
MCBS Advanced Tutorial on Pooled Cross-Sectional Analysis Using MCBS Data

Version Control Log

Date	Version	Revisions
3/28/23	1.0	Initial version released.



Tutorial Outline

- Section 1: Introduction
- Section 2: Performing Pooled Cross-Sectional Analysis with MCBS Data
- Section 3: Analytic Example

Section 1: Introduction





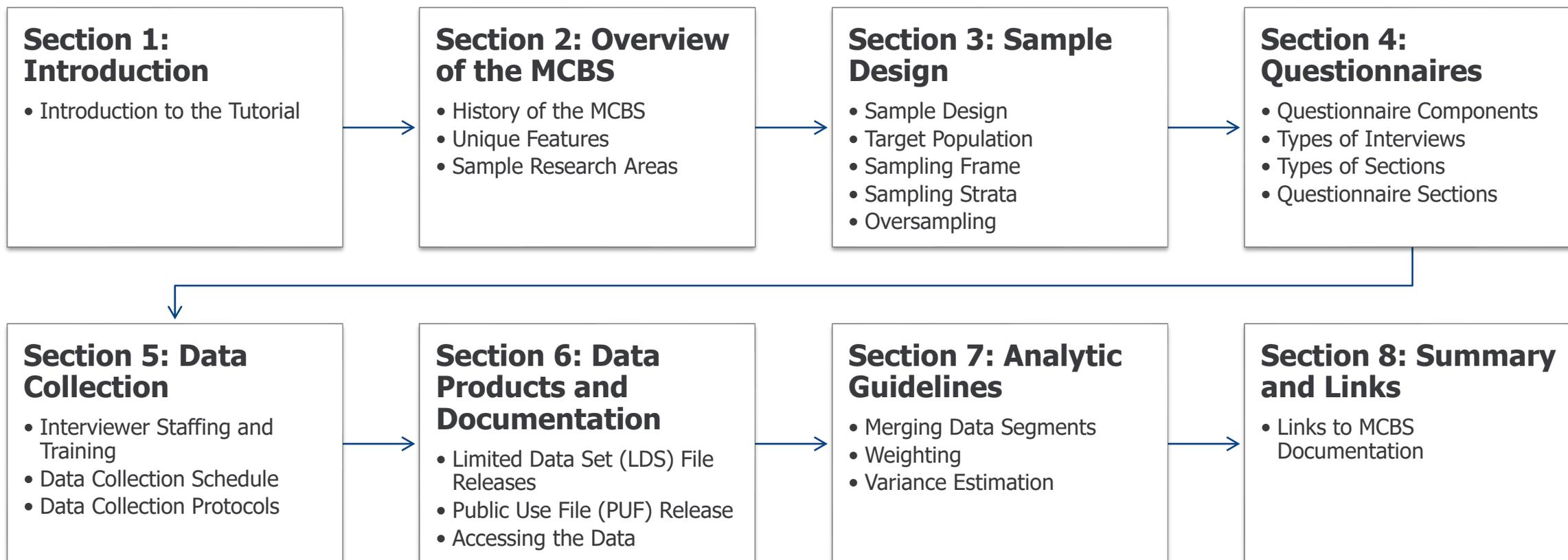
Learning Objectives

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

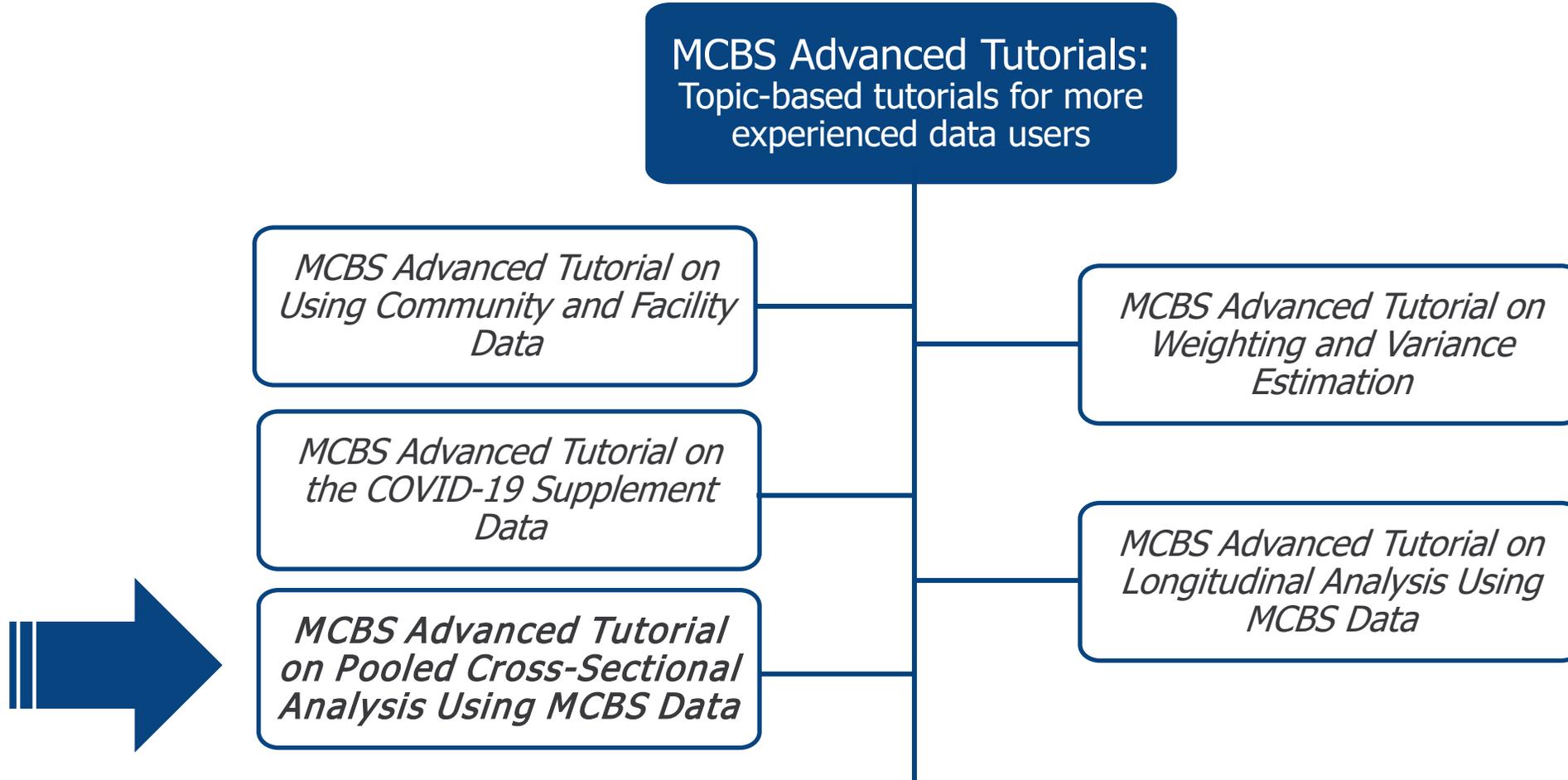
- Why conduct pooled cross-sectional analyses with MCBS data?
- How does the MCBS support pooled cross-sectional analysis?
- What are some considerations for developing a pooled cross-sectional analysis using MCBS data?
- What types of weights should be used for pooled cross-sectional analysis and variance estimation using MCBS data?

MCBS New User Tutorial Outline

- It is recommended that new MCBS data users refer to the [New User Tutorial](#) before consulting this or other Advanced Tutorials. Below is a graphic displaying the information found in the *New User Tutorial*.



Overview of MCBS Advanced Tutorials



MCBS Documentation and Resources

- CMS provides a wide array of MCBS documentation 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](#)
 - 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, data briefs, 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 Pooled Cross-Sectional Analysis with MCBS Data



Why Conduct Pooled Cross-Sectional Analysis of MCBS data?

- The MCBS is a longitudinal survey with a rotating panel design that supports both longitudinal¹ and cross-sectional analyses. Longitudinal analysis can be used to analyze beneficiary-level changes over time while cross-sectional analysis can be used to analyze population-level characteristics at a particular point in time (e.g., a data year).
- Cross-sectional analyses of sub-populations for which MCBS sample sizes are small may not provide sufficient statistical power to draw conclusions in any one data year.
- Pooled cross-sectional analysis allows the data user to combine data across years, therefore increasing sample sizes and statistical power.
- Pooled cross-sectional analysis yields estimates that are, in effect, a moving average of nationally representative year-specific estimates. The pooled estimates can be interpreted as being representative of the midpoint of the pooled period.

¹For more information on using MCBS data to conduct longitudinal analysis, please see the [MCBS Advanced Tutorial on Longitudinal Analysis Using MCBS Data](#).



Design Considerations for Pooled Cross-Sectional Analysis

- To account for the complex sample design features of the MCBS, important adjustments to variance estimation are required.
- 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 is both serial and intra-cluster correlation in MCBS data.
 - Many records are retained from one year to the next (serial correlation).
 - The same set of Primary Sampling Units (PSUs) and Secondary Sampling Units (SSUs) are used for each year (intra-cluster correlation).
- This tutorial includes guidance for using the Balanced Repeated Replication (BRR) method of variance estimation to account for these correlations when conducting pooled cross-sectional analysis.

Comparability Considerations for Pooled Cross-Sectional Analysis

- For all years included in the analysis, data users should confirm that:
 - The MCBS variables of interest are available, comparable, and consistent across all years.
 - Pooling the data across years results in sample sizes sufficient for the analysis.
- Both the Survey File and the Cost Supplement cross-sectional population definitions are consistent from year to year, so the data are generally comparable between years.
- Questionnaire revisions or data collection changes may impact the comparability and availability of specific variables between data years. Data users should consult the [MCBS Data User's Guides](#), [Methodology Report](#), and annual [Questionnaire Specifications](#) for each data year of interest to confirm data are comparable.
- The [MCBS Variable Crosswalks](#) show the availability of variables for each data year and can be used to identify variables that are good candidates for a pooled cross-sectional analysis.

Select Considerations for Pooled Cross-Sectional Analysis with MCBS Data

Consideration	Example	Next Steps
MCBS variables of interest are <u>available</u> for all pooled data years.	<p>Research question: To what extent did beneficiaries with diabetes have trouble paying medical bills from 2016-2019?</p> <p>Concern: Variable of interest, <i>PAYPROB</i>, was not available prior to the 2017 data year.</p>	Consider whether analysis can be conducted over a different time period (e.g., 2017-2019).
MCBS variables of interest are <u>consistent</u> across pooled data years. ¹	<p>Research question: Did beneficiaries with Medicare Advantage have difficulty getting referrals to a certain type of specialist from 2017-2020?</p> <p>Concern: The LDS Survey File Variable Crosswalk indicates that the variable of interest, <i>MADVSPCL</i>, was renamed as <i>MPSPCLTY</i> in 2019. The construction of the variable, however, did not change.</p>	Standardize the naming convention of the variable of interest across pooled data years.
Sample size should be sufficient for the analysis. ²	<p>Research question: What is the relationship between race/ethnicity and bi-annual mammograms among women aged 50-74 from 2016-2019?</p> <p>Concern: 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 pooling additional year(s) of data (e.g., 2016-2020).

1. Data users should confirm that MCBS variables of interest are consistent in terms of naming convention, construction, and administration schedule across years of interest.

2. The absolute sample size that is “sufficient” depends upon the statistical power or other statistical reliability requirements for a given analysis.

Examples in the Literature of Pooled Cross-Sectional Research Questions Using MCBS Data

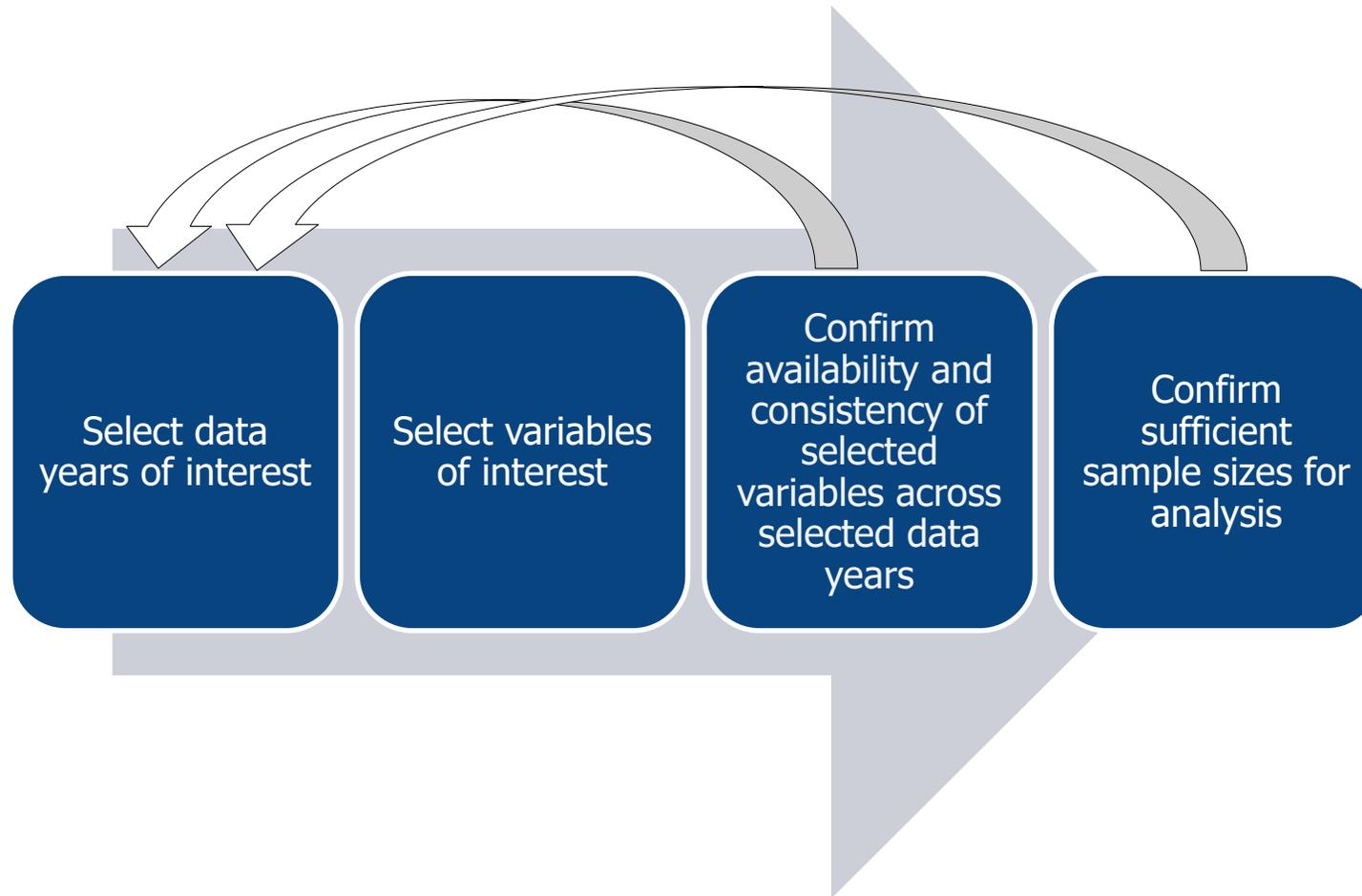
- How does the use of heart failure medication impact expenditures by Medicare and other payers from 2010-2011?¹
- What is the relationship between asthma-chronic obstructive pulmonary disease comorbidity and medication non-adherence due to cost among Medicare Fee-for-Service beneficiaries aged 65 and over who lived in the community from 2006-2013?²
- What is the relationship between residential settings/services and total healthcare expenditures among Medicare beneficiaries aged 65 and over with instrumental activities of daily living (IADL) impairment from 2001-2013?³

¹McGee, B. T., Parikh, R., & Phillips, V. (2021). Cost implications of patient spending on heart failure medications in the US Medicare program. *Journal of Pharmaceutical Health Services Research*, 12(2), 101-108. <https://doi.org/10.1093/jphsr/rmaa018>.

²Nili, M., Adelman, M., Madhavan, S. S., LeMasters, T., Dwibedi, N., & Sambamoorthi, U. (2021). Asthma-chronic obstructive pulmonary disease overlap and cost-related medication non-adherence among older adults in the United States. *Journal of Asthma*, 1-15. <https://doi.org/10.1080/02770903.2020.1868497>.

³Akincigil, A., & Greenfield, E. A. (2020). Housing Plus Services, IADL Impairment, and Healthcare Expenditures: Evidence From the Medicare Current Beneficiaries Survey. *The Gerontologist*, 60(1), 22-31. <https://doi.org/10.1093/geront/gny181>.

Selecting MCBS Variables for Pooled Cross-Sectional Analysis



Note: This process is iterative. If variables are unavailable or inconsistent across years or sample size is insufficient, researchers may return to earlier steps to add data years or select different variables.

Constructing an MCBS Dataset

The graphic below displays the steps to construct a pooled cross-sectional analytic dataset using MCBS data, using the Health Insurance Summary (HISUMRY) segment as an example.



- Step 1: Variables representing the outcome of interest should be merged with the cross-sectional weights file (CENWGTS) for each year in the analysis.
 - While merging, all observations in the weights file should be preserved.
- Step 2: The year-specific files should be concatenated to produce the analytic dataset.

Variance Estimation for Pooled Cross-Sectional Analysis

- Recall that there is both serial and intra-cluster correlation in MCBS data due to the complex sample design.
- Use the **BRR method** to account for various correlations.
 - This method uses 100 sets of replicate weights along with each full-sample weight to estimate variances.
 - For example, the Survey File full-sample weight CEYRSWT is accompanied by the replicate weights CEYRS001-CEYRS100.
 - PROC SURVEYFREQ and PROC SURVEYMEANS in SAS can estimate pooled variances; to use these SAS procedures, specify:
 - The full-sample weight in the 'weight' statement
 - The replicate weights in the 'repweight' statement
 - The 'Fay' value, which is 0.30 for the MCBS¹
 - No need to specify sampling strata or clusters or define domains or subgroups when using BRR method
 - See example PROC SURVEYFREQ code in the Analytic Example section of this tutorial

¹Fay's Method is a generalization of BRR and is used to provide estimates of variance when there are only two PSUs in some sampling strata. Fay's coefficient can take values from 0.00 to 0.50; the recommended value of 0.30 was chosen as a standard for the MCBS to provide a balance between the stability of variance estimates and the slight increases in variance seen with larger values of Fay's coefficient.

Additional Notes for Estimation by Pooled Analysis

- Estimates of population means using pooled data will reflect estimated values as of the midpoint of the pooled data period.
- Estimates of population totals using pooled data will reflect a total across all years included in the pooled data period.
- Pooled cross-sectional analysis is not used to estimate change over time.
 - Variances of estimates of year-to-year change are more complicated to estimate and are more appropriately obtained by a longitudinal or repeated cross-sectional analysis.
 - See [*Advanced Tutorial on Weighting and Variance Estimation*](#) and [*Advanced Tutorial on Longitudinal Analysis*](#) for details on estimating variances for differences over time.
- The same weights should be used for both single-year cross-sectional analysis and pooled cross-sectional analysis; no further adjustments are needed for weighting pooled cross-sectional analyses.
- For more information and guidance on appropriate use of weights, please reference Exhibit 9.4.1 of the data year-specific [*MCBS Survey File Data User's Guide*](#).



A Note Regarding Pooled Cross-Sectional Analyses with MCBS Data Spanning 2014

- To accommodate changes in sampling and data collection methodologies, the 2014 MCBS data were not released.
 - 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 pooled cross-sectional analysis spanning 2014 of ever enrolled beneficiaries is not possible due to lack of 2014 MCBS data.

Section 3: Analytic Example



Analyzing MCBS Data



Analyzing MCBS Data (continued)

- This section walks 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 BRR method (Fay's method).
- For more information on these topics, see the [MCBS Data User's Guide](#).

Step 1: Define Your Research Question



Step 1: Define Your Research Question (continued)

Among Medicare beneficiaries aged 80 and over and living in the community with high-impact chronic pain (HICP) in 2020, what is the association between dual eligibility status and trouble getting needed care?

- Objective of this example:
 - Demonstrate how to use MCBS data in a pooled cross-sectional analysis when one year of data is insufficient
 - Demonstrate how to combine Survey File data across years for pooled cross-sectional analysis

Note: High-impact chronic pain (HICP) is defined as beneficiaries who report having pain on most/every day in the previous 3 months with ≥ 1 concomitant activity limitation/participation restriction.

Step 2: Create Your Analytic File



Step 2: Create Your Analytic File (continued)

- 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 and recode variables as necessary.

Note: Steps in this process may be iterative (as demonstrated in this example).

Step 2: Create Your Analytic File (continued)

- *Identify the MCBS data file(s) that your research question requires.*
- *Identify the data segments and variables that your research question requires.*

Measure	File	Segment	Variable
Age	2020 Survey File	Demographics (DEMO)	D_STRAT
Type of interview	2020 Survey File	Demographics (DEMO)	INT_TYPE
Frequency of pain	2020 Survey File	Chronic Pain (CHRNPAIN)	PAINOFTN
Pain limitations	2020 Survey File	Chronic Pain (CHRNPAIN)	PAINLIMT
Trouble getting needed health care	2020 Survey File	Access to Care (ACCESSCR)	HCTROUBL
Dual eligibility status	2020 Survey File	Health Insurance (HISUMRY)	H_OPMDCD

Step 2: Create Your Analytic File (continued)

3. *Identify considerations specific to coding, data collection, and/or processing of variables of interest.*
- Estimate should be suppressed when the unweighted frequency for a category of the denominator variable is less than 50 due to small sample size.
 - When using MCBS data, the unweighted frequency for the dual eligibility status variable requires suppression due to a denominator of less than 50.
 - Due to lack of sufficient sample size for analysis, researchers should consider modifying the research question to pool multiple years of MCBS data.

Dual Eligibility Status	Unweighted Frequency	Suppression
Full-/partial-benefit dually eligible	<50	Yes
Not dually eligible	269	No

Step 1: (Re)define Your Research Question

Among Medicare beneficiaries aged 80 and over and living in the community with HICP in 2018-2020, what is the association between dual eligibility status and trouble getting needed care?

- Objective of this example:
 - Demonstrate how to use MCBS data in a pooled cross-sectional analysis when one year of data is insufficient
 - Demonstrate how to combine Survey File data across years for pooled cross-sectional analysis

Step 2: (Re)create Your Analytic File

1. *Identify the MCBS data file(s) that your research question requires.*
 2. *Identify the data segments and variables that your research question requires.*
- This updated research question requires the Survey File LDS from three data years.

Measure	File	Segment	Variable
Age	2018-2020 Survey Files	Demographics (DEMO)	D_STRAT
Type of interview	2018-2020 Survey Files	Demographics (DEMO)	INT_TYPE
Frequency of pain	2018-2020 Survey Files	Chronic Pain (CHRNPAIN)	PAINOFTN
Pain limitations	2018-2020 Survey Files	Chronic Pain (CHRNPAIN)	PAINLIMT
Trouble getting needed health care	2018-2020 Survey Files	Access to Care (ACCESSCR)	HCTROUBL
Dual eligibility status	2018-2020 Survey Files	Health Insurance (HISUMRY)	H_OPMDCD

Step 2: (Re)create Your Analytic File (continued)

3. *Identify considerations specific to coding, data collection, and/or processing of variables of interest.*
- Estimate should be suppressed when the unweighted frequency for a category of the denominator variable is less than 50 due to small sample size.
 - When using 2018-2020 MCBS data, the unweighted frequency for the dual eligibility status variable does not require suppression due to a denominator of greater than 50.
 - With the updated pooled analysis research question, sample sizes are sufficient for the analysis to continue.

Dual Eligibility Status	Unweighted Frequency	Suppression
Full-/partial-benefit dually eligible	155	No
Not dually eligible	762	No

Step 2: Create Your Analytic File

4. *Identify the study population and use the corresponding weights.*

- The population is restricted to Medicare beneficiaries living in the community who reported sufficient information to determine their age and if they were living with or without HICP from 2018-2020.
- The estimates will be generated using data from a topical questionnaire section (CHRNPAIN), therefore, the special non-response adjusted weights included in the CHRNPAIN topical segment should be used (CPSEWT, CPSE1-CPSE100) from 2018, 2019, and 2020.

Step 2: Create Your Analytic File (continued)

5. *Merge segments to create your analytic dataset and recode variables as necessary.*

- The code on the next two slides shows how Survey File segments from multiple years can be merged with the appropriate weights segment using BASEID as the key variable.
- While merging, all observations in the weights file should be preserved.
- This example creates year-specific datasets using 2018, 2019, and 2020 data and then concatenates those files to produce an analytic dataset, `pain_merged`, containing data from all three years of interest.

Step 2: Create Your Analytic File (continued)

```
data pain_merged_2018;  
    merge    survey18.DEMO (keep=BASEID D_STRAT INT_TYPE)  
            survey18.CHRNPAIN (keep=BASEID PAINOFTN PAINLIMIT CPSEWT CPSE1-CPSE100)  
            survey18.ACCESSCR (keep=BASEID HCTROUBL)  
            survey18.HISUMRY (keep=BASEID H_OPMDCD)  
            ;  
    by BASEID;  
run;
```

```
data pain_merged_2019;  
    merge    survey19.DEMO (keep=BASEID D_STRAT INT_TYPE)  
            survey19.CHRNPAIN (keep=BASEID PAINOFTN PAINLIMIT CPSEWT CPSE1-CPSE100)  
            survey19.ACCESSCR (keep=BASEID HCTROUBL)  
            survey19.HISUMRY (keep=BASEID H_OPMDCD)  
            ;  
    by BASEID;  
run;
```

Step 2: Create Your Analytic File (continued)

```
data pain_merged_2020;  
    merge    survey20.DEMO (keep=BASEID D_STRAT INT_TYPE)  
            survey20.CHRNPAIN (keep=BASEID PAINOFTN PAINLIMIT CPSEWT CPSE1-CPSE100)  
            survey20.ACCESSCR (keep=BASEID HCTROUBL)  
            survey20.HISUMRY (keep=BASEID H_OPMDCD)  
            ;  
    by BASEID;  
run;  
  
data pain_merged;  
    set pain_merged_2018 pain_merged_2019 pain_merged_2020;  
    where INT_TYPE = "C";  
run;
```

Note: INT_TYPE = "C" limits the analysis to only beneficiaries living in the community.

Step 2: Create Your Analytic File (continued)

- This example creates four recoded variables: age, HICP status, health care access trouble, and dual eligibility status in the concatenated dataset pain_merged.

```
data pain_recode;
  set pain_merged;
  if D_STRAT in (1:5) then AGE2 = 1; /*Beneficiary under 80*/
  else if D_STRAT in (6:7) then AGE2 = 2; /*Beneficiary 80 or over*/
  else AGE2 = .;
  if PAINOFTN in (3, 4) AND PAINLIMT in (3, 4) then HICP = 1; /*Beneficiary has HICP*/
  else if PAINOFTN in (1, 2) OR PAINLIMT in (1, 2) then HICP = 0; /*Beneficiary does not have HICP*/
  else HICP = .;
  if HCTROUBL = . then ACCESS_TROUBLE = .;
  else if HCTROUBL = 1 then ACCESS_TROUBLE = 1; /*Beneficiary has health care access trouble*/
  else if HCTROUBL = 2 then ACCESS_TROUBLE = 2; /*Beneficiary does not have trouble*/
  if H_OPMDCD = . then ANYDUAL = .;
  else if H_OPMDCD in (1,3,4) then ANYDUAL = 1; /*Beneficiary is dually eligible*/
  else if H_OPMDCD = 2 then ANYDUAL = 2; /*Beneficiary is not dually eligible*/
run;
```

Step 2: Create Your Analytic File (continued)

Measure	Original Variable	Recoded Variable
Age	D_STRAT 1 0-44 2 45-64 3 65-69 4 70-74 5 75-79 6 80-84 7 85+	AGE2 1 Less than 80 years 2 80+ years
HICP	PAINOFTN 1 Never 2 Some days 3 Most days 4 Every day PAINLIMT 1 Never 2 Some days 3 Most days 4 Every day	HICP 1 Yes, high-impact chronic pain 0 No, not high-impact chronic pain

Step 2: Create Your Analytic File (continued)

Measure	Original Variable	Recoded Variable
Health Care Access Trouble	HCTROUBL 1 Yes 2 No	ACCESS_TROUBLE 1 Yes 2 No
Dual Eligibility Status	H_OPMDCD 1 Full Dual 2 Non-Dual 3 Partial Dual 4 QMB Dual	ANYDUAL 1 Full-/partial-benefit dually eligible 2 Not dually eligible

Step 3: Conduct Analyses Using Appropriate Variance Estimation Methods



SAS Code: Trouble getting needed health care among Medicare beneficiaries aged 80 and over and living in the community with HICP, by dual eligibility status, 2018-2020

BRR Method

The following code requests the frequency, weighted frequency, percent, and standard errors of dual eligibility status (ANYDUAL) by trouble getting needed health care (ACCESS_TROUBLE) from 2018 through 2020 for Medicare beneficiaries aged 80 and over and living in the community with HICP.

```
proc surveyfreq data=pain_recode varmethod=brr (fay=.30);  
  weight CPSEWT;  
  repweights CPSE1-CPSE100;  
  tables ACCESS_TROUBLE*ANYDUAL / col cl(type=cp) chisq1 nocellpercent ;  
  where HICP=1 and AGE2=2;  
run;
```

Results: Trouble getting needed health care among Medicare beneficiaries aged 80 and over and living in the community with HICP, by dual eligibility status, 2018-2020

What is the association between dual eligibility status and trouble getting needed health care from 2018 through 2020 among Medicare beneficiaries aged 80 and over living in the community with HICP?

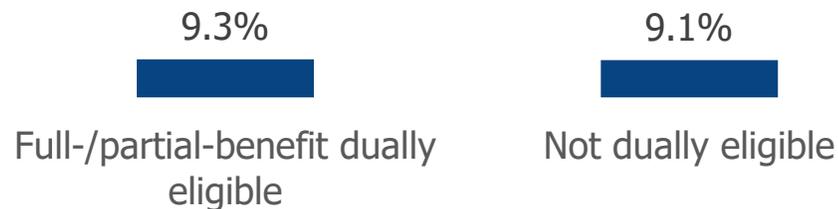
Dual Eligibility Status	Unweighted N	Weighted N	Estimate - % (St. Error - %)
Full-/partial-benefit dually eligible	13	56,343	9.3 (2.87)
Not dually eligible	66	287,713	9.1 (1.19)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2018-2020.

NOTES: Estimates for 2018-2020 were based on Medicare beneficiaries living in the community who reported sufficient information to determine their age and if they were living with or without HICP.

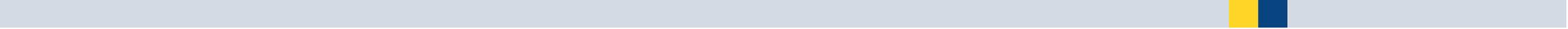
Results: Trouble getting needed health care among Medicare beneficiaries aged 80 and over and living in the community with HICP, by dual eligibility status, 2018-2020

Among beneficiaries with HICP, the percentage of beneficiaries who had trouble getting needed care does not differ much by dual eligibility status.



SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2018-2020.

NOTES: Estimates for 2018-2020 were based on Medicare beneficiaries living in the community who reported sufficient information to determine their age and if they were living with or without HICP.



Please refer to the [*MCBS Data User's Guides*](#) and other [*MCBS Advanced Tutorials*](#) or 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.

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.

