MCBS Advanced Tutorial on Using Community and Facility Data

MCBS
Medicare Current Beneficiary Survey

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES
OFFICE OF ENTERPRISE DATA & ANALYTICS

Version Control Log

Date	Version	Revisions
1/6/20	1.0	Initial version released based on data year 2017.

Section 1: Introduction

MCBS
Medicare Current Beneficiary Survey

Learning Objectives

- After completing this Medicare Current Beneficiary Survey (MCBS) advanced tutorial, you will be able to answer the following questions:
  - What are the similarities and differences in sample design, questionnaire instruments, data collection, and data processing methods for the Community and Facility questionnaire components?
  - Where can you find data for the Community component and the Facility component in the MCBS Limited Data Set (LDS) data files?
  - Which LDS variables can you use to identify a beneficiary's residence status?
  - What are the analytic guidelines for combining data collected by the Community and Facility components?

Advanced Tutorial Outline

- Section 1: Introduction
- Section 2: Overview of Community and Facility Data in the MCBS
- Section 3: Analytic Guidelines for Combining Community and Facility Data
- Section 4: Analytic Examples for Cross-sectional Analysis Using Community and Facility Data
- Section 5: Analytic Examples for Longitudinal Analysis Using Community and Facility Data
- Appendix: SAS Code for Analytic Examples

Introduction to the MCBS

- The Medicare Current Beneficiary Survey (MCBS) is a continuous, in-person, multi-purpose longitudinal survey of a nationally representative sample of the Medicare population.
- The MCBS consists of a sample of beneficiaries aged 65 and over and beneficiaries aged 64 and below with disabilities, residing 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. A leading source of information on Medicare and its impact on beneficiaries, the MCBS provides important information on beneficiaries that is not available in CMS administrative data and plays an essential role in monitoring and evaluating beneficiary health status and health care policy.

Inclusion of Facility Beneficiaries in the MCBS

- To obtain an accurate representation of all Medicare beneficiaries, the MCBS sample includes all beneficiaries regardless of residence status. The MCBS follows beneficiaries into and out of long-term care facilities to maintain a comprehensive profile of their health care utilization and expenditures.
- About 5-8% of the sample are beneficiaries who live in a long-term care facility (henceforth referred to as Facility) or who alternate between living in the Community and living in a Facility.
- Unlike beneficiaries living in the Community, beneficiaries living in Facilities or their proxies do not complete an interview. Instead, an interviewer conducts the interview with appropriate facility staff and abstracts some information from medical records to reduce burden on facility staff.
  - For example, data on health status and functioning are abstracted from the Long-Term Care Minimum Data Set (MDS) when it is available.

MCBS Definition of Facility

- A facility interview is conducted when the beneficiary lives in a long-term care or other residential facility that meets the following MCBS definition of a facility:
  - 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** is 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, walking, preparing meals, etc.).
- For the purposes of this tutorial, data collection for beneficiaries living in facilities is referred to as the "Facility component" and data collection for beneficiaries living in the community is referred to as the "Community component." Further, data collected from the Facility component are referred to as "Facility data" and data collected from the Community component are referred to as "Community data."

Introduction to the Advanced Tutorial on Using Community and Facility Data

- This tutorial is intended to provide an overview of the differences between MCBS data collected from beneficiaries living in the community and those living in facilities and analytic guidance on when and how these data should and should not be combined.
- Analytic decisions about whether to include all beneficiaries regardless of residence status, or those living only in the community or only in facilities, are driven by both the research question and data limitations. However, since the underlying data collection approach differs based on the beneficiary's residence status, caution must be observed when combining data across these populations to address questions requiring analysis of all Medicare beneficiaries.
- In addition to reviewing the *MCBS New User's Tutorial*, this advanced tutorial builds on information provided in the MCBS *Data User's Guides, Methodology Reports*, *Questionnaire User Documentation*, and other documentation, which provide documentation of the differences between Community and Facility data.

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 *Questionnaire User Documentation*
  - Data documentation including *Data User's Guides*, *Methodology Reports* and codebooks for the Limited Data Set files
  - Annual *Chartbooks* and data tables
  - Annual bibliographies
  - Topical data briefs and tutorials, including the *New User Tutorial*

CMS Website:
https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index

Section 2: Overview of Community and Facility Data in the MCBS

MCBS
Medicare Current Beneficiary Survey

Key Differences in How Community and Facility Data are Collected

Sampling

No differences in sampling methods. Sampling for the MCBS is person-based, not setting-based.

Data Collection

Key Difference	Community Component	Facility Component
Survey Respondent	Administered to the beneficiary or a proxy.	Administered to facility staff. Interviewers abstract some information from medical records to reduce burden on facility staff.
Questionnaire Scope	Core sections collect data on socio-demographics, health insurance, utilization, cost, experiences with care, and health status. Topical sections collect data on housing characteristics, health behaviors, and knowledge and decision-making about Medicare.	Core sections collect data on socio-demographics, health insurance, utilization, cost, and health status. Facility questionnaire does not include topical sections. Questionnaire is designed to align with the information available to facility staff. For example, it collects fewer details about the beneficiary's income and assets and supplemental insurance compared to Community.
Questionnaire Reference Periods	Because of the differences in questionnaire administration, the reference periods for similar items can differ between the Community and Facility components. For example, questions about flu shot and pneumonia shot have different reference periods for the Community and Facility components in the 2017 Survey File.	
Utilization Data Collection	Collects event-level utilization data for all service types the beneficiary received.	Collects counts per month of utilization for services provided inside and outside the facility.
Costs Data Collection	Collects event-level cost data for all services types, which is linked to the utilization events.	Collects per-diem costs associated with each facility stay; costs for services outside of the facility are not collected.

Key Differences in How Community and Facility Data are Processed

Data Processing

Key Difference	Community Component	Facility Component
Administrative Sources	The administrative data sources available for use in data processing for both components are the same, however they are used differently depending on component.	
Claims Matched to Survey Data	All FFS claims	Only Skilled Nursing Facility (SNF) and Inpatient FFS claims
Claims Added to Survey Data	Data from all types of FFS claims are used to fill in gaps in the survey-collected data.	
Imputation	Conducted for items each respective questionnaire would have collected. For example Facility only collects total income so only total income is imputed. The more detailed income and assets variables collected for Community only are imputed for Community only.	

Takeaway:

- The Community component collects costs and utilization for all events, whereas the Facility component collects utilization for events inside and outside the Facility but only collects costs for the Facility inpatient stay. Therefore, the lack of non-FFS claims results in larger gaps in the Facility data than in the Community data.

Overview of MCBS Limited Data Set (LDS) Segments by Data Source

- Because of the differences in data collection and data processing, it is important to know which LDS file segments contain data sourced from the Community component, Facility component, or administrative data.

● **Community Component** ■ **Facility Component** ◆ **Administrative Data**

Survey File

- ACCESSCR ●
- ACCSSMED ●
- ADMNUTLS ◆
- ASSIST ●
- CHRNCOND ●
- CHRNCDFL ◆
- DEMO ● ■ ◆
- DIABETES ●
- FACASMNT ■
- FACCHAR ■
- FALLS ●
- FOODINS ●
- GENHLTH ●
- HISUMRY ● ◆
- HITLINE ● ■ ◆
- HHCHAR ●
- INTERV ● ■
- INCASSET ●
- MAPLANQX ●
- MCREPLNQ ●
- MDS3 ◆
- MENTHLTH ●
- MOBILITY ●
- NAGIDIS ●
- NICOALCO ●
- OASIS ◆
- PMUSE ●
- PNTACT ●
- PREVCARE ●
- RESTMLN ● ■
- RXMED ●
- RXPARTD ●
- SATWCARE ●
- USCARE ●
- VISHEAR ● ■

Cost Supplement

- DUE ● ■ ◆
- FAE ■ ◆
- IPE ● ■ ◆
- IUE ● ■ ◆
- MPE ● ■ ◆
- OPE ● ■ ◆
- PME ● ◆
- PS ● ■ ◆
- SS ● ■ ◆

For full definitions of the LDS segments, please reference the *Data User's Guides*.

Overview of MCBS Limited Data Set (LDS) Segments by Universe

- As shown in the previous slide, for some topics in the Survey File LDS (such as health and functional status) both Community and Facility data are available but included in separate LDS segments. The table below provides an overview by topic of the LDS segments that include data for each beneficiary universe.
- If the analysis includes all Medicare beneficiaries and the data are in separate Survey File LDS segments, data users will need to appropriately merge segments before analyzing the data. Question text, response categories, and reference period may differ between Community and Facility variables, requiring recoding in order to combine the variables for analysis.

Topic	Segments with Community Data	Segments with Facility Data	Segments with Data for All Beneficiaries
Health Status	GENHLTH; FALLS; CHRNCOND; MENTHLTH; OASIS	FACASMNT; MDS3	
Functional Status & Assistance with Long-Term Care Needs	ASSIST; NAGIDIS; OASIS; MOBILITY	FACASMNT; MDS3	
Demographics and Socio-Economic Status	INCASSET		DEMO
Health Insurance Coverage			HISUMRY; HITLINE; ADMNUTLS
Cost and Utilization		FAE*	ADMNUTLS; DUE; IPE; IUE; MPE; OPE; PME; PS; SS

*Costs and utilization inside the facility are summarized on the SS and PS LDS segments. Utilization outside of the facility is available on the LDS segment FAE only.

Identifying Beneficiaries by Residence Status in the Survey File LDS

- There are two ways to identify a beneficiary's residence status:

	Interview Type	Beneficiary Classification Type
Survey File LDS Segment	DEMO	RESTMLN
Variable Name	INT_TYPE	D_TYPE
Code Frame	C = Completed only Community interviews in the data year F = Completed only Facility interviews in the data year B = Completed both Community and Facility interviews in the data year	C = Lived in the Community the full year F = Lived in a Facility the full year B = Spent at least one day in both settings
Definition	Interview type as defined by the questionnaire completed for the beneficiary.	Beneficiary classification as defined by the number of days the beneficiary spent in each setting.

Differences between INT_TYPE and D_TYPE

- For most beneficiaries, INT_TYPE and D_TYPE are consistent. However, because INT_TYPE reflects the type of **interviews** conducted in a year and D_TYPE reflects the **days** spent in the two settings, it is possible for the INT_TYPE and D_TYPE to be different if a beneficiary was administered only one type of interview but spent days in both settings.
  - A beneficiary could have spent a week in a facility but was only administered Community interviews, resulting in INT_TYPE=C and D_TYPE=B.

Which Definition of Community and Facility should I use?

- Whether you use INT_TYPE or D_TYPE in your analysis depends on your research question.
  - If the research question requires distinguishing between the type of interviews administered, INT_TYPE provides this information.
  - If the research question requires knowing the number of days spent in a setting, D_TYPE provides this information.
- Because the Community and Facility components have differences in type and extent of data collected, analysts need to account for these differences and identify the best universe for analysis. The way to identify these universes is by using INT_TYPE. Therefore, the rest of the tutorial is going to focus on using **INT_TYPE** to define the universe.
  - For more information on using INT_TYPE, see Section 10, “Data File Notes” in the *MCBS Survey File Data User’s Guide*.

Section 3: Analytic Guidelines for Combining Community and Facility Data

MCBS
Medicare Current Beneficiary Survey

Analytic Scenarios for Combining Data for the Community and Facility Populations

Data users may encounter the following analytic scenarios when conducting an analysis using MCBS data:

1. Data are available from the Community or Facility components but not both
2. Data are available from both Community and Facility components and there are no differences in the LDS variable coding, data collection, and processing
3. Data are available from both the Community and Facility components but the LDS variable constructs and/or coding are different
4. Data are available from both the Community and Facility components and the LDS variable constructs and coding are similar, but data collection and/or processing methods are different

The steps on the next slide provide guidance on the recommended step-by-step procedures for combining data from Community and Facility components.

Recommended Steps for Combining Community and Facility Data

Define the Population	Define the population based on research questions and identify Community and Facility populations
Assess Availability of Data	Identify the LDS segments and variables associated with each of the study's domains to determine what data are available for the Community and Facility components
Determine if Recoding is Needed	Assess whether the "universe", level of measurement, and response categories are similar for both Community and Facility components
Recode to Align Community and Facility	Recode LDS variables to align the coding between Community and Facility components and create analytic variables
Merge Data from Community and Facility	Merge the Community and Facility segments with the appropriate weights segments. Assess preliminary estimates for variation between Community and Facility
Review Documentation to Determine Limitations	Review documentation to determine if there are underlying differences in data collection and processing
Conduct Analysis	Conduct analysis and document potential limitations

Use LDS variable DEMO:INT_TYPE

Review the *Data User's Guide*

Review the LDS Codebooks

Review Questionnaire User Documentation, *Methodology Report*, and *Data User's Guides*.

Section 4: Analytic Examples for Cross-sectional Analysis Using Community and Facility Data

MCBS
Medicare Current Beneficiary Survey

Combining Data from the Community and Facility Components to Conduct Analysis

The analytic examples* in the slides to follow show how the step-by-step guidance introduced earlier can be applied to the following analytic scenarios:

- Data are available for the Community or Facility component but not for both **(Example 1)**
- Data are available for both Community and Facility components and there are no differences in the LDS variable coding, data collection, and processing **(Example 2)**
- Data are available for both Community and Facility components but LDS variable constructs and coding are different **(Example 3)**
- Data are available for both Community and Facility components, and the LDS variable constructs and coding are similar, but data collection and processing methods are different **(Examples 4)**

***SAS code for all examples can be found in the Appendix at the end this tutorial.**

Example 1: Data are available for the Community or Facility component but not for both

Research Question: What is the level of satisfaction with the amount paid for prescription drugs for Medicare beneficiaries in calendar year 2017 (CY2017)?

Define the Population

The research question is applicable to all beneficiaries. The target population includes beneficiaries enrolled in Medicare Part D who are also continuously enrolled in 2017 and still alive, and entitled in Summer 2018.

Assess Availability of Data

`Satisfaction with Rx costs' (**MCAMTPAY**) – the outcome variable - is in the **RXMED** segment, which is a topical segment collected for Community only

Found in the *2017 Survey File Data User's Guide*

Example 1: Data are available for the Community or Facility component but not for both

Review documentation to apply proper weights

Applying topical weights:

- Special non-response adjustment weights are included in the **RXMED** segment to account for survey non-response from the Fall to Summer data collection period. This includes the weight **RXCWT** and replicate weights **RXC1-RXC100**
- When merging with appropriate weights, all observations in the weights segment file should be preserved

Found in the *2017 Survey File Data User's Guide*

Limit the analysis based on availability of data

RXMED is a topical segment administered to Community in the Summer round. There is no satisfaction measure for Facility because the measure asks for beneficiary's opinions, which are not accessible to the facility staff. The *Data User's Guide* explains that there could be a small number of **INT_TYPE**=F cases in RXMED because of when Summer topical sections are conducted versus released for analysis.

As a result, conduct analysis with data where **INT_TYPE**=C and document that the study is limited to beneficiaries with data for the Community component who had Part D insurance coverage during CY2017.

Found in the *2017 Survey File Data User's Guide*

Example 1: Data are available for the Community or Facility component but not for both

Conduct Analysis and Document Limitations

Level of Satisfaction with the Amount Paid for Prescription Drugs Among Beneficiaries with Community Component Data and Enrolled in Medicare Part D (CY2017)

MCAMTPAY	Description	Unweighted N	Weighted N (Standard Error)	Percent - % (Standard Error)
1	Very Satisfied	1,815	10,673,632 (305,778)	28.0 (0.8)
2	Satisfied	3,225	19,680,583 (420,711)	51.5 (0.9)
3	Dissatisfied	627	3,950,941 (170,721)	10.3 (0.5)
4	Very dissatisfied	182	1,280,666 (110,770)	3.4 (0.3)
5	No experience	418	2,596,153 (162,606)	6.8 (0.4)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2017.

Analytic Note: Beneficiaries with Facility component data were excluded. Data collection occurred during the Summer of 2018. Only beneficiaries with Community component data who were enrolled in Medicare Part D in the past year (H_PARTD) were included.

Please refer to Appendix, Example 1 for detailed guidance on constructing the analytic file and the SAS code for producing estimates.

Example 2: Data are available for both Community and Facility components and there are no differences in the LDS variable coding, data collection, and processing

Research Question: What is the average Part-D deductible for Medicare beneficiaries enrolled in Part-D during CY2017?

Define the Population

The research question applies to all Medicare beneficiaries, including beneficiaries living in a facility

Found in the *2017 Survey File Data User's Guide*

Assess Availability of Data

- Part D Contract IDs are flagged in the **HISUMRY** segment by month (**H_PRTD01-H_PRTD12**)
- Deductible payments are located in the **HISUMRY** segment by month (**H_DDED01-H_DDED12**)

Found in the *LDS Codebooks*

Assess LDS Variable Coding

- Variables **H_DDED01-H_DDED12** capture information about deductible payments for beneficiaries in both the Community and Facility components
- Therefore, recoding of the LDS variable is not required for analysis

Example 2: Data are available for both Community and Facility components and there are no differences in the LDS variable coding, data collection, and processing

Assess LDS Variable Coding

- Variables **H_DDED01-H_DDED12** capture information about deductible payments for beneficiaries in both the Community and Facility components
- Therefore, recoding of the LDS variable is not required for analysis

Found in the *LDS Codebooks*

Review Documentation to Assess Underlying Differences in Data Collection and Processing

Information on Part-D enrollment and deductible payments is sourced from administrative data, which covers all beneficiaries. Therefore, there are no differences in data collection and processing between the Community and Facility components.

Found in the *2017 Survey File Data User's Guide*

Example 2: Data are available for both Community and Facility components and there are no differences in the LDS variable coding, data collection, and processing

Conduct Analysis and Document Limitations

Average Annual Part-D Deductible Payments for Medicare Beneficiaries Enrolled in Medicare Part-D (CY2017)

Unweighted N	Mean Deductible - $ (Standard Error)
9,368	187.6 (3.1)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2017.

Analytic Note: Information on Medicare Part-D enrollment and deductible payments is sourced from administrative data.

Please refer to Appendix, Example 2, found at the end of this tutorial, for detailed guidance on constructing the analytic file and the SAS code for producing estimates

Example 3: Data are available for both Community and Facility components; LDS variable constructs are different

Research Question: What is the prevalence of cognitive impairment among all beneficiaries aged 65 and over in CY2017?

Define the Population

Cognitive impairment impacts beneficiaries living in both the Community and Facilities. Excluding beneficiaries with Facility component data would result in substantial underestimation of the prevalence of cognitive impairment in the Medicare population. Stratification by age group is recommended when the research question involves health status because beneficiaries under 65 are in a distinctly different health situation from beneficiaries over 65.

Assess Availability of Data

- Information on cognitive impairment for beneficiaries with Community component data is available in the **NAGIDIS** and **MENTHLTH** segments
- Information on cognitive impairment for beneficiaries with Facility component data is available in the **FACASMNT** segment

Found in the *2017 Survey File Data User's Guide*

Example 3: Data are available for both Community and Facility components; LDS variable constructs are different

Assess LDS Variable Coding

- The **NAGIDIS** and **MENTHLTH** segments contain questions related to the beneficiary's ability to concentrate and include data from the Patient Health Questionnaire (PHQ-9)
- The **FACASMNT** segment contains questions related to the beneficiary's memory, recall, decision-making, capacity to understand, and overall mental status as assessed through the Brief Interview for Mental Status (BIMS)
- In order to properly restrict and analyze the data, information on demographics and weights for all beneficiaries is needed:
  - The **DEMO** segment contains information on interview type (**INT_TYPE**) and beneficiary age (**H_AGE**)

Found in the *LDS Codebooks*

Example 3: Data are available for both Community and Facility components; LDS variable constructs are different

Recode Community variables to create binary indicator of cognitive impairment

Cognitively impaired (=1) if:

- beneficiary has “difficulty concentrating/remembering/deciding”

OR

- has “trouble concentrating for more than half the days or nearly every day” as per the Patient Health Questionnaire (PHQ-9)

Recode Facility variables to create binary indicator of cognitive impairment

Cognitively impaired (=1) if:

- beneficiary has issue with memory, recall, or decision-making

OR

- has a diagnosis related to cognitive impairment

OR

- is moderately to severely cognitively impaired as per the BIMS (0 – 12)

Example 3: Data are available for both Community and Facility components; LDS variable constructs are different

Found in the *2017 Survey File Data User's Guide*

Review documentation to properly restrict, weight, and analyze

- When merging with appropriate weights, all observations in the weights segment file should be preserved
- When generating estimates:
  - Based on the research question, restrict data to all beneficiaries aged 65 and over
  - Use appropriate survey commands to generate preliminary estimates stratified by residence status and compare the rates to external benchmarks

Example 3: Data are available for both Community and Facility components; LDS variable constructs are different

Conduct Analysis and Document Limitations

Prevalence of Cognitive Impairment among Medicare Beneficiaries, Aged 65 and Over, by Residence Status (CY2017)

Residence Status	COGIMP	Unweighted N	Weighted N (Standard Error)	Percent - % (Standard Error)
Community	Yes	1,788	6,432,108 (195,141)	13.9 (0.4)
	No	9,312	39,897,059 (235,392)	86.1 (0.4)
Facility	Yes	677	1,032,621 (48,027)	75.9 (1.8)
	No	183	327,490 (26,921)	24.1 (1.8)
Total	Yes	2,528	7,627,326 (213,410)	15.9 (0.4)
	No	9,533	40,318,464 (233,988)	84.1 (0.4)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2017.
NOTE: Estimates are not presented for the residence status of "Both", although they are included in the total.

Limitations: Data collection methods are substantially different for the Community and Facility components. Cognitive status is self-reported or administered to a proxy for beneficiaries with Community component data. Cognitive status is abstracted from medical records or reported by facility staff for beneficiaries with Facility component data.

Please refer to Appendix, Example 3, found at the end of this tutorial, for detailed guidance on constructing the analytic file and the SAS code for producing estimates

Example 4: Data are available for both Community and Facility components, and the LDS variable constructs and coding are similar, but data collection and processing methods are different

Research Question: What was the average annual spending per beneficiary on inpatient and outpatient care for all Medicare beneficiaries enrolled in the program during CY2017?

Define the Population

The research question applies to all Medicare beneficiaries, including those living in Facilities.

Example 4: Data are available for both Community and Facility components, and the LDS variable constructs and coding are similar, but data collection and processing methods are different

Assess Availability of Data

- The LDS segment **HISUMRY** has information for all beneficiaries on whether they had some or no enrollment in a group health plan, such as Medicare Advantage, during the year (**H_GHPSW**)
- The LDS segment **PS** contains information for all beneficiaries on the adjusted sum of inpatient (**PAMTIP**) and outpatient events (**PAMTOP**)

Found in the *2017 Survey File Data User's Guide*

Found in the *2017 Cost Supplement Data User's Guide*

Assess LDS Variable Coding

- The **HISUMRY** segment shows that the variable **H_GHPSW** captures information on group health participation for both the Community and Facility components
- Therefore, recoding of the LDS variable is not required for analysis

Found in the *LDS Codebooks*

Example 4: Data are available for both Community and Facility components, and the LDS variable constructs and coding are similar, but data collection and processing methods are different

Review Documentation to Assess Underlying Differences in Data Collection and Processing

The Facility component does not collect costs associated with events occurring outside of the facility. FFS claims are used to fill this gap, however since claims for other payers are not available, any costs and utilization associated with other payers for events occurring outside of the facility are mostly unavailable.

Preliminary assessment of **H_GHPSW** stratified by type of interview (**INT_TYPE**) reveals that the estimates for Facility may be systematically underestimating the costs and utilization of beneficiaries with Facility component data.

Found in the *2017 Cost Supplement Data User's Guide*

Group Health Participation by Residence Status (CY2017)

Residence Status	Medicare Advantage participation – Unweighted N	Medicare Advantage participation – Weighted N (Standard Error)	Fee-for-service only – Unweighted N	Fee-for-service only – Weighted N (Standard Error)
Community	3,106	20,176,661 (2,036,782)	5,204	36,885,169 (2,142,222)
Facility	190	398,479 (245,379)	599	1,360,338 (499,555)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2017.

Analytic Note: Whenever the Data User's Guide points to a substantial difference in data collection and processing methods between Community and Facility components, it is helpful to stratify output for your variable of interest by residence status in order to quantify these differences and guide further analysis.

Example 4: Data are available for both Community and Facility components, and the LDS variable constructs and coding are similar, but data collection and processing methods are different

Conduct Analysis and Document Limitations

Average Annual Spending per Beneficiary on Inpatient and Outpatient Care for All Medicare Beneficiaries Enrolled in the Program (CY2017)

Residence Status	Medicare Advantage / Fee-For-Service	Adjusted Sum for Inpatient Events – Weighted Mean ($)	Adjusted Sum for Outpatient Events – Weighted Mean ($)
Community	MA	2,439	1,239
	FFS	3,443	2,196
Facility	MA	290	309
	FFS	5,754	2,733

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2017.

Limitations: Data collection and processing methods are substantially different for the Community and Facility components. For Community, all survey data on MA and FFS events are included in the data and FFS events are matched with claims data. The Facility component does not collect costs for services outside of the facility. Therefore, the data underrepresents cost and utilization for beneficiaries living in Facilities.

Please refer to Appendix, Example 4, found at the end of this tutorial, for detailed guidance on constructing the analytic file and the SAS code for producing estimates

Section 5: Analytic Examples for Longitudinal Analysis Using Community and Facility Data

MCBS
Medicare Current Beneficiary Survey

Combining Data from the Community and Facility Components to Conduct Longitudinal Analysis

Data users may need to take additional steps when data from the Community and Facility components are combined to conduct longitudinal analyses. The analytic examples in the following slides illustrate some of the additional steps that data users will need to take to appropriately combine the data, conduct the analysis, and document potential limitations.

For beneficiaries who transition between living in the Community and living in a Facility during the analytic period, LDS variables related to the research question:

- may only be available for the Community or Facility component but not for both **(Example 5)**; or
- may be available for both Community and Facility components, but data collection and/or processing methods are different **(Example 6)**.

Example 5: Data are available for the Community or Facility component but not for both

Research Question: Was there a change in satisfaction with availability of care from specialists for beneficiaries who switched from FFS to MA between CY2016 and CY2017 compared to those who were enrolled in MA during both years?

Define the Population

The research question is applicable to all beneficiaries because satisfaction with availability of care from specialists is relevant to both populations.

Assess Availability of Data

- `Satisfaction with care available from specialists' (**MCSPECAR**) – the outcome variable – is in the **SATWCARE** segment, which is collected by the Community component only
- `Participation in MA' (**H_GHPSW**) – the variable required to assess change in coverage during the time period – is in the **HISUMRY** segment, which includes all beneficiaries

Found in the 2016 and 2017 *Survey File Data User's Guide*

Limit the analysis based on availability of data

- Since **SATWCARE** only includes Community data, retain only those beneficiaries in the 2017 Survey File who are in the two-year longitudinal weights segment (**LNG2WGTS**) AND have Community component data throughout the analytic period (i.e., CY2016-CY2017)
- Use type of interview (**INT_TYPE**) in the 2016 and 2017 DEMO segments to identify beneficiaries with Community component data throughout the analytic period

Example 5: Data are available for the Community or Facility component but not for both

Conduct Analysis and Document Limitations

Change in Satisfaction with Availability of Care from Specialists Among Beneficiaries Living in the Community – MA Beneficiaries who Switch from FFS

	Unweighted N	Weighted N	Satisfaction with Availability of Care from Specialists	
			Dissatisfied / Very Dissatisfied in 2016 Percent - % (Standard Error)	Dissatisfied / Very Dissatisfied in 2017 Percent - % (Standard Error)
Beneficiaries enrolled in FFS in 2016 and MA in 2017	213	1,812,295	5.0 (1.8)	6.2 (1.7)
Beneficiaries enrolled in MA in 2016 and 2017	2,463	15,823,385	5.3 (0.6)	4.9 (0.5)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File, 2017.

Analytic Note: Beneficiaries with Facility component data during CY2016-CY2017 were excluded from the analysis.

Please refer to Appendix Example 5, found at the end of this tutorial, for detailed guidance on constructing the analytic file and the SAS code for producing estimates

Example 6: Data are available for the Community and Facility component, but data collection and/or processing methods are different

Research Question: What is the median percent change in out-of-pocket payments for Medicare beneficiaries between CY2016 and CY2017?

Define the Population

The research question is applicable to all beneficiaries, including those who lived in a facility at any time during CY2016-CY2017.

Assess Availability of Data

`Adjusted sum of out-of-pocket payments' (**PAMTOOP**) – the outcome variable – is in the **PS** segment of the Cost Supplement file, which includes information for all beneficiaries.

Found in the 2016 and 2017 *Survey File Data User's Guide*

Assess LDS Variable Coding

- The LDS variable– **PAMTOOP** in the **PS** segment – captures information on out-of-pocket payments for all beneficiaries
- Therefore, no additional recoding steps are required to analyze the LDS variable

Example 6: Data are available for the Community and Facility component, but data collection and/or processing methods are different

Found in the 2016 and 2017 *Survey File Data User's Guide*

Review Documentation to Assess Underlying Differences in Data Collection and Processing

The Community component collects out-of-pocket payments for all health care services rendered to the beneficiary and survey responses are reconciled with all FFS claims. The Facility component does not collect costs associated with events occurring outside of the facility so the only source of data on these events are FFS claims and imputed payers.

As expected, out-of-pocket payments are substantially higher for beneficiaries with Facility component data compared to those with Community component data. Therefore, stratifying the estimates based on whether beneficiaries remained in the same setting or had a change in their residence status is recommended.

Median Out-of-Pocket Payment for Medicare Beneficiaries by Residence Status (CY2016-CY2017)

Residence Status	Median Out-of-Pocket Payment in CY2016 - $ (Standard Error)	Median Out-of-Pocket Payment in CY2017 - $ (Standard Error)
Community	1,000.0 (37.7)	1,021.7 (35.2)
Facility or Both	15,197.0 (1,338.8)	12,700.0 (1,024.4)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2017.
NOTE: Median income in CY2016 and CY2017 are shown in 2017 dollars.

Example 6: Data are available for the Community and Facility component, but data collection and/or processing methods are different

Conduct Analysis and Document Limitations

Median Percent Change in Out-of-Pocket Payments for Medicare Beneficiaries between CY2016-CY2017

Residence Status during CY2016-CY2017	Weighted N (Unweighted N)	Median Out-of-Pocket Payment in CY2016 - $ (Standard Error)	Median Change in Out-of-Pocket Payment between CY2016-CY2017 - $ (Standard Error)	Median Percent Change in Out-of-Pocket Payment between CY2016-CY2017 - % (Standard Error)
Remained in Community	527,638,845 (4,548)	999.7 (37.2)	3.9 (9.2)	1.3 (2.6)
Remained in Facility	1,716,455 (427)	15,493.0 (1,329.6)	-293.3 (215.3)	-2.9 (2.1)
Transitioned from Community to Facility	517,401 (61)	1,135.5 (380.7)	6,234.0 (3,859.0)	500.0 (239.5)

SOURCE: Centers for Medicare & Medicaid Services, Medicare Current Beneficiary Survey, Survey File and Cost Supplement File, 2017.
NOTE: Median out-of-pocket payment in CY2016 is shown in 2017 dollars. Change in out-of-pocket payments and percent change in out-of-pocket payments are inflation-adjusted. Due to small sample sizes, beneficiaries who transitioned from the Facility to the Community component between CY2016-CY2017 are excluded.

Limitations: Since the Facility component does not collect costs associated with events occurring outside of the facility and the only source of data for these events are FFS claims, out-of-pocket costs associated with such services are likely underestimated for beneficiaries in the Facility component data.

Please refer to Appendix Example 6, found at the end of this tutorial, for detailed guidance on constructing the analytic file and the SAS code for producing estimates

Key Takeaways

- About **5-8% of the MCBS sample** are beneficiaries who live in a Facility or who alternate between living in the Community and living in a Facility
- Refer to the *Data User's Guides* to **ascertain whether differences in data collection and processing methods between** the Community and Facility components will influence analysis and interpretation
- Follow the **recommended steps for combining and weighting data** from the Community and Facility components
- Whenever there are substantial differences in data collection and processing methods between components, it is helpful to **stratify** output for your variable of interest **by interview type** in order to quantify these differences and guide further analysis
- Always **document** limitations!

Thank you!

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

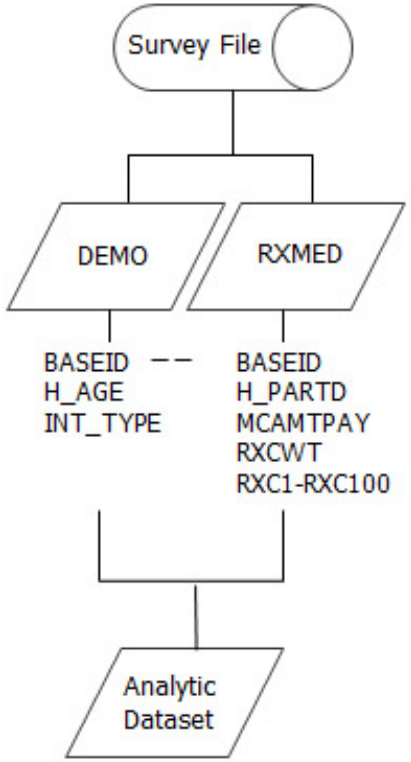
Appendix:
SAS Code for Analytic Examples

MCBS
Medicare Current Beneficiary Survey

Example 1: Flow Chart for Merging Segments

Data Segments

Variables & Topical Weights



Example 1: SAS code for Setup and Analysis

- Create dataset of beneficiaries with Community component data using variables related to satisfaction

```
data rx_satis;
     merge survey17.DEMO (keep=BASEID H_AGE INT_TYPE)
           survey17.RXMED (keep=BASEID H_PARTD MCAMTPAY RXCWT RXC1-RXC100);
     by BASEID;
run;
```

- Use topical weights and BRR method to produce estimates only for beneficiaries with Community component data in the weight file. Beneficiaries with Facility component data at any point during CY2016-CY2017 were excluded from the analysis.

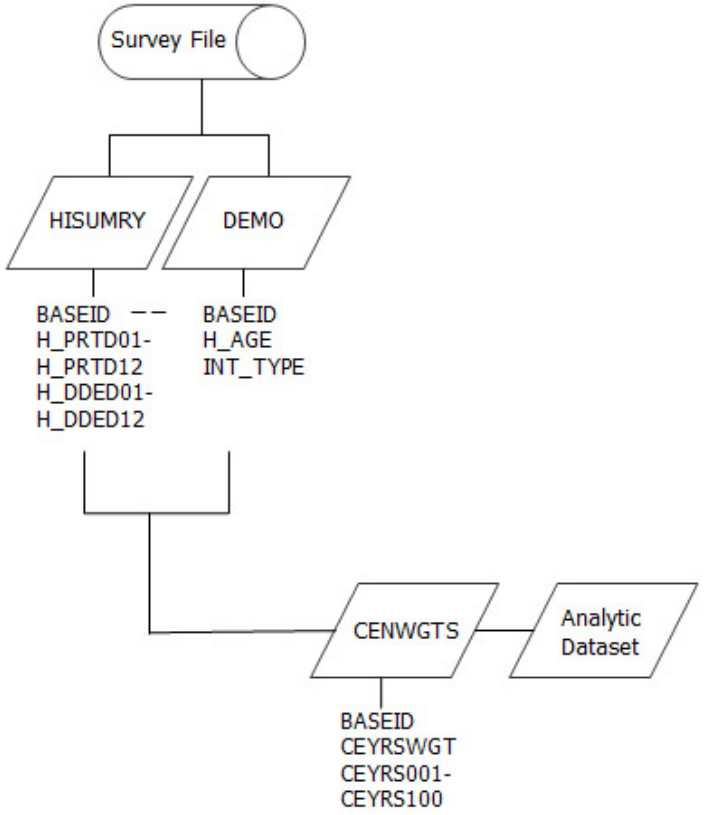
```
proc surveyfreq data=rx_satis varmethod=brr (fay=.30);
     tables MCAMTPAY;
     weight RXCWT;
     repweights RXC1-RXC100;
     where INT_TYPE="C" and RXCWT ne . and H_PARTD = 1;
run;
```

Example 2: Flow Chart for Merging Segments

Data Segments

Variables

Weights



Example 2: SAS code for Setup and Recode

- Create dataset of all beneficiaries using variables related to Part-D deductibles by month in CY2017

```
data partd;
     merge survey17.CENWGTS (in=a)
           survey17.DEMO (keep=BASEID H_AGE INT_TYPE)
           survey17.HISUMRY (keep=BASEID H_PRTD01-H_PRTD12 H_DDED01-H_DDED12);
     if a then output;
     by BASEID;
run;
```

- Create variable that averages monthly Part D deductible, based on 2017 MCBS Codebook

```
data partd_final;
  set partd;
     DEDUC=mean(of H_DDED01-H_DDED12);
run;
```

Example 2: SAS code for Analysis

- Use weights and BRR method to produce estimates of mean Part-D deductible in CY2017

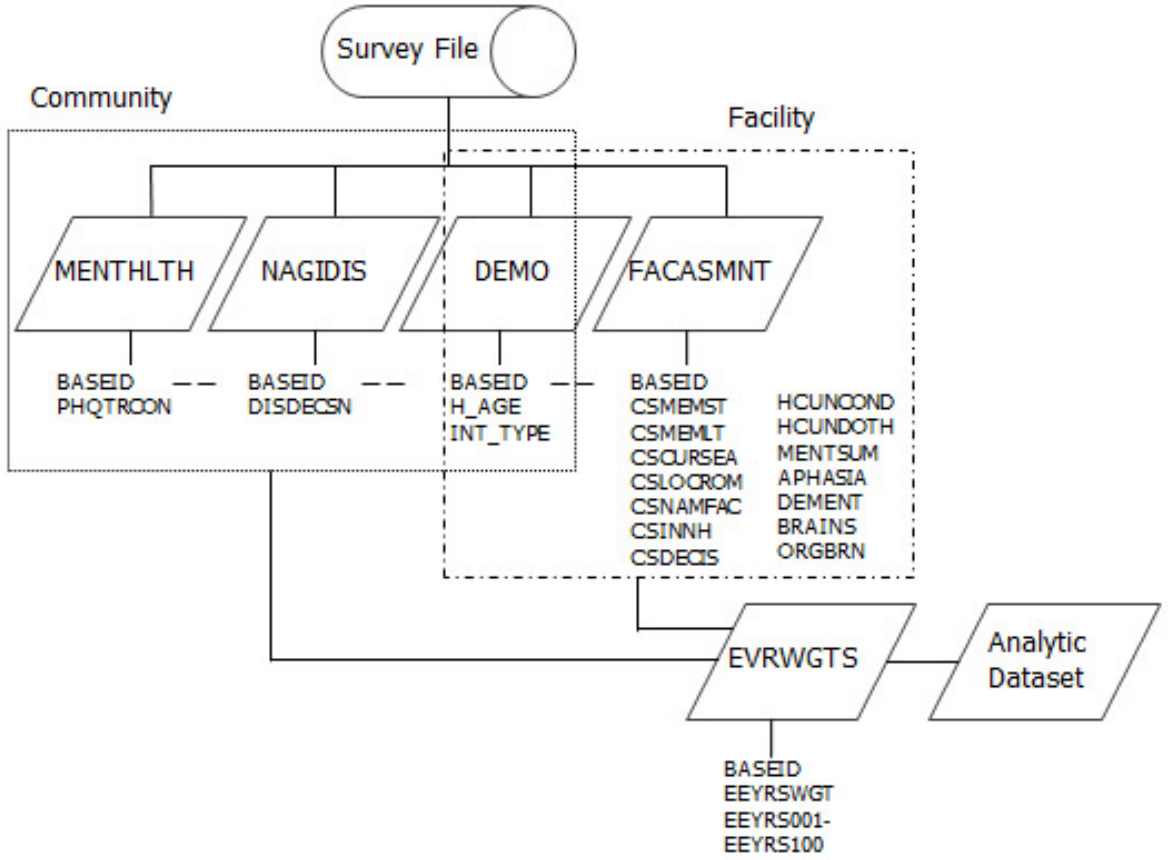
```
proc surveymeans data=partd_final varmethod=brr (fay=.30) plots=none;
     var DEDUC;
     weight CEYRSWGT;
     repweights CEYRS001-CEYRS100;
run;
```

Example 3: Flow Chart for Merging Segments

Data Segments

Variables

Weights



Example 3: SAS code for Setup

- Create dataset of beneficiaries with Community component data using variables related to cognitive impairment

```
data comm;
     merge survey17.DEMO (keep=BASEID H_AGE INT_TYPE in=a)
           survey17.NAGIDIS (keep=BASEID DISDECSN in=b)
           survey17.MENTHLTH (keep=BASEID PHQTRCON in=c);
     if a and b or a and c then output;
     by BASEID;
run;
```

- Create dataset of beneficiaries with Facility component data using variables related to cognitive impairment

```
data facl;
     merge survey17.DEMO (keep=BASEID H_AGE INT_TYPE in=a)
           survey17.FACASMNT (keep=BASEID CSMEMST CSMEMLT CSCURSEA CSLOCROM CSNAMFAC CSINNH
                              HCUNCOND HCUNDOTH CSDECIS MENTSUM APHASIA DEMENT BRAINS ORGBRN in=b);
     if a and b then output;
     by BASEID;
run;
```

Example 3: SAS code for Recode

- Recode Community variables as a binary indicator of cognitive impairment (Yes=1; No=0) based on responses to survey or PHQ-9

```
data comm_recode;
  set comm;
     if DISDECSN = 1 or PHQTRCON in (3:4) then COGIMP = 1;
          else if DISDECSN = 2 or PHQTRCON in (1:2) then COGIMP = 0;
run;
```

- Recode Facility variables as a binary indicator of cognitive impairment (Yes=1; No=0) based on condition, diagnosis, or BIMS scale

```
data facl_recode;
  set facl;
COGIMP_ABILITY = 0;
            if CSMEMST = 1 then COGIMP_ABILITY + 1; if CSMEMLT = 1 then COGIMP_ABILITY + 1;
            if CSCURSEA = 0 then COGIMP_ABILITY + 1; if CSLOCROM = 0 then COGIMP_ABILITY + 1;
            if CSNAMFAC = 0 then COGIMP_ABILITY + 1; if CSINNH = 0 then COGIMP_ABILITY + 1;
            if HCUNCOND in (2,3) then COGIMP_ABILITY + 1; if HCUNDOTH in (2,3) then COGIMP_ABILITY + 1;
            if CSDECIS in (2,3) then COGIMP_ABILITY + 1;
COGIMP_DX = 0;
            if APHASIA = 1 then COGIMP_DX + 1; if BRAINS = 1 then COGIMP_DX + 1;
            if DEMENT ge 1 then COGIMP_DX + 1; if ORGBRN = 1 then COGIMP_DX + 1;
            if MENTSUM in (0:12) or COGIMP_ABILITY ge 1 or COGIMP_DX ge 1 then COGIMP = 1;
            else COGIMP = 0;
run;
```

Example 3: SAS code for Merge and Analysis

- Merge Community and Facility data to weights file, preserving only BASEIDs that exist in the weights file

```
data ment_merged;
     merge survey17.EVRWGTS (in=a)
               work.comm_recode
               work.facl_recode;
     if a then output;
     by BASEID;
run;
```

- Restrict file to beneficiaries aged 65 and over

```
data ment_final;
  set ment_merged;
     where H_AGE ge 65;
run;
```

- Use weights and BRR method to produce COGIMP prevalence estimates stratified by INT_TYPE

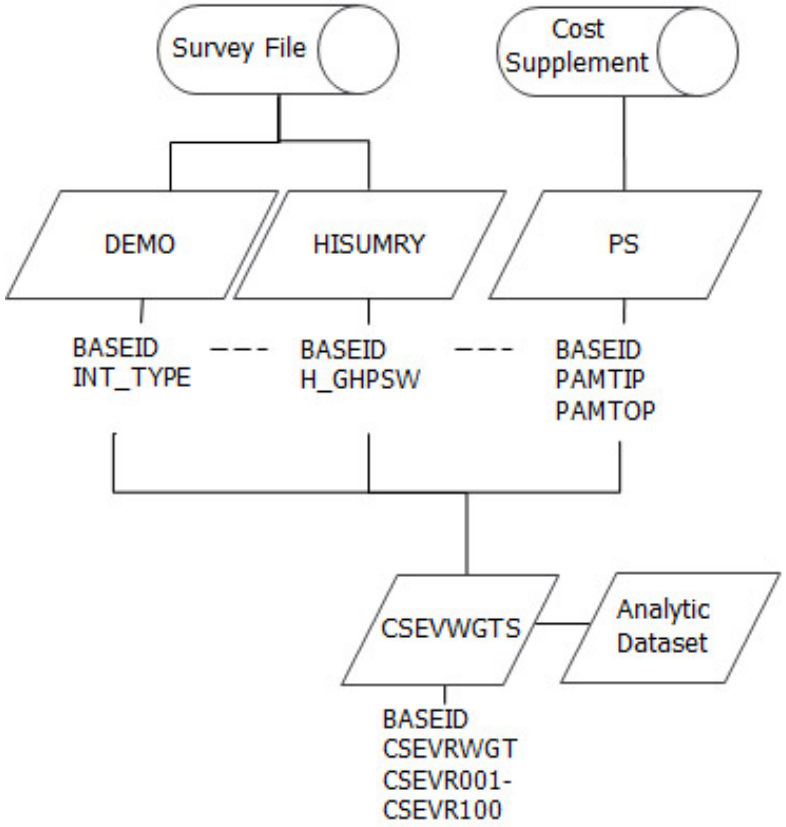
```
proc surveyfreq data=ment_final varmethod=brr (fay=.30);
     tables INT_TYPE*COGIMP / row;
     weight EEYRSWGT;
     repweights EEYRS001-EEYRS100;
run;
```

Example 4: Flow Chart for Merging Segments

Data Segments

Variables

Weights



Example 4: SAS code for Setup and Recode

- Create CY2017 dataset for all beneficiaries using variables related to group health plan participation and inpatient and outpatient cost events

```
data cost_merged;
    merge survey17.DEMO (keep=BASEID INT_TYPE)
          survey17.HISUMRY (keep=BASEID H_GHPSW)
          cost17.PS (keep=BASEID PAMTIP PAMTOP);
    by BASEID;
run;
```

- Merge dataset with weights for Cost Supplement File

```
data cost_final;
    merge cost17.CSEVWGTS (in=a)
          work.cost_merged;
          if a then output;
    by BASEID;
run;
```

Example 4: SAS code for Analysis

- Make a preliminary assessment of the H_GHPSW variable stratified by interview type

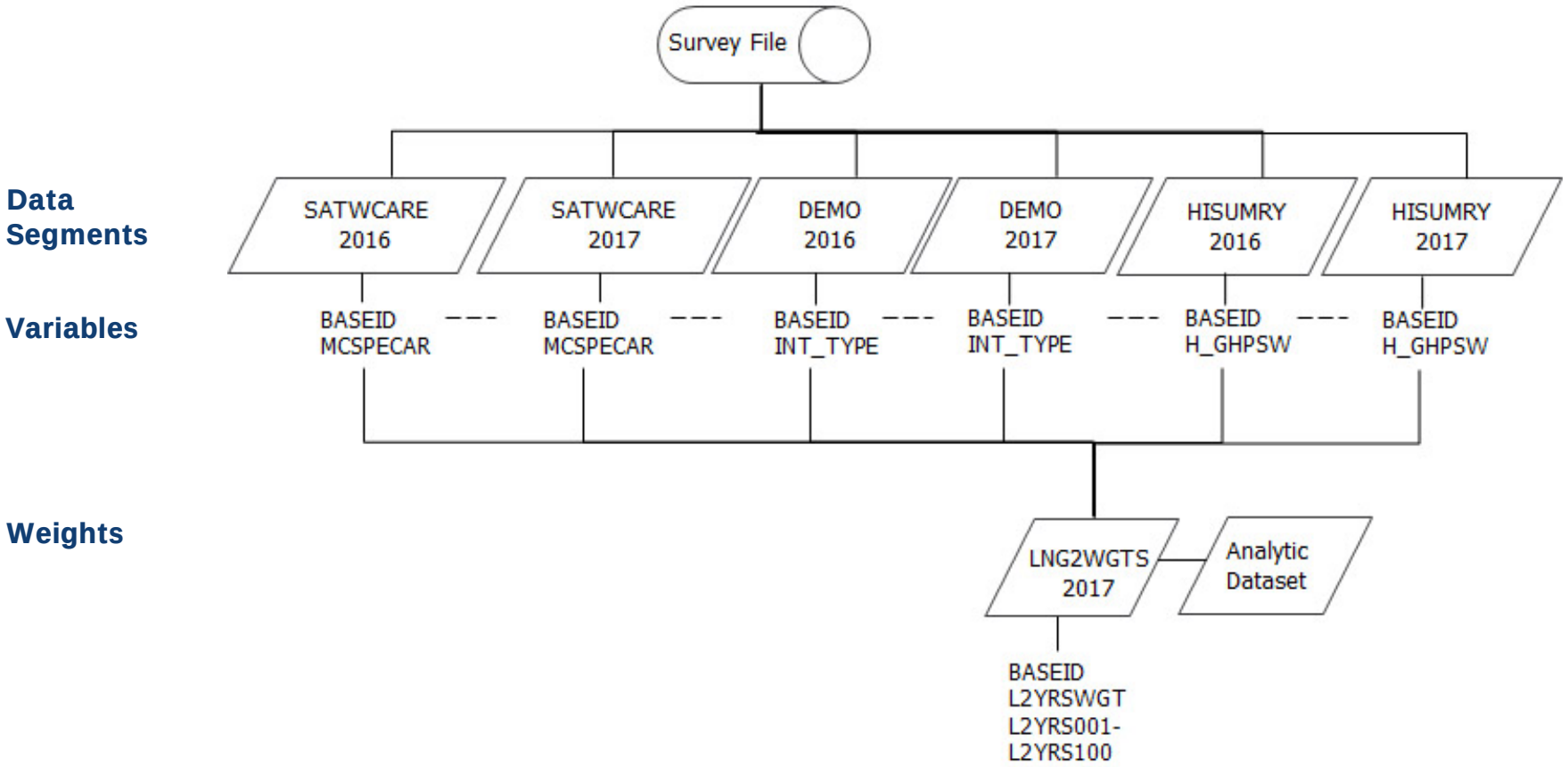
```
proc surveyfreq data=cost_final;
    tables INT_TYPE*H_GHPSW / row;
    weight CSEVRWGT;
    repweights CSEVR001-CSEVR100;
run;
```

- Use weights and BRR method to produce estimates stratified by interview type and group health participation

```
proc sort data = cost_final;
    by INT_TYPE H_GHPSW;
run;

proc surveymeans data=cost_final varmethod=brr (fay=.30) plots=none;
    var PAMTIP PAMTOP;
    weight CSEVRWGT;
    repweights CSEVR001-CSEVR100;
    by INT_TYPE H_GHPSW;
run;
```

Example 5: Flow Chart for Merging Segments



Example 5: SAS code for Setup

Create longitudinal analytic dataset by merging the 2016 and 2017 Survey File segments to the 2017 two-year backward longitudinal weights segment. Subset the analytic file to beneficiaries belonging to the Community component in both years, enrolled in MA during CY2017, and included in the 2017 two-year backward longitudinal weights file.

```
data lex5;
     merge sf17.Lng2wgts (in=a)
           sf17.SATWCARE (keep=BASEID MCSPECAR rename=(MCSPECAR=MCSPECAR17))
           sf16.SATWCARE (keep=BASEID MCSPECAR rename=(MCSPECAR=MCSPECAR16))
           sf17.DEMO (keep=BASEID INT_TYPE rename=(INT_TYPE=INT_TYPE17))
           sf16.DEMO (keep=BASEID INT_TYPE rename=(INT_TYPE=INT_TYPE16))
           sf17.HISUMRY (keep = BASEID H_GHPSW rename=(H_GHPSW=H_GHPSW17))
           sf16.HISUMRY (keep = BASEID H_GHPSW rename=(H_GHPSW=H_GHPSW16));
     by BASEID;
     if a and INT_TYPE16 = "C" and INT_TYPE17 = "C" and MCSPECAR16 in (1,2,3,4,5) and
          MCSPECAR17 in (1,2,3,4,5) and H_GHPSW17 = 1
     then output;
run;
```

Example 5: SAS code for Recode and Analysis

- Recode `Satisfaction with care available from specialists' (MCSPECAR) to create a dichotomous variable to identify beneficiaries who were dissatisfied or very dissatisfied with care available from specialists (Yes=1; No=0).

```
data lex5;
     set lex5;
         dsatspc16 = (MCSPECAR16 in (3,4));
         dsatspc17 = (MCSPECAR17 in (3,4));
run;
```

- Utilize two-year backward longitudinal weights and BRR variance estimation method to estimate the proportion of beneficiaries who are dissatisfied or very dissatisfied with care available from specialists during CY2016 and CY2017. Stratify the estimates by beneficiaries who transition into MA from FFS between CY2016 and CY2017 and beneficiaries who are enrolled in MA during CY2016 and CY2017.

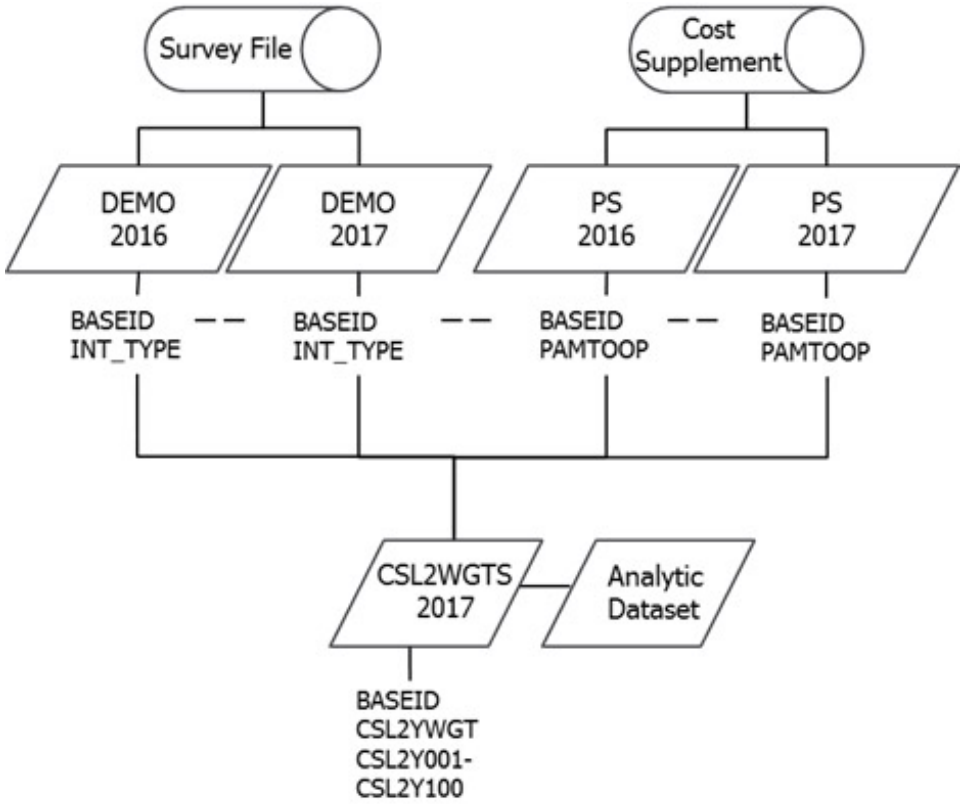
```
proc surveymeans data=lex5 varmethod=brr (fay=.30) plots=none;
            var dsatspc16  dsatspc17;
            domain H_GHPSW16;
            weight L2YRSWGT;
            repweights L2YRS001-L2YRS100;
run;
```

Example 6: Flow Chart for Merging Segments

Data Segments

Variables

Weights



Example 6: SAS code for Setup and Preliminary Assessment

- Create longitudinal analytic dataset by merging the 2016 and 2017 Survey File and Cost Supplements segments to the 2017 two-year, cost-supplement, backward longitudinal weights file. Subset the analytic file to beneficiaries included in the 2017 two-year backward longitudinal weights file.

```
data lex6;
     merge cs17.csl2wgts (in=a)
           sf17.DEMO (keep=BASEID INT_TYPE rename=(INT_TYPE=INT_TYPE15) )
           sf16.DEMO (keep=BASEID INT_TYPE rename=(INT_TYPE=INT_TYPE16) )
           cs17.ps (keep=BASEID PAMTOOP rename=(PAMTOOP = OOP15))
           cs16.ps (keep=BASEID PAMTOOP rename=(PAMTOOP = OOP16));
     by BASEID;
     if a  then output;
run;
```

- Calculate change in out-of-pocket payments and percent change in out-of-pocket payments between CY2016-CY2017 for each beneficiary in the sample. Inflation-adjust CY2016 dollars to CY2017 dollars using a CPI conversion factor.

```
data lex6;
     set lex6;
         OOP16a=OOP16*(242.839/236.916);
         CHNG_OOP= OOP17 - OOP16a;
     if OOP16a > 0 then PCHNG_OOP= (CHNG_OOP/OOP16a) * 100;
         COMM16 = ( INT_TYPE16 in ("C"));
         COMM17 = ( INT_TYPE17 in ("C"));
run;
```

Example 6: SAS code for Preliminary Assessment and Analysis

- Conduct preliminary assessment of the extent to which out-of-pocket payments differ across Community and Facility components

```
proc surveymeans data=lex6 varmethod=brr (fay=.30) plots=none median;
            var OOP16a OOP17;
            domain COMM16 COMM17;
            weight CSL2YWGT;
            repweights CSL2Y001-CSL2Y100;
run;
```

- Utilize two-year backward longitudinal weights and BRR variance estimation method to estimate the median change in out-of-pocket payments for Medicare beneficiaries in the Community and Facility components during CY2015-CY2016.

```
proc surveymeans data=lex6 varmethod=brr (fay=.30) plots=none mean median;
            var OOP16a CHNG_OOP PCHNG_OOP;
            domain COMM16 * COMM17;
            weight CSL2YWGT;
            repweights CSL2Y001-CSL2Y100;
run;
```