional Review Board
LDS	Limited Data Set(s)
MA	Medicare Advantage
MAPLANQX	LDS Survey File Medicare Advantage Plan Questions segment
MCREPLNQ	LDS Survey File Medicare Plan Beneficiary Knowledge segment
MCBS	Medicare Current Beneficiary Survey
NAGIDIS	LDS Survey File Nagi Disability segment
NICOALCO	LDS Survey File Nicotine and Alcohol segment
NORC	NORC at the University of Chicago
OMB	Office of Management and Budget
PHI	Protected Health Information
PII	Personally Identifiable Information
PREVCARE	LDS Survey File Preventive Care segment
PS	LDS Cost Supplement File Person Summary segment
PSU	Primary Sampling Units
PUF	Public Use File
RXMED	LDS Survey File RX Medications segment
SAS	Statistical Analysis System
SATWCARE	LDS Survey File Satisfaction with Care segment
SS	LDS Cost Supplement File Service Summary segment
SSU	Secondary Sampling Units
TELEMED	LDS Survey File Telemedicine segment
USCARE	LDS Survey File Usual Source of Care segment
USU	Ultimate Sampling Unit
VISHEAR	LDS Survey File Vision and Hearing segment

1. INTRODUCTION

CMS provides users with multiple ways to access Medicare Current Beneficiary Survey (MCBS) data, and a wide array of documentation is publicly available on the CMS MCBS website. MCBS data are made available via three Limited Data Set (LDS) releases and two annual Microdata Public Use File (PUF) releases, an MCBS Survey File Microdata PUF based on the Survey File LDS and an MCBS Cost Supplement File Microdata PUF based on the Cost Supplement File LDS.¹

The content of the MCBS Microdata PUFs is governed by their central focus of serving as unique sources of information on beneficiaries' health/well-being and cost/utilization that cannot be obtained through CMS administrative sources alone.

- Available beginning with the 2013 data year, the **Survey File Microdata PUF** includes data on Medicare beneficiaries' access to care, health status and functioning, knowledge of, attitudes toward, and satisfaction with their health care, demographics, and all types of health insurance coverage.
- Available beginning with the 2018 data year, the **Cost Supplement File Microdata PUF** includes data that links Medicare claims to survey-reported health care events and provides summarized expenditure and source of payment data on all health care services, including those not covered by Medicare.

MCBS Microdata PUFs expand access to survey data for researchers using the MCBS by offering free downloadable files that align with the format and standards of other public use survey files from the U.S. Department of Health and Human Services (HHS).

Disclosure protections have been applied to the files, including de-identification and other methods; as a result, the MCBS Microdata PUFs do not require a Data Use Agreement (DUA). In contrast, the MCBS LDS releases contain beneficiary-level protected health information (PHI) and therefore require a DUA. The MCBS Microdata PUFs are not intended to replace the more detailed LDS files; rather, they make available general-use public alternatives that provide the highest degree of protection to the Medicare beneficiaries' PHI.

This user guide contains general information about the MCBS Survey File and Cost Supplement File Microdata PUFs to help data users understand and analyze these files.² Each year, the PUFs are reviewed for updates to incorporate newly available or revised data. As a result, the number of variables included may fluctuate from year to year. For information specific to each release, researchers should refer to the respective *MCBS Microdata PUF Summary of Changes* provided with each MCBS Microdata PUF release.

Data users interested in understanding or analyzing MCBS data should also familiarize themselves with the content of the *Data Year Release Notes*, the *MCBS Data User's Guide*, and the *MCBS Methodology Report* for an overview of the survey, questionnaires, sample design, and other topics relevant to the MCBS. Data users can access these documents along with other data documentation at: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>. To access topical PUF tables with estimates that were prepared from MCBS data, researchers can visit: <https://data.cms.gov/medicare-current-beneficiary-survey-mcbs>.

¹ The MCBS Cost Supplement File PUF is based on the MCBS Cost Supplement File LDS; however, this PUF also includes select demographic and health factor variables from the MCBS Survey File LDS.

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

2. OVERVIEW OF THE MCBS

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 MCBS is sponsored by CMS and contains data provided by a representative national sample of the Medicare population. 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 from the Medicare program and plays an essential role in the monitoring and evaluation of 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 benefits such as Medicare's Part D prescription drug benefit.

This *Data User's Guide* uses the following definitions for beneficiary and respondent:

- *Beneficiary* refers to a person receiving Medicare services who may or may not be participating in the MCBS. Beneficiary may also refer to an individual selected from the MCBS sample about whom the MCBS collects information.³
- *Respondent* refers to a person who answers questions for the MCBS; for Community interviews, this person can be the beneficiary or a proxy. If the respondent is a proxy, they answer questions about the beneficiary rather than themselves.
- 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.

For questions or suggestions on this document or other MCBS data-related questions, please email MCBS@cms.hhs.gov.

³ <https://www.cms.gov/about-cms/what-we-do/medicare>.

3. TECHNICAL AND PROGRAMMING INFORMATION

3.1 General Information

The MCBS Survey File Microdata PUF is divided into three segments (Fall, Winter, and Summer) that are based on a variety of Survey File LDS segments, which allows for the release of data collected in all rounds. The samples for the Winter and Summer segments are subsets of the Fall segment sample. All three segments include survey weights that allow for analysis that is nationally representative of the population of beneficiaries ever enrolled in Medicare at any point during the data collection year.

The MCBS Cost Supplement File Microdata PUF is a single segment that is primarily based on the Person Summary (PS) segment of the MCBS Cost Supplement File LDS. This file also includes survey weights that allow for analysis that is nationally representative of the population of beneficiaries who were ever enrolled in Medicare at any point during the data collection year.

All records begin with a PUF_ID, a unique number for each beneficiary in the Microdata PUF:

- The Survey File Microdata PUF_ID serves to identify records in the three MCBS Survey File Microdata PUF segments only and cannot be used for linking to MCBS data files other than the MCBS Survey File Microdata PUF. The PUF_ID does link a beneficiary between the three MCBS Survey File Microdata PUF segments in the file.
- The MCBS Cost Supplement File Microdata PUF_ID can only be used to identify records in the MCBS Cost Supplement File Microdata PUF and likewise cannot be used for linking to MCBS data files other than the MCBS Cost Supplement File Microdata PUF.

Each beneficiary's PUF_ID is randomly generated each year, so it is not possible to link a beneficiary's data between years, and the value of the PUF_ID does not provide any information about the beneficiary or their year of enrollment.

All variables in the MCBS Microdata PUFs are in numeric or integer formats. Formats and values for each variable are available in the MCBS Microdata PUF codebook.

In the MCBS Survey File Microdata PUF, variable groups contain prefixes to help users identify these groups by topic area. Exhibits 3.1.1 – 3.1.3 include information about these variable prefixes and the locations of the corresponding variables in the LDS data segments.

3.2 Data File Information

Detailed information about variables in the MCBS Survey File and Cost Supplement File Microdata PUFs can be found in the PUF codebooks. The codebook includes SAS® variable names, labels, and any applicable notes. Certain variables in the PUFs were recoded due to disclosure concerns, so the categories in the PUF codebook may differ from the categories in the questionnaire specifications. Other variables were created by combining multiple variables, and their variable label indicates a recoded variable.

For each variable, the formats and format values are included in the codebook:

- Values of .R indicate "refused" and .D indicate "don't know."
- All values of "inapplicable" have been combined with missing values.
- Unweighted frequencies of most variables included in the MCBS Microdata PUFs are provided in the accompanying codebook file.

The MCBS Microdata PUF datasets are saved as SAS export files. Directions and sample SAS code are given below and in Appendix A and Appendix B to help users read the datasets into SAS.

Assume the MCBS Survey File Microdata PUF export files (e.g., PUF20YY_1_FALL.xpt) are downloaded into the folder "C:\MCBS\DOWNLOAD". The "YY" in "PUF20YY" refers to the data year of the MCBS Survey File Microdata PUF. The following SAS code can then be used to import the Survey File Microdata PUF Fall segment into SAS:

```
LIBNAME PUFLIB 'C:\MCBS\SASDATA';
FILENAME F_SFPUF "C:\MCBS\DOWNLOAD\PUF20YY_1_FALL.XPT";
PROC CIMPORT LIBRARY=PUFLIB INFILE=F_SFPUF;
RUN;
```

Likewise, the SAS code above could be altered to import the Survey File Microdata PUF Winter segment (SAS export file PUF20YY_2_WINTER.xpt) or Survey File Microdata PUF Summer segment (SAS export file PUF20YY_3_SUMMER.xpt) into SAS.

Assume the MCBS Cost Supplement File Microdata PUF export file (e.g., PUF20YY_1.xpt) is downloaded into the folder "C:\MCBS\DOWNLOAD". The "YY" in "PUF20YY" refers to the data year of the MCBS Cost Supplement File Microdata PUF. The following SAS code can then be used to import the Cost Supplement File Microdata PUF segment into SAS:

```
LIBNAME PUFLIB 'C:\MCBS\SASDATA';
FILENAME F_CSPUF "C:\MCBS\DOWNLOAD\PUF20YY_1.XPT";
PROC CIMPORT LIBRARY=PUFLIB INFILE= F_CSPUF;
RUN;
```

A text file with SAS programming code to import the .xpt files, create formats, and apply SAS labels is provided for users. Additionally, a comma-separated values (CSV) file is available for use with other statistical software packages such as R® and STATA®.

3.3 Comparison to the LDS

The MCBS Microdata PUFs differ from the MCBS LDS files because they have been evaluated for disclosure risk and additional steps were taken to protect beneficiary confidentiality. Many LDS variables that posed a disclosure risk were dropped or recoded to create the variable set for the MCBS Microdata PUFs.

Due to disclosure concerns, the MCBS Survey File Microdata PUF includes only beneficiaries living in the community for at least one interview during the data collection year, and thus excludes all beneficiaries who were in a health care facility during all interviews that year. Additionally, unlike the MCBS Survey File LDS, the MCBS Survey File Microdata PUF contains no health care utilization, cost, or payment data (including Medicare claims data) for individual beneficiaries. The MCBS Cost Supplement File Microdata PUF includes only beneficiaries living in the community the entire year. The file excludes beneficiaries who had a Facility interview during the year or who incurred any facility, hospice, or institutional events or costs during the year. In addition, unlike the MCBS Cost Supplement File LDS, which must be linked to the MCBS Survey File LDS for analysis, the MCBS Cost Supplement File Microdata PUF is a stand-alone file that cannot be linked to the MCBS Survey File PUF for analysis. Variables that were only created for beneficiaries living in Facilities are excluded from both Microdata PUFs.⁴

⁴ Facilities are defined as nursing homes, retirement homes, domiciliary or personal care facility, distinct long-term units in a hospital complex, mental health facility and centers.

MCBS Cost Supplement File LDS data are provided at three different levels of summarization: Event level, Service Summary (SS) level, and PS level. However, the MCBS Cost Supplement File PUF only provides data for the PS level, which summarizes utilization and expenditures by type of service and expenditures by payer, resulting in one record per beneficiary. In addition, unlike the Cost Supplement File LDS, the PUF only includes payment amounts that were adjusted to compensate for Medicare covered days that were not covered by the interview reference periods (as opposed to both the unadjusted and adjusted payment amounts available in the Cost Supplement File LDS). As in the Cost Supplement File LDS, the adjusted totals also include an upward ratio adjustment to Non Prescription Medicine (Non PM) utilization and expenditure data for beneficiaries enrolled in Medicare Advantage (MA) plans.

The MCBS Microdata PUFs are free and available for download on the CMS website. For users interested in the MCBS Survey File and Cost Supplement File LDS, more information on the LDS process can be found at: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey>.

A summary of the differences between the MCBS Survey File Microdata PUF and Survey File LDS is presented in Exhibits 3.3.1.a. A comparison between the MCBS Cost Supplement File Microdata PUF and Cost Supplement File LDS is shown in Exhibit 3.3.1.b.

Exhibit 3.3.1.a: Comparison between the MCBS Survey File Microdata PUF and MCBS Survey File LDS

Domain	MCBS Survey File PUF	MCBS Survey File LDS
Population	Community	Community and facility
ID	PUF_ID; Randomly generated, can't be linked back to BASEID, changes each year	BASEID; Randomly generated, can't be linked back to Medicare Beneficiary Identifier (MBI), consistent between years
Date fields	NO	YES
Geographic identifiers	NO	YES
Cost/payment data	NO	YES
Demographic data	YES; All variables are categorical, limited age categories	YES; Continuous, all age variables available
Insurance coverage	YES; Summarized to annual level	YES; Monthly level
Identifiable plan-related information for MA or Part D	NO	YES
Population weights	Ever enrolled weights only	Both ever enrolled and continuously enrolled weights

Exhibit 3.3.1.b: Comparison between the MCBS Cost Supplement File Microdata PUF and MCBS Cost Supplement File LDS

Domain	MCBS Cost Supplement File PUF	MCBS Cost Supplement File LDS
Population	Community only	Community and facility
ID	PUF_ID; Randomly generated, can't be linked back to BASEID, changes each year	BASEID; Randomly generated, consistent between years
Date fields	NO	YES
Geographic identifiers	NO	NO
Cost/payment data	YES	YES

Domain	MCBS Cost Supplement File PUF	MCBS Cost Supplement File LDS
Demographic data	YES; All variables are available in the MCBS Survey File LDS	NO
Population weights	Ever enrolled weights only	Ever enrolled weights only
Linkable to other MCBS data products	NO	YES; Can be linked to MCBS Survey File LDS and over time to other MCBS Survey File LDS and Cost Supplement File LDS data years

3.4 Top- and Bottom-coding in the MCBS Cost Supplement File Microdata PUF

All of the service- and payer-specific costs and events variables are top-coded at the 99.5 percent level. For each of these variables, beneficiaries who have a value in the highest 0.5 percent of the unweighted sample are separated out and have their mean value calculated. Their values are then replaced by the mean value of the top 0.5 percent. This process is performed separately for each of the eight variables for cost by service type, the eight variables for number of events by service type, and the eight variables for payments by payer type. Additionally, the variable PAMTOTH, which is the only variable for which negative values are possible, is bottom-coded; all negative values in the unweighted sample are averaged and then replaced with that mean value. The variable PAMTTOT is calculated as the sum of the seven top- and bottom-coded constituent variables for payments by payer type. The variable PEVENTS is calculated as the sum of the eight bottom-coded constituent variables for number of events by service type.

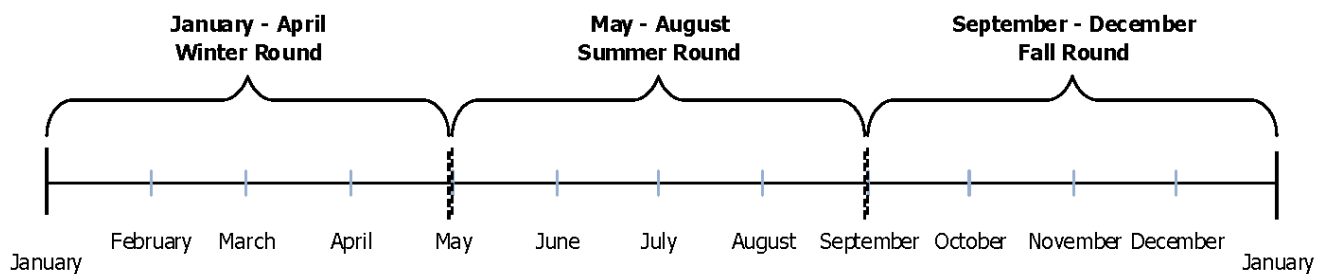
4. SURVEY OVERVIEW

4.1 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 recall periods during the year to capture more complete and accurate health care costs and utilization for beneficiaries.

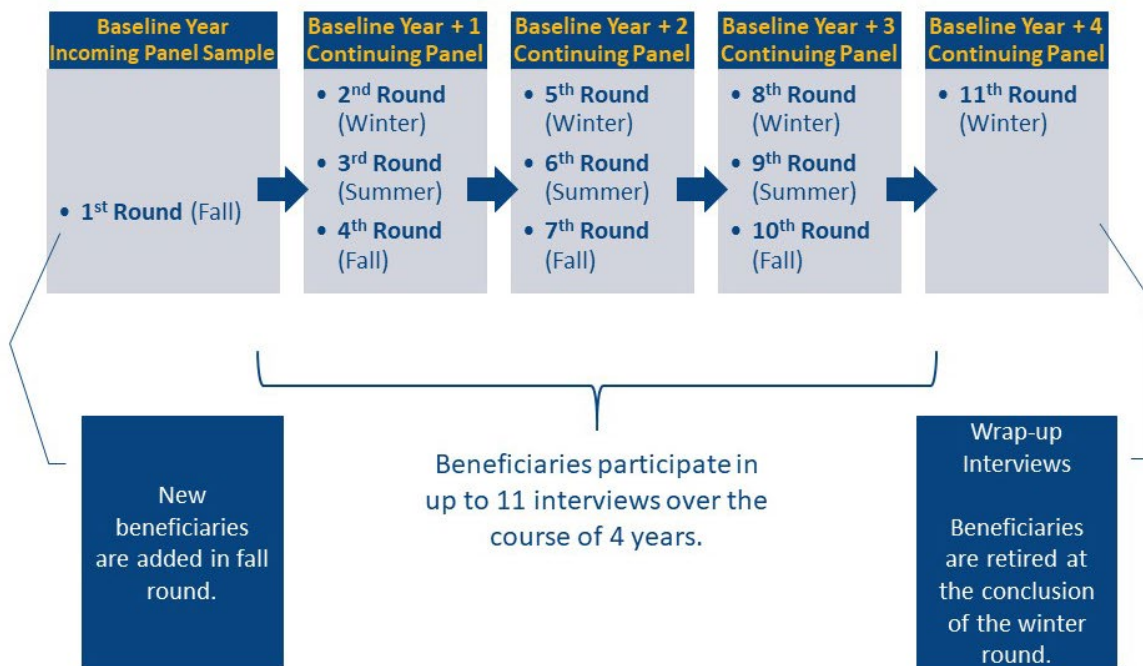
The MCBS data releases for each data year reflect data collected from January through December of that data collection year (see Exhibit 4.1.1), as well as data on income and assets, access to care, usual source of care, preventive care, COVID-19, beneficiary knowledge and information needs, drug coverage, and chronic pain information collected through the winter and summer rounds of the following data collection year.⁵ Exhibit 4.1.1 depicts an MCBS data collection year and the typical span of the rounds.

Exhibit 4.1.1: 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) until they have completed four years of participation (up to 11 interviews in total). Exhibit 4.1.2 depicts the timeline of participation for beneficiaries selected to be in the MCBS sample.

⁵ Due to the nature of some survey items, PUF data for each data year may include data pulled forward from a prior data collection year and/or data added from a subsequent data collection year due to the specific reference period.

Exhibit 4.1.2: MCBS Beneficiary Participation Timeline**4.2 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.⁶ Each MCBS panel, an annual statistical sample of all Medicare enrollees, is interviewed up to three times a year over a four-year period creating a continuous profile of selected beneficiaries' health care experiences.⁷ One panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round (see Exhibit 4.2.1). The size of the new panel is designed to provide a stable number of beneficiaries across all panels participating in the survey annually.

⁶ 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.

⁷ 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.

Exhibit 4.2.1: MCBS Rotating Panel Design

Interview Number (per Panel)	Year	Year 1Year 2Year 3Year 4											
	Round	F	W	S	F	W	S	F	W	S	F	W	
	Panel 1	←	7th	8th	9th	10th	11th						
	Panel 2	←	4th	5th	6th	7th	8th	9th	10th	11th			
	Panel 3		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
	Panel 4					1st	2nd	3rd	4th	5th	6th	7th	8th
Panel 5								1st	2nd	3rd	4th	5th	→

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

The MCBS employs a three-stage cluster sample design. Primary sampling units (PSUs) are made up of major geographic areas consisting of metropolitan areas or groups of rural counties. Secondary sampling units (SSUs) are made up of census tracts or groups of tracts within the selected PSUs. Medicare beneficiaries, the ultimate sampling units (USUs), are then selected from within the selected SSUs. The final annual MCBS Panel was drawn from 104 PSUs, which contained 685 SSUs. The MCBS sample is annually “supplemented” during the fall round to account for attrition (deaths, dis-enrollments, refusals) and current-year enrollees. Each annual supplement is referred to as the Incoming Panel sample.

Beneficiaries for the MCBS are sampled from the Medicare Administrative enrollment data. The beneficiaries included in the MCBS Microdata PUFs represent a randomly selected 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.⁸ The MCBS Microdata PUFs represent four separate MCBS panels identified by the year in which the panel was selected and first interviewed (i.e., for the Calendar Year (CY) 4 MCBS Cost Supplement File Microdata PUF, the CY 1, CY 2, CY 3, and CY 4 Panels).

For more information on the sample design, please see the *Methodology Report* at:

<https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

4.3 Eligibility

4.3.1 Medicare Population Covered by the LDS and Microdata PUFs

Beneficiaries who became eligible for Medicare Part A or B and enrolled anytime during the sampling year were eligible to be sampled as part of the annual panel.

⁸ While beneficiaries included in the MCBS Survey File LDS release represent both the ever enrolled and continuously enrolled Medicare population, the MCBS Survey File Microdata PUF solely represents the ever enrolled population.

4.4 Case Types

MCBS beneficiaries are classified by their phase of survey participation (i.e., Incoming or Continuing) and interview participation (i.e., Community or Facility), which is determined by residence status. Although they appear in the MCBS LDS releases, beneficiaries for whom only Facility interviews were conducted during the data collection period are not included in the MCBS Survey File Microdata PUF and beneficiaries for whom any Facility interview was conducted during the data collection period are not included in the MCBS Cost Supplement File PUF. Researchers interested in the population of beneficiaries living in facilities will need to use the MCBS LDS, as discussed in Section 3.3.

- **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:** 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.

4.5 Interviewing and Training Procedures

4.5.1 Overview of Data Collection

CMS contracts with NORC at the University of Chicago (NORC) to administer the MCBS. A national team of specially trained and certified NORC field interviewers conduct either interviews with MCBS beneficiaries or their designated proxies or they conduct interviews with Facility administrators on behalf of beneficiaries. The first interview conducted for an Incoming Panel beneficiary is relatively short, as it does not collect health care utilization or cost data. Continuing interviews are longer, as field interviewers collect information about the beneficiary's health care utilization and associated costs.

MCBS data collection includes both in-person and phone outreach and interviewing. Mode is determined by efficiency in outreach, beneficiary preference, and local interviewer availability, among other factors. Data quality is monitored to ensure high quality data are collected, regardless of mode.

4.5.1.1 Overview of Recruitment of Beneficiaries and Scheduling Procedures

Medicare beneficiaries selected to participate in the MCBS receive a letter and a brochure in the mail, introducing the study and explaining that an interviewer from NORC will contact them to schedule an appointment. For Incoming Panel interviews, initial contact is typically made in person; for Continuing interviews, outreach to set an appointment is most often made by phone. If beneficiaries are unable to answer questions or require language assistance, they can enlist the help of an assistant, such as a family member, to help complete the interview; a proxy can also respond on behalf of the beneficiary if the beneficiary is incapacitated or unable to complete the interview. For Spanish speaking respondents, a Spanish version of the Community Questionnaire is available, and bilingual interviewers conduct the interview.

4.5.1.2 Computer-Assisted Personal Interviewing (CAPI)

Field interviewers complete MCBS interviews using a CAPI instrument loaded on a laptop. The CAPI program automatically guides the field interviewer through the questions, records the answers, and contains logic and skip flows that increase the output of timely and high quality data. The CAPI also contains follow-up questions where data were missing from the previous interview. When the interview is completed, the CAPI system allows the field interviewer to transmit the data electronically to the NORC central office in a secure manner.

4.5.2 Interviewer Training

Nationally, the MCBS employs field interviewers who participate in a combination of several targeted training initiatives and careful coaching and monitoring activities throughout data collection. The MCBS Training Program consists of remote and in-person trainings which vary based on the level of experience of the interviewer (new to MCBS or MCBS-experienced), the type of interview component (Community or Facility), the sample type (Incoming Panel or Continuing), and the season-specific requirements (new or changing questionnaire sections or data collection protocols). The program is structured to expose all field staff to the same training content, ensuring that the performance of data collection responsibilities is standardized, methodical, and measurable.

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

4.5.3 Privacy and Data Security

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

NORC and CMS are committed to protecting respondent confidentiality and privacy, and both organizations diligently uphold provisions established under the Privacy Act of 1974, the NORC Institutional Review Board (IRB), the Office of Management and Budget (OMB), and the Federal Information Security Management Act of 2002. As stated in the MCBS OMB documentation, the information collected for MCBS is protected by NORC and by CMS. Respondent data are used only for research and statistical purposes. As required under the Privacy Act of 1974, identifiable information is not disclosed or released without the consent of the individual or the establishment, except to those involved in research (Public Law 93-579). 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.

5. QUESTIONNAIRES

5.1 Overview

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 vary by interview type and data collection season (fall, winter, or summer). There are two types of interviews (Baseline and Continuing) containing two types of questionnaire sections (Core and Topical). See Exhibit 3.1 within the *Methodology Report* for a depiction of the MCBS Questionnaire structure:

<https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

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

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

- **Baseline:** The initial questionnaire administered in the fall round of the year the beneficiary is selected into the sample (interview #1).
- **Continuing:** The questionnaire administered as beneficiaries progress through the study (interviews #2-11).

Depending on the interview type and data collection season (fall, winter, or summer), the MCBS Questionnaire includes Core and Topical sections:

- **Core:** These sections collect data central to policy goals of CMS. Core sections used for the MCBS Survey File Microdata PUF collect information related to socio-demographics, health insurance coverage, health care utilization, beneficiary health status, and experiences with care, as well as operational and procedural data. The questionnaire sections may be asked each round or seasonally (fall, winter, summer). Core sections used for the MCBS Cost Supplement File Microdata PUF collect data about beneficiaries' health insurance coverage, health care utilization and costs, and operational management data such as locating information.
- **Topical:** These sections collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, health behaviors, knowledge about Medicare, and health-related decision making. Data from the Topical sections do not appear on the MCBS Cost Supplement File Microdata PUF, just the Survey File Microdata PUF.

See the *Methodology Report* for additional detail on the Core and Topical sections: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

6. SAMPLING

6.1 Medicare Population Covered by the MCBS Microdata PUFs

The MCBS data releases are a reflection of 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. Residents of foreign countries and U.S. possessions and territories are excluded. The MCBS Survey File PUF further excludes Medicare beneficiaries who only provided facility-based interviews during the year, while the MCBS Cost Supplement File Microdata PUF further excludes Medicare beneficiaries who provided any facility interviews during the year, or who incurred any facility, hospice, or institutional costs or events. The MCBS data releases include two overlapping but differing populations:

- 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.⁹ The ever enrolled population includes beneficiaries who were enrolled in Medicare for at least one day at any point during the calendar 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 their fall interview.

The MCBS Survey File Microdata PUF includes weights that represent the ever enrolled population, whereas the MCBS Survey File LDS releases include separate sets of weights that represent both the ever enrolled and continuously enrolled Medicare population.

The MCBS Cost Supplement File Microdata PUF and the MCBS Cost Supplement File LDS releases include weights that represent only the ever enrolled population. The MCBS Cost Supplement File Microdata PUF and MCBS Cost Supplement File LDS ever enrolled population represents a subset of beneficiaries with complete cost and utilization data for the year.

6.2 Targeted Population and Sampling Strata

The targeted population for the MCBS consists of persons enrolled in one or both parts of the Medicare program, that is, Part A or Part B, as of December 31 of the applicable sample-selection year, and whose address on the Medicare files is in one of the 48 contiguous states (excludes Alaska and Hawaii) or the District of Columbia. For example, for the fall rounds of CY1, CY2, CY3, and CY4 (the four rounds in which the CY1, CY2, CY3, and CY4 Panels included in the CY4 MCBS data are selected), the targeted population includes individuals enrolled as of December 31 of CY1, CY2, CY3, and CY4 respectively.

Additionally, in these panels, beneficiaries residing within the U.S. who are Hispanic (based on a Hispanic ethnicity classification code in the Medicare enrollment data) are oversampled to improve precision of estimates for this group.¹⁰ For more information on the sampling strata, please see Exhibit 2.1.3 of the *Methodology Report*: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

⁹ 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.

¹⁰ Oversampling of Hispanic beneficiaries has been conducted throughout the MCBS and has evolved over time. See prior *MCBS Methodology Reports* for more information: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

6.3 Primary and Secondary Sampling Units

All of the panels in the reference year's data releases are distributed across the subset of 104 PSUs from the redesigned sample of 107 PSUs selected in 2001. These PSUs are a representative, national sample of beneficiaries who are geographically dispersed throughout metropolitan areas and groups of non-metropolitan counties. Recall that SSUs are census tracts or groups of contiguous tracts within the selected PSUs.

6.4 Sample Selection

The MCBS sampling design provides nearly self-weighting (i.e., equal probabilities of selection) samples of beneficiaries within each of the 14 sampling strata. Within the selected PSUs and SSUs, a systematic sampling scheme with random starts is employed for selecting beneficiaries.¹¹ For each Continuing beneficiary, the survey questions corresponding to the data release are administered in all three rounds of the collection year. For beneficiaries new to the MCBS, the survey questions are administered as part of the initial fall Baseline interview.

¹¹ The MCBS Panels are drawn by systematic random sampling with probability proportional to probabilities of selection with an independently selected random start within each PSU. For more information on this sampling method, please see the *MCBS Methodology Report*, available at: <https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks>.

7. TECHNICAL NOTES ON USING THE DATA

7.1 Weights and Variance Estimation

The sample design of MCBS includes stratification, clustering, multiple stages of selection, and disproportionate sampling. Furthermore, the MCBS sampling weights reflect adjustments for survey nonresponse. These survey design and estimation complexities require special consideration when analyzing MCBS data (i.e., it is not appropriate to assume simple random sampling).

To obtain accurate estimates from MCBS data, for either descriptive statistics or more sophisticated analyses based on multivariate models, the survey design complexities need to be taken into account by applying MCBS weights to produce estimates and using an appropriate technique to derive standard errors associated with the weighted estimates.

Most commercial software packages today include techniques to accommodate the complex design, through replicate weight approaches. Among these are STATA®, SUDAAN®, R®, and the complex survey procedures in SAS®. When using the replicate weight approach to variance estimation, the variance estimation method of balanced repeated replication (BRR) using Fay's adjustment of 0.3 is recommended. Sample code in SAS, STATA, and R for estimating statistics can be found in Appendix A. Analysis of subgroups should utilize the domain functions within the statistical package of the data user's choice (e.g., the DOMAIN statement in SAS, or the OVER function in STATA); restricting the sample to the subgroup and then performing an analysis would lead to slightly biased point estimates and estimates of variance.

For more information on weighting and variance estimation as well as other analytic guidance, please refer to the *MCBS LDS Data User's Guide*.

7.1.1 Survey File Microdata PUF

Each segment of the MCBS Survey File Microdata PUF includes ever enrolled, full sample cross-sectional weights (Fall: PUFFWGT; Winter: PUFWWGT; Summer: PUFSWGWT). The MCBS Survey File Microdata PUF segments do not include the continuously enrolled cross-sectional weights in order to protect the confidentiality of the beneficiaries. The continuously enrolled cross-sectional weights are available, however, in the LDS.

For the MCBS Survey File Microdata PUF for CY4 (for example), the ever enrolled cross-sectional weights apply to both the Continuing sample (beneficiaries sampled in CY1, CY2, or CY3) and to the Incoming Panel sample (beneficiaries sampled in CY4). These weights are intended for use in cross-sectional statistics involving the total (combined) Fall CY4 sample. Each weight is greater than zero for all beneficiaries on the file. The ever enrolled cross-sectional weights should be used to make estimates of parameters for the Medicare population who were enrolled at any point in CY4 (i.e., the ever enrolled population).

To generate estimates using the data from merged seasonal segments, the data user must always use the weights that correspond to the segment that is the smaller subset of the other. The samples for the Winter and Summer segments are subsets of the Fall segment sample. Thus, when generating estimates from a merged Summer segment and Fall segment analytic file, the data user must use the Summer segment weights. There are no weights that support joint analysis between the Summer segment and Winter segment, as each segment has a different set of beneficiaries included. To permit the calculation of random errors due to sampling, a series of replicate weights were computed. Unless the complex nature of the MCBS is taken into account, estimates of the variance of a survey statistic may be biased downward. The replicate weights included in the MCBS Survey File Microdata PUF can be used to calculate standard errors of the sample-based estimates. The replicate cross-sectional weights in the Fall segment are labeled PUFF001 through PUFF100

corresponding to the ever enrolled weight PUFFWGT. The replicate cross-sectional weights in the Winter segment are labeled PUFW001 through PUFW100 corresponding to the ever enrolled weight PUFFWGT. The replicate cross-sectional weights in the Summer segment are labeled PUF001 through PUF100 corresponding to the ever enrolled weight PUF0WGT.

7.1.2 Cost Supplement File Microdata PUF

The MCBS Cost Supplement File PUF for CY4 (for example) includes ever enrolled cross-sectional weights (CSPUFWGT) which apply to both the Continuing sample (beneficiaries sampled in CY1, CY2, or CY3) and to the Incoming Panel sample (beneficiaries sampled in CY4). These weights are intended for use in cross-sectional statistics involving the total (combined) Fall CY4 sample. Each weight is greater than zero for all beneficiaries on the file. The ever enrolled cross-sectional weights should be used to make estimates of parameters for the Medicare population who were enrolled at any point in CY4 (i.e., the ever enrolled population). To permit the calculation of random errors due to sampling, a series of replicate weights were computed. Unless the complex nature of the MCBS is taken into account, estimates of the variance of a survey statistic may be biased downward. The replicate weights included in the MCBS Cost Supplement File PUF can be used to calculate standard errors of the sample-based estimates.

7.2 Item Non-Response

As in any other survey, some respondents could not, or would not, supply answers to some questions.¹² 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.¹³ The calculation of the study-wide response rates generally follows the guidelines specified in the American Association for Public Opinion Research (AAPOR) and OMB.

7.3 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 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.

¹² 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*, Section 9.

¹³ In the LDS files, item non-response types are indicated by missing type codes in SAS, including refusal to answer, don't know the answer, and invalid skip. The code .D represents a "don't know" response, the code .R represents a "refused" response, and .N represents an "invalid skip" response.

7.4 Example Research Questions

Exhibit 7.4.1 presents example research questions by topic, differentiating between those that can be addressed by the MCBS Survey File Microdata PUF or MCBS Survey File LDS, and those that can only be addressed by the MCBS Survey File LDS. These research questions are intended to illustrate the types of analyses researchers can perform using either the MCBS Survey File Microdata PUF or MCBS Survey File LDS, and are not meant to be a comprehensive list of possible research questions that can be answered with these data. The variables that are unique to the MCBS Survey File LDS and not contained in the MCBS Survey File Microdata PUF are italicized below.

Exhibit 7.4.1: Example Research Questions That Can be Answered Using the MCBS Survey File Microdata PUF or MCBS Survey File LDS

Topic	Example Research Questions Addressed Using the MCBS Survey File Microdata PUF or MCBS Survey File LDS	Example Research Questions Addressed Using the MCBS Survey File LDS
Quality of Patient Experience	Are there differences in Medicare beneficiaries' quality of patient experience across socio-demographic characteristics?	Are there differences in Medicare beneficiaries' quality of patient experience between beneficiaries with <i>limited English proficiency</i> and beneficiaries who are proficient in English?
Access to Care	Among Medicare beneficiaries, are there differences in access to care by income (below or above \$25,000) and level of education?	Among Medicare beneficiaries, are there differences in access to care by <i>employment status</i> ?
Preventive Care	Are there differences in receipt of preventive care (e.g., pneumococcal vaccination) by age?	Are there differences in receipt of preventive care (e.g., pneumococcal vaccination) by <i>patient activation</i> ; that is, the degree to which beneficiaries actively participate in their own health care and decisions concerning that health care?
Cost and Utilization	Among Medicare beneficiaries, are there differences in the average number of inpatient hospital stays between 2019 and 2021?	Are there changes in Medicare beneficiaries' <i>out-of-pocket costs over the last 20 years</i> ?
Health Behaviors or Socio-demographics	Are there differences in the percentage of Medicare beneficiaries who smoke cigarettes or consume excessive amounts of alcohol by socio-demographics?	Are there differences in the percentage of Medicare beneficiaries who have <i>used e-cigarettes</i> by socio-demographics?
Health Status and Functioning	Which disease conditions are more common among Medicare beneficiaries who had falls requiring medical help?	Are there differences in the disease conditions associated with falls <i>between 2010 and 2020</i> ?
Housing Characteristics	What is the profile (e.g., socio-demographic characteristics and disease conditions) of Medicare beneficiaries with accessibility modifications (e.g., a ramp, bathroom modification and/or railing) in the house?	Are <i>specific Instrumental Activities of Daily Living (IADL) or Activities of Daily Living (ADL) limitations</i> associated with accessibility modifications among Medicare beneficiaries?

APPENDICES

8. APPENDICES

Appendix A: Technical Appendix – Survey File Microdata PUF Sample Code and Output

Please note that the code examples below use the Survey File Microdata PUF Fall segment weights, which begin with the prefix "PUFF." The data user should use the Survey File PUF Fall segment weights only if they are using data from the Survey File PUF Fall segment that have not been merged with data from any other segment. If the data user is analyzing data from the winter, summer, or a combination of Survey File PUF segments, please see the discussion in section 7.1.1 of this document as to which weights should be used.

SAS Analysis Statements

Cross-tabulations

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
    table <Var name> / row chisq lrchisq;
    weight PUFFWGT;
    repweight PUFF001 - PUFF100;
run;
```

Subgroup Analysis

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
    table <Var name> * <Subgroup variable> / row chisq lrchisq;
    weight PUFFWGT;
    repweight PUFF001 - PUFF100;
run;
```

Stata Analysis Statements

Declare dataset as survey sample with replicate weights

```
svyset _n [pweight= PUFFWGT ], brrweight(PUFF001 - PUFF100) fay(.3) vce(brr) singleunit(missing)
```

For categorical variables, use:

```
svy brr, fay(.3) : tabulate <Var name> <Var name>
```

For subgroup analysis use:

```
svy brr, subpop(if <Subgroup>) fay(.3) : tabulate <Var name>, over(<Var name>)
```

R Analysis Statements

Declare MCBS survey design object with replicate weights

```
mcbs <- svrepdesign(
  weights = ~PUFFWGT,
  repweights = "PUFF[001-100]+",
  type = "Fay",
  rho = 0.3,
  data = <Source dataset>,
  combined.weights = TRUE
)
```

For categorical variables, use:

```
svytable(~<Var name>, design=mcbs)
```

For subgroup analysis use:

```
mcbs_subgrp <- subset(mcbs, <Subgroup criteria>)  
svytable(~<Var name>, design=mcbs_subgrp)
```

Appendix B: Technical Appendix – Cost Supplement File Microdata PUF Sample Code and Output

Please note that the code examples below use the Cost Supplement File PUF weights, which begin with the prefix "CSPUF."

SAS Analysis Statements

Frequencies

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
    table <Var name>;
    weight CSPUFWGT;
    repweight CSPUF001 - CSPUF100;
run;
```

Cross-tabulations

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
    table <Subgroup variable> *<Var name> / row chisq lrchisq;
    weight CSPUFWGT;
    repweight CSPUF001 - CSPUF100;
run;
```

Means

```
proc surveymeans data=<Analytic dataset> VARMETHOD = brr (fay=.30);
    var <Var name>;
    weight CSPUFWGT;
    repweight CSPUF001 - CSPUF100;
run;
```

Stata Analysis Statements

Declare dataset as survey sample with replicate weights

```
svyset _n [pweight= CSPUFWGT ], brrweight(CSPUF001 - CSPUF100) fay(.3) vce(brr) singleunit(missing)
```

For categorical variables, use:

```
svy brr, fay(.3) : tabulate <Var name> <Var name>
```

For means of continuous variables, use:

```
svy brr, subpop(if <Subgroup>) fay(.3) : mean <Continuous var name>
```

For subgroup analysis of categorical variables, use:

```
svy brr, subpop(if <Subgroup>) fay(.3) : tabulate <Var name> <Var name>
```

For subgroup analysis of continuous variables, use:

```
svy brr, subpop(if <Subgroup>) fay(.3) : mean <Continuous name>, over(<Categorical var name>)
```

*R Analysis Statements***Declare MCBS survey design object with replicate weights**

```
mcbs <- svrepdesign(  
  weights = ~CSPUFWGT,  
  repweights = "CSPUF[001-100] +",  
  type = "Fay",  
  rho = 0.3,  
  data = <Source dataset>,  
  combined.weights = TRUE  
)
```

For categorical variables, use:

```
svytable(~<Var name>, design=mcbs)
```

For means of continuous variables, use:

```
svymean(~<Var name>, design=mcbs)
```

For subgroup analysis of categorical variables, use:

```
mcbs_subgrp <- subset(mcbs, <Subgroup criteria>)  
svytable(~<Var name>, design=mcbs_subgrp)
```

For subgroup analysis of continuous variables, use:

```
mcbs_subgrp <- subset(mcbs, <Subgroup criteria>)  
svymean(~<Var name>, design=mcbs_subgrp)
```