

---

# MCBS Advanced Tutorial on Longitudinal Analysis Using MCBS Data

## Version Control Log

<b>Date</b>	<b>Version</b>	<b>Revisions</b>
3/4/22	1.0	Initial version released.

# Section 1: Introduction



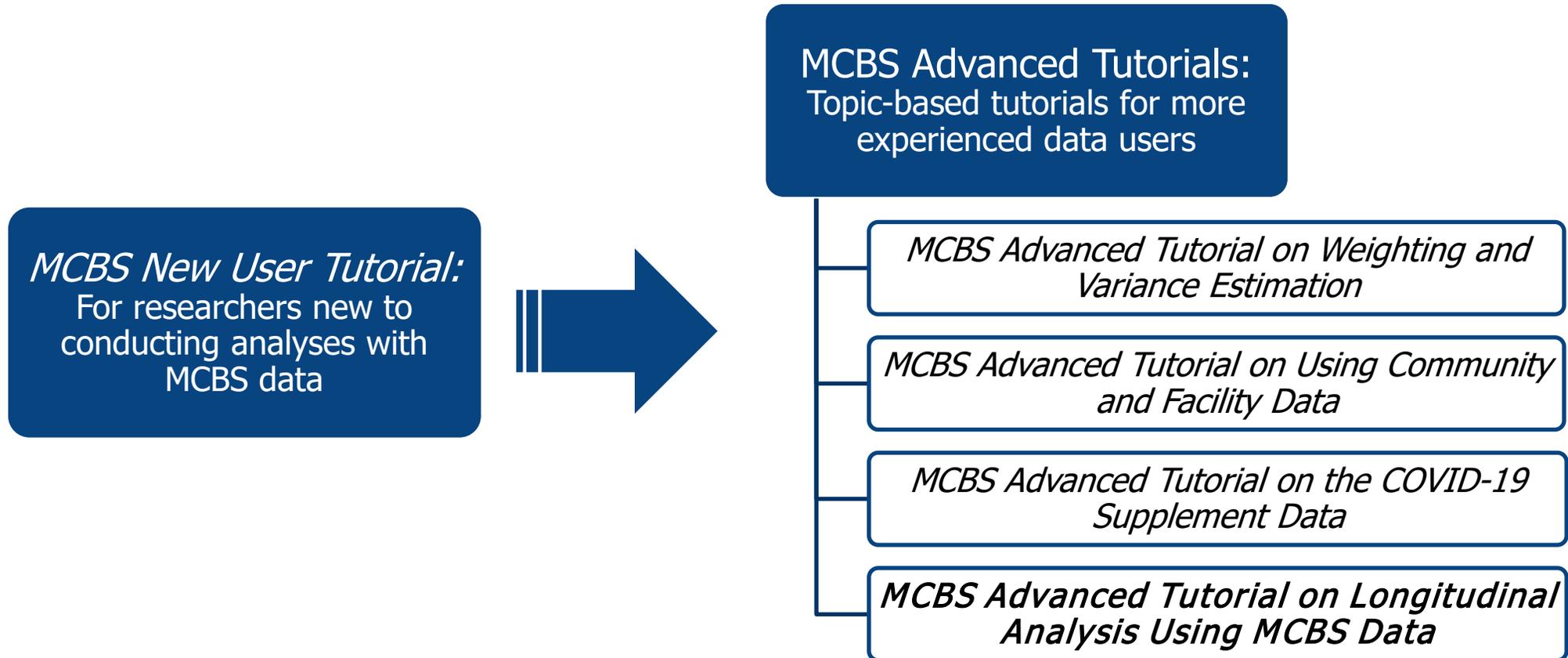


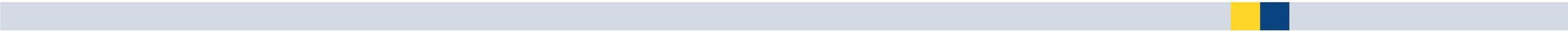
# Learning Objectives

After completing this Medicare Current Beneficiary Survey (MCBS) Advanced Tutorial, a data user will be able to answer the following questions:

- How does the MCBS support longitudinal analysis?
- What are some considerations to account for when developing a longitudinal analysis using MCBS data?
- What types of weights should be used for longitudinal analysis and variance estimation using MCBS data?

# MCBS Instructional Materials





# Tutorial Outline

- Section 1: Introduction
- Section 2: Performing Longitudinal Analysis with MCBS Data
- Section 3: Analytic Examples

# MCBS Overview

- The MCBS is a continuous, multi-purpose longitudinal survey covering a representative national sample of the Medicare population. Interviews are usually conducted in-person using computer-assisted personal interviewing (CAPI); however, conducting interviews by telephone has also been permitted on a limited basis since the origin of the MCBS.
- 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 (U.S.).
- 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).



## MCBS Overview (continued)

- The MCBS serves as the leading source of information on the Medicare program and its impact on beneficiaries, including health care utilization, barriers to care, health care expenditures, and physical, emotional, and financial wellbeing.
- The MCBS is based on a rotating panel design, which allows for longitudinal analysis of up to four years.
- The MCBS collects data in two main components:
  - Community Survey of beneficiaries residing in non-institutional, residential settings.
  - Facility Survey of beneficiaries residing in facilities such as long-term care nursing homes.

## MCBS Overview (continued)

The MCBS sample is based on Medicare Enrollment data, so individual beneficiaries are selected, not addresses or households.

- The MCBS follows beneficiaries who move from the community to a facility, from a facility back to the community, or move from one facility to another during their four years of participation.
- If beneficiaries are unable to answer questions or require language assistance, they can enlist the help of an assistant, such as a family member, to help complete the interview; a proxy can also respond on behalf of the beneficiary if the beneficiary is incapacitated or unable to complete the interview.
- In facilities, interviews are conducted with facility administrators or caregivers on behalf of beneficiaries living in the facility.

# Overview of the Limited Data Set (LDS) Files

- Each year, the MCBS data are made available to users via two annual LDS files.
  - Survey File – includes demographic information, health insurance coverage, self-reported health status and conditions, and responses regarding access to care and satisfaction with care.
  - Cost Supplement File – includes comprehensive accounting of health care use, expenditures, and sources of payment.
- The LDS files contain multiple segments that can be merged using a unique beneficiary key ID (the BASEID variable).
- The LDS files include several sets of weights that can be used to conduct both cross-sectional and longitudinal analysis.
- For each LDS file, CMS provides technical documentation with the following resources for data users: codebooks, *data user's guides*, *questionnaires and questionnaire user documentation*, data files (SAS<sup>®</sup>, CSV), research claims (SAS, CSV), format control files, and sample SAS code to apply the formats and labels for those not using SAS.
- Researchers must have a Data Use Agreement (DUA) in place with CMS to obtain LDS files.

# Overview of the MCBS Public Use File (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, comprising the COVID-19 Summer and Fall 2020 Public Use Files and the COVID-19 Winter 2021 Public Use File.
- 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 publicly-available alternatives that do not require a DUA.
- Longitudinal analysis using MCBS PUF data is not possible. In order to protect the confidentiality of the beneficiaries, the MCBS PUF segments do not include longitudinal weights. Longitudinal weights are only available with the LDS files. For this reason, this tutorial focuses on analysis with the LDS files.

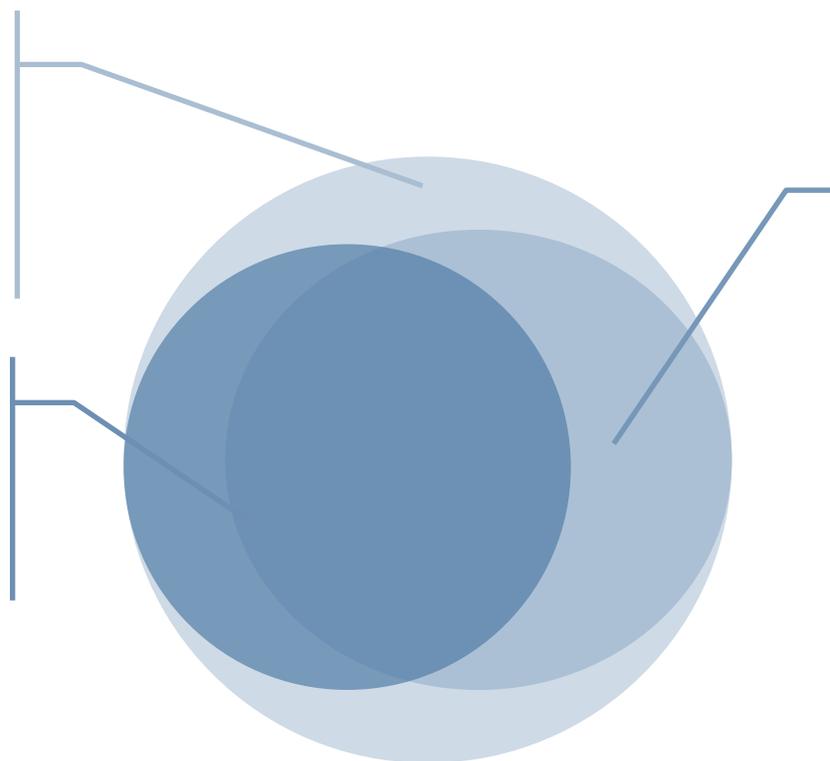
# Medicare Population Represented by the MCBS Data

## 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.

# 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.

- [Tutorials, including the \*New User Tutorial\* and topical Advanced Tutorials](#)
- It is recommended that new MCBS data users refer to the *New User Tutorial* before consulting this or other Advanced Tutorials.
- Annual [Questionnaires and Questionnaire User Documentation](#)
- Data documentation including [Data User's Guides, Methodology Reports, Variable Crosswalks, and codebooks](#) for the LDS files
- Annual [Chartbooks and data tables](#)
- Annual [Bibliographies](#), which include annotations starting in 2020
- Annual [Early Looks and topical infographics](#)

## CMS Website:

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

### Medicare Current Beneficiary Survey (MCBS)

[Questionnaires](#)

[Data Documentation and Codebooks](#)

[Data Tables](#)

[Bibliography](#)

[Data Briefs and Tutorials](#)

## Section 2: Performing Longitudinal Analysis with MCBS Data



# Why Conduct Longitudinal Analysis?

- The study objective in longitudinal analysis is to assess change over time for beneficiaries who completed multiple interviews over multiple years.
- Longitudinal analysis can also be used to assess repeated events or activities over multiple years (i.e., consistency of preventive care use over four years or total out-of-pocket costs over three years).
- Longitudinal analysis can also be used to assess the causal effect of continuous exposure to a risk factor on an outcome of interest on the individual-level.
  - However, analyses that focus on continuous exposure/treatment on outcomes should not be performed using MCBS data.
- Longitudinal analysis allows researchers to evaluate long-term patterns among or between groups or analyze outcomes and utilization over time in relation to specific risk factors or parameters.
- The MCBS data also accommodates cross-sectional trending and pooled cross-sectional trending, however this tutorial focuses on longitudinal analysis.

# Longitudinal Research Questions

## Examples using MCBS data from Published Literature

Is patient activation associated with preventive care utilization (flu vaccine, mammography screening, annual wellness visit, and colorectal screening) among Medicare Fee-for-Service beneficiaries living in the community between 2011 and 2013?

- Ashford, N. C. (2020). *Clinical preventive service utilization among Medicare beneficiaries: An examination of the role of patient activation* (Publication No. 27740896) [Doctoral dissertation, George Washington University]. ProQuest Dissertations and Theses Global.

How are physical and functional disabilities and health status associated with health care utilization (hospitalization, rehospitalization) and mortality rates among Medicare beneficiaries aged 65 and older who are cancer survivors between 2008 and 2013?

- Chavan, P. P., Kedia, S. K., Mzayek, F., Ahn, S., & Yu, X. (2020). Impact of self-assessed health status and physical and functional limitations on healthcare utilization and mortality among older cancer survivors in US. *Aging Clinical and Experimental Research*. Advance online publication.

<https://doi.org/10.1007/s40520-020-01654-5>

Does antidepressant medication use increase risk for falls among Medicare beneficiaries aged 65 and older living in the community between 2009 and 2013?

- Haddad, Y. K., Luo, F., Bergen, G., Legha, J. K., & Atherly, A. (2020). Special report from the CDC antidepressant subclass use and fall risk in community-dwelling older Americans. *Journal of Safety Research*. Advance online publication. <https://doi.org/10.1016/j.jsr.2020.11.008>

# Longitudinal Design of the MCBS

- The MCBS is based on a rotating panel design, which allows for longitudinal analysis of up to four years (for Survey File data) when appropriate longitudinal weights are used.
  - **Note that longitudinal analysis with Cost Supplement File data can only be conducted for up to three years.** Utilization and cost data are not collected for the first year in which beneficiaries are sampled, only in the three subsequent survey years.
- The Survey File cross-sectional and longitudinal population definitions are consistent from year to year, so the data are comparable between years. The Cost Supplement cross-sectional population definition is also consistent and comparable from year to year.<sup>1</sup>

1. The Cost Supplement two-year longitudinal population changed slightly in 2016 from what was defined the last time the two-year longitudinal weights were supplied (i.e., in 2013). In 2013, the two-year longitudinal (i.e., one-year backward longitudinal weight) Cost Supplement weights represented the population that enrolled on or before 1/1/2011 and were still enrolled in 2013 (i.e., enrollees after 1/1/2011 were not included). Beginning in 2016, the two-year longitudinal weights represent a true two-year ever enrolled population (i.e., the 2019 two-year longitudinal weights represent the population of beneficiaries that were ever enrolled in both 2018 and 2019).

# MCBS Data Collection Year

MCBS data collection is continual throughout the year with three distinct seasons (i.e., rounds) of data collection per year.



# MCBS Rotating Panel Design

Each MCBS panel, an annual statistical sample of all Medicare enrollees, is interviewed up to three times a year over the four-year period, creating a continuous profile of selected beneficiaries' health care experiences. One panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round.



<b>Baseline Year</b> <i>Incoming Panel Sample</i>				<b>1</b>	Baseline Interview	
<b>Baseline Year + 1</b> <i>Continuing Panel</i>	<b>2</b>	Second Interview	<b>3</b>	Third Interview	<b>4</b>	Fourth Interview
<b>Baseline Year + 2</b> <i>Continuing Panel</i>	<b>5</b>	Fifth Interview	<b>6</b>	Sixth Interview	<b>7</b>	Seventh Interview
<b>Baseline Year + 3</b> <i>Continuing Panel</i>	<b>8</b>	Eight Interview	<b>9</b>	Ninth Interview	<b>10</b>	Tenth Interview
<b>Baseline Year + 4</b> <i>Continuing Panel</i>	<b>11</b>	Eleventh (Exit) Interview				

# Who Can Be Included In a Longitudinal Analysis with MCBS Data?

- Longitudinal analysis with the MCBS can include beneficiaries who transition between community and facility settings during their participation.
- Respondents remain in the survey as long as they do not skip two consecutive interviews. A respondent who missed a round of data collection can still be captured in a longitudinal analysis provided they complete follow-up interviews.
- Longitudinal weights are only available on the Survey File for surviving beneficiaries, appropriately weighted longitudinal analyses that include death endpoints cannot be conducted with Survey File weights.<sup>1</sup> Decedents are included in the longitudinal weights on the Cost Supplement File.
- MCBS longitudinal weights are representative of the *ever enrolled* population (see slide 12). Beneficiaries may move in and out of Medicare enrollment and still be included in the longitudinal sample. Weighted longitudinal analyses of the *continuously enrolled* population cannot be conducted with MCBS data and it is, thus, important to design research questions appropriately.

1. [Methodological Issues in Using Multiple Years of the Medicare Current Beneficiary Survey \(nih.gov\)](#)

# Considerations for Longitudinal Analysis with MCBS Data

In order to properly account for the complex longitudinal design of the MCBS, there are noteworthy considerations to ensure analysis integrity.

- Data users should confirm that their MCBS variables of interest are available and comparable for all years for which the analysis is being conducted.
  - Questionnaire revisions or data collection changes may impact the comparability and availability of specific variables between data years. Researchers should consult the *MCBS Data User's Guides, Methodology Report*, and annual *Questionnaires and Questionnaire User Documentation* for each data year of interest to confirm whether data are comparable.
  - The *MCBS Variable Crosswalks* can also be used as tools to support researchers in determining availability of variables for each data year.
- Data users should also assess sample sizes for the entire longitudinal sample and confirm that they are sufficient for their needs. This is particularly true when conducting longitudinal analysis with small subpopulations such as racial/ethnic minorities or persons with rare chronic diseases.

# Considerations for Longitudinal Analysis with MCBS Data (continued)

Consideration	Example	Analysis Next Steps
MCBS variables of interest are <u>available</u> for all years for which the analysis is being conducted.	<p><b>Research question:</b> Were beneficiaries who had trouble paying medical bills in 2016 continue to have trouble paying medical bills in 2019?</p> <p><b>Concern:</b> Variable of interest, <i>PAYPROB</i>, was not available prior to the 2017 MCBS data year.</p>	Consider whether analysis can be conducted over a different time period (e.g., 2017 to 2019).
MCBS variables of interest are <u>comparable</u> for all years for which the analysis is being conducted.	<p><b>Research question:</b> Did beneficiaries who changed residential settings (from the community to a facility or vice versa) between 2018 and 2019 receive flu shots as recommended in both years?</p> <p><b>Concern:</b> The variables for receipt of flu shot for beneficiaries living in the community (<i>FLUSHOT</i>) and beneficiaries living in facilities (<i>FLUSHOTF</i>) have different reference periods and are collected in different seasons.</p>	Consider whether the target population can be subset to a population of beneficiaries for whom data are comparable or whether a different variable or analysis time frame may be suitable.
Sample size should be sufficient <sup>1</sup> for all years for which the analysis is being conducted.	<p><b>Research question:</b> How does race/ethnicity influence whether women aged 50-74 received bi-annual mammograms from 2016 to 2019?</p> <p><b>Concern:</b> The cell sizes are very small for each race/ethnicity category for women in the target age range who have mammography data for all four years of interest.</p>	Consider conducting analysis by aggregating subgroups or whether a different variable or analysis time frame may be suitable.

<sup>1</sup>. The absolute sample size that is “sufficient” depends upon the statistical power or other statistical reliability requirements for a given analysis.

# Considerations for Longitudinal Analysis with Expenditures Data

Note that when conducting longitudinal analysis with expenditures data, especially if comparing expenditures across data years, it may be necessary to account for inflation by converting dollars from an earlier data year to inflation-adjusted dollars for the most recent data year. The Consumer Price Index (CPI)<sup>1</sup> can be used for adjustment when comparing consumer spending between two or more years (i.e., beneficiary out-of-pocket spending). The Gross Domestic Product (GDP)<sup>2</sup> can be used for adjustment when comparing trends in expenditures generally.

- Adjusted costs can be calculated as follows using CPI as an example and assuming year  $y1$  is a year preceding year  $y2$  :

$$y1\_AdjustedCost = y1\_Cost * \frac{y2\_CPI}{y1\_CPI}$$

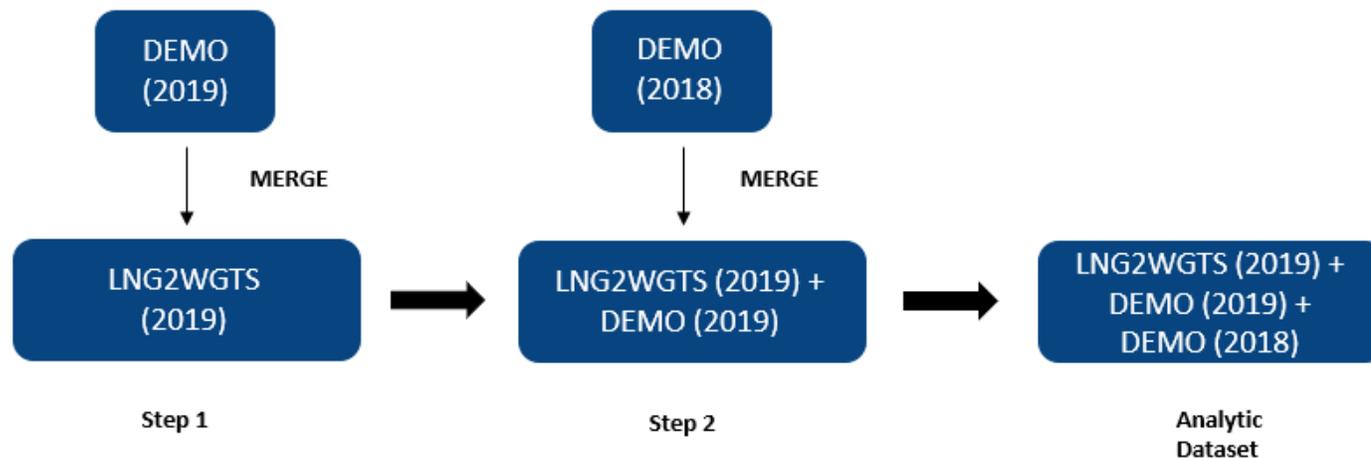
1. U.S. Bureau of Labor Statistics. <https://data.bls.gov/cgi-bin/surveymost?cu>
2. Bureau of Economic Analysis. [Gross Domestic Product | U.S. Bureau of Economic Analysis \(BEA\)](#)

# MCBS Longitudinal Weights

- Longitudinal weights allow for the study of beneficiaries across data years.
- Weights that are part of the most recent annual LDS file release among the years of interest should be used.
  - For example, if a researcher is conducting longitudinal analysis spanning 2016 – 2018, the researcher should use longitudinal weights from the 2018 LDS.
- Data users wishing to conduct joint analysis of both Survey File and Cost Supplement data must use the Cost Supplement weights as the Cost Supplement File includes a smaller subset of the Survey File beneficiaries with complete cost and utilization data for the year.
- Weights should be used to identify the analytic population for a longitudinal analysis. MCBS longitudinal weights are only non-zero for beneficiaries with data for a given longitudinal time frame.
- Weights must be used in order to make any population-level longitudinal inferences with MCBS data.

# Constructing an MCBS Longitudinal Data Set

To construct a longitudinal analytic dataset, the first step is to use the appropriate longitudinal weights file. For example, as shown in the diagram, to assess changes over time for beneficiaries who have been in the sample for at least two years – from 2018 to 2019 – the two-year longitudinal weights should be used.



- Step 1: Variables from current year files representing the outcome of interest should then be merged with the current year's longitudinal weights file.
- Step 2: The same variables from the prior year's files should be merged with the current year's longitudinal weights file.
- While merging in both steps, all observations in the weights file should be preserved.

# MCBS Longitudinal Weights – Survey File LDS

Longitudinal weights provided with the annual Survey File LDS.

Description	Variable Name	Population for 2019 Survey File LDS
Survey File Two-Year Longitudinal Weights	LNG2WGTS	Beneficiaries enrolled on or before 1/1/2018 and surviving and entitled as of Fall 2019. These two-year longitudinal weights are populated only for beneficiaries who are in the 2018 and 2019 Survey File data and were continuously enrolled for two years.
Survey File Three-Year Longitudinal Weights	LNG3WGTS	Beneficiaries who enrolled on or before 1/1/2017 and were still alive and entitled as of Fall 2019. These three-year longitudinal weights are populated only for beneficiaries who were continuously enrolled during all of the years 2017-2019 and had Survey File data in 2017 and 2019.
Survey File Four-Year Longitudinal Weights	LNG4WGTS	Beneficiaries who enrolled on or before 1/1/2016 and were still alive and entitled as of Fall 2019. These four-year longitudinal weights are populated only for beneficiaries who were continuously enrolled during all of the years 2016-2019.

# MCBS Longitudinal Weights – Cost Supplement File LDS

Longitudinal weights provided with the annual Cost Supplement File LDS.

Description	Variable Name	Population for 2019 Cost Supplement File LDS
Cost Supplement Two-Year Longitudinal Weights	CSL2WGTS	For the 2019 Cost Supplement File LDS, these two-year longitudinal weights are populated for members of the 2016, 2017, and 2018 panels who were ever enrolled in Medicare at any time during both 2018 and 2019 and provided utilization and cost data for both years.
Cost Supplement Three-Year Longitudinal Weights	CSL3WGTS	For the 2019 Cost Supplement File LDS, these three-year longitudinal weights are populated for members of the 2016 and 2017 panels who were ever enrolled in Medicare at any time during 2017, 2018, and 2019, and provided utilization and cost data for all three years.

As noted on slide 15, four-year longitudinal analysis is not possible with data from the Cost Supplement File as utilization and cost data are not collected for the first year in which beneficiaries are sampled. Therefore, four-year longitudinal weights are not available on the Cost Supplement File.

## MCBS Longitudinal Weights (continued)

**Note on the population represented in the three-year longitudinal weights:** The three-year longitudinal Survey File weights only require (for the 2019 data year) that 2017 and 2019 be complete, while 2018 can be incomplete. That is in contrast with the Cost Supplement three-year weights, which do require complete data in all three years.

- Because most Survey File items are “snapshot” questions asked at a single point in time each year, the Survey File longitudinal weights are designed to allow for comparison at two points in time (e.g., fall 2017 and fall 2019), which is still possible when data for intervening years are missing (e.g., 2018); most beneficiaries with longitudinal weights do have data for every year, so rates of missingness for intervening years are small.
- Cost Supplement data on utilization and costs, however, are collected for continuous time periods spanning calendar years, so the Cost Supplement longitudinal weights are designed to facilitate analysis over all or part of the span of the included years (e.g., calendar years 2017 through 2019); beneficiaries with longitudinal weights will have Cost Supplement data for the entire time they were enrolled during the included years.

## Variance Estimation for Longitudinal Analysis

Just as there are full-sample longitudinal weights, there are corresponding sets of replicate weights. The replicate weights included in the longitudinal weights data files can be used to calculate standard errors of the sample-based estimates. Data users can consult the *MCBS Advanced Tutorial on Weighting and Variance Estimation* or the *Data User's Guides* for more information on conducting variance estimation.

LDS File	Longitudinal Period	Segment	Replicate Weights
Survey File	Two-year Longitudinal	LNG2WGTS	L2YRS001 – L2YRS100
Survey File	Three-year Longitudinal	LNG3WGTS	L3YRS001 – L3YRS100
Survey File	Four-year Longitudinal	LNG4WGTS	L4YRS001 – L4YRS100
Cost Supplement File	Two-year Longitudinal	CSL2WGTS	CSL2Y001 – CSL2Y100
Cost Supplement File	Three-year Longitudinal	CSL3WGTS	CSL3Y001 – CSL3Y100

# A Note Regarding Longitudinal Analyses with MCBS Data Spanning 2014

Note that to accommodate changes in sampling and data collection methodologies, the 2014 MCBS data were not released.

- In order to perform a longitudinal analysis spanning 2014 of continuously enrolled beneficiaries in the Survey File, utilize 2013 and 2015 MCBS data.
  - The Survey File and Cost Supplement File structure for the LDS was introduced in 2015. In 2013 and years prior, data were released via an Access to Care (ATC) file and a Cost and Use (CAU) file.
- A true longitudinal analysis spanning 2014 of ever enrolled beneficiaries is not possible.
  - Instead, perform a repeated cross-sectional analysis using the cross-sectional weights in the Survey File or Cost Supplement File for the years of interest.
  - For example, an analysis of utilization and cost data for the combined calendar years 2013-2015 is not possible. The user would need to analyze 2013 and 2015 separately, using cross-sectional weights, and compare.

## Section 3: Analytic Examples



# Analyzing MCBS Data



## Analyzing MCBS Data (continued)

- This section will walk 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, 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), weights files, 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 that allow researchers to obtain weighted estimates and estimated standard errors using two approaches: 1) the Taylor-series linearization method and 2) the balanced repeated replication (BRR) method (Fay's method).
- For more information regarding any of the topics discussed, see the *MCBS Data User's Guide*.

## Example 1 Step 1: Define Your Research Question



## Example 1 Step 1: Defining Your Research Question

**What is the association between gender and yearly receipt of a flu shot from 2016 through 2019 among Medicare beneficiaries living in the community?**

- Objectives of this example:
  - Demonstrate how to link Survey File data across years
  - Demonstrate how to apply longitudinal weights to produce nationally representative estimates

## Example 1 Step 2: Create Your Analytic File



## Example 1 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.

## Example 1: Identifying Data File(s), Data Segments, and Variables

- This research question requires the following variables:

<b>Measure</b>	<b>File</b>	<b>Segment</b>	<b>Variable</b>
Gender	2016 Survey File	Demographics (DEMO)	ROSTSEX
Self-reported receipt of flu shot for the last flu season	2016-2019 Survey File	Preventive Care (PREVCARE)	FLUSHOT

## Example 1: Identifying Population and Weights

- As a reminder, the Survey File is representative of the “ever enrolled” population, which includes anyone enrolled at any time during the calendar year. However, the Survey File longitudinal weights are populated only for beneficiaries continuously enrolled for the given longitudinal time period. Users conducting longitudinal analysis of Survey File data combined across multiple years must use the longitudinal weights in the Survey File of the last year of interest, and the weights should be corresponding to the study duration (n years). This analysis requires the 4-year longitudinal weights from the 2019 Survey File LNG4WGTS segment to produce estimates for Medicare beneficiaries living in the community from 2016 to 2019. Thus, this analysis represents beneficiaries who were continuously enrolled in Medicare from the beginning of the first year (2016 / Year 1) to the fall of the last year (2019 / Year n, n=2,3,4).
- To limit this analysis to beneficiaries with complete data for each year between 2016 and 2019, the population is restricted to those living in the community who have flu shot data available on the preventive care segment of the Survey File for each data year. For reference, the questions regarding receipt of flu shots in the community and facility questionnaires have different reference periods, so the data for beneficiaries living in long-term care facilities are therefore not directly comparable.

## Example 1: Merge and Append Segments to Create Analytic File

- The below code shows how Survey File segments from multiple years can be merged with the longitudinal weight segment in SAS using BASEID as the key variable. When merging segments, all observations in the longitudinal weight segment should be preserved. This example utilizes the LNG4WGTS segment for the 4-year longitudinal weights.

```
data flu_merged;  
  merge survey19.LNG4WGTS (in=a)  
    survey16.DEMO (keep=BASEID H_AGE ROSTSEX)  
    survey16.PREVCARE (keep=BASEID FLUSHOT rename=(FLUSHOT=FLUSHOT16))  
    survey17.PREVCARE (keep=BASEID FLUSHOT rename=(FLUSHOT=FLUSHOT17))  
    survey18.PREVCARE (keep=BASEID FLUSHOT rename=(FLUSHOT=FLUTHOST18))  
    survey19.PREVCARE (keep=BASEID FLUSHOT rename=(FLUSHOT=FLUTHOST19));  
  if a then output;  
  by BASEID;  
  
run;
```

## Example 1: Recoding Variables

- There are numerous ways to convert raw variables into analytic variables. This example recodes FLUSHOT from multiple years to one variable measuring the yearly receipt of a flu shot over the 2016-2019 period.

```
data flu_recode;  
  set flu_merged;  
  if FLUSHOT16 = 1 and FLUSHOT17 = 1 and FLUSHOT18 = 1 and FLUSHOT19 = 1 then  
    FLUSHOT_ANNUAL=1;  
  else if FLUSHOT16 = 2 or FLUSHOT17 = 2 or FLUSHOT18 = 2 or FLUSHOT19 = 2 then  
    FLUSHOT_ANNUAL= 0;  
  else if FLUSHOT16 = . or FLUSHOT17 = . or FLUSHOT18 = . or FLUSHOT19 = . then  
    FLUSHOT_ANNUAL= .;  
run;
```

## Example 1: Recoded Variables

Measure	Original variable	Recoded variable
Annual Flu Shot	FLUSHOT16/17/18/19 1 Received flu shot 2 Did not receive flu shot	FLUSHOT_ANNUAL 1 Received flu shot annually from 2016-2019 0 Did not receive flu shot annually from 2016-2019

## Example 1 Step 3: Conduct Analyses Using Appropriate Variance Estimation Methods



# Example 1 SAS Code: Annual receipt of flu shot, Medicare beneficiaries living in the community, by gender, 2016-2019

## Balanced Repeated Replication (BRR) Method

The following code requests the frequency, weighted frequency, percent and standard errors of yearly receipt of a flu shot (FLUSHOT\_ANNUAL) from 2016 through 2019 for Medicare beneficiaries living in the community, by gender, using the BRR method of variance estimation.

```
proc surveyfreq data=flu_recode varmethod=brr (fay=.30);  
  weight L4YRSWGT;  
  repweights L4YRS001-L4YRS100;  
  tables ROSTSEX*FLUSHOT_ANNUAL;  
run;
```

## Example 1 Results: Annual receipt of flu shot, Medicare beneficiaries living in the community, by gender, 2016-2019

What is the association between gender and yearly receipt of a flu shot from 2016 through 2019 among Medicare beneficiaries living in the community?

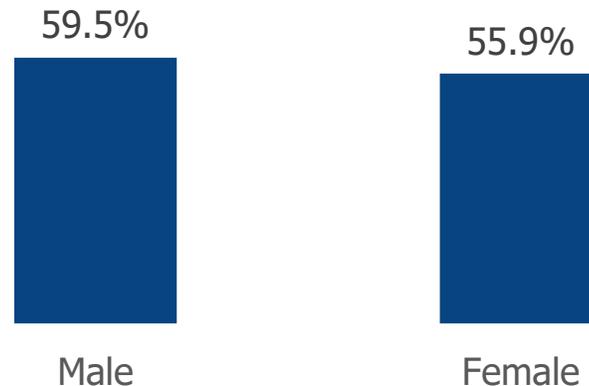
Gender	Unweighted N	Weighted N	Estimate - % (St. Error - %)
Male	912	20,189,004	59.5 (1.86)
Female	1,250	25,318,338	55.9 (1.79)
Total	2,162	45,507,342	

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2016-2019.

NOTES: Estimates are representative of beneficiaries who were continuously enrolled in Medicare from 1/1/2016 and were still alive, living in the community, and eligible and enrolled in Medicare as of Fall 2019. Beneficiaries living in the community answered questions themselves or by proxy.

# Example 1 Results: Annual receipt of flu shot, Medicare beneficiaries living in the community, by gender, 2016-2019

**What is the association between gender and yearly receipt of a flu shot from 2016 through 2019 among Medicare beneficiaries living in the community?**



SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2016-2019.

NOTES: Estimates are representative of beneficiaries who were continuously enrolled in Medicare from 1/1/2016 and were still alive, living in the community, and eligible and enrolled in Medicare as of Fall 2019. Beneficiaries living in the community answered questions themselves or by proxy.

## Example 2 Step 1: Define Your Research Question



## Example 2 Step 1: Defining Your Research Question

**What is the association between year-to-year obesity and total health care expenditures among Medicare beneficiaries aged 65 and over living in the community during 2018 and 2019?**

- Objectives of this example:
  - Demonstrate how to link Survey File and Cost Supplement File segments across data years
  - Demonstrate how to apply longitudinal weights to produce nationally representative estimates

## Example 2 Step 2: Create Your Analytic File

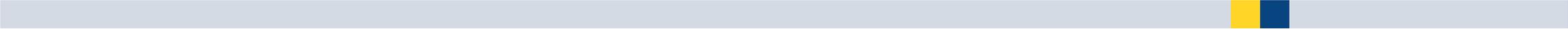


## Example 2: Identifying Data File(s), Data Segments, and Variables

Obesity and total health expenditures for beneficiaries living in the community aged 65 and over

- This research question requires the following variables:

Measure	File	Segment	Variable
Age	2019 Survey File	Demographics (DEMO)	H_AGE
Residence status	2019 Survey File	Demographics (DEMO)	INT_TYPE
Body mass index	2018 & 2019 Survey File	General Health (GENHLTH)	BMI_CAT
Total expenditures	2018 & 2019 Cost Supplement File	Person Summary (PS)	PAMTTOT



## Example 2: Identifying Population and Weights

- The Survey File and Cost Supplement File are representative of the “ever enrolled” population, which includes anyone enrolled at any time during the calendar year. However, the 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 two-year longitudinal analysis of both Survey File and Cost Supplement File data must use the weights in the Cost Supplement File for the most recent data year.

## Example 2: Merge and Append Segments to Create Analytic File

- The SAS code below shows how multiple Survey File segments can be merged with the CSL2WGTS segment in SAS using BASEID as the key variable. When merging segments from 2018, all observations in the CSL2WGTS segment of the 2019 Cost Supplement File should be preserved.

```
data bmi19_merged;  
  merge cost19.CSL2WGTS (in=a)  
        survey19.DEMO (keep=BASEID H_AGE INT_TYPE)  
        survey19.GENHLTH (keep=BASEID BMI_CAT)  
        cost19.PS (keep=BASEID PAMTTOT);  
  by BASEID;  
  rename BMI_CAT=BMI_CAT19 PAMTTOT=COST19;  
  if a then output;  
run;
```

```
data bmi18_merged;  
  merge survey18.DEMO (keep=BASEID)  
        survey18.GENHLTH (keep=BASEID BMI_CAT)  
        cost18.PS (keep=BASEID PAMTTOT);  
  by BASEID;  
  rename BMI_CAT=BMI_CAT18 PAMTTOT=COST18;  
run;
```

```
data bmi_all;  
  merge bmi19_merged (in=b)  
        bmi18_merged (in=c);  
  by BASEID;  
  if b and c then output;  
run;
```

## Example 2: Converting Real Dollars to Inflation-Adjusted Dollars

- Prior to comparing costs between 2018 and 2019, one needs to account for inflation by converting dollars to inflation-adjusted dollars. The Consumer Price Index (CPI)<sup>1</sup> can be used for adjustment when comparing spending between two or more years.

```
data bmi_adj;  
  set bmi_all;  
  COST18_ADJ = COST18* (251.712/247.867);  
run;
```

1. U.S. Bureau of Labor Statistics. <https://data.bls.gov/cgi-bin/surveymost?cu>

## Example 2: Recoding Variables

- There are numerous ways to convert raw variables into analytic variables. This example creates COSTDIFF and recodes BMI\_CAT from 2018 and 2019 into a two-category variable as such:

```
data bmi_recode;
  set bmi_adj;
  BMI18=.;
  BMI19=.;
  if BMI_CAT18 in (1,2,3) then BMI18=0; else if BMI_CAT18 in (4,5) then BMI18=1;
  if BMI_CAT19 in (1,2,3) then BMI19=0; else if BMI_CAT19 in (4,5) then BMI19=1;
  COSTDIFF=COST19-COST18_ADJ;
run;

data bmi_final;
  set bmi_recode;
  BMI=.;
  if BMI18=0 and BMI19=0 then BMI=1; /* Beneficiary was not obese in 2018 or 2019 */
  if BMI18=1 and BMI19=1 then BMI=2; /* Beneficiary remained obese in 2018 and 2019 */
  if BMI18=1 and BMI19=0 then BMI=3; /* Beneficiary was no longer deemed obese in 2019 */
  if BMI18=0 and BMI19=1 then BMI=4; /* Beneficiary was deemed obese in 2019 */
run;
```

## Example 2: Recoded Variables

Measure	Original variable	Recoded variable
Body Mass Index	BMI_CAT18/19 1 Underweight 2 Healthy 3 Overweight 4 Obese 5 Extreme or high risk obesity	BMI18/19 1 Obese 0 Not obese
Change in Body Mass Index	BMI18/19 1 Obese 0 Not obese	BMI 1 Beneficiary was not obese in 2018 or 2019 2 Beneficiary remained obese in 2018 and 2019 3 Beneficiary was no longer deemed obese in 2019 4 Beneficiary was deemed obese in 2019 . Data not available

## Example 2: Segmenting the Population

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

```
data bmi_seg;  
  set bmi_final;  
  where INT_TYPE='C' and H_AGE GE 65;  
run;
```

## Example 2 Step 3: Conduct Analyses Using Appropriate Variance Estimation Methods



## Example 2 SAS Code: Total health care expenditures, Medicare beneficiaries aged 65 and over living in the community in 2018 and 2019, by change in BMI

### Balanced Repeated Replication (BRR) Method

The following code requests the 2018 and 2019 total health care expenditures among beneficiaries aged 65 and over living in the community by their change in BMI using the BRR method for variance estimation.

```
proc surveymeans data= bmi_seg varmethod=brr (fay=.30) missing;  
  var COSTDIFF COST18_ADJ COST19;  
  weight CSL2YWGT;  
  repweights CSL2Y001-CSL2Y100;  
  domain BMI;  
run;
```

## Example 2 Results: Total health care expenditures, Medicare beneficiaries aged 65 and over living in the community in 2018 and 2019, by change in BMI

**What is the association between year-to-year obesity and total health care expenditures among Medicare beneficiaries aged 65 and over living in the community during 2018 and 2019?**

<b>Change in BMI</b>	<b>2018 Estimate - \$ (St. Error- \$)</b>	<b>2019 Estimate - \$ (St. Error- \$)</b>
Beneficiary was not obese in 2018 or 2019	11,075 (484)	13,849 (619)
Beneficiary remained obese in 2018 and 2019	14,037 (913)	16,088 (828)
Beneficiary was no longer deemed obese in 2019	15,787 (3,201)	18,432 (2,232)
Beneficiary was deemed obese in 2019	16,361 (3,368)	18,034 (3,651)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2018 & 2019.

NOTES: Estimates are representative of beneficiaries who were continuously enrolled in Medicare from 1/1/2018 and were still alive, living in the community, and eligible and enrolled in Medicare as of Fall 2019. Beneficiaries living in the community answered questions themselves or by proxy. 2018 estimates are shown in 2019 dollars.



Please refer to the *MCBS Data User's Guides* and other *MCBS Advanced Tutorials* for more information on conducting analyses with MCBS data.

# Thank you!

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

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.

**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.

