

---

# MCBS Tutorial

Introduction for New Users

## Version Control Log

Date	Version	Revisions
12/08/2017	1.0	Initial version released based on data year 2015.
05/30/2018	1.1	Added Analytic Example 2 using 2015 Cost Supplement data. Corrected standard errors for Analytic Example 1.1.
01/17/2019	2.0	Content updated with information for data year 2016.
01/15/2020	3.0	Content updated with information for data year 2017.
11/04/2020	4.0	Content updated with information for data year 2018 and other updates through calendar year 2020.
11/23/2021	5.0	Content updated with information for data year 2019 and other updates through calendar year 2021.

# Section 1: Introduction





## Introduction to the MCBS

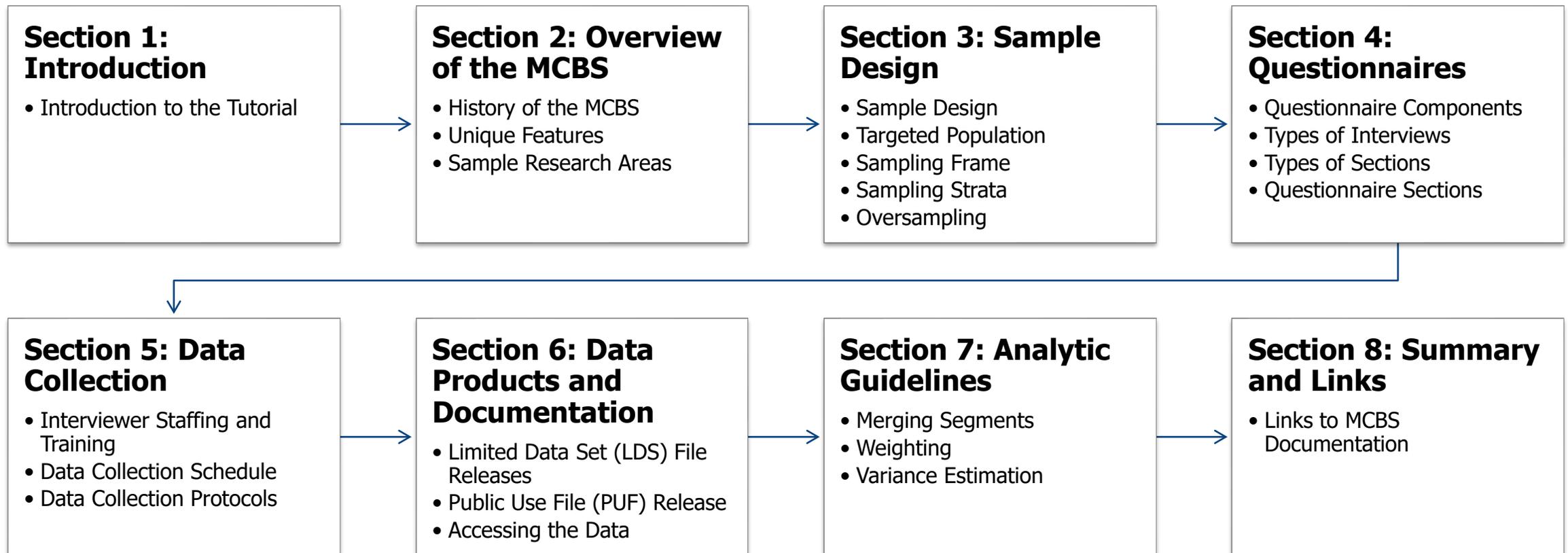
- The Medicare Current Beneficiary Survey (MCBS) is a continuous, multi-purpose longitudinal survey.
- The MCBS represents the population of Medicare beneficiaries aged 65 and over and beneficiaries aged 64 and under with certain disabling conditions living in the United States.
- The MCBS is sponsored by the Office of Enterprise Data and Analytics (OEDA) of the Centers for Medicare & Medicaid Services (CMS) and is conducted through a contract with NORC at the University of Chicago (NORC).
- The MCBS is designed to aid CMS in administering, monitoring, and evaluating the Medicare program. The MCBS is the most comprehensive and complete survey available on the Medicare population and is essential in providing important information on beneficiaries that is not otherwise collected through operational or administrative data from the Medicare program.



# Introduction to the MCBS New User Tutorial

- This tutorial is intended to be a starting point for new users who are unfamiliar with the MCBS. It provides an overview of MCBS history, policy relevance, survey design, data products, and best practices for analysis.
- This tutorial is organized into eight sections, grouped by topic area, which are outlined on the next slide. Although the tutorial is designed to be read in sequence, each section is relatively independent and self-contained.
- More detailed descriptions of the topics discussed in this tutorial can be found in the *MCBS Data User's Guides* that accompany the Limited Data Set (LDS) file releases, the *MCBS Methodology Report*, and other survey documentation provided online, such as the *Questionnaires and Questionnaire User Documentation*.

# MCBS New User Tutorial Outline



# MCBS Documentation and Resources

- CMS provides a wide array of MCBS documentation that is publically available on the CMS MCBS website. This documentation contains more in-depth descriptions of the topics covered in this tutorial.
  - Annual [Questionnaires and Questionnaire User Documentation](#)
  - Data documentation including [Data User's Guides, Methodology Reports, and codebooks](#) for the LDS files
  - Annual [Chartbooks and data tables](#)
  - Annual [Bibliographies](#), which include annotations starting in 2020
  - [Early Looks, topical data briefs, infographics, and tutorials](#), including the *New User Tutorial* and *Advanced Topic-Based Tutorials*
- More information on MCBS documentation and resources can be found in Section 8 of this tutorial.

## CMS Website:

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS>

<b>Medicare Current Beneficiary Survey (MCBS)</b>
<a href="#"><u>Questionnaires</u></a>
<a href="#"><u>Data Documentation and Codebooks</u></a>
<a href="#"><u>Data Tables</u></a>
<a href="#"><u>Bibliography</u></a>
<a href="#"><u>Data Briefs and Tutorials</u></a>

# MCBS Documentation and Resources (continued)

- Additionally, data documentation for the Survey File Public Use File (PUF) is available free for download.
- The PUF is available for 2013 and every data year since 2015.
- The full package available for download includes the following:

**MCBS PUF Website:**  
<https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File>

- Data File
- Codebook
- Documentation
- SAS code

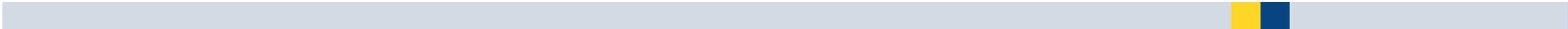


Data File Year	Complete PUF Package	Data File	Codebook	Documentation	SAS® code
2019 Medicare Current Beneficiary Survey File PUF	<a href="#">MCBSPUF19 (ZIP)</a>	<a href="#">PUF2019 (ZIP)</a>	<a href="#">2019MCBSPUF Codebooks (ZIP)</a>	<a href="#">2019MCBSPUF DataUserGuide (PDF)</a>	<a href="#">MCBSPUF19.txt (PLAIN)</a>
2018 Medicare Current Beneficiary Survey File PUF	<a href="#">MCBSPUF18 (ZIP)</a>	<a href="#">PUF2018_XPT (ZIP)</a>	<a href="#">2018MCBSPUFCodebooks (ZIP)</a>	<a href="#">2018MCBSPUFUserGuide (PDF)</a>	<a href="#">MCBSPUF18.txt (PLAIN)</a>
2017 Medicare Current Beneficiary Survey File PUF	<a href="#">MCBSPUF17 (ZIP)</a>	<a href="#">PUF2017_XPT (ZIP)</a>	<a href="#">2017MCBSPUFCodebooks (ZIP)</a>	<a href="#">2017MCBSPUFUserGuide (PDF)</a>	<a href="#">MCBSPUF17.txt (PLAIN)</a>

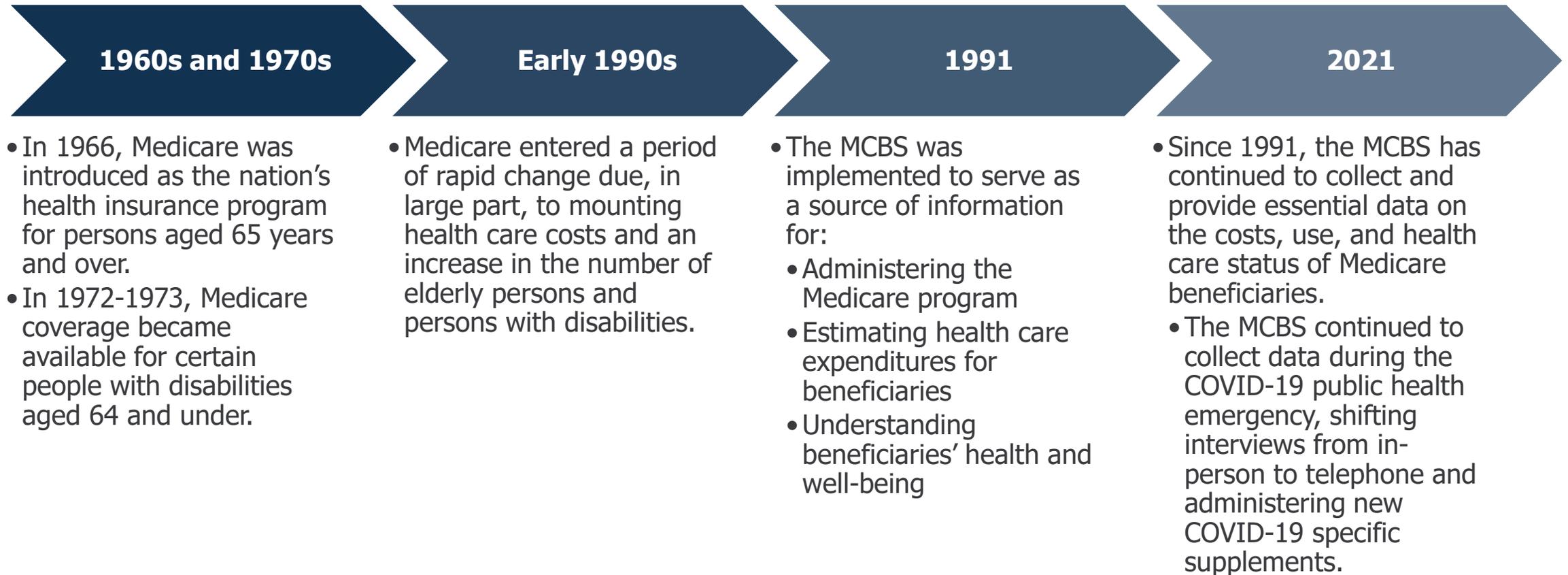
# Key Takeaways

- The MCBS is a continuous, longitudinal survey, representing the population of Medicare beneficiaries in the United States. The MCBS is designed to aid CMS in administering, monitoring, and evaluating the Medicare program.
- This tutorial is intended to be a starting point for new users who are unfamiliar with the MCBS. It provides an overview of MCBS history, policy relevance, survey design, data products, and best practices for analysis.
- In addition to this tutorial, CMS provides a wide range of documentation for researchers and data users that is publically available on the CMS website, including annual *Data User's Guides*, *Methodology Reports*, *Chartbooks*/data tables, *Bibliographies*, *Early Looks*, infographics, and questionnaires and questionnaire user documentation.

# Section 2: Overview of the MCBS



# History of the MCBS



# Unique Features of the MCBS

- The MCBS is distinguished by several unique features, including:
  - 1. Linked self-report and administrative claims data:** The MCBS collects information directly from a sample of beneficiaries to understand their health care needs, costs, and experiences, and then links their responses to administrative claims, which provides researchers and policy analysts with more accurate and complete estimates of beneficiaries' total health care costs and utilization.
  - 2. Complete source of payment information:** The MCBS collects survey-reported data on all sources of payment for health care costs including those not covered by Medicare, such as co-payments, deductibles, and uncovered services. MCBS data also reveal the extent to which retiree benefits, Medigap, VA benefits, and other sources of coverage address costs.
  - 3. Rotating panel design:** The MCBS features a rotating panel design that enables users to conduct longitudinal analysis. The MCBS rotating panel is designed to be representative of the population of all Medicare beneficiaries for the survey year. Each MCBS panel, an annual statistical sample of all Medicare enrollees, is interviewed up to three times a year over a four year period. Beginning in 2018, one panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round.<sup>1</sup>

<sup>1</sup> Prior to 2018, the MCBS included 12 rounds; beginning in 2018, the MCBS design was modified to 11 rounds to reduce respondent burden.

## Unique Features of the MCBS (continued)

- The MCBS is distinguished by several unique features, including:
  4. **Inclusion of beneficiaries living in a facility:** The MCBS follows beneficiaries into and out of long-term care facilities to maintain a comprehensive profile of their health care utilization and expenditures.
  5. **Oversampling:** Hispanic beneficiaries, beneficiaries with disabilities (aged 64 and under), and beneficiaries aged 85 and over are oversampled to ensure sufficient sample for analysis purposes due to interest in their special health care needs.
  6. **Medicare Advantage (MA) beneficiary data:** The MCBS collects comprehensive survey-reported health care utilization and cost data for beneficiaries covered by MA, allowing researchers to analyze MA beneficiary cost and utilization data in the absence of administrative claims data for MA beneficiaries. In addition, starting with data year 2019, the MCBS began to use MA encounter data in imputation procedures to account for medical events and costs.
  7. **Rapid data collection for emerging needs:** In response to the COVID-19 public health emergency, MCBS COVID-19 Rapid Response Supplements were developed and conducted in 2020 and 2021 over the telephone to quickly collect vital information on how the pandemic impacted the Medicare population.

## Where are MCBS Data Used?

- MCBS data and estimates are vital to informing the production of highly visible publications, including:
  - The Medicare Payment Advisory Commission's annual data book "Health Care Spending and the Medicare Program"
  - The Federal Interagency Forum on Aging Related Statistics' chartbook "Older Americans: Key Indicators of Well-Being"
  - The Centers for Medicare & Medicaid Services' Annual Trustees Report
- Articles based on MCBS data are published in a wide array of peer-reviewed journals, including *Health Services Research*, *Journal of General Internal Medicine*, *Health Economics*, *American Journal of Managed Care*, *Health Affairs*, and the *New England Journal of Medicine*.
- MCBS data are also used to produce CMS data briefs on diverse topics, including dental health, diabetes, heart disease, preventive care, and hearing problems, which are made publically available on the CMS website.



## Sample Research Areas for Prospective Data Users

- Given the breadth and depth of MCBS data, the MCBS can be used to support a wide range of analyses.
- Sample research areas include:
  - Impact of payment and delivery system reforms
  - Access to care and care experience
  - Beneficiary awareness and knowledge of Medicare
  - Beneficiary engagement and activation
  - Cognitive and functional status
  - Assessment of behavioral risk factors
  - Trends in out-of-pocket spending
  - Caregiver burden
  - Social determinants of health
  - Supplemental health insurance coverage
  - Preventive care and health screenings
  - Experiences during the COVID-19 pandemic

## Sample Research Areas for Prospective Data Users (continued)

- Populations represented in the data include:
  - Beneficiaries enrolled in traditional Medicare and MA
  - Dual-eligible beneficiaries (beneficiaries eligible for both Medicare and Medicaid)
  - Newly enrolled beneficiaries and decedents
  - Beneficiaries living in the community and in long-term care facilities
- Different types of data include:<sup>1</sup>
  - Survey-reported data
  - Medicare Fee-for-Service (FFS) claims
  - Medicare Part D claims
  - Assessments [Minimum Data Set (MDS); Outcome and Assessment Information Set (OASIS)]
  - Medicare Administrative enrollment data

<sup>1</sup> MA encounter data are not included in the MCBS but are used for the imputation of medical events and costs for beneficiaries enrolled in MA, starting with the 2019 data year.

## Overview of the MCBS COVID-19 Supplements

- With the emergence of COVID-19 in the U.S., CMS was uniquely positioned to quickly collect vital information on how the pandemic impacted the Medicare population by using the MCBS as a vehicle to collect data.
- The MCBS COVID-19 Rapid Response Supplements collected data on topics such as the availability of telemedicine visits, deferred medical care, social distancing and other preventive health behaviors, COVID-19 testing and vaccination, and the consequences for social, emotional, and financial well-being.
  - Three COVID-19 Community Supplements were administered as standalone surveys in Summer 2020, Fall 2020, and Winter 2021 for beneficiaries living in the community. After Winter 2021, a subset of COVID-19 Community Supplement items were integrated into the main MCBS questionnaire instrument.
  - Two COVID-19 Facility Supplements were administered in Fall 2020 and Winter 2021 within the main MCBS Facility Instrument for beneficiaries living in facilities.
- This tutorial focuses on the main MCBS; for more information on the COVID-19 Supplements, please see the corresponding *Data User's Guides* and the *Advanced Tutorial on the COVID-19 Supplement Data*.



## Key Takeaways

- The MCBS was implemented in 1991 to serve as a source of information for administering the Medicare program, estimating health care expenditures for beneficiaries, and better understanding the health and well-being of Medicare beneficiaries.
- The MCBS is distinguished by several unique features, including linked self-reported and administrative claims data, complete source of payment information, and the inclusion of beneficiaries living in a facility.
- MCBS estimates are used in producing a variety of highly-regarded publications, peer-reviewed articles, and data briefs.
- MCBS data can be used to support a wide range of analyses, including the impact of payment and delivery system reforms, access to care and care experience, beneficiary awareness and knowledge, and beneficiary engagement and activation.

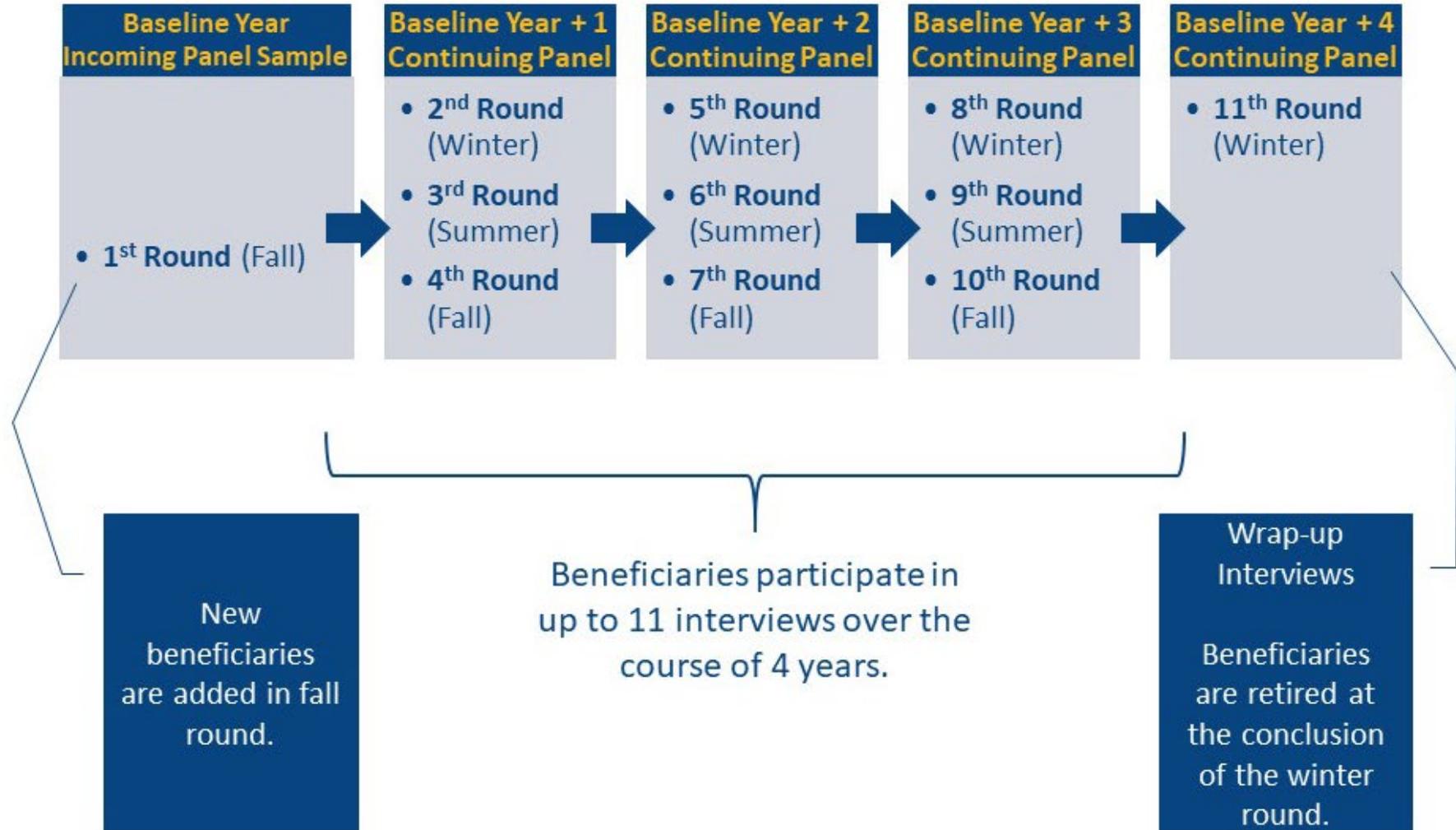
# Section 3: Sample Design



## Sample Design Overview

- The MCBS uses a rotating panel sample design, and is designed to be representative of the population of all Medicare beneficiaries for the survey year. The MCBS sample is selected from locations throughout the continental U.S. and the District of Columbia (excludes Alaska and Hawaii).
  - Although MCBS data are nationally representative, they are not representative at the regional or state level and cannot be used to produce regional or state-level estimates.
- Each MCBS panel, an annual statistical sample of all Medicare enrollees, is interviewed up to three times per year over a four year period creating a continuous profile of selected beneficiaries' health care experiences.
- The three rounds of data collection per year are referred to seasonally (fall, winter, summer). One panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round. The new panel is referred to as the Incoming Panel.
- The MCBS sample is designed to yield about 14,500 completed cases annually in the MCBS Survey File LDS release and about 9,000 completed cases annually in the MCBS Cost Supplement File LDS release.

# MCBS Participation Timeline



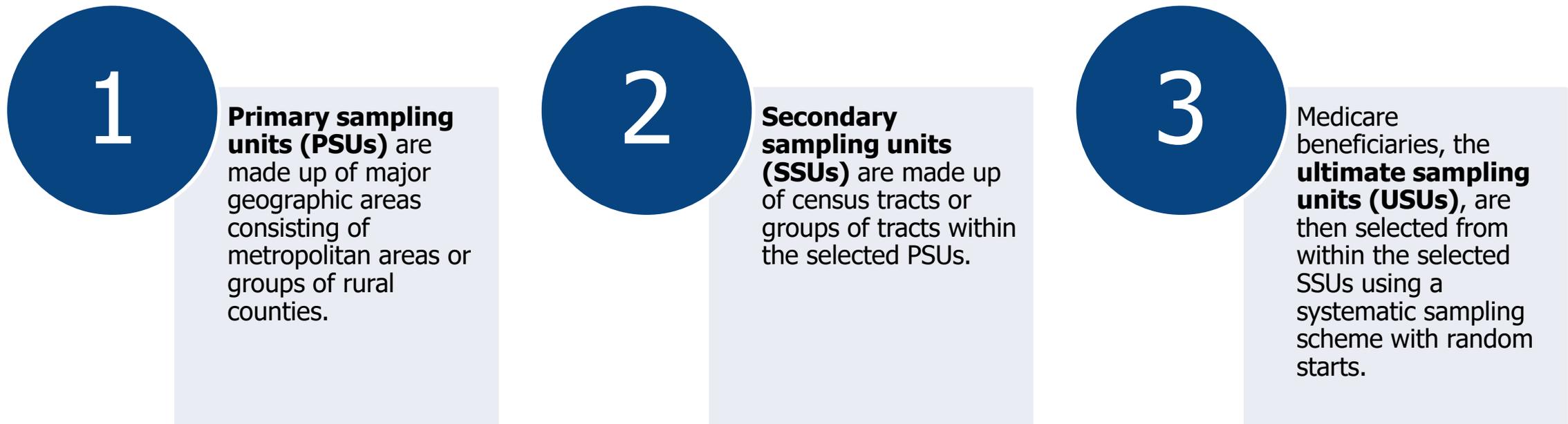
# MCBS Data Year Cycle, 2015 – 2019

Data Collection Schedule			Panel				
Calendar Year	Season	Round#	2015	2016	2017	2018	2019
2015	Winter/Summer*	71/72					
	Fall	73	Yellow				
2016	Winter	74	Yellow				
	Summer	75		Blue			
	Fall	76		Blue			
2017	Winter	77	Yellow	Blue			
	Summer	78		Blue	Light Blue		
	Fall	79		Blue	Light Blue		
2018	Winter	80	Yellow	Blue	Light Blue		
	Summer	81		Blue	Light Blue		
	Fall	82		Blue	Light Blue	Green	
2019	Winter	83	Yellow	Blue	Light Blue	Green	
	Summer	84		Blue	Light Blue	Green	
	Fall	85		Blue	Light Blue	Green	Yellow

\*The Summer and Winter Rounds in 2015 were combined due to contract transition.

# Sample Design

The MCBS employs a three-stage cluster sample design:





## Targeted Population

- The targeted population for the MCBS consists of Medicare Part A and/or Part B enrollees as of December 31 of the sample-selection year.
  - This allows the MCBS to include current-year enrollees (i.e., those who enroll in Medicare between January 1 and December 31 of the sample-selection year) in that year's Incoming Panel.



## Sampling Frame

- The sampling frame for the Medicare beneficiaries utilizes Medicare Administrative enrollment data.
- Medicare Administrative enrollment data contain information on all individuals entitled to Medicare, including demographic information, enrollment dates, third party buy-in information, and Medicare managed care enrollment.
- Every year, CMS specifies a 5-percent sample of the administrative enrollment data that is used for selecting the MCBS sample. The first extract of the 5-percent file is made available in the spring, and the bulk of the sample is selected from that extract. Additional extracts of the 5-percent file are utilized to support sampling of current-year enrollees. These extracts contain new enrollees who were not included in the initial extract. The combination of these extracts constitutes the full frame from which the Incoming Panel is selected.



## Sampling Strata

- The universe of beneficiaries for the MCBS is divided into seven sampling strata based on age as of December 31 of the sampling year in order to include all beneficiaries enrolling during the sampling year. The age categories are: under 45, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 or over. The strata were expanded to separate Hispanic and non-Hispanic beneficiaries by age group.
- Beginning in 2017, Puerto Rico was removed from the MCBS sample. All data collection in Puerto Rico was discontinued beginning in 2018. Hispanic sampling rates in the continental U.S. have been increased to make Hispanic estimates more representative of the Hispanic population residing outside of Puerto Rico.
- The MCBS sampling design provides nearly self-weighting (i.e., equal probabilities of selection) samples of beneficiaries within each of the fourteen sampling strata.

## Sampling Strata

- Since the 2017 data year, the MCBS sample is selected within the following fourteen strata.

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



## Oversampling

- Elderly beneficiaries (aged 85 and over) and beneficiaries with disabilities (aged 64 and under) are oversampled to permit more detailed analysis of these subpopulations due to interest in their special health care needs.
- Oversampling of Hispanics is also conducted as part of the beneficiary selection process. Beginning in 2015 through 2018, an additional oversample of Hispanic beneficiaries was implemented to allow for more precise estimates of health disparities experienced by these populations and to increase the proportion of MCBS Hispanic beneficiaries, with the goal of achieving 1,500 annual completed interviews with Hispanic beneficiaries across all panels by the 2018 Survey File LDS. Oversampling of Hispanics relative to their non-Hispanic counterparts continues into 2019 and beyond, but the additional oversample is discontinued.

## Eligibility for Participation in the MCBS

- Eligibility for the MCBS varies overall by date of death, lost entitlement, and location.
  - **Date of Death:** For beneficiaries who become deceased after their second interview, proxy or facility administrators are interviewed to collect data from the date of the last interview up to the date of death. However, Incoming Panel beneficiaries and second-round Continuing beneficiaries who become deceased may be ineligible for the MCBS depending on their date of death and the year they enrolled in Medicare.
  - **Lost Entitlement:** CMS provides updates when a beneficiary enrolled in the MCBS has lost entitlement. If an Incoming Panel beneficiary loses entitlement in the year in which they are sampled into the MCBS, the case is ineligible and will not be fielded. Continuing cases that lose entitlement will be fielded after the date entitlement was lost to allow for the collection of data for the entire period for which the beneficiary had Medicare coverage.
  - **Location:** If a beneficiary has permanently moved more than 30 miles outside of MCBS sampled PSUs, the case will no longer continue to be fielded.



## Key Takeaways

- The MCBS sample is designed to produce national estimates of the Medicare population.
- The MCBS sample is selected from Medicare Administrative enrollment data.
- Using a rotating panel design, a new Incoming Panel is selected each fall and remains in the survey for 11 interviews over four years. Each winter, a panel completes its 11th interview and retires.
- The MCBS sample is designed to yield about 14,500 completed cases annually in the MCBS Survey File and about 9,000 completed cases annually in the MCBS Cost Supplement File.

# Section 4: Questionnaire



# Questionnaire Overview

- The MCBS Questionnaire enables CMS to collect detailed longitudinal data on a wide range of topics for both beneficiaries living in the community and in a facility.
- The MCBS Questionnaire is broadly structured as follows:
  - Two components: Community and Facility
  - Two types of interviews within each component: Baseline and Continuing
  - Two types of sections within each type of interview: Core and Topical
- The questionnaire flow varies based on the component and type of interview, as well as other factors such as the season of data collection (fall, winter, or summer), whether the beneficiary is alive or deceased, and whether the interview is conducted with the beneficiary or a proxy.

# Questionnaire Components

- The MCBS Questionnaire includes a Community and Facility component. The beneficiary's residence status determines which questionnaire component is used and how it is administered.
  - **Community component:** Survey administered to 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.

# Types of Interviews

- Within each questionnaire component, there are two types of MCBS interviews.
  - **Baseline Interview (#1):** The initial questionnaire administered in the fall round of the year the beneficiary is selected into the sample. This interview serves as a reference interview and gathers information on the health status, insurance coverage, and demographic information for the beneficiary.
  - **Continuing Interview (#2-11):** The questionnaire administered as beneficiaries progress through the study. Unlike the Baseline interview, Continuing interviews collect health care utilization and cost data. Continuing interviews also update information collected at baseline.

# Types of Sections

- Depending on the interview type and data collection season (fall, winter, or summer), the respondent is asked a set of Core and Topical sections during each interview.
  - **Core sections:** These sections are of critical purpose and policy relevance to the MCBS. Core sections collect information on beneficiaries' health insurance coverage, health care utilization and costs, access to care, and satisfaction with care.
  - **Topical sections:** 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.
- The Community Questionnaire includes both Core and Topical sections, while the Facility Instrument only includes Core sections.

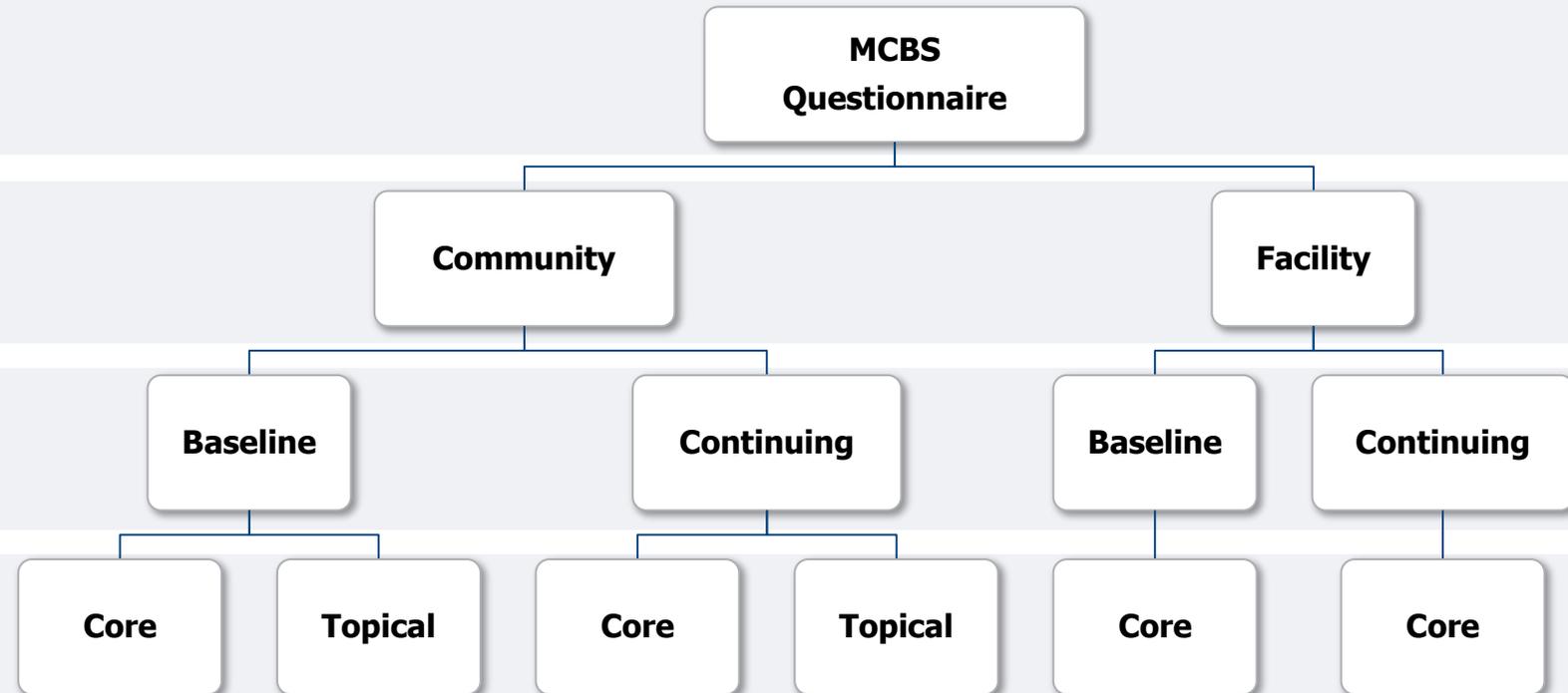
# Questionnaire Summary

MCBS Questionnaire

Components

Interview Types

Section Types



# Community Questionnaire Sections

- On the next slides, the MCBS Community Questionnaire sections are presented by season of administration (fall, winter, summer) and data release (Survey File or Cost Supplement File).
- The annual *MCBS Data User's Guide* and *Methodology Report* contain detailed questionnaire flow diagrams, a description of each section, and descriptions of changes to the questionnaire for a given data year.
- The Questionnaires and Questionnaire User Documentation are posted annually each fall for that year's questionnaire content: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires>



## Community Questionnaire Sections (continued)

### **Core Section Content**

- Socio-Demographics
- Health Insurance
- Health Status
- Experiences with Care
- Health Care Utilization
- Health Care Cost

### **Topical Section Content**

- Housing Characteristics
- Social Determinants of Health or Health Behaviors
- Knowledge and Decision Making

## Community Questionnaire Sections: Socio-Demographics

<b>Section Abbr.</b>	<b>Section Description</b>	<b>Administration Schedule</b>	<b>Section Type</b>	<b>Survey File</b>	<b>Cost Supplement File</b>
IAQ	Income and Assets	Summer	Core	X	
DIQ	Demographics/Income	Fall, Baseline interview	Core	X	

# Community Questionnaire Sections: Health Insurance

<b>Section Abbr.</b>	<b>Section Description</b>	<b>Administration Schedule</b>	<b>Section Type</b>	<b>Survey File</b>	<b>Cost Supplement File</b>
HIQ	Health Insurance	Fall, winter, summer	Core	X	

## Community Questionnaire Sections: Health Care Utilization

Section Abbr.	Section Description	Administration Schedule	Section Type	Survey File	Cost Supplement File
DVH	Dental, Vision, and Hearing Care Utilization	Fall, winter, summer	Core		X
ERQ	Emergency Room Utilization	Fall, winter, summer	Core		X
IPQ	Inpatient Hospital Utilization	Fall, winter, summer	Core		X
OPQ	Outpatient Hospital Utilization	Fall, winter, summer	Core		X
IUQ	Institutional Utilization	Fall, winter, summer	Core		X
HHS, HHQ	Home Health Summary and Utilization	Fall, winter, summer	Core		X
MPQ	Medical Provider Utilization	Fall, winter, summer	Core		X
OMQ	Other Medical Expenses Utilization	Fall, winter, summer	Core		X
PMQ	Prescribed Medicine Utilization	Fall, winter, summer	Core		X

## Community Questionnaire Sections: Health Care Costs

<b>Section Abbr.</b>	<b>Section Description</b>	<b>Administration Schedule</b>	<b>Section Type</b>	<b>Survey File</b>	<b>Cost Supplement File</b>
STQ	Statement Cost Series	Fall, winter, summer	Core		X
PSQ	Post-Statement Charge	Fall, winter, summer	Core		X
NSQ	No Statement Charge	Fall, winter, summer	Core		X
CPS	Charge Payment Summary	Fall, winter, summer	Core		X

# Community Questionnaire Sections: Health Status and Experiences with Care

<b>Section Abbr.</b>	<b>Section Description</b>	<b>Administration Schedule</b>	<b>Section Type</b>	<b>Survey File</b>	<b>Cost Supplement File</b>
HFQ	Health Status and Functioning	Fall	Core	X	
ACQ	Access to Care	Winter	Core	X	
SCQ	Satisfaction with Care	Fall	Core	X	
USQ	Usual Source of Care	Winter	Core	X	

# Community Questionnaire Sections: Housing Characteristics, Social Determinants of Health or Health Behaviors, and Knowledge and Decision Making

Section Abbr.	Section Description	Administration Schedule	Section Type	Survey File	Cost Supplement File
HAQ	Housing Characteristics	Fall	Topical	X	
CPQ	Chronic Pain	Summer	Topical	X	
MBQ	Mobility of Beneficiaries	Fall, winter, summer	Topical	X	
NAQ	Nicotine and Alcohol Use	Fall	Topical	X	
PVQ	Preventive Care	Fall, winter, summer	Topical	X	
IAQ	Food Insecurity items	Summer*	Topical	X	
RXQ	Drug Coverage	Summer	Topical	X	
KNQ	Beneficiary Knowledge and Information Needs	Winter	Topical	X	

\*The Food Insecurity items are included within the Income and Assets Questionnaire (IAQ).

## Facility Instrument Sections

- On the next slides, the MCBS Facility Instrument sections are presented by season of administration (fall, winter, summer) and data release (Survey File or Cost Supplement File).
- The sections in the Facility Instrument parallel those in the Community Questionnaire and are administered according to the same rotation schedule as the Community Questionnaire sections.
- Unlike the Community Questionnaire, the Facility Instrument is tailored to be administered to Facility staff instead of the actual beneficiary, and is modeled after the MDS assessment. The beneficiary does not answer questions during a Facility interview; instead, facility administrators and staff answer questions about the beneficiary.
- The annual MCBS Data User's Guide and Methodology Report contain detailed questionnaire flow diagrams, a description of each section, and descriptions of changes to the questionnaire for a given data year.
- The annual Questionnaires and Questionnaire User Documentation can be found here: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires>



## Facility Instrument Sections (continued)

### **Core Section Content**

- Socio-Demographics
- Health Insurance
- Health Care Utilization
- Health Care Cost
- Health Status
- Facility Characteristics

### **Topical Section Content**

- None

## Facility Instrument Sections (continued)

Section Abbr.	Section Description	Administration Schedule	Section Type	Survey File	Cost Supplement File
FQ	Facility Questionnaire	Fall, winter, summer	Core	X	
BQ	Background	Fall, Baseline interview*	Core	X	
RH	Residence History	Fall, winter, summer	Core	X	
IN	Health Insurance	Fall*	Core	X	
US	Use of Health Services	Fall, winter, summer	Core		X
EX	Expenditures	Fall, winter, summer	Core		X
HS	Health Status	Fall*	Core	X	

\*The BQ section is also administered to Community-to-Facility crossover cases each season. The IN and HS sections are also administered to Community-to-Facility and Facility-to-Facility crossover cases each season. The Prescribed Medicines section was permanently removed from the Facility Instrument in Fall 2019.

## Facility Instrument Redesign and Blended Facility Data

- In Fall 2019, the Facility Instrument was redesigned to skip items redundant with Certification and Survey Provider Enhanced Reports (CASPER) and MDS administrative data for interviews conducted at Medicare- or Medicaid-certified facilities with a CMS Certification Number (CCN).
  - For interviews conducted at facilities with a CCN, the redesigned Facility Instrument skips more than 100 questionnaire items in the Facility Questionnaire and Health Status sections.
  - For interviews conducted at other facilities, the full Facility Instrument continues to be administered.
- As a result of the redesign, survey-reported data for beneficiaries living in facilities with a CCN are merged with CASPER and MDS data during data processing, resulting in a blended data product.

# Crossovers

- Beneficiaries may move to a new setting over the course of their participation in the survey. Crossovers are cases that have moved to a new interview setting requiring a new interview component or transferred to a different facility since the last round interview.
- There are three types of crossovers:
  - Community-to-Facility
  - Facility-to-Facility
  - Facility-to-Community
- In a crossover situation, because the beneficiary has spent part of the reference period in more than one setting, interviewers must complete two separate questionnaires to collect data from both locations.

# Key Takeaways

- The MCBS Questionnaire includes two components, two types of interviews, and two types of sections.
  - The MCBS Questionnaire includes a Community and Facility component. The beneficiary's residence status determines which questionnaire component is used and how it is administered. The questionnaires for each component are parallel, but there are some key differences.
  - Incoming Panel beneficiaries receive a Baseline interview that collects information about their health insurance, health status, and more. As beneficiaries progress through the survey, they receive a Continuing interview. Unlike the Baseline interview, Continuing interviews collect health care utilization and cost data.
  - Core sections collect data that are of critical purpose and policy relevance to the MCBS, such as health insurance, health status, utilization, and costs. Topical sections collect data on special interest topics.

# Section 5: Data Collection





## Data Collection Overview

- CMS contracts with NORC at the University of Chicago (NORC) to conduct the MCBS.
- A national team of trained and certified field interviewers conduct interviews with either MCBS beneficiaries and their designated proxies for beneficiaries living in the community, or facility administrators and staff for beneficiaries living in a facility.
- All survey and data collection protocols for the MCBS are cleared by both the Office of Management and Budget (OMB) under the Paperwork Reduction Act (PRA) and by NORC's Institutional Review Board (IRB). 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.

## Interviewer Staffing and Training

- Nationally, NORC employs an average of approximately 200 interviewers who collect MCBS data throughout the year.
- All interviewers participate in training and certification and are paired with ongoing field management for coaching and monitoring.
- Each training is customized to the level of experience of the interviewer (new to MCBS or MCBS-experienced), the type of interview (Community or Facility), the type of sample (Incoming Panel or Continuing), and the unique requirements of each round (changing questionnaire sections or data collection protocols).
- Interviewers also participate in continuous refresher training.



## Data Collection Schedule

- Data collection for the MCBS is continuous throughout the year.
- There are three distinct seasons (i.e., rounds) of data collection per year: fall, winter and summer.
- The winter and summer rounds typically last 16 weeks, while the fall round is slightly longer and lasts approximately 18 to 24 weeks. The fall round is scheduled as a longer data collection period to accommodate contacting, locating, and interviewing efforts for the Incoming Panel.
- Within a given round, data collection is structured around several case releases due to the cyclical nature of fielding. For Continuing beneficiaries, questionnaire data from the prior round are cleaned, edited, and preloaded before a case is released into production for the next round. Continuing cases are staged and released in batches scheduled throughout the data collection round.

## Contacting and Outreach

- Given the longitudinal panel design of the MCBS, significant efforts are made to maintain respondent engagement in order to minimize non-response bias and lessen the impact of attrition over time.
- Materials are used to initiate and maintain respondents' commitment to the survey. They are disseminated to beneficiaries, proxies, and facility administrators to raise awareness of the survey and educate about the purposes of the survey.
- Interviewers are trained to establish contact with respondents. Greater effort is invested in contacting the Incoming Panel beneficiaries for their first and second interviews due to the fact that activities such as locating, gaining cooperation, and establishing familiarity with the MCBS are often required. Contacting efforts for the remaining interviews typically require fewer attempts to make contact with respondents and schedule appointments.



## Questionnaire Administration

- Whether interviews are conducted in-person or by telephone, the MCBS questionnaire is administered using a Computer-Assisted Personal Interviewing (CAPI) instrument loaded on a laptop computer.
- The CAPI system automatically guides interviewers through the questions, records the answers, and compares them to edit specifications. The CAPI also guides interviewers through complex skip patterns and inserts follow-up questions where data were missing from the previous interview.
- When the interview is completed, the CAPI system allows the interviewers to securely transmit the data electronically to the NORC Central Office.

# Interviewing Languages

- The Community Questionnaire can be administered in English or Spanish.
  - Community interviews with respondents who speak Spanish as their primary language are conducted by bilingual interviewers who have been certified to conduct the Spanish version of the MCBS Community Questionnaire.
  - If an interviewer encounters a beneficiary who speaks a primary language other than English or Spanish, the interviewer will attempt to identify an English-speaker to either serve as a proxy or assistant for the beneficiary.
  - Historically, approximately 5% of Community interviews have been conducted in Spanish.
  - A variable for interview language is available in the data for analysis.
- The Facility Instrument is administered only in English.
  - The beneficiary is not interviewed directly. Instead, data in the Facility interview are collected by speaking with facility staff and by abstracting from medical records.

## Community vs. Facility Data Collection

- The MCBS includes both a Community and Facility component, which is determined by the beneficiary's residence status. Data collection and questionnaire administration protocols vary based on component.
  - For beneficiaries living in the community, the Community Questionnaire is administered and interviews are conducted with the beneficiary or a designated proxy.
  - For beneficiaries living in a facility that meets the definitions set forth by CMS, the Facility Instrument is administered and interviews are conducted with facility administrators and staff.
- Specially-trained interviewers conduct interviews for beneficiaries living in facilities.
- Historically, approximately 8% of interviews have been conducted using the Facility component.
- Most interviews were traditionally conducted in-person in households and facilities using CAPI. However, due to the COVID-19 pandemic, data collection switched to phone-only interviews in March 2020 and throughout 2021.

# Community Data Collection Protocols

- The Community Questionnaire is traditionally administered in person to establish survey legitimacy, build rapport with the respondent, provide context for future rounds, and facilitate the review of beneficiaries' health care utilization and cost documentation.
- After the initial Baseline interview, interviewers provide respondents with a planner and instructional aid to document medical events and save any hardcopy statements received after the date of the current interview. During the subsequent round, interviewers review calendars with respondents, as well as sort and match any hardcopy documentation associated with past reported medical events.
- Interviews are usually conducted in person; administration by phone is also permitted. In response to the COVID-19 pandemic, the mode of administration changed from in-person to phone-only starting in March 2020.
- Community interviews are conducted with the beneficiary or a designated proxy.

## Proxy Interviews and Use of Assistants

- Beneficiaries living in the community may need assistance with providing the detailed information needed to accurately respond to survey items. During the course of data collection, beneficiaries may designate a proxy to participate in the interview on their behalf or an assistant to provide help in answering the survey questions.
  - A **proxy** is a person who is more familiar with the beneficiary's health care events and costs than the beneficiary, or is more physically or mentally able to respond to the Community survey questions. A proxy can respond for the beneficiary in his/her absence. In addition, a proxy will be utilized if a beneficiary has been reported as deceased during the current round's reference period or if the beneficiary has entered the Facility component.
  - An **assistant** helps the beneficiary answer factual questions such as provider dates, but cannot respond for the beneficiary in his/her absence. The assistant is chosen by the beneficiary to help in situations where the beneficiary can respond to the interview as long as he/she receives some help from another person.
  - Variables for whether a proxy or assistant were used are available in the data for analysis.

## Facility Data Collection Protocols

- If a beneficiary resides in a location that meets the MCBS definition of a facility, then a Facility interview is conducted. If a facility does not meet the MCBS definition, or if the beneficiary is not residing in the section of the facility that provides long term care, a Community interview is conducted.
- Facility interviews are conducted with staff at the facility, not the beneficiary. It is common for interviewers to collect data from more than one person at the facility because different staff have the most complete information for specific sections of the questionnaire.
- Much of the content of the Facility interview can be found in medical documentation. As such, facility interviewers are trained to abstract interview responses directly from medical records.
- More information about the Facility interview and combining Community and Facility data can be found in the *MCBS Advanced Tutorial on Using Community and Facility Data*.

## Facility Definition

- A Facility component is conducted when the beneficiary is living in a long-term care or other residential facility that meets the following conditions:
  - Is a place or unit of a larger place with three or more beds, and either:
    - Is certified by Medicare as a Skilled Nursing Facility (SNF);
    - **or** certified by Medicaid as a Nursing Facility or an Intermediate Care Facility for the Mentally Challenged;
    - **or** is licensed as a Personal Care Home, Board and Care Home, Assisted Living Facility, Domiciliary Care Home or Rest Home by a state, or local government agency;
    - **or** provides 24 hours a day, 7 days a week supervision by a person willing and able to provide personal care;
    - **or** provides personal care services to residents (personal care may include assistance with eating, dressing, preparing meals, etc.).
- Most beneficiaries who reside in a place that meets the MCBS definition of a facility live in a type of nursing home.

## Key Takeaways

- Data collection for the MCBS is continuous throughout the year and divided into three seasons: fall, winter, and summer.
- A national team of trained and certified interviewers conduct interviews with MCBS beneficiaries, proxies, and facility administrators using a CAPI instrument loaded on a laptop computer.
- Given the longitudinal panel design of the MCBS, significant efforts are made to maintain respondent engagement with the survey to minimize non-response bias and lessen the impact of attrition over time.
- Data collection and questionnaire administration protocols are tailored to each questionnaire component (Community or Facility).

# Section 6: Data Products and Documentation



# Data Products and Documentation Overview

- CMS provides users with multiple ways to access MCBS data and a wide array of documentation that is publically available on the CMS MCBS website.
- Data users have several options to access MCBS data:
  - MCBS data are made available via **two annual Limited Data Set (LDS) releases** and **Public Use Files (PUFs)**. These releases are accompanied by detailed *Data User's Guides* as well as other technical documentation.
  - CMS also releases an annual *Chartbook* based on the LDS files that identifies key estimates and survey findings and includes free downloadable charts for use by the research community.
  - In addition, CMS releases an annual *Early Look* that presents key preliminary estimates prior to the MCBS data releases, and periodic **infographics** that highlight new MCBS questionnaire content or other topics of interest.
- Most MCBS documentation (including *Data User's Guides*, *Methodology Reports*, tutorials on a variety of topics, codebooks, and questionnaires and questionnaire user documentation) is publically available on the CMS website.
- More detailed information on MCBS documentation can be found in Section 8 of this tutorial.



## MCBS Data Files Overview

- MCBS data are made available to users via two annual LDS files and PUFs.
  - **Limited Data Set (LDS)** files contain beneficiary-level protected health information. However, selected variables within the files are encrypted, blanked, or provided as a range to protect and limit the potential of beneficiary re-identification. LDS files require a Data Use Agreement (DUA).
  - **Public Use Files (PUFs)**, also called Non-Identifiable Data Files, have been edited and stripped of all information that could be used to identify individuals. As such, PUFs do not require a DUA.

## Overview of the MCBS LDS Releases

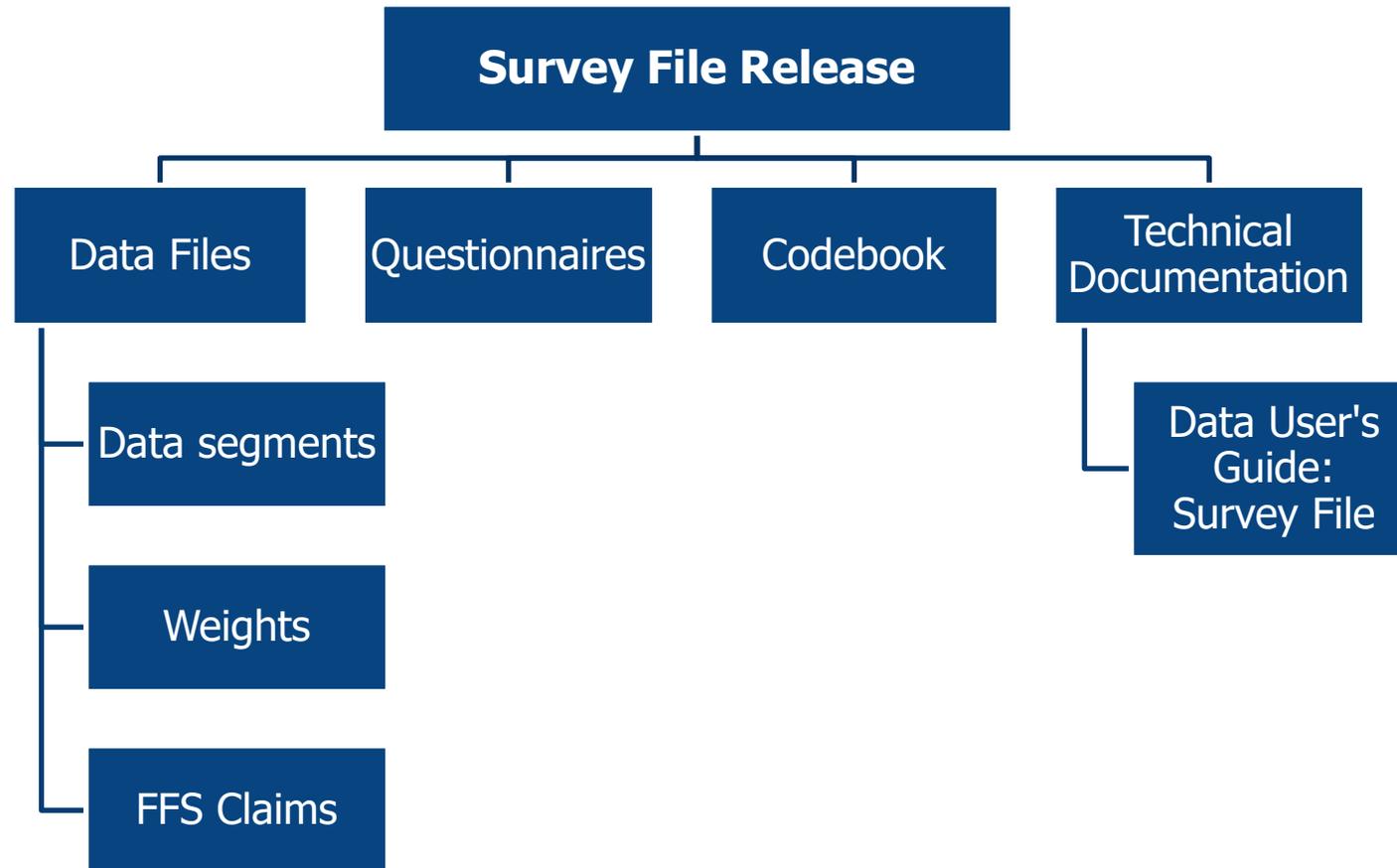
- The MCBS LDS releases contain the following data:
  1. The **Survey File** contains beneficiaries' demographic information, health insurance coverage, self-reported health status and conditions, and responses regarding access to care and satisfaction with care.
  2. The **Cost Supplement File** contains a comprehensive accounting of beneficiaries' health care use, expenditures, and sources of payment.
- Each of the two LDS releases contains multiple files, called segments, which are easily linkable through a common beneficiary key ID.
- Depending on their research questions, data users have the option of obtaining the Survey File alone or obtaining both the Survey File and Cost Supplement File.

# Survey File

- The Survey File contains beneficiaries' demographic information, information about health insurance coverage, self-reported health status and conditions, and responses regarding access to care and satisfaction with care.
- The Survey File contains survey data collected directly from respondents. Survey data are supplemented by facility (non-cost) information, Medicare administrative data, and Medicare FFS claims to facilitate analysis.
- The 2019 Survey File contains over 6,000 variables across 41 segments.<sup>1</sup>
- The subsequent slides show the contents of the 2019 Survey File release and describe the data files included.

<sup>1</sup> Note that the 2018 Survey File LDS included over 4,000 variables. The large increase from 2018 to 2019 is due to the addition of two topic-specific LDS segments on data collected from the MCBS COVID-19 Summer 2020 and Fall 2020 Community Supplements. These segments were released with the 2019 Survey File LDS because the population that was administered those COVID-19 Community Supplements aligns with the 2019 Survey File population.

# Contents of Survey File Release



# Survey File Data Files

## Data Segments

- Data segments organized by content area, including health status and limitations, access to care, and health insurance coverage.
- Facility-related data including a residence timeline, facility characteristics, and assessment [MDS] measures.
- Topical questionnaire section data including beneficiary knowledge and drug coverage.

## Medicare FFS Claims

- FFS claims data that provide CMS administrative information on medical services and payments paid by Medicare claims for beneficiaries enrolled in original Fee-for-Service Medicare.
  - NOTE: Approximately 1/3 of Medicare beneficiaries are enrolled in MA and are therefore not represented by these claims.

## Survey File Weights

- General-purpose cross-sectional weights.
- A series of replicate weights.
- Longitudinal weights.
- Topical weights.

# Survey File Data Segments

## Demographics

- Demographics
- Income and Assets
- Household Characteristics

## Health Insurance

- Health Insurance Summary
- Health Insurance Timeline

## Residence Status

- Residence Timeline
- Interview Characteristics

## Facility Information

- Facility Characteristics

## Health Status

- General Health
- Chronic Conditions
- Chronic Pain
- Diabetes
- Mental Health
- Vision and Hearing
- Facility Assessments

## Functional Status

- Assistance
- Falls
- Mobility
- NAGI Disability

## Related Health Data

- Nicotine and Alcohol Use
- Preventive Care
- Food Insecurity

## Access to Care

- Access to Care
- Access to Care, Medical Appointments
- Usual Source of Care
- Patient Activation

## Satisfaction with Care

- Satisfaction with Care

## Plan Knowledge/Experience

- Medicare Plan Beneficiary Knowledge
- MA Plan Questions
- RX Medications

## COVID-19 Topic-Specific

- COVID-19 Summer 2020 Community Supplement
- COVID-19 Fall 2020 Community Supplement

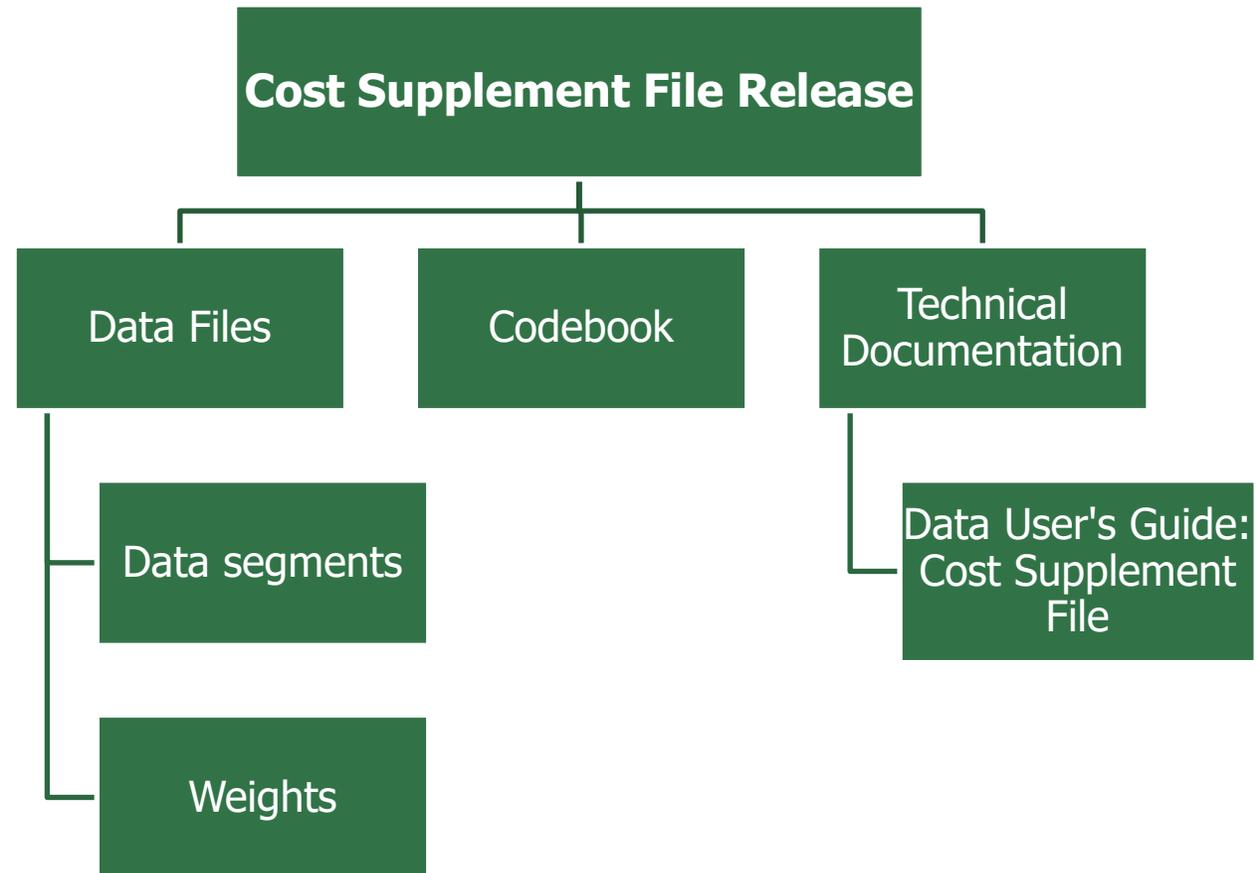
## Cost Supplement File

- The Cost Supplement File contains survey-reported data, linked with administrative data, on beneficiaries' health care utilization, expenditures, and sources of payment, including out-of-pocket costs and payments.
- The Cost Supplement File can be linked to the Survey File to conduct analyses on health care cost and utilization. By linking survey-reported data in the Cost Supplement File to Medicare FFS claims data in the Survey File, data users can obtain a comprehensive picture of health services received, amounts paid, and sources of payment, including those not covered by Medicare.
- Starting with the 2019 Cost Supplement File, MA encounter data were used during imputation to improve estimation of medical events and costs for beneficiaries enrolled in MA.
- The 2019 Cost Supplement File contains over 600 variables across 14 segments.

## Cost Supplement File Structure

- The Cost Supplement File data are provided at three different levels of summarization. The tri-level structure allows researchers to fit the research problem they are addressing to the available file summary levels, and potentially avoid having to process all the detailed event records in the file when summaries may suffice.
- The Cost Supplement segments are assembled at three levels:
  1. The **Event level** reports all payers, costs, and utilization at the most detailed level available (one observation per event per person).
  2. The **Service Summary level** summarizes all payers, costs, and utilization for a person at the service level (one observation per service type per person).
  3. The **Person Summary level** summarizes all payers and costs across service categories and summarizes type of service amounts (one observation per person).

# Contents of Cost Supplement File Release



# Cost Supplement File Data Files

## Data Segments

- Survey-reported data on the use and cost of medical services, including all sources of payment and out-of-pocket costs.
- Medicare FFS and prescription drug claims data including administrative and billing information on the cost and utilization of medical services.
- MA cost and utilization information.

## Cost Supplement Weights

- General-purpose cross-sectional weights.
- A series of replicate and longitudinal weights.

# Cost Supplement File Data Segments

## Event-level segments

- Dental utilization events
- Vision utilization events
- Hearing utilization events
- Facility events
- Inpatient hospital events
- Institutional events
- Medical provider events
- Outpatient hospital events
- Prescribed medicine events

## Summary segments

- Person summary
- Service summary

## Medicare Population Represented by the MCBS Data

- The MCBS data are representative of two overlapping but different populations:
  1. The “**ever enrolled**” population represents individuals who were enrolled in Medicare at any time during the calendar year. This population includes beneficiaries who were enrolled throughout or who were current-year enrollees during the calendar year as well as those who dis-enrolled or died prior to their fall interview. The ever enrolled population includes beneficiaries who were enrolled in Medicare for at least one day at any point during the calendar year.
  2. The “**continuously enrolled**” population represents only those individuals continuously enrolled in Medicare from January 1 of the calendar year up to and including their fall interview. This population specifically excludes beneficiaries who were current-year enrollees during the calendar year and those who dis-enrolled or died prior to their fall interview. The continuously enrolled population is representative of the "stable" Medicare population and therefore can affect inferences from the data.

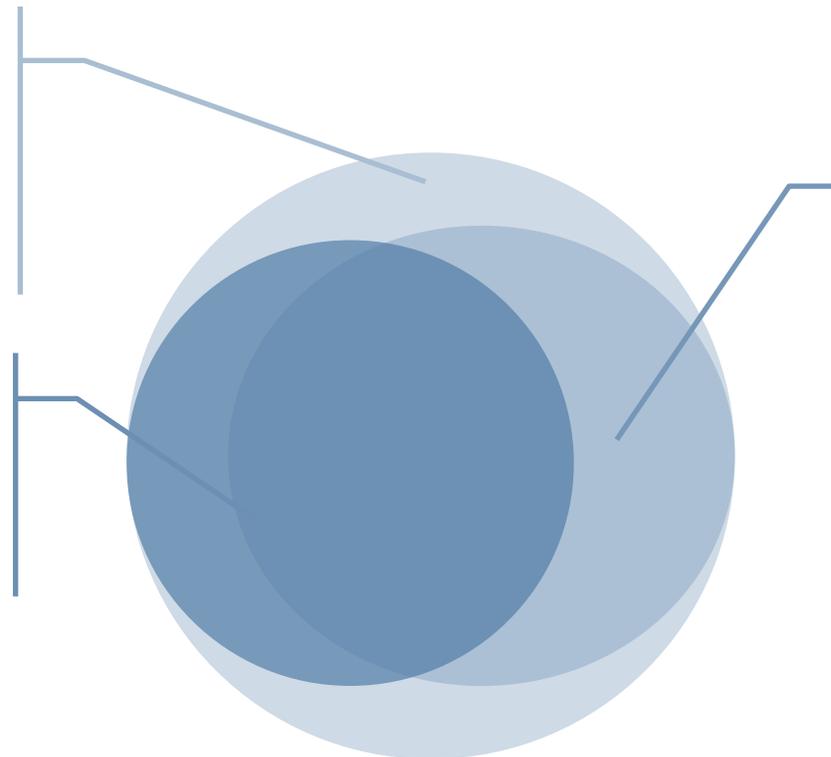
# Medicare Population Represented by the MCBS Data (continued)

## Survey File

The Survey File is representative of the **ever enrolled** population, which is the largest population represented across the MCBS data files and includes anyone enrolled at any time during the calendar year.

The Survey File is also representative of the **continuously enrolled** population, which includes beneficiaries enrolled in Medicare from January 1 up to and including their fall interview.

*Note that this graphic is not drawn to scale.*



## Cost Supplement File

The Cost Supplement File is representative of the **ever enrolled** population. Although this population is identical to the **ever enrolled** population represented by the Survey File, it is derived from a smaller subset of sampled beneficiaries with complete cost and utilization data for the year.

## Which File Do I Need?

	Survey File Only	Survey File + Cost Supplement File
<i>Types of Research Questions</i>	<ul style="list-style-type: none"> <li>• Demographic characteristics</li> <li>• Health status</li> <li>• Cognitive and functional status</li> <li>• Access to care</li> <li>• Satisfaction with care</li> <li>• Care experience</li> <li>• Cost and utilization for services covered by traditional FFS Medicare</li> </ul>	<ul style="list-style-type: none"> <li>• Total health care spending and utilization, including tracking trends over time</li> <li>• Estimating health care spending by source of payment, including assessing out-of-pocket payment burden</li> </ul>
<i>Study Population</i>	<ul style="list-style-type: none"> <li>• “Continuously enrolled”</li> <li>• “Ever enrolled”</li> </ul>	<ul style="list-style-type: none"> <li>• “Ever enrolled”</li> </ul>
<i>Types of Analysis</i>	<ul style="list-style-type: none"> <li>• Cross-sectional analysis</li> <li>• Pooled cross-sectional analysis</li> <li>• Longitudinal analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Cross-sectional analysis</li> <li>• Pooled cross-sectional analysis</li> <li>• Longitudinal analysis</li> </ul>

## Which File Do I Need? (continued)

Example research question	Survey File Only	Survey File + Cost Supplement File
Among Medicare beneficiaries, are there differences in the quality of patient experience by income-to-poverty ratio?	X	
Are there changes in Medicare beneficiaries' out-of-pocket costs between 2009 and 2019?		X
What is the difference in skilled nursing facility utilization between beneficiaries with traditional Fee-for-Service (FFS) Medicare and those with MA?		X
Are there differences in food insecurity; that is, a beneficiary's ability to obtain sufficient food, by socio-demographics?	X	

## Accessing the LDS Data

- The MCBS LDS files must be obtained from CMS.
- Administrative processing fees for obtaining the files are \$300 for the 2019 Survey File alone, and \$600 for the 2019 Survey File with the 2019 Cost Supplement.
- Requests for the MCBS LDS files must be made through the CMS Data Use Agreement (DUA) tracking system known as the Enterprise Privacy Policy Engine (EPPE). EPPE can be used to initiate a new LDS DUA request or to amend/update an existing LDS DUA.
- Upon request and formal approval, CMS releases the data on DVD or via the CMS Virtual Research Data Center (VRDC). Each data release contains multiple data sets that are linkable through a key identification variable (BASEID).
- For each survey year, CMS provides technical documentation along with the following resources for data users: codebooks, questionnaires, data files (SAS, CSV), research claims (SAS, CSV), format control files, and sample SAS code to read in files and apply the formats.

## Overview of the MCBS PUF Releases

- In addition to the LDS files, MCBS data are made available to users through two types of PUFs:
  - The **Survey File Public Use File**, released annually.
  - The **COVID-19 Public Use Files**, released for Summer 2020, Fall 2020, and Winter 2021.
- Like the LDS files, the PUFs serve as unique sources of information on beneficiaries' health and well-being that cannot be obtained through CMS administrative sources alone.
- The MCBS PUFs are not intended to replace the more detailed LDS files; rather, they are publically-available alternatives that do not require a DUA.

## Overview of the MCBS Survey File Public Use File

- The MCBS Survey File PUF is a publically available downloadable data file that allows researchers to conduct analyses on access to care, health status, knowledge of, attitudes toward, and satisfaction with health care, and as well as demographic data and information on health insurance coverage for Medicare beneficiaries living in the community.
- The first edition of the MCBS Survey File PUF included data from the 2013 MCBS, and starting with 2015 data, an MCBS Survey File PUF has been released for each data year.
- Starting with the 2017 Survey File PUF, data are released in three segments, one for Fall, Winter, and Summer, which allows for the release of data collected in each round. The samples for the Winter and Summer segments are subsets of the Fall segment sample. All three segments include survey weights that allow researchers to produce nationally representative estimates for beneficiaries enrolled in Medicare at any point during the calendar year (ever enrolled).
  - In the 2019 Survey File PUF, the Fall segment includes data for 14,167 sampled beneficiaries. The Winter segment includes data for 10,933 sampled beneficiaries, and the Summer segment includes data for 8,843 sampled beneficiaries.

## Comparison of the 2019 MCBS PUF and 2019 MCBS LDS

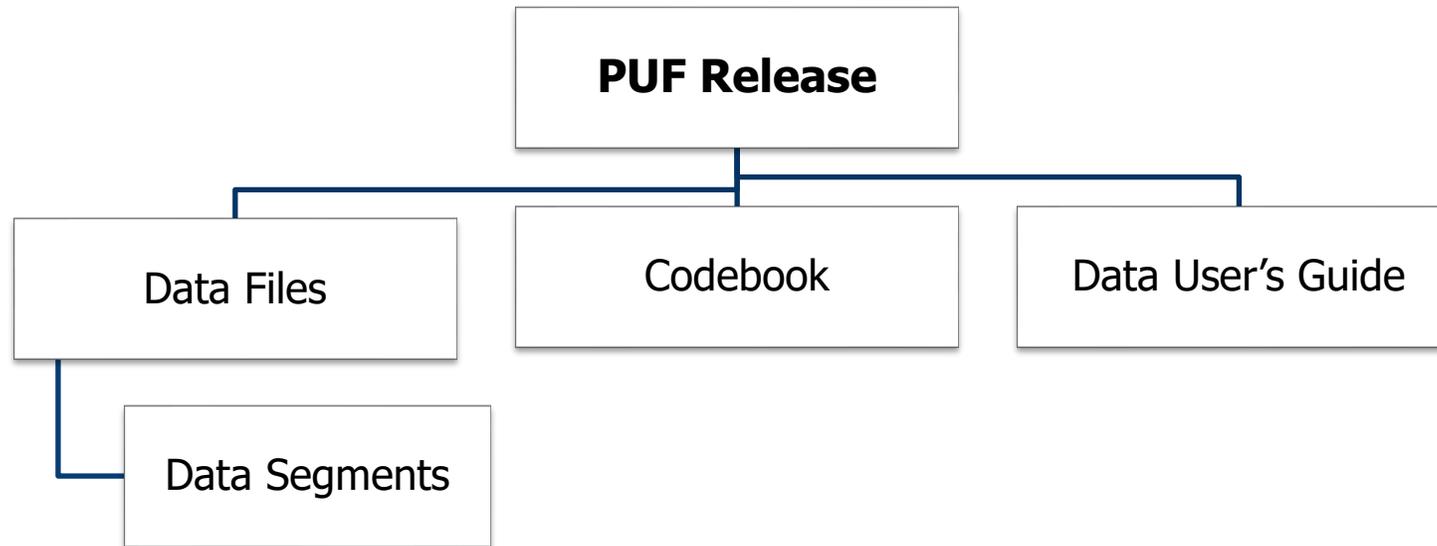
Domain	2019 MCBS Survey File PUF	2019 MCBS Survey File LDS
Population	Community	Community and facility
Number of Variables	662 across three data segments	Over 6,000 across 41 data segments
ID	PUF_ID; Randomly generated, can't be linked back to BASEID, changes each year	BASEID; Randomly generated, can't 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 variables, 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



## Overview of the MCBS COVID-19 Public Use Files

- To enable researchers to conduct analyses to understand the impact of the pandemic on Medicare beneficiaries, data from the COVID-19 Community Supplements were made publicly available through three MCBS COVID-19 PUFs as well as the 2019 Survey File LDS. In addition, data from the COVID-19 Community Supplements will be released in the 2020 Survey File LDS.
- Data from the COVID-19 Facility Supplements will be released in the 2020 and 2021 Survey File LDS.
- For more information on how to access and analyze the data the COVID-19 Supplements, please see the corresponding *Data User's Guides* and the *Advanced Tutorial on the COVID-19 Supplement Data*.

# Contents of PUF Releases



## Links to Access MCBS Data Products and Documentation

Resource	Link
CMS MCBS website	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index</a>
MCBS LDS files	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA - NewLDS">https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA - NewLDS</a>
MCBS PUFs	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index</a>
CMS Virtual Research Data Center (VRDC)	<a href="https://www.resdac.org/cms-data/request/cms-virtual-research-data-center">https://www.resdac.org/cms-data/request/cms-virtual-research-data-center</a>
MCBS codebooks and data file documentation	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks</a>
<i>MCBS Chartbook</i> (based on the LDS files)	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables</a>

## Key Takeaways

- MCBS data are made available via two annual LDS releases, the Survey File and the Cost Supplement File, and two types of PUFs, the annual Survey File PUF and the COVID-19 PUFs. The MCBS LDS files require that users pay an administrative processing fee and require a DUA, while the MCBS PUFs are free and available for download on the CMS website.
- CMS also releases an annual *Chartbook* based on the LDS files which includes free downloadable charts for use by the research community, an annual *Early Look* with key preliminary MCBS estimates, and a series of infographics highlighting new questionnaire content or other topics of interest.
- Most MCBS documentation is publically available on the CMS website, including codebooks, tutorials on a variety of topics for new and advanced data users, questionnaires and questionnaire user documentation, *Data User's Guides*, and *Methodology Reports*.

# Section 7: Analytic Guidelines



# Analytic Guidelines Overview

- This section outlines the process of using MCBS data. The following two example research questions illustrate how MCBS data can be analyzed.
  - Example 1: To what extent does out-of-pocket spending vary by health status for Medicare beneficiaries living in the community aged 65+?
  - Example 2: To what extent does food insecurity vary by income among Medicare beneficiaries living in the community aged 65+?
- The following slides walk through each step of the process (with sample SAS code) of how to construct an analytic dataset and how to produce basic descriptive statistics using appropriate variance estimation methods.
- For more detailed information on any of these topics, and for example code using other statistical packages such as R and Stata, see the *MCBS Data User's Guide*.



## Analytic Guidelines Overview (continued)

- Each year, the MCBS data are made available to users via two annual LDS files and the Survey File PUF, which are described in more detail in Section 6: Data Products and Documentation.
- This section will focus on analytic guidelines for using the LDS files. Additional guidance is available in the *MCBS Advanced Tutorial on Weighting and Variance Estimation*.
- The LDS data releases contain multiple, topic-related segments that can be merged using the BASEID variable. Segments can be merged within and/or across data releases.

# Analyzing Survey File Data and Cost Supplement File Data Separately or in Combination

- Data users may choose to conduct analyses of the Survey File or joint analysis of survey and cost and utilization data using both files.
- Depending on their research question, data users may need to merge segments within and/or across the Survey File and Cost Supplement File to construct the appropriate analytic file.
- The Survey File and Cost Supplement File both contain weight segments. However, it is important to note that users conducting joint analysis of both Survey File and Cost Supplement File data must use the Cost Supplement File weights. Users analyzing Survey File data separately should use the Survey File weights.
- Further, to generate estimates using the data from one of the Topical Questionnaire sections on their own or merged with another segment, researchers must always use the special non-response adjustment general and replicate weights included in the Topical segment **INSTEAD** of using the weights that appear in the separate weight segments.

# Medicare Population Represented by the MCBS Data

- Recall that the MCBS data are representative of two overlapping but differing populations:
  1. The “**ever enrolled**” population represents individuals who were enrolled in Medicare at any time during the calendar year. This population includes beneficiaries who were enrolled throughout or who were current-year enrollees during the calendar year as well as those who dis-enrolled or died prior to their fall interview. The ever enrolled population includes beneficiaries who were enrolled in Medicare for at least one day at any point during the calendar year.
  2. The “**continuously enrolled**” population represents only those individuals continuously enrolled in Medicare from January 1 of the calendar year up to and including their fall interview. This population specifically excludes beneficiaries who were current-year enrollees during the calendar year and those who dis-enrolled or died prior to their fall interview. The continuously enrolled population is representative of the "stable" Medicare population and is likely therefore to have inferential consequences for the population of interest.
- Before using the data, it is important to understand which population(s) are represented in the LDS files, as depicted on the next slide.

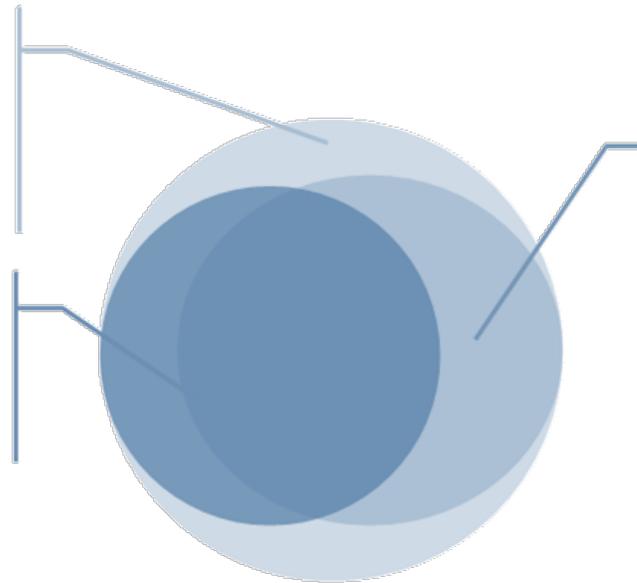
# Medicare Population Represented by the MCBS Data

## Survey File

The Survey File is representative of the **ever enrolled** population, which is the largest population represented across the MCBS data files and includes anyone enrolled at any time during the calendar year.

The Survey File is also representative of the **continuously enrolled** population, which includes beneficiaries enrolled in Medicare from January 1 up to and including their fall interview.

*Note that this graphic is not drawn to scale.*



## Cost Supplement File

The Cost Supplement File is representative of the **ever enrolled** population. Although this population is identical to the **ever enrolled** population represented by the Survey File, it is derived from a smaller subset of sampled beneficiaries with complete cost and utilization data for the year.

The use of **analytic weights** ensures that each population is appropriately represented.

## Full Sample Weights in the Survey File

- For analysis of Survey File data, the ever enrolled and continuously enrolled populations can be obtained through the use of two distinct weights:
  - The **ever enrolled Survey File weight** is greater than zero for all beneficiaries in the Survey File. This weight segment is EVRWGTS, and the name of the weight is EEYRSWGT. The sum of this weight represents the population of beneficiaries who were entitled and enrolled in Medicare for at least one day at any time during the calendar year.
  - The **continuously enrolled Survey File weight** is greater than zero for the subset of beneficiaries in the Survey File who were continuously enrolled in Medicare from January 1 of the calendar year, through completion of their fall interview. This weight segment is CENWGTS, and the weight is named CEYRSWGT. The population represented by the sum of this weight is the continuously enrolled population of Medicare beneficiaries who were enrolled from the first of the year through the Fall round.

## Full Sample Weights in the Cost Supplement File

- Analyses of the Cost Supplement File data should be done with the **Cost Supplement ever enrolled weight**, which represents an ever enrolled population of Medicare beneficiaries enrolled in Medicare on at least one day at any time during the calendar year.
- The Cost Supplement weights segment is named CSEVWGTS.
- The population represented by the sum of this weight is identical to the population represented by the sum of the ever enrolled Survey File weight, but it is populated for a smaller subset of respondents with complete cost and utilization data.

## Using Weights with Data from Topical Questionnaires

- As shown in the *Data User's Guide*, the following 2019 Survey File segments include content from the Topical Questionnaires:
  - ACCSSMED, CHRNPAIN, COVIDSUM, COVIDFAL, FOODINS, INCASSET, MCREPLNQ, RXMED, PNTACT, US CARE
- As mentioned, to generate estimates using the data from one of the Topical Questionnaire sections on their own or merged with another segment, researchers must **always** use the special non-response adjustment general and replicate weights included in the Topical segment **INSTEAD** of using the general and replicate weights that appear in the separate weight segments (CENWGTS, EVRWGTS; CSEVWGTS).
- The “continuously enrolled” Topical weights represent the Survey File continuously enrolled population. The “ever enrolled” Topical weights represent the Cost Supplement File ever enrolled population. Weights for the Survey File ever enrolled population are not calculated.



## Using Weights for Cross-Sectional Analyses of Topical Content (continued)

- Topical data from questionnaire sections collected in the same round include the same set of beneficiaries and the same weights, so joint analysis of multiple topical section data is possible in this case. There are no Topical weights that allow joint analysis of questionnaire section data collected in different rounds.
  - Please refer to the *Data User's Guide* for information on the timing of data collection for each section.

## Process Overview: Example 1

- This section walks through the process of constructing an analytic file and conducting analyses using appropriate variance estimation methods.



# Step 1: Define Your Research Question





## Defining Your Research Question

- This section examines the following question:

**To what extent does out-of-pocket spending vary by health status for Medicare beneficiaries living in the community aged 65+?**

## Step 2: Create Your Analytic File



## Process for Creating Your Analytic File

- After you have defined your research question, you are ready to create your analytic file.
- It is always a good idea to review the *Data User's Guide*, codebooks, *Methodology Report*, *Chartbook*, and other MCBS documentation resources prior to creating an analytic file.
- Creating your analytic file requires five steps:
  1. Identify the MCBS data file(s) that your research question requires.
  2. Identify the data segments and variables that your research question requires.
  3. Identify considerations specific to coding, data collection, and/or processing of variables of interest.
  4. Identify the study population and use the corresponding general purpose weights.
  5. Merge segments to create your analytic dataset.
- The next slides walk through sample SAS code for the construction of an analytic file.

# Identifying Data File(s), Data Segments, and Variables

To what extent does out-of-pocket spending vary by health status for Medicare beneficiaries living in the community aged 65+?

- Our research question requires the following variables:

<b>Measure</b>	<b>File</b>	<b>Segment</b>	<b>Variable</b>
Age	Survey File	Demographics (DEMO)	H_AGE
Residence status	Survey File	Demographics (DEMO)	INT_TYPE
Self-reported health status	Survey File	General Health (GENHLTH)	GENHELTH
Out-of-pocket costs	Cost Supplement File	Person Summary (PS)	PAMTOOP

## Identifying Data File(s), Data Segments, and Variables: Notes on Age and Residence Status

- **Age:** The Survey File contains demographic variables including age. The variable for age is H\_AGE, and it is included in the Demographics segment (DEMO). The MCBS sample includes beneficiaries who enrolled in Medicare upon turning 65 as well as beneficiaries under 65 who qualified for Medicare due to a disability. Given the differences between these two groups, it is recommended to segment or stratify the MCBS population by age. This example restricts the population to beneficiaries aged 65 and over.
- **Residence status:** The Survey File contains a variable for residence status that denotes whether beneficiaries lived in the community for all or part of the year, or lived in a facility for the entire year. The variable for residence status is INT\_TYPE, and it is included in the Demographics segment (DEMO). Due to differences in expenditures and health status between these two groups, this example will also restrict the population to beneficiaries living in the community for all or part of the year.

## Identifying Data File(s), Data Segments, and Variables: Notes on Out-of-Pocket Spending

- Recall that the Cost Supplement File contains data segments with health care events reported at three levels of summarization: the individual event level, person summary level, and service summary level.
- To analyze out-of-pocket costs, use the adjusted total out-of-pocket spending variable (PAMTOOP) in the Person Summary (PS) segment.
- The PAMTOOP variable summarizes spending for 11 types of events (dental, facility, hearing, home health, hospice, inpatient hospital, institutional, medical provider, outpatient hospital, prescribed medicine, and vision), for each beneficiary, including by source of payment.



## Identifying General Purpose Weights

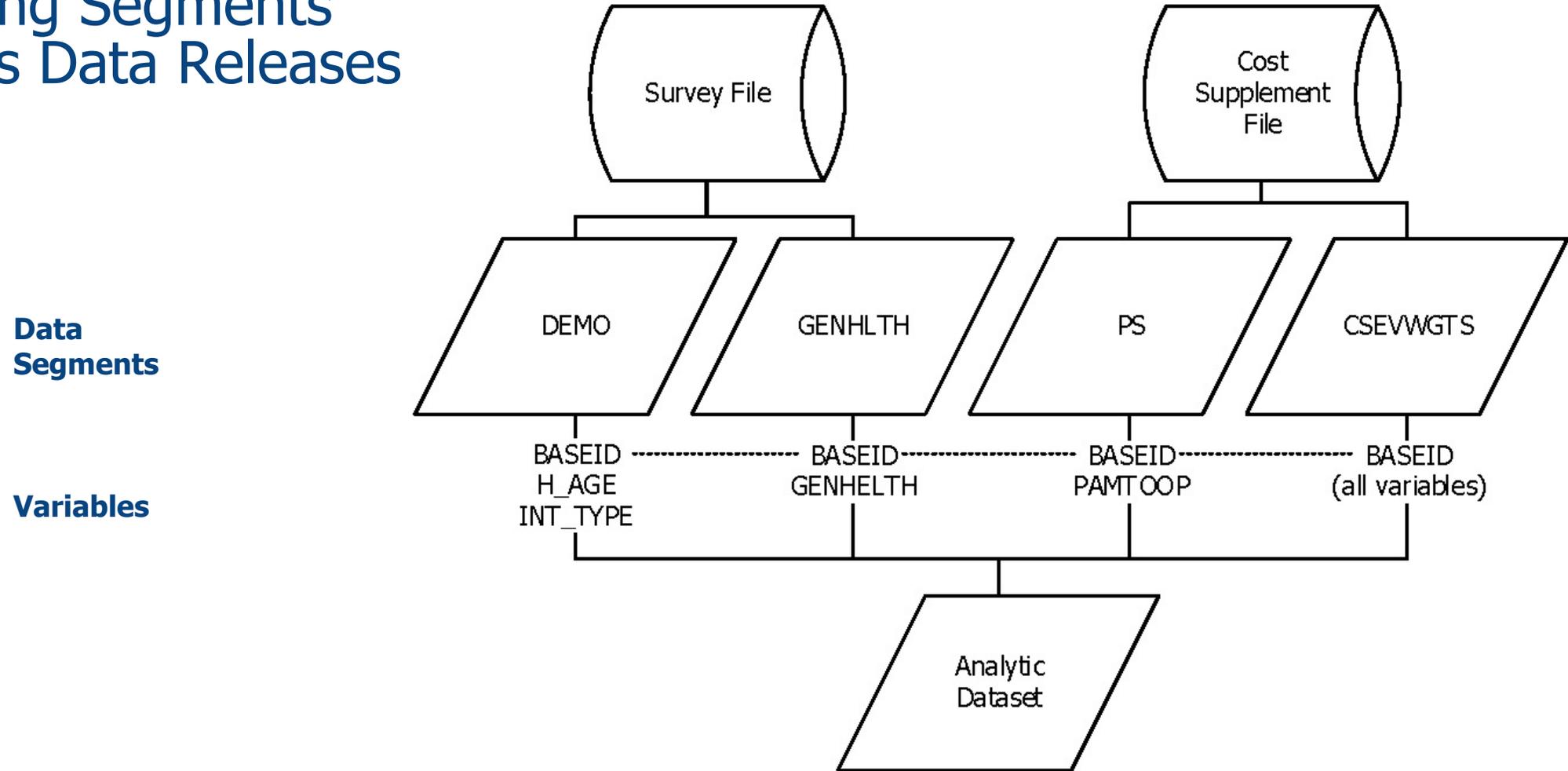
- This example uses the Survey File and the Cost Supplement File.
- The Survey File and Cost Supplement File are both representative of the ever enrolled population. However, the ever enrolled population represented in the Cost Supplement File includes a smaller subset of beneficiaries with complete cost and utilization data for the year.
- As such, users conducting joint analysis of both Survey File and Cost Supplement File data must use the ever enrolled weights in the Cost Supplement File (not the ever enrolled weights in the Survey File).



## Merging Segments to Create an Analytic Dataset

- Data users can merge segments within and/or across the Survey File and Cost Supplement File using the BASEID variable.
- The next slide illustrates how multiple Survey File and Cost Supplement File segments can be merged with the CSEVWGTS segment in the Cost Supplement File using BASEID as the key variable.
- When merging segments, all observations in the CSEVWGTS segment should be preserved. Your final dataset should be restricted to the observations included in the CSEVWGTS segment.

# Merging Segments Across Data Releases



## Sample SAS Code for Merging Segments Across Data Releases

- Below is the SAS code you need to create an analytic file that includes the Survey File and Cost Supplement variables, and Cost Supplement ever enrolled general purpose weights, required to analyze out-of-pocket costs for Medicare beneficiaries by self-reported health status.

```
data oop_merged;  
  merge costYY.CSEVWGTS (in=a)  
        surveyYY.DEMO (keep=BASEID H_AGE INT_TYPE)  
        surveyYY.GENHLTH (keep=BASEID GENHELTH)  
        costYY.PS (keep=BASEID PAMTOOP);  
  by BASEID;  
  if a then output;  
run;
```

**NOTE:** To produce estimates that are representative of the ever enrolled population, you need to limit your dataset to the population included in the weights segment (CSEVWGTS). Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that all observations in the weights segment are preserved, regardless of whether or not the other segments include the entire sample, and that your dataset is restricted to the observations included in the weights file. The "YY" in "costYY" and "surveyYY" refers to the data year of the Cost Supplement File and Survey File, respectively.

# Creating Your Analytic File Summary

To what extent does out-of-pocket spending vary by health status for Medicare beneficiaries living in the community aged 65+?



## Identify the MCBS file(s) required

In this example, you are analyzing both survey data (demographics, health status) and cost data (out-of-pocket spending), so you need to use the Survey File and Cost Supplement File.



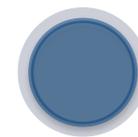
## Identify data segments and variables

You will need the Survey File segments containing the variables for age and residence status (DEMO) and health status (GENHLTH). You will also need the Cost Supplement File segment (PS) containing the variable for annual out-of-pocket costs at the person summary level.



## Identify general purpose weights

You are combining data in the Survey File and the Cost Supplement File, therefore you must use the ever enrolled weights in the Cost Supplement File (segment CSEVWGTS).



## Merge segments to create analytic dataset

Merge the identified Survey File and Cost Supplement segments, and all variables in the ever enrolled Cost Supplement weights segment, using BASEID as the key variable.

## Additional Steps: Segmenting the Population

- In order to restrict the file to beneficiaries living in the community aged 65 and over, segment the file using the variables INT\_TYPE and H\_AGE. Then, create a new analytic variable HEALTH, that correlates to the GENHEALTH variable.

```
data oop_final;  
  set oop_merged;  
  where INT_TYPE='C' and H_AGE GE 65;  
  HEALTH=GENHEALTH;  
run;
```

## Step 3: Conduct Analyses Using Appropriate Variance Estimation Methods



## Variance Estimation Methods

- After you have defined your research question and constructed your analytic file, you are ready to conduct analyses using the appropriate variance estimation methods.
- Basic procedures for calculating sampling errors (e.g., variances, standard errors, etc.) typically assume that the data were collected in a simple random sample.
- However, the MCBS utilizes a stratified, unequal-probability, multi-stage sample. As such, procedures for calculating sampling errors (variances, standard errors) that assume the data were collected in a simple random sample are not appropriate for the MCBS.
- There are two methods that are appropriate for conducting variance estimation using MCBS data: the Taylor-series linearization method and the balanced repeated replication (BRR) method (recommended).

## Variance Estimation Methods for the MCBS

- The MCBS includes variables to obtain weighted estimates and estimated standard errors using two approaches:
  - 1. Taylor-series linearization method:** the variables SUDSTRAT (sampling strata) and SUDUNIT (primary sampling unit) are included for variance estimation using this method.
  - 2. Balanced repeated replication (BRR) method (Fay's method):** a series of replicate weights are included in the Survey File release for variance estimation using this method.
- Most commercial software packages today (including R, STATA, SUDAAN, and complex survey procedures in SAS) include techniques to accommodate the complex design of the MCBS.

## Notes About Subgroup Analysis

- When analyzing survey data, researchers are often interested in focusing on specific subgroups of the full population sample (e.g., Medicare beneficiaries aged 65 and over, Hispanics or females).
  - **Taylor-series linearization method:** A common pitfall when performing sub-group analysis of survey data using methods such as Taylor-series is to delete or exclude observations not relevant to the subgroup of interest. When using this method, standard errors for MCBS estimates are most accurate when the analytic file includes all sample persons.
  - **Balanced repeated replication (BRR) method (Fay's method):** Unlike the Taylor series method, deleting observations not relevant to the subgroup of interest prior to analyzing the subgroup will produce accurate standard errors when replicate weights are used for variance estimation.
- Almost all statistical packages provide the capability to limit the analysis to a subgroup of the population.

## Analyzing a Single Year of MCBS Data

- The following slides walk through an example of variance estimation in SAS using the example analytic file created in the first example.
- The example that follows will employ **balanced repeated replication (BRR) variance estimation**.
- More information on variance estimation methods for the MCBS is included in the *Data User's Guides*.

# Example 1 Code: Out-of-pocket spending among Medicare beneficiaries living in the community aged 65+ by self-reported health status

## Balanced Repeated Replication (BRR) Method

The following code requests the mean value and standard errors of out-of-pocket spending (PAMTOOP) for beneficiaries living in the community aged 65 and over by self-reported health status.

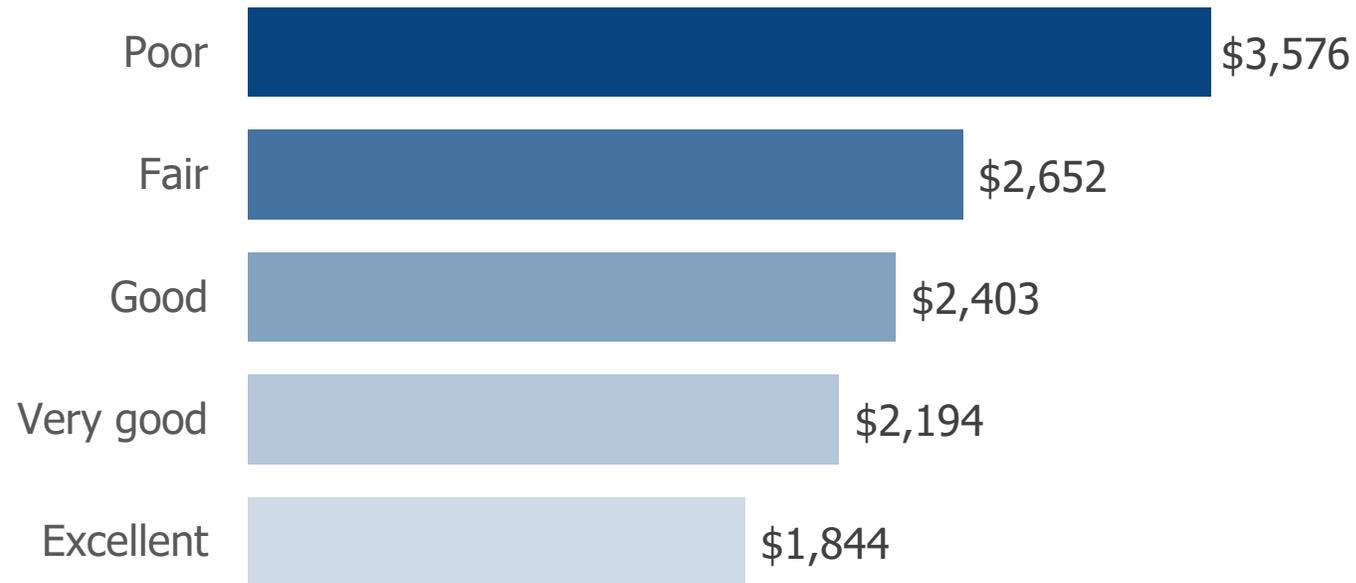
```
proc surveymeans data=oop_final varmethod=brr (fay=.30);  
    var PAMTOOP;  
    weight CSEVRWGT;  
    repweights CSEVR001-CSEVR100;  
    domain HEALTH;  
  
run;
```

## Example 1 Output: Out-of-pocket spending among Medicare beneficiaries living in the community aged 65+ by self-reported health status, 2019

	Estimate - \$ (St. Error - \$)
Self-Reported Health Status	Out-of-Pocket Spending [Average]
Poor	3,576 (590)
Fair	2,652 (162)
Good	2,403 (119)
Very Good	2,194 (108)
Excellent	1,844 (97)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2019.

## Example 1 Figure: Out-of-pocket spending among Medicare beneficiaries living in the community aged 65+ by self-reported health status, 2019



SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2019.

## Process Overview: Example 2





## Step 1: Defining Your Research Question

- Let's continue by examining a second example:

**To what extent does food insecurity vary by income among Medicare beneficiaries living in the community aged 65+?**

## Step 2: Create Your Analytic File



## Step 2: Creating Your Analytic File

- After you have defined your research question, you are ready to create your analytic file.
- Creating your analytic file requires five steps:
  1. Identify the MCBS data file(s) that your research question requires.
  2. Identify the data segments and variables that your research question requires.
  3. Identify considerations specific to coding, data collection, and/or processing of variables of interest.
  4. Identify the study population and use the corresponding weights.
  5. Merge segments to create your analytic dataset.

# Identifying Data File(s), Data Segments, and Variables

To what extent does food insecurity vary by income among Medicare beneficiaries living in the community aged 65+?

- This research question requires the following variables:

Measure	File	Segment	Variable
Age	Survey File	Demographics (DEMO)	H_AGE
Residence status	Survey File	Demographics (DEMO)	INT_TYPE
Income	Survey File	Demographics (DEMO)	INCOME
Food insecurity	Survey File	Food Insecurity (FOODINS)	AFFDMEAL, SKIPMEAL

## Identifying Data File(s), Data Segments, and Variables: Notes on Food Insecurity

- The FOODINS segment in the Survey File contains variables related to food security and food insecurity.
- These questions are part of the Income and Assets Questionnaire and are based upon the USDA Economic Research Service (ERS) Six-Item Short Form of the Food Security Survey Module found at <https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/survey-tools>.
- The examples focus on the following two variables:
  - AFFDMEAL: Adults couldn't afford to eat balanced meals (often/sometimes/never true)
  - SKIPMEAL: Adults ever cut the size of meals or skipped meals because there wasn't enough money for food (yes/no)

## Identifying Weights

- When analyzing data from one of the Topical Questionnaire sections, one must use the special non-response adjustment full-sample and replicate weights included in the Topical segment. There are three sets of weights for each Topical section:
  - The topical “ever enrolled” weights based on the Survey File ever enrolled population.
  - The topical “continuously enrolled” weights based on the Survey File continuously enrolled population.
  - The topical “ever enrolled” weights based on the Cost Supplement File ever enrolled population.
- The example uses the **continuously enrolled** population represented in the Food Insecurity (FOODINS) segment in the Survey File. The continuously enrolled population represents only those individuals continuously enrolled in Medicare from January 1 of the calendar year up to and including their fall interview. This population specifically excludes beneficiaries who were current-year enrollees during the calendar year and those who dis-enrolled or died prior to their fall interview.

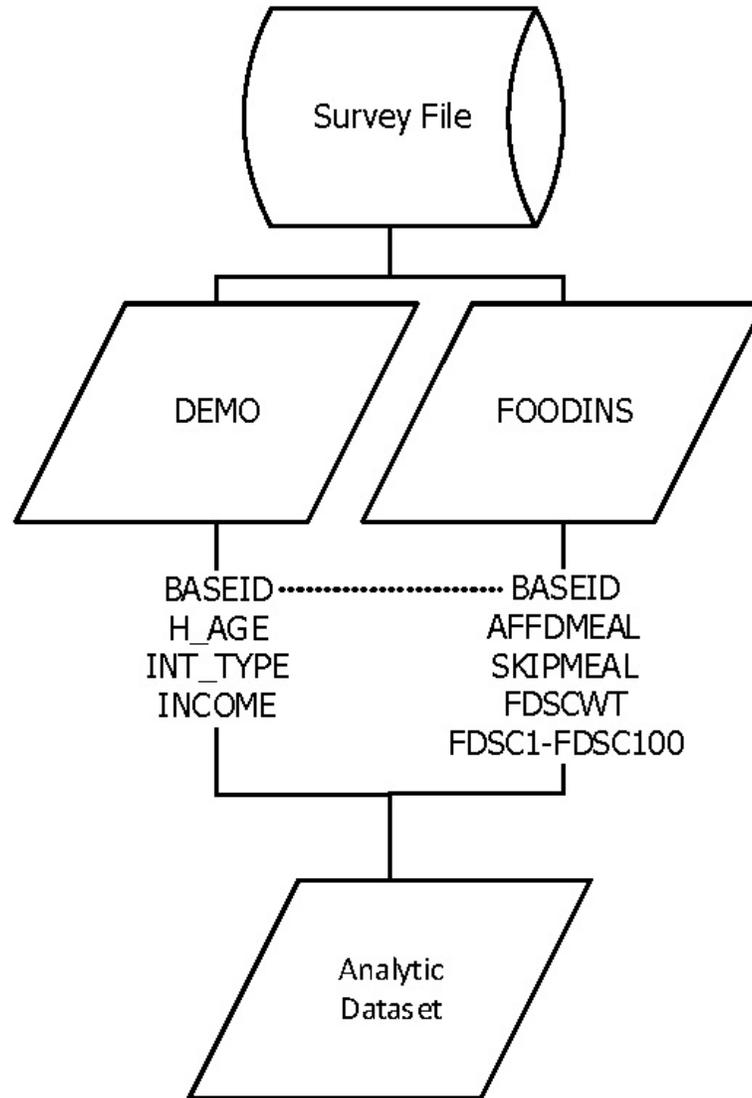
## Merging Segments Within Data Releases

- Data users can merge segments **within** and/or across the Survey File and Cost Supplement File using the BASEID variable.
- The next slide-illustrates how the Survey File DEMO segment can be merged with the FOODINS segment using BASEID as the key variable.
- To produce estimates that are representative of the continuously enrolled population, you need to limit your dataset to this population in the FOODINS segment. When merging segments, all observations with a continuously enrolled weight (FDSCWT) in the Food Insecurity segment should be preserved. Your final dataset should be restricted to the observations included in the FOODINS segment with a continuously enrolled weight.

# Merging Segments Within Data Releases

**Data Segments**

**Variables**



## Sample SAS Code for Merging Segments Within Data Releases

- Below is the SAS code you need to create an analytic file that includes the Survey File variables, as well as the topical continuously enrolled weights and replicate weights, required to analyze food insecurity for Medicare beneficiaries living in the community aged 65 and over by income.

```
data food_inc_merged;  
  merge surveyYY.FOODINS (in=a keep=BASEID  
    AFFDMEAL SKIPMEAL FDSCWT FDSC1-FDSC100)  
  surveyYY.DEMO (keep=BASEID INCOME INCOME_H  
    INT_TYPE H_AGE);  
  by BASEID;  
  if a and FDSCWT^=. then output;  
run;
```

**NOTE:** To produce estimates that are representative of the continuously enrolled population, you need to limit your dataset to the continuously enrolled population in the FOODINS segment. Inclusion of "(in=a)" and "if a and FDSCWT^=." in the SAS code statement ensures that all observations in the FOODINS segment are preserved, regardless of whether or not the other segments include the entire sample, and that your dataset is restricted to the observations with a continuously enrolled weight. The "YY" in "surveyYY" refers to the data year of the Survey File.

# Creating Your Analytic File Summary

To what extent does food insecurity vary by income among Medicare beneficiaries living in the community aged 65+?



## Identify the MCBS file(s) required

In this example, you are analyzing survey data only, so you need to use the Survey File.



## Identify your data segments and variables

You will need the Survey File segments containing the variables for age, residence status, income, and food insecurity (DEMO and FOODINS).



## Identify your general purpose weights

Topical segments in the Survey File are representative of the ever enrolled and the continuously enrolled Medicare population. The example uses the continuously enrolled weights in the FOODINS segment.



## Merge segments to create analytic dataset

Merge the identified Survey File segments using BASEID as the key variable, ensuring that all observations with a continuously enrolled weight in the FOODINS segment are preserved.

## Additional Steps: Recoding Variables

- There are numerous ways to convert raw variables into analytic variables. The example recodes AFFDMEAL and INCOME into two-category variables, as shown below and on the subsequent slide.

```
data food_inc_recode;  
  set food_inc_merged;  
  if INCOME in (1,2,3,4,5) then INCOME_GROUP=1; else INCOME_GROUP=2;  
  if AFFDMEAL in (1,2) then AFFORD=1; else if AFFDMEAL=3 then AFFORD=2;  
run;
```

## Recoded Variables

Measure	Original variable	Recoded variable
Food insecurity – can afford healthy meals	AFFDMEAL 1 Often true 2 Sometimes true 3 Never true	AFFORD 1 Often or sometimes true 2 Never true
Income	INCOME 1 Less than \$5000 2 \$5,000-\$9,999 3 \$10,000-\$14,999 4 \$15,000-\$19,999 5 \$20,000-\$24,999 6 \$25,000-\$29,999 7 \$30,000-\$39,999 8 \$40,000-\$49,999 9 \$50,000 +	INCOME_GROUP 1 Less than \$25,000 2 Greater than or equal to \$25,000

## Additional Steps: Segmenting the Population

- In order to restrict the file to beneficiaries living in the community ages 65 and over, segment the file on the variables INT\_TYPE and H\_AGE.

```
data food_inc_final;  
  set food_inc_recode;  
  where INT_TYPE='C' and H_AGE GE 65;  
run;
```

## Step 3: Conduct Analyses Using Appropriate Variance Estimation Methods



## Step 3: Variance Estimation

- Recall that the MCBS includes variables to obtain weighted estimates and estimated standard errors using two approaches:
  - 1. Taylor-series linearization method:** the variables SUDSTRAT (sampling strata) and SUDUNIT (primary sampling unit) are included for variance estimation using this method.
  - 2. Balanced repeated replication (BRR) method (Fay's method):** a series of replicate weights are included for variance estimation using this method.
- Most commercial software packages today (including R, STATA, SUDAAN, and complex survey procedures in SAS) include techniques to accommodate the complex design of the MCBS.

## Example 2 Code: Food insecurity among Medicare beneficiaries living in the community aged 65+ by income

### Balanced Repeated Replication (BRR) Method

The following code requests the frequencies of beneficiaries living in the community aged 65 and over who experience food insecurity by income using the BRR method for variance estimation.

```
proc surveyfreq data= food_inc_final varmethod=brr (fay=.30);  
  table INCOME_GROUP*(AFFORD SKIPMEAL) / row;  
  weight FDSCWT;  
  repweights FDSC1-FDSC100;  
run;
```

## Example 2 Output: Food insecurity among Medicare beneficiaries living in the community aged 65+ by income, 2019

### Couldn't afford to eat balanced meals

	Estimate - % (St. Error)
Income	Often or sometimes true
Less than \$25,000	22.6 (1.0)
\$25,000 or greater	3.7 (0.3)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2019.

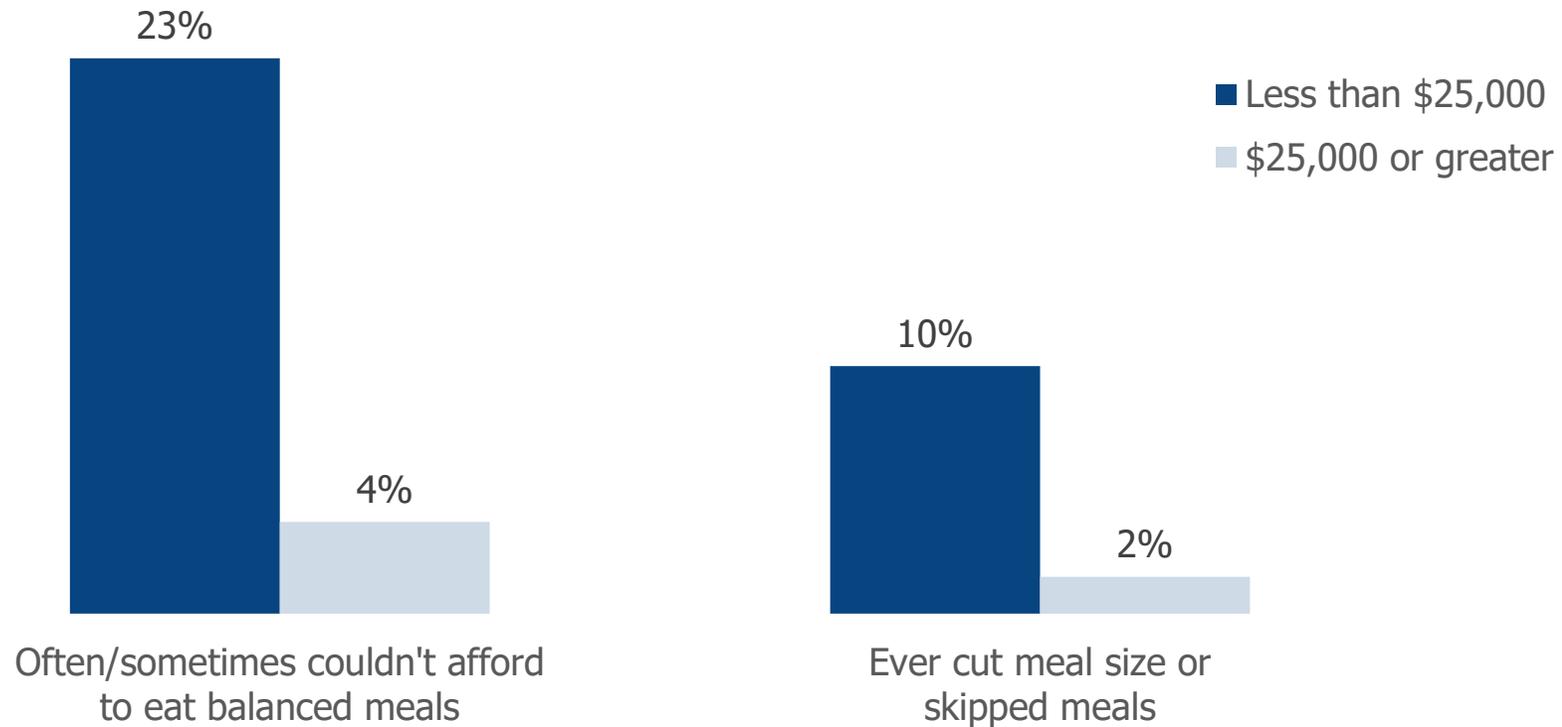
## Example 2 Output: Food insecurity among Medicare beneficiaries living in the community aged 65+ by income, 2019 (continued)

### Ever cut meal size or skipped meals

	Estimate - % (St. Error)
Income	Yes
Less than \$25,000	10.1 (0.7)
\$25,000 or greater	1.5 (0.2)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2019.

## Example 2 Figure: Food insecurity among Medicare beneficiaries living in the community aged 65+ by income, 2019



SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2019.

## Other Technical Notes on Using the Data: Item Non-Response

- Beyond what has already been discussed, data users should be aware of some additional points.
- **Item Non-Response:** Item non-response rates are generally low in the MCBS data release, but data users still need to be aware of the missing data and be cautious about patterns of non-response. There are different ways to address item non-response, such as excluding observations with missing values or recoding these values into an “other” category.
- Beginning in 2015, item non-response types are indicated by missing type codes in SAS.
  - **.D** represents a “don’t know” response.
  - **.R** represents a “refused” response.
  - **.N** represents an “invalid skip” response.
  - **.** represents a “valid skip”.

## Other Technical Notes on Using the Data: Imputation

- **Imputation:** Missing values are imputed for income and assets data, payment data for prescription medicine events, and payment data for non-prescription medicine events (e.g., durable medical equipment bills, inpatient hospital events).
- Flags are available for each imputed variable denoting whether or not the corresponding value was edited in the imputation process. If the record was imputed, imputation flags indicate which method was used.
- Starting with data year 2019, MA encounter data were used to improve estimation of medical events and costs for beneficiaries enrolled in MA, with the goal of accounting for unreported MA medical events. The results of these adjustments are summarized in the service level summary and person level summary files and are not applied at the event level. See the *MCBS Methodology Report* for more information.

## Key Takeaways

- This section walked through the process of producing basic descriptive statistics using MCBS data.
  - **Step 1: Define your research question.** The MCBS can support a broad range of analyses on the health and health care of the Medicare population. For examples, see Section 2: Overview of the MCBS, and refer to the annual *MCBS Bibliographies* on the CMS website.
  - **Step 2: Create your analytic file.** Once you have defined your research question, identify the MCBS data file(s), general purpose weights, and data segments and variables that your research question requires. Merge segments within or across files to create your analytic dataset, and recode variables, as necessary.
  - **Step 3: Conduct analyses using appropriate variance estimation methods.** The MCBS includes variables to obtain weighted estimates and estimated standard errors using two approaches: 1) the Taylor-series linearization method and 2) the balanced repeated replication method (Fay's method).
- For more information any of the topics discussed, see the *MCBS Data User's Guide*.

# Section 8: Summary and Links to Documentation





## MCBS Tutorial Summary

- This tutorial is intended to provide an introduction to the MCBS for users who are unfamiliar with the survey, covering the uses and applications of the survey data, survey sample design, questionnaire structure and content, data collection protocols, data products and documentation, and analytic guidelines.
- Beyond this tutorial, CMS provides a wide range of documentation for researchers and data users that is publically available on the CMS website. This documentation contains more in-depth descriptions of the topics covered in this tutorial, as well as more detailed descriptions of annual changes to the MCBS survey design, operations, and/or data products and documentation.
- This section contains detailed descriptions of the additional MCBS resources available.

# MCBS Documentation and Resources: Quick Links

- The [CMS MCBS website](#) contains links to the following documentation:

<b>Medicare Current Beneficiary Survey (MCBS)</b>
<a href="#">Questionnaires</a>
<a href="#">Data Documentation and Codebooks</a>
<a href="#">Data Tables</a>
<a href="#">Bibliography</a>
<a href="#">Data Briefs and Tutorials</a>

- Questionnaires and Questionnaire User Documentation
- Data Documentation (*Data User's Guides* and *Methodology Report*) and Codebooks
- Data Tables (*Chartbook*)
- *Bibliographies*, which include annotations starting in 2020
- *Early Look*, Data Briefs, Infographics, and Tutorials

## MCBS Documentation and Resources: Quick Links (continued)

- Information about obtaining the MCBS LDS files is available on the CMS website, including instructions for accessing and using EPPE to make a request:
  - Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA - NewLDS>
- The MCBS PUFs and accompanying documentation are available free for download from the CMS website:
  - Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index>

## *MCBS Data User's Guide*

- The annual *MCBS Data User's Guide* offers a publically available, easily searchable resource for data users. It is updated for each new data year to ensure that users have current documentation on the survey design, methods, and estimation, and MCBS data products. For each MCBS data year, two stand-alone *Data User's Guides* are released to help data users in understanding and analyzing the data.
  - The *MCBS Data User's Guide: Survey File* documents the key features of the study and MCBS data products overall. It also provides technical information on the Survey File LDS including the derivation of variables and any significant changes in the variables and/or file structure.
  - The *MCBS Data User's Guide: Cost Supplement File* provides technical information on the Cost Supplement File LDS and also describes the derivation of variables and any significant changes in the variables and/or file structure. In addition, it contains detailed information about matching survey and administrative data as well as imputation. Annual MCBS codebooks are also publically available online.
- Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>

## *MCBS Methodology Report*

- Starting with data year 2015, the annual *MCBS Methodology Report* provides an operational perspective on the collection of survey data for each data year. Like the *Data User's Guide*, an updated version is released after each year's data release.
- The *Methodology Report* complements the *Data User's Guide* with an overview of all activities carried out in support of the annual data files, including sampling, instrument design, interviewer training, data collection, data processing and delivery, and weighting and imputation.
- Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>



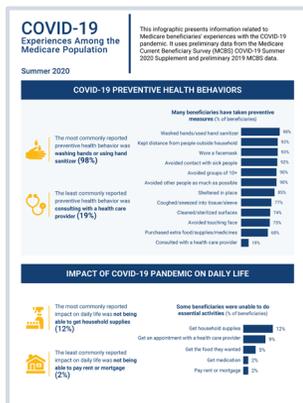
# MCBS Early Looks, Data Briefs, and Infographics



- Starting with data year 2017, the *MCBS Early Looks* are published annually by CMS. They summarize key preliminary estimates prior to the MCBS data release and the MCBS Chartbook publication, facilitating the timely dissemination of MCBS data to the public.

- MCBS data briefs are published periodically by CMS. They summarize key findings using MCBS data on current topics of interest, including depression, patient activation, and the quality of home-based care among Medicare beneficiaries.

- Starting with data year 2018, the MCBS infographics and data snapshots are also published periodically by CMS. They introduce or highlight new MCBS questionnaire content or other topics of interest, such as experiences with chronic pain among Medicare beneficiaries living in the community and the impact of the COVID-19 pandemic on Medicare beneficiaries.



- Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Briefs>

## MCBS Advanced Topic-Based Tutorials

- Additional tutorials are periodically published by CMS as resources for advanced data users.
- These tutorials build on information provided in the *Data User's Guides* and *Methodology Report*, and may provide analytic guidance and examples using MCBS data. Topics of the advanced tutorials include:
  - *Using Community and Facility Data*
  - *Weighting and Variance Estimation*
  - *COVID-19 Supplement Data*
- Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Briefs>

## *MCBS Bibliographies* and Questionnaires

- Annual *MCBS Bibliographies* list all publications that analyze MCBS data in a given year, including journal articles, reports, issue briefs, dissertations, and working papers.
  - Starting with calendar year 2020, the *MCBS Bibliographies* include annotations for each source, which describe the research question, analysis, and findings in order to give data users more information about how MCBS data can be studied.
  - Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Bibliography>
- The annual MCBS Questionnaires and Questionnaire User Documentation include general information on the structure and content of the questionnaire, a description of changes to the questionnaire that year, and detailed specifications for each section of the questionnaire including question text, response options, and skip logic.
  - Link: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires>

# Thank you!

If you have any questions, please contact CMS at the following email address: [MCBS@cms.hhs.gov](mailto:MCBS@cms.hhs.gov).



**Copyright Information:** This communication was produced, published, and disseminated at U.S. taxpayer expense. All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.