MCBS
Medicare Current Beneficiary Survey

2022 | DATA USER'S GUIDE: COST SUPPLEMENT FILE PUBLIC USE FILE

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

Version Control Log

Date	Version	Revisions
01/16/2025	1.0	Initial version published.

OVERVIEW OF MCBS DOCUMENTATION

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

- **Data User's Guides**: A Data User's Guide is produced for each MCBS Limited Data Set (LDS) and Microdata Public Use File (PUF) data release. There are three broad categories of Data User's Guides.
  - *Survey File Data User's Guide*: Updated annually for each new data year, the *Survey File Data User's Guide* supports researchers in understanding and analyzing Survey File LDS data. This Data User's Guide contains detailed information about the Survey File LDS, including changes between years, important data user considerations, and sample code, as well as basic background information on the MCBS, including sampling, questionnaires, data collection, and data processing. Along with the *New User Tutorial* (see below), this Data User's Guide is the recommended starting point, particularly for researchers new to studying MCBS data.
  - *Cost Supplement File Data User's Guide*: Updated annually for each new data year, the *Cost Supplement File Data User's Guide* functions as a supplement to the corresponding *Survey File Data User's Guide* and supports researchers in understanding and analyzing Cost Supplement File LDS data. This Data User's Guide focuses on providing detailed information about the Cost Supplement File LDS, including changes between years, important data user considerations, and sample code.
  - *Public Use File Data User's Guides* (this document): A Data User's Guide is also produced for each MCBS Microdata PUF release, including the annual Survey File PUF, the annual Cost Supplement File PUF, and the three COVID-19 Supplement PUFs. These Data User's Guide provide detailed, focused information to support researchers in understanding and analyzing PUF data.
- **Methodology Report**: Updated annually for each new data year, the *Methodology Report* provides detailed background information on the methods used to conduct the MCBS and process MCBS data. This includes information on sampling methodology, questionnaire development and programming, interviewer recruitment and training, data collection procedures, data processing and editing, including weighting and imputation, and response rates.
- Data User Tutorials:
  - *New User Tutorial*: Aimed at new data users who are unfamiliar with the MCBS, the *New User Tutorial* provides an overview of MCBS history, policy relevance, survey design, data products, and best practices for analysis. Along with the *Survey File Data User's Guide* (see above), the *New User Tutorial* is the recommended starting point for researchers.
  - *Advanced Topic-Based Tutorials*: In addition to the *New User Tutorial*, CMS has released a series of tutorials on more advanced topics, with the goal of supporting researchers in better understanding how to analyze and interpret MCBS data by providing detailed analytic guidance and examples. Topics of these tutorials include the differences between MCBS Community and Facility data, weighting and variance estimation, using data from the MCBS COVID-19 Supplements, conducting longitudinal analysis, and conducting pooled cross-sectional analysis with MCBS data.
- **Glossary**: Formerly included as an appendix in MCBS documentation products, this new standalone resource provides the definitions for common key terms used by the MCBS.

TABLE OF CONTENTS

LIST OF EXHIBITS

ACRONYM LIST

AAPOR	American Association for Public Opinion Research
CAPI	Computer-Assisted Personal Interviewing
CHRNCOND	LDS Survey File Chronic Conditions segment
CMS	Centers for Medicare & Medicaid Services
CSV	Comma-separated values file
CSEVWGTS	LDS Cost Supplement File Ever Enrolled weights
DEMO	LDS Survey File Demographics segment
DUA	Data Use Agreement
HHS	U.S. Department of Health and Human Services
IRB	Institutional Review Board
LDS	Limited Data Set(s)
MA	Medicare Advantage
MCBS	Medicare Current Beneficiary Survey
NORC	NORC at the University of Chicago
OMB	Office of Management and Budget
PHI	Protected Health Information
PII	Personally Identifiable Information
PS	LDS Cost Supplement File Person Summary segment
PSU	Primary Sampling Units
PUF	Public Use File
SAS	Statistical Analysis System
SS	LDS Cost Supplement File Service Summary segment
SSU	Secondary Sampling Units
USU	Ultimate Sampling Unit

1. INTRODUCTION

Over the past several years, the Centers for Medicare and Medicaid Services (CMS) has made it a priority to make more data available, including releasing to the public an unprecedented amount of information on services and procedures provided to Medicare beneficiaries. 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 two annual Limited Data Set (LDS) releases, and two annual Microdata Public Use File (PUF) releases, an MCBS Survey File PUF based on the Survey File LDS and an MCBS Cost Supplement File PUF based on the Cost Supplement File LDS.[1] In addition, in response to the emergence of "coronavirus disease 2019" ("COVID-19") in the United States in 2020, the MCBS has also released three out-of-cycle topic-specific Microdata PUFs (MCBS COVID-19 Summer 2020 PUF, MCBS COVID-19 Fall 2020 PUF, and MCBS COVID-19 Winter 2021 PUF).

The content of the MCBS Cost Supplement File PUF is governed by its central focus of serving as a unique source of information on beneficiaries' cost and utilization that cannot be obtained through CMS administrative sources alone. The MCBS Cost Supplement File 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. Disclosure protections have been applied to the file, including de-identification and other methods. As a result, the MCBS Cost Supplement File PUF does 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 Cost Supplement File PUF is not intended to replace the more detailed LDS files but, rather, it makes available a general-use publicly-available alternative that provides the highest degree of protection to the Medicare beneficiaries' PHI.

The main benefits of the MCBS Cost Supplement File PUF are:

1. Increased data access for researchers of the MCBS through a free file download that is consistent with other U.S. Department of Health and Human Services (HHS) public-use survey files.
2. Enhanced potential for policy-related analyses by attracting new researchers and policymakers. Accessing the MCBS LDS can be a significant deterrent due to the associated costs and time, but the MCBS Cost Supplement PUF mitigates these barriers to encourage broader utilization.

This user guide contains information about the 2022 MCBS Cost Supplement File PUF with detailed information about the MCBS and background to help data users understand and analyze the PUF. This guide is updated each time a new set of PUF data are released.[2]

Readers interested in understanding or analyzing the 2022 MCBS data should also familiarize themselves with the content of the *2022 MCBS Data User's Guide: Cost Supplement File,* the *2022 MCBS Data User's Guide: Survey File,* and the *2022 MCBS Methodology Report* documents in order to obtain 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 (including the new standalone MCBS Glossary) at: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks. To access topical public use file (PUF) table packages based on MCBS data, data users can visit: https://data.cms.gov/medicare-current-beneficiary-survey-mcbs. Data users interested in a collection of charts

[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.

and tables presenting estimates from the LDS releases can access the *MCBS Chartbook* at: https://chartbook.mcbs.org/.

What's New in 2022?

1.1 What's New in the+ 2022 MCBS Cost Supplement File PUF?

There are no updates to highlight for the 2022 MCBS Cost Supplement File PUF. Detailed information about changes to the 2022 MCBS, sampling, questionnaires, documentation, and data processing is available in the *2022 MCBS Data User's Guide: Survey File.*

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 below 65 with certain disabling conditions, residing in the United States. Most interviews were traditionally conducted in-person in households and facilities using computer-assisted personal interviewing (CAPI). However, due to the COVID-19 pandemic, data collection switched to phone-only interviews in March 2020 and throughout most of 2021 with a gradual return to some in-person interviewing beginning in November 2021. MCBS data collection will include both in-person and phone interviewing going forward. Fieldwork for the first round of data collection began in September 1991; since then, the MCBS has continued to collect and provide essential data on the costs, use, and health care status of Medicare beneficiaries. The MCBS has been continuously conducted 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.

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.

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 2022 MCBS Cost Supplement File PUF is primarily based on the Person Summary (PS) segment of the 2022 MCBS Cost Supplement File LDS and includes data for 6,621 sampled beneficiaries. This file also includes survey weights that allow for analysis that is nationally representative of the population of beneficiaries living in the community who were ever enrolled in Medicare at any point in 2022.

All records begin with a PUF_ID, a unique number for each beneficiary in the PUF. This PUF_ID serves to identify records in the 2022 MCBS Cost Supplement File PUF and cannot be used for linking to MCBS data files other than the MCBS Cost Supplement File 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 Cost Supplement File PUF are in numeric or integer formats. Formats and values for each variable are available in the 2022 MCBS Cost Supplement File PUF codebook.

Exhibit 3.1.1 includes information about the variables grouped by topic area and the locations of the corresponding variables in the LDS data segments (i.e., the 2022 LDS Cost Supplement File or the 2022 LDS Survey File individual files).

Exhibit 3.1.1: 2022 MCBS Cost Supplement File PUF Variable Group, Number of Variables, and Related LDS Segments

MCBS Cost Supplement File PUF Variable Group	Number of PUF Variables	LDS Files Data Segments
Demographics and health factors	5	Survey File- DEMO, CHRNCOND
Adjusted utilization (number of events) by service type	9	Cost Supplement File- PS
Adjusted cost by service type	8	Cost Supplement File- PS
Adjusted cost by payer	8	Cost Supplement File- PS
Cost Supplement File PUF ever enrolled weights	101	Cost Supplement File- CSEVWGTS

3.2 Data File Information

Detailed information about variables in the MCBS Cost Supplement File PUF can be found in the Cost Supplement File PUF codebook. The codebook includes SAS® variable names, labels, and any applicable notes. Certain variables in the 2022 MCBS Cost Supplement File PUF were recoded due to disclosure concerns so the categories in the 2022 MCBS Cost Supplement File 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 (e.g., CSP_NCHRNCND).

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 Cost Supplement File PUF are provided in the accompanying codebook file.

The MCBS Cost Supplement File PUF dataset is saved as a SAS export file. Directions and sample SAS code are given below and also in Appendix A to help users read the dataset into SAS.

Assume the MCBS 2022 Cost Supplement File 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 PUF. The following SAS code can then be used to import the Cost Supplement File 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 2022 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.[4] In addition, unlike the 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. Detailed information about the contents of each level of summarization and the payment adjustments is available in the *2022 MCBS Data User's Guide: Cost Supplement File*.

In addition, the 2022 Cost Supplement File PUF differs from the MCBS Cost Supplement File LDS because it has been evaluated for disclosure risk and additional steps were taken to protect beneficiary confidentiality. The 2022 MCBS Cost Supplement File PUF contains data for 6,621 sampled beneficiaries and 134 variables, which is similar to the number of beneficiaries contained in the PS segment of the 2022 MCBS Cost Supplement File LDS, with a few additional demographic variables from the 2022 MCBS Survey File LDS to support analysis. LDS variables that posed a disclosure risk were dropped or recoded to create the variable set for the MCBS Cost Supplement File PUF. In addition, unlike the MCBS Cost Supplement File LDS, which must be linked to the MCBS Survey File LDS for analysis, the 2022 MCBS Cost Supplement File PUF is a stand-alone file that cannot be linked to the 2022 MCBS Survey File PUF for analysis.

Due to disclosure concerns, the 2022 MCBS Cost Supplement File PUF includes only beneficiaries living in the community the entire year. The file excludes 1,263 beneficiaries who had a Facility interview during the year

[4] The 2022 MCBS Cost Supplement File PUF includes utilization and expenditure data on 8 of the 11 service types included in the 2022 MCBS Cost Supplement LDS: Dental (DU), home health (HH), hearing (HU), inpatient hospital (IP), medical provider (MP), outpatient hospital (OP), prescribed medicine (PM), and vision (VU). The MCBS Cost Supplement PUF excludes facility (FA), hospice (HP), and institutional (IU) services. The MCBS Cost Supplement PUF also includes expenditure data on all 11 payers included in the MCBS Cost Supplement LDS: all payers, Medicare Fee-for-Service (FFS), Medicaid, Medicare Managed Care Organization (MCO)/Health Maintenance Organization (HMO), private MCO/HMO, employer-sponsored private insurance, self-purchased private insurance, private insurance (unknown purchase), out-of-pocket, uncollected liability/discounted amount, and other sources (including the Veteran's Administration).

or who incurred any facility, hospice, or institutional events or costs during the year. Variables that were only created for Facility residents are excluded.[5]

The MCBS Cost Supplement File PUF is 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 two data products is presented in Exhibits 3.3.1.a and 3.3.1.b.

Exhibit 3.3.1.a: Comparison between the 2022 MCBS Cost Supplement File PUF and 2022 MCBS Cost Supplement File LDS

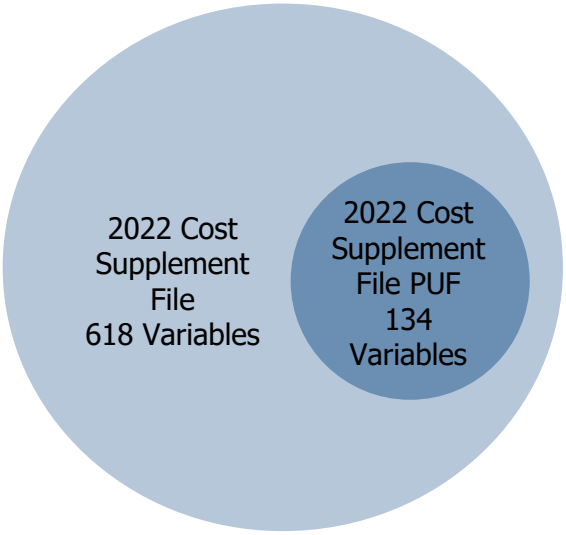


Exhibit 3.3.1.b: Comparison between the 2022 MCBS Cost Supplement File PUF and 2022 MCBS Cost Supplement File LDS

Domain	MCBS Cost Supplement File PUF	MCBS Cost Supplement File LDS
Population	Community only	Community and facility
Number of variables	134 variables	618 variables across 14 data segments
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
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

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

Exhibit 3.3.2 details the socio-demographic and health factor variables along with the values available in the MCBS Cost Supplement File PUF and the MCBS Survey File LDS for comparison purposes. Variables in the MCBS Survey File LDS without an equivalent in the MCBS Cost Supplement File PUF are noted below in the exhibit.[6,7]

Exhibit 3.3.2: Comparison of Socio-demographic and Health Factor Variables and Values in the MCBS Cost Supplement File PUF and MCBS Survey File LDS

Socio-demographic Characteristic or Health Factor	MCBS Cost Supplement File PUF Variables and Values	MCBS Survey File LDS Variables and Values
Gender	CSP_SEX (Gender): Male; Female	ROSTSEX (Gender): Male; Female
Age	CSP_AGE (Age group): <65 years; 65-74 years; ≥75 years	D_STRAT (MCBS Sample age stratum): 0-44 years; 45-64 years; 65-69 years; 70-74 years; 75-79 years; 80-84 years; ≥ 85 years H_AGE (Age of beneficiary): Age of beneficiary in years
Race/Ethnicity	CSP_RACE (Race/ethnicity group): White non-Hispanic; Black non-Hispanic; Hispanic; Other	D_RACE2 (Race of beneficiary): Asian; African American; Native Hawaiian or Pacific Islander; White; American Indian or Alaska Native; More than one RACEAS: Asian; RACEASAI: Asian Indian; RACEASCH: Chinese; RACEASFI: Filipino; RACEASJA: Japanese; RACEASKO: Korean; RACEASVI: Vietnamese; RACEASOT: Other Asian; RACEAA: Black or African-American; RACENH: Native Hawaiian or Pacific Islander; RACEPIHA: Native Hawaiian; RACEPIGU: Guamanian Chamorro; RACEPISA: Samoan; RACEPIOT: Other Pacific Islander; RACEWH: Caucasian; RACEAI: American Indian or Alaska Native HISPORIG (Is beneficiary of Hispanic or Latino origin?): Yes; No HISPORMA: Mexican/Mex American/Chicano; HISPORPR: Puerto Rican; HISPORCU: Cuban; HISPOROT: Other Hispanic/Latino/Spanish origin

[6] The MCBS Survey File LDS contains additional socio-demographic information, including educational attainment, metropolitan area residence status and location of residence, rural-urban commuting area details, marital status, the number of living children the beneficiary has, employment status, veteran status, status of Social Security Administration (SSA) check, interview language, language spoken at home, and English proficiency, Area Deprivation Index (ADI), and Income Poverty Ratio (IPR), which do not have corresponding variables available in the MCBS Cost Supplement File PUF. Please note that additional race/ethnicity variables from administrative sources are included in the MCBS Survey File LDS.

[7] As well as additional socio-demographic and health factor variables, the Survey File LDS contains data on a wide variety of other topics that can be combined with payer, cost, and utilization data in the Cost Supplement File LDS for analysis, including health status and functioning, health insurance coverage, health behaviors, preventive care, financial resources, and access to, knowledge of, attitudes towards, and satisfaction with care. See the LDS Data User's Guides for additional information.

Socio-demographic Characteristic or Health Factor	MCBS Cost Supplement File PUF Variables and Values	MCBS Survey File LDS Variables and Values
Household Income	CSP_INCOME (Income of beneficiary and spouse): <$25,000; ≥$25,000	INCOME (Income range of beneficiary and spouse): <$5,000; $5,000 - $9,999; $10,000 - $14,999; $15,000 - $19,999; $20,000 - $24,999; $25,000 - $29,999; $30,000 - $39,999; $40,000 - $49,999; $50,000 - $59,999; $60,000 - $79,999; $80,000 - $99,999; $100,000 - $119,999; $120,000 - $139,999; ≥$140,000 INCOME_H (Beneficiary and spouse total income last year): Range of values
Chronic Conditions	CSP_NCHRNCND (Number of chronic conditions): 0-1, 2-3, 4+ conditions	OCHBP (High blood pressure), OCMYOCAR (Myocardial infarction), OCCHD (Coronary heart disease (CHD)/Angina pectoris), OCCFAIL (Congestive heart failure), OCHRTCND (Valve/rhythm/other heart condition), OCSTROKE (Stroke), OCCANCER (Cancer), OCBETES (Diabetes/High blood sugar), OCALZMER (Alzheimer's disease), OCDEMENT (Dementia), OCDEPRSS (Depression), OCPSYCHO (Mental disorder), OCMENTAL (Intellectual disability), OCARTHRH (Rheumatoid arthritis), OCOSARTH (Osteoarthritis), OCARTHOT (Other arthritis), OCOSTEOP (Osteoporosis), OCBRKHIP (Broken hip), OCPARKIN (Parkinson's disease), OCEMPHYS (Emphysema/Asthma/Chronic obstructive pulmonary disease (COPD)), OCCHOLES (High cholesterol): Yes; No

3.4 Top- and Bottom-coding

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

4. SURVEY OVERVIEW

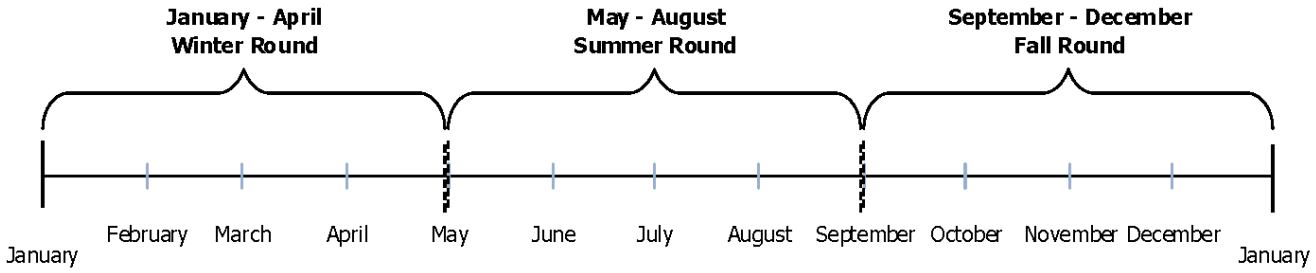
4.1 Design of MCBS

In its initial design, the MCBS was intended to serve as a traditional longitudinal survey of the Medicare population. There was no predetermined limit to the duration of time a beneficiary, once selected to participate, was to remain in the sample. However, beginning in 1994, participation of beneficiaries in the MCBS was limited to no more than four years.

Although participation in the survey is limited to four years, MCBS data collection is continuous 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 to round design is to create shorter recall periods during the year to capture more complete and accurate health care costs and utilization for beneficiaries.

The 2022 MCBS data releases reflect data collected from January 2022 through December 2022 (see Exhibit 4.1.1), as well as Topical sections, income and assets data, and chronic pain information collected through the Winter and Summer 2023 rounds.[8] 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

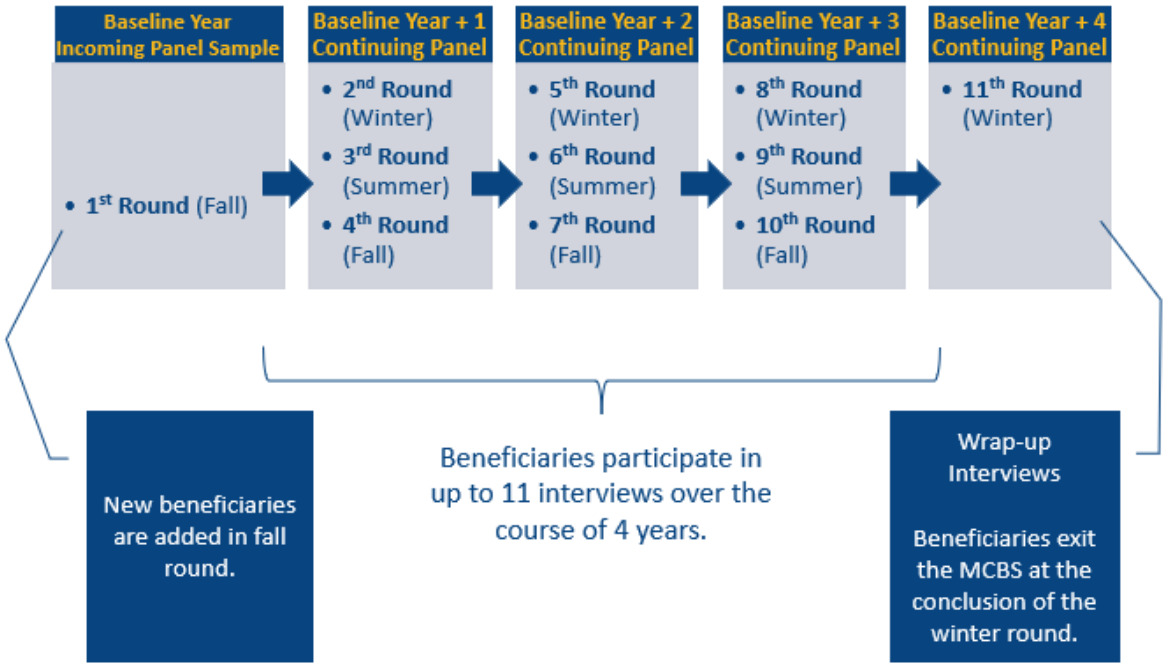


The initial interview of newly-selected beneficiaries takes place in the fall round. Since 2016, the fall round begins early (in late July or early August) to allow more time to locate and conduct outreach to the new survey respondents.

Subsequent rounds, which occur every four months, involve the re-interviewing of the same beneficiary (or appropriate proxy respondents or facility staff) over a four year period (up to 11 interviews in total). Exhibit 4.1.2 depicts the timeline of participation for beneficiaries selected to be in the MCBS sample.

[8] 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 future 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 survey year.[9] 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.[10] 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.

[9] 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. In 2017, sampling from Puerto Rico was discontinued. In 2018, all data collection in Puerto Rico was discontinued.

[10] 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: 2019-2022 MCBS Rotating Panel Design

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

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 2022 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 Cost Supplement File PUF 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 2022.[11] The MCBS Cost Supplement File PUF represents four separate MCBS panels identified by the year in which the panel was selected and first interviewed (i.e., for the 2022 MCBS Cost Supplement File PUF, the 2019, 2020, 2021, and 2022 Panels). Exhibit 4.2.2 shows the distribution of each of the four panels included in the 2022 MCBS Cost Supplement File PUF.

For more information on the sample design, please see the *Cost Supplement File Data User's Guide* at: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

Exhibit 4.2.2: 2022 MCBS Composition of Panels in the MCBS Cost Supplement File PUF

Data Year (Fall)	Number of Beneficiaries Selected
2019	1,676
2020	2,097
2021	2,629
2022	219

[11] While beneficiaries included in the LDS releases represent both the ever enrolled and continuously enrolled Medicare population, the MCBS Cost Supplement File PUF solely represents the ever enrolled population.

4.3 Eligibility

4.3.1 Medicare Population Covered by the 2022 LDS and MCBS Cost Supplement File PUF

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.

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 any Facility interviews were conducted during the data collection period, or who accrued any Facility events or costs during the data year, 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.

4.4.1 Incoming and Continuing Cases

Every fall round of data collection, a new panel of sampled beneficiaries is added to the total sample to replace the panel of sampled beneficiaries completing a final interview and exiting the MCBS in the prior winter round. Newly selected MCBS beneficiaries begin in the fall round and are referred to as Incoming Panel cases. After the initial interview, they are referred to as Continuing cases.

4.4.2 Community Interviews

Approximately 93 percent of the interviews are conducted with beneficiaries or proxies in their own residence or in a neutral interview location, such as a library or public venue. These interviews are called Community interviews; the remaining 7 percent of the interviews are administered for beneficiaries living in a facility, and these beneficiaries are not included in the MCBS Cost Supplement File PUF.

Over the course of a four-year period, however, it is not uncommon for beneficiaries to enter long-term care facilities (e.g., nursing homes) or to go back and forth between community and facility settings. In order to obtain an accurate representation of the experiences of all Medicare beneficiaries, the MCBS includes beneficiaries wherever they reside, even if they enter or reside in a facility for the duration of their four years with the study. The MCBS Cost Supplement File PUF excludes beneficiaries who were in a facility at any point in 2022 due to disclosure concerns.

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.

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 an average of approximately 170 field interviewers[12] who participate in a combination of several targeted training initiatives and careful coaching and monitoring activities throughout data collection.

The 2022 MCBS Training Program consisted of remote and in-person trainings which varied based on the level of experience of the interviewer (new to MCBS or MCBS-experienced), the type of interview component (Community or Facility), the sample type (Incoming Panel or Continuing), and the season-specific requirements (new or changing questionnaire sections or data collection protocols). The program was structured to expose all field staff to the same training content, ensuring that the performance of data collection responsibilities was standardized, methodical, and measurable. In addition to formal trainings, throughout data collection, the MCBS Training Program emphasized proper protocols through continuous quality improvement, featuring skill specialization, reinforcement of key behavior, and targeted messaging to boost interviewer performance. To meet all interviewers' skill-building and training needs, NORC continued to work with field managers to ensure interviewers received additional training 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 covered important data collection tips, provided answers to interviewer questions, and offered 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.

[12] The fall round starts with a higher number of field interviewers which, over the course of the year, is reduced due to staff turnover. Each summer, a small cohort of new interviewers is hired for the MCBS.

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 139511) of the Social Security Act and is conducted by NORC at the University of Chicago for the U.S. Department of Health and Human Services. The OMB Number for this survey is 0938-0568.

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 varies by interview type and data collection season (fall, winter, or summer). There are two types of interviews (Baseline and Continuing) containing two types of questionnaire sections (Core and Topical). See Exhibit 5.1 within the *Survey File Data User's Guide* 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 are of critical purpose and policy relevance to the MCBS, regardless of season of administration. Core sections collect information on beneficiaries' health insurance coverage, health care utilization and costs, and operational management data such as locating information.
- Topical: These sections collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

See Sections 5.2 and 5.3 in the *Survey File Data User's Guide* for additional detail on the 2022 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 2022 MCBS Cost Supplement File PUF

The MCBS data releases are a reflection of enrolled Medicare beneficiaries residing in the continental United States.[13] The sample for the MCBS is drawn from a subset of the Medicare enrollment data, which is a list of all Medicare beneficiaries. Excluded are residents of foreign countries and U.S. possessions and territories. The MCBS Cost Supplement File PUF further excludes Medicare beneficiaries who provided any Facility interviews during the data 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 2022 as well as beneficiaries who dis-enrolled or died prior to their fall interview.[14] The ever enrolled population includes beneficiaries who were enrolled in Medicare for at least one day at any point during 2022.
- The continuously enrolled population represents only individuals continuously enrolled in Medicare from January 1, 2022 up to and including their fall interview; this specifically excludes beneficiaries who enrolled during the calendar year 2022 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 Cost Supplement File PUF and the MCBS Cost Supplement File LDS releases include weights that represent only the ever enrolled population. The 2022 MCBS Cost Supplement File PUF and MCBS Cost Supplement File LDS ever enrolled population represents a subset of beneficiaries with complete cost and utilization data for the year.

Exhibits 6.1.1 and 6.1.2 present estimates of the size of the ever enrolled Medicare population living in the community by race/ethnicity[15] and age (as of December 31, 2022), by sex in the 2022 MCBS Cost Supplement File PUF. Exhibit 6.1.3 presents the aggregated estimates of the size of the ever enrolled Medicare population living in the community overall and by sex and race/ethnicity.

Exhibit 6.1.1: Estimated Number of Male Community Medicare Beneficiaries by Race/Ethnicity and Age in the 2022 MCBS Cost Supplement File PUF*

Race/Ethnicity	Age as of 12/31/2022	Weighted Count
White non-Hispanic	Under 65 years	2,292,378
	65-74 years	10,808,309
	75+ years	7,449,602

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

[14] Note that data collection for beneficiaries who enrolled during 2022 and died in 2022 after enrollment but before their fall interview was still pursued through attempts at conducting proxy interviews.

[15] Hispanic origin and race are two separate and distinct categories. Persons of Hispanic origin may be of any race or combination of races. Hispanic origin includes persons of Mexican, Puerto Rican, Cuban, Central and South American, or Spanish origin. For the MCBS, responses to beneficiary race and ethnicity questions are reported by the respondent. More than one race may be reported. For conciseness, the text, tables, and figures in this document use shorter versions of the terms for race and Hispanic or Latino origin specified in the Office of Management and Budget 1997 Standards for Data on Race and Ethnicity. Beneficiaries reported as White and not of Hispanic origin were coded as White non-Hispanic; beneficiaries reported as Black/African American and not of Hispanic origin were coded as Black non-Hispanic; beneficiaries reported as Hispanic, Latino/Latina, or of Spanish origin, regardless of their race, were coded as Hispanic. The "Other" race category includes other single races not of Hispanic origin (including American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander), Two or More Races, or Unknown Races.

Race/Ethnicity	Age as of 12/31/2022	Weighted Count
Black non-Hispanic	Under 65 years	666,318
	65-74 years	1,232,669
	75+ years	626,294
Hispanic	Under 65 years	435,843
	65-74 years	1,073,686
	75+ years	572,152
Other†	Under 65 years	284,244
	65-74 years	907,649
	75+ years	474,219

SOURCE: 2022 MCBS Cost Supplement File PUF, weighted counts.

*Weighted counts may not sum to the total of beneficiaries living in the community in the U.S. due to missingness.
†The "Other" race/ethnicity category includes other single races not of Hispanic origin, Two or More Races, and Unknown Races.

Exhibit 6.1.2: Estimated Number of Female Community Medicare Beneficiaries by Race/Ethnicity and Age in the 2022 MCBS Cost Supplement File PUF*

Race/Ethnicity	Age as of 12/31/2022	Weighted Count
White non-Hispanic	Under 65 years	1,943,053
	65-74 years	12,679,555
	75+ years	9,671,196
Black non-Hispanic	Under 65 years	692,374
	65-74 years	1,883,969
	75+ years	1,008,915
Hispanic	Under 65 years	481,965
	65-74 years	1,522,531
	75+ years	945,740
Other†	Under 65 years	265,384
	65-74 years	916,353
	75+ years	654,813

SOURCE: 2022 MCBS Cost Supplement File PUF, weighted counts.

*Weighted counts may not sum to the total of beneficiaries living in the community in the U.S. due to missingness.
†The "Other" race/ethnicity category includes other single races not of Hispanic origin, Two or More Races, and Unknown Races.

Exhibit 6.1.3: Estimated Number of Community Medicare Beneficiaries by Race/Ethnicity and Sex in the 2022 MCBS Cost Supplement File PUF*

Group	Subgroup	Weighted Count
Overall Total		59,489,209
Sex	Male Total	26,823,361
	Female Total	32,665,848
Race/Ethnicity	White non-Hispanic Total	44,844,092
	Black non-Hispanic Total	6,110,539
	Hispanic Total	5,031,917
	Other Total†	3,502,661

SOURCE: 2022 MCBS Cost Supplement File PUF, weighted counts.

*Weighted counts may not sum to the total of beneficiaries living in the community in the U.S. due to missingness.
†The "Other" race/ethnicity category includes other single races not of Hispanic origin, Two or More Races, and Unknown Races.

6.2 Targeted Population and Sampling Strata

The targeted population for the MCBS consisted 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 Fall Rounds 2019, 2020, 2021, and 2022 (the four rounds in which the 2019, 2020, 2021, and 2022 Panels, included in the 2022 MCBS data, were selected), the targeted population included individuals enrolled as of December 31 of 2019, 2020, 2021, and 2022 respectively.

Additionally, in the 2019, 2020, 2021, and 2022 Panels, beneficiaries residing within the U.S. who were Hispanic (based on a Hispanic ethnicity classification code in the Medicare enrollment data; see Eicheldinger[16] for more details) were oversampled to improve precision of estimates for this group.[17] For more information on the sampling strata, please see Section 6 of the *Survey File Data User's Guide*: https://www.cms.gov/data-research/research/medicare-current-beneficiary-survey/data-documentation-codebooks.

Exhibit 6.2.1 displays the number of beneficiaries included in the 2022 MCBS Cost Supplement File PUF, by age and ethnicity.

Exhibit 6.2.1: 2022 Panel of Selected Beneficiaries by U.S. Hispanic and U.S. Non-Hispanic Ethnicity Classification and Age Category*

Age Category as of 12/31/2022	TOTAL Sample Size	TOTAL Weighted	Hispanic Sample Size	Hispanic Weighted	Non-Hispanic Sample Size	Non-Hispanic Weighted
Under 65 years	1,093	7,061,558	135	917,808	958	6,143,749
65-74 years	2,378	31,024,721	247	2,596,217	2,131	28,428,504
75+ years	3,150	21,402,930	296	1,517,891	2,854	19,885,039
Total	6,621	59,489,209	678	5,031,917	5,943	54,457,292

SOURCE: 2022 MCBS Cost Supplement File PUF.

*Weighted counts may not sum to the total of beneficiaries living in the community in the U.S. due to missingness.

6.3 Primary and Secondary Sampling Units

All of the panels in the 2022 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.

[16] Celia Eicheldinger and Arthur Bonito, "More Accurate Racial and Ethnic Codes for Medicare Administrative Data," *Health Care Financing Review* 29, no. 3 (2008): 27-42.

[17] 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.

6.4 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.[18] For each Continuing beneficiary, the survey questions corresponding to the Cost Supplement File data release are administered in all three rounds of the data collection year. For beneficiaries new to the MCBS, the survey questions are administered as part of the initial fall Baseline interview.

[18] The MCBS 2022 Panel was 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.

The MCBS Cost Supplement File PUF includes ever enrolled cross-sectional weights (CSPUFWGT) which apply to both the Continuing sample (beneficiaries sampled between 2019-2021) and to the Incoming Panel sample (beneficiaries sampled in 2022). These weights are intended for use in cross-sectional statistics involving the total (combined) Fall 2022 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 2022 (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.

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.

7.2 Item Non-Response

As in any other survey, some respondents could not, or would not, supply answers to some questions.[19] 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.[20] The calculation of the study-wide response rates generally follows the guidelines specified in the American Association for Public Opinion Research (AAPOR) and OMB. For the ever enrolled sample represented by the MCBS 2022 Cost Supplement File, the calculated overall response rate was 64.0 percent. This rate includes non-response for persons in facilities, as the response rates are not calculated separated by questionnaire component. Therefore, this may not reflect exactly the response rate for the sample represented in the 2022 MCBS Cost Supplement File PUF, which excludes beneficiaries for whom any Facility interview was conducted during the data collection year.

[19] 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.

[20] 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.

7.3 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.

8. REFERENCES

Eicheldinger, Celia, and Arthur Bonito. "More accurate racial and ethnic codes for Medicare administrative data." *Health care financing review* 29, no. 3 (2008): 27-42.

Lo, A, A Chu, and R Apodaca. "Redesign of the Medicare Current Beneficiary Survey Sample." *Proceedings of the Survey Research Section of the American Statistical Association* (2002): 2139-44.

APPENDICES

9. APPENDICES

Appendix A: Technical Appendix – 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)
```