MCBS
Medicare Current Beneficiary Survey

2016 | DATA USER'S GUIDE: PUBLIC USE FILE

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

Published: December, 2018

Version Control Log

Date	Version	Revisions
12/6/2018	1.0	Initial version published.

TABLE OF CONTENTS

ACRONYM LIST v

1. INTRODUCTION 1

2. OVERVIEW OF THE MCBS 2

3. TECHNICAL AND PROGRAMMING INFORMATION 3

3.1 General Information 3

3.2 Data File Information 3

3.3 Comparison to the LDS 4

4. SURVEY OVERVIEW 6

4.1 Design of MCBS 6

4.2 Sample Design 7

4.3 Eligibility 9

4.3.1 Medicare Population Covered by the 2016 LDS and MCBS PUF 9

4.4 Case Types 9

4.4.1 Incoming and Continuing Cases 9

4.4.2 Community Interviews 9

4.5 Interviewing and Training Procedures 10

4.5.1 Overview of Data Collection 10

4.5.2 Interviewer Training 10

4.5.3 Privacy and Data Security 11

5. QUESTIONNAIRES 12

5.1 Overview 12

6. SAMPLING 13

6.1 Medicare Population Covered by the 2016 MCBS PUF 13

6.2 Targeted Population and Sampling Strata 14

6.3 Primary and Secondary Sampling Units 15

6.4 Sample Selection 15

7. TECHNICAL NOTES ON USING THE DATA 16

7.1 Weights and Variance Estimation 16
7.2 Item Non-Response 17
7.3 Subgroup Analysis 17

8. REFERENCES 18

9. APPENDICES 20
Appendix A: Glossary 20
Appendix B: Technical Appendix – Sample Code and Output 23
SAS Analysis Statements 23
STATA Analysis Statements 23
R Analysis Statements 23

LIST OF EXHIBITS

Exhibit 3.1.1: 2016 MCBS PUF Variable Prefixes, Number of Variables, Descriptions and Related LDS Survey File Segments 3

Exhibit 3.3.1: Comparison between the 2016 MCBS PUF and 2016 MCBS Survey File LDS 5

Exhibit 4.1.1: Typical MCBS Data Collection Year 6

Exhibit 4.1.2: MCBS Beneficiary Participation Timeline 7

Exhibit 4.2.1: 2011-2016 MCBS Rotating Panel Design 8

Exhibit 4.2.2: 2016 MCBS Composition of Panels in the MCBS PUF 9

Exhibit 6.1.1: Estimated Male Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF 13

Exhibit 6.1.2: Estimated Female Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF 13

Exhibit 6.1.3: Estimated Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF 14

Exhibit 6.2.1: 2016 Panel of Selected Beneficiaries by U.S. Hispanic and U.S. Non-Hispanic Ethnicity Classification and Age Category 15

ACRONYM LIST

AAPOR	American Association for Public Opinion Research
ACCESSCR	LDS Survey File Access to Care segment
ACO	Accountable Care Organization
ADMNUTLS	LDS Survey File Administrative Utilization Summary segment
CAPI	Computer-Assisted Personal Interviewing
CHRNCOND	LDS Survey File Chronic Conditions segment
CMS	Centers for Medicare & Medicaid Services
CSV	Comma-separated values file
DEMO	LDS Survey File Demographics segment
DUA	Data Use Agreement
EEYRSWGT	MCBS PUF Ever Enrolled Full Sample Weight
EVRWGTS	LDS Survey File Ever Enrolled Population Weight Segment
FALLS	LDS Survey File Falls segment
GENHLTH	LDS Survey File General Health segment
HHCHAR	LDS Survey File Household Characteristics segment
HHS	U.S. Department of Health and Human Services
HIC	Health Insurance Claim
HISUMRY	LDS Survey File Health Insurance Summary
HITLINE	LDS Survey File Health Insurance Timeline segment
INTERV	LDS Survey File Interview Characteristics segment
IRB	Institutional Review Board
LDS	Limited Data Set(s)
MA	Medicare Advantage
MAPLANQX	LDS Survey File Medicare Advantage Plan Questions segment
MCBS	Medicare Current Beneficiary Survey
NAGIDIS	LDS Survey File NAGI Disability segment
NICOALCO	LDS Survey File Nicotine and Alcohol segment
NORC	NORC at the University of Chicago
OMB	Office of Management and Budget
PHI	Protected Health Information
PII	Personally Identifiable Information
PMUSE	LDS Survey File Prescription Medicine Usage segment
PREVCARE	LDS Survey File Preventive Care segment
PSU	Primary Sampling Units
PUF	Public Use File
SAS	Statistical Analysis System
SATWCARE	LDS Survey File Satisfaction with Care segment
SSU	Secondary Sampling Units
USCARE	LDS Survey File Usual Source of Care segment
USU	Ultimate Sampling Unit
VISHEAR	LDS Survey File Vision and Hearing segment

1. INTRODUCTION

Over the past several years, the Centers for Medicare and Medicaid Services (CMS) has made it a priority to make more data available, including releasing to the public an unprecedented amount of information on services and procedures provided to Medicare beneficiaries. CMS provides users with multiple ways to access Medicare Current Beneficiary Survey (MCBS) data and a wide array of documentation is publically available on the CMS MCBS website. MCBS data are made available via two annual Limited Data Set (LDS) releases and a MCBS Public Use File (MCBS PUF) based on the Survey File LDS.

The content of the MCBS PUF is governed by its central focus of serving as a unique source of information on beneficiaries' health and well-being that cannot be obtained through CMS administrative sources alone. The file includes data related to Medicare beneficiaries' access to care, health status, other information regarding beneficiaries' knowledge of, attitudes toward, and satisfaction with their health care, as well as demographic data and information on all types of health insurance coverage. Disclosure protections have been applied to the file including de-identification and other methods to ensure individuals cannot be identified; as a result the MCBS PUF does not require a Data Use Agreement (DUA). In contrast, the MCBS LDS releases contain beneficiary-level protected health information (PHI) and therefore require a DUA. The MCBS PUF is not intended to replace the more detailed LDS files; rather, it provides a general-use publically-available alternative that provides the highest degree of protection to the Medicare beneficiaries' PHI.

The main benefits of the MCBS PUF are:

1. Increased data access for researchers of the MCBS through a free file download that is consistent with other U.S. Department of Health and Human Services (HHS) public-use survey files;

2. Increased policy-relevant analyses, by attracting new researchers and policy-makers, for whom the cost and time associated with accessing the MCBS LDS can pose significant deterrents to use.

This user guide contains information about the 2016 MCBS PUF. It contains detailed information about the MCBS and specific background information to help data users understand and analyze the PUF. This guide will be updated each time a new set of PUF data are released.

Readers interested in understanding or analyzing the 2016 MCBS data should also familiarize themselves with the content of the 2016 Data User's Guide: General Information; the 2016 Data User's Guide: Survey File; and the 2016 MCBS Methodology Report documents in order to obtain an overview of the survey, questionnaires, sample design, and other topics relevant to the MCBS. Data users can access these documents along with other data documentation at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks.html. Data users interested in a collection of charts and tables presenting estimates from the LDS releases can access the MCBS Chartbook at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables.html.

2. OVERVIEW OF THE MCBS

Medicare is the nation's health insurance program for persons 65 years and over and for persons younger than 65 years who have a qualifying disability. The MCBS is sponsored by CMS and contains data provided by a representative national sample of the Medicare population. The MCBS is designed to aid CMS in administering, monitoring, and evaluating Medicare programs. A leading source of information on Medicare and its impact on beneficiaries, the MCBS contains important beneficiary data that is not available in CMS administrative data and plays an essential role in the monitoring and evaluation of beneficiary health status and health care policy.

The MCBS is a continuous, in-person, multi-purpose longitudinal survey covering a representative national sample of the Medicare population, including the population of beneficiaries aged 65 and over and beneficiaries aged 64 and below with disabilities. Fieldwork for the first round of data collection began in September 1991; since then, it has continued to collect and provide essential data on the costs, use, and health care status of Medicare beneficiaries. Recently celebrating its 25th anniversary of continuous data collection, the MCBS has completed more than one million interviews provided by thousands of respondents.

The MCBS primarily focuses on economic and beneficiary topics including health care use and health care access barriers, health care expenditures, and factors that affect health care utilization. As a part of this focus, the MCBS collects a variety of information about the beneficiary, including demographic characteristics, health status and functioning, access to care, insurance coverage and out of pocket expenses, financial resources, and potential family support. The MCBS collects this information in three data collection periods, or rounds, per year. Over the years, data from the MCBS have been used to inform many advancements in the Medicare program, including the creation of new benefits such as Medicare's Part D prescription drug benefit.

For questions or suggestions on this document or other MCBS data-related questions, please email MCBS@cms.hhs.gov.

3. TECHNICAL AND PROGRAMMING INFORMATION

3.1 General Information

The 2016 MCBS PUF includes data for 12,852 survey respondents. All records begin with a PUF_ID, a unique number for each beneficiary in the public use file. This PUF_ID serves to identify records in the MCBS PUF only and cannot be used for linking to other sources of data from the MCBS. Each beneficiary's PUF_ID is randomly generated each year, so it is not possible to link a beneficiary's data between years, and the value of the PUF_ID does not provide any information as to when the beneficiary first entered the MCBS.

All variables in the MCBS PUF are numeric or integer. Formats and values for each variable are available in the MCBS PUF codebook.

Variable groups contain prefixes to help users identify these groups by topic area. Exhibit 3.1.1 includes information about these variable prefixes and the locations of the corresponding variables in the Survey File LDS data segments (i.e. the 2016 LDS Survey File individual files).

Exhibit 3.1.1: 2016 MCBS PUF Variable Prefixes, Number of Variables, Descriptions and Related LDS Survey File Segments

MCBS PUF Variable Prefix	Description	Number of PUF Variables in Grouping	LDS Survey File Data Segments
ADM_	Administrative Data	24	ADMNUTLS, HISUMRY, HITLINE
INS_	Insurance status, coverage, and type	15	HISUMRY, HITLINE
INT_	Interview characteristics	3	INTERV
DEM_	Age, sex, and race groups	10	DEMO
ACC_	Access to, use, and satisfaction with health care	115	ACCESSCR, SATWCARE, PMUSE, USCARE, ASSIST
HLT_	Health conditions and limitations in activities of daily living	65	GENHLTH, VISHEAR, NAGIDIS, CHRNCOND
PRV_	Preventive care and physical activity	22	PREVCARE, NAGIDIS,
RSK_	Health behavior risk factors	5	NICOALCO
FAL_	Falls	12	FALLS
HOU_	Housing characteristics	30	HHCHAR
MA_	Medicare Advantage supplement	8	MAPLANQX
EE	Survey weights	101	EVRWGTS

3.2 Data File Information

Detailed information about variables in the MCBS PUF can be found in the codebook. The codebook includes SAS variable names, labels, a note to indicate which respondents were eligible for the question, the question number for the question that was asked in the survey, and a label which summarizes the question text. For certain variables some of the questionnaire categories do not match those provided in the PUF because certain categories were recoded due to disclosure concerns (e.g. "no usable vision" for variable HLT_ECTROUB and "deaf" for HLT_HCTROUB were collapsed into existing categories). Other variables were created by combining two questions and their variable label indicates a recoded variable (e.g. HLT_ALZDEM). In cases where data

were collected from both survey and administrative sources, especially for variables related to insurance coverage, administrative data supersede survey data (e.g. ADM_OP_MDCD).

For each variable, the formats and format values are included in the codebook:

- Values of .R indicate "refused" and .D indicate "don't know."
- All values of "inapplicable" have been combined with missing values.
- Unweighted frequencies of most variables included in the MCBS PUF are provided in the accompanying codebook file.

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

Assume the MCBS 2016 PUF export (PUF2016.xpt) file is downloaded into the folder "C:\MCBS\DOWNLOAD". The following SAS code can then be used to import it into SAS:

```
LIBNAME PUFLIB 'C:\MCBS\SASDATA';
FILENAME F "C:\MCBS\DOWNLOAD\PUF2016.XPT";
PROC CIMPORT LIBRARY=PUFLIB INFILE=F;
RUN;
```

Additionally, a comma-separated values (CSV) file is available for use with other statistical software packages such as R® and STATA®.

A text file with SAS programming code to create formats and to apply SAS labels is provided for users.

3.3 Comparison to the LDS

The MCBS PUF differs from the MCBS Survey File LDS because it has been evaluated for disclosure risk and additional steps were taken to protect respondent confidentiality. The 2016 MCBS PUF contains data for 12,852 beneficiaries and 412 variables, which is similar to the number of beneficiaries contained in the community-only 2016 MCBS Survey File LDS segments, but with fewer variables. Many Survey File LDS variables that were of limited use or posed a disclosure risk were dropped or recoded to create the variable set for the MCBS PUF.

Due to disclosure concerns, the MCBS PUF includes only those beneficiaries interviewed in the community, and thus excludes all beneficiaries who were in a health care facility during all interviews that year (n=1,368). Variables that were only created for facility residents are excluded.[1] Additionally, the MCBS PUF contains no health care utilization, cost or payment data (including Medicare claims data) for individual beneficiaries.

The MCBS PUF is free and available for download on the CMS website. For users interested in the MCBS Survey File and Cost Supplement File LDS, more information on the LDS process can be found at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/.

A summary of the differences between the two data products is presented in Exhibit 3.3.1.

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

Exhibit 3.3.1: Comparison between the 2016 MCBS PUF and 2016 MCBS Survey File LDS

Domain	MCBS PUF	MCBS Survey File LDS
Population	Community	Community and facility
Number of variables	412	4,043 variables across 37 data segments
ID	PUF_ID; Randomly generated, can't be linked back to BASEID, changes each year	BASEID; Randomly generated, can't be linked back to health insurance claim (HIC) number, consistent between years
Date fields	NO	YES
Geographic identifiers	NO	YES
Cost/payment data	NO	YES
Demographic data	YES; All variables are categorical, limited age categories	YES; Continuous, all age variables available
Insurance coverage	YES; Summarized to annual level	YES; Monthly level
Identifiable plan-related information for MA or Part D	NO	YES
Population weights	Ever enrolled weights only	Both ever enrolled and continuously enrolled weights

4. SURVEY OVERVIEW

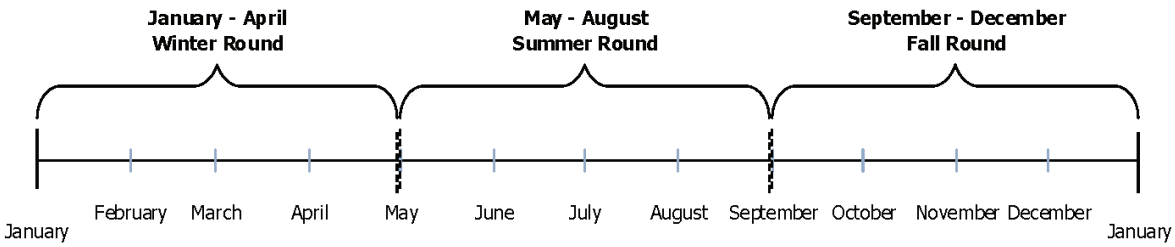
4.1 Design of MCBS

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

Although participation in the survey is limited to four years, MCBS data collection is continuous throughout the year with three distinct seasons (i.e., rounds) of data collection per year. In general, the three rounds are: winter (January through April); summer (May through August); and fall (September through December). The primary reason for the round to round design is to create shorter recall periods during the year to capture more complete and accurate health care costs and utilization from beneficiaries.

The 2016 MCBS data releases reflect data collected from January 2016 through early January 2017 (see Exhibit 4.1.1).

Exhibit 4.1.1: Typical MCBS Data Collection Year

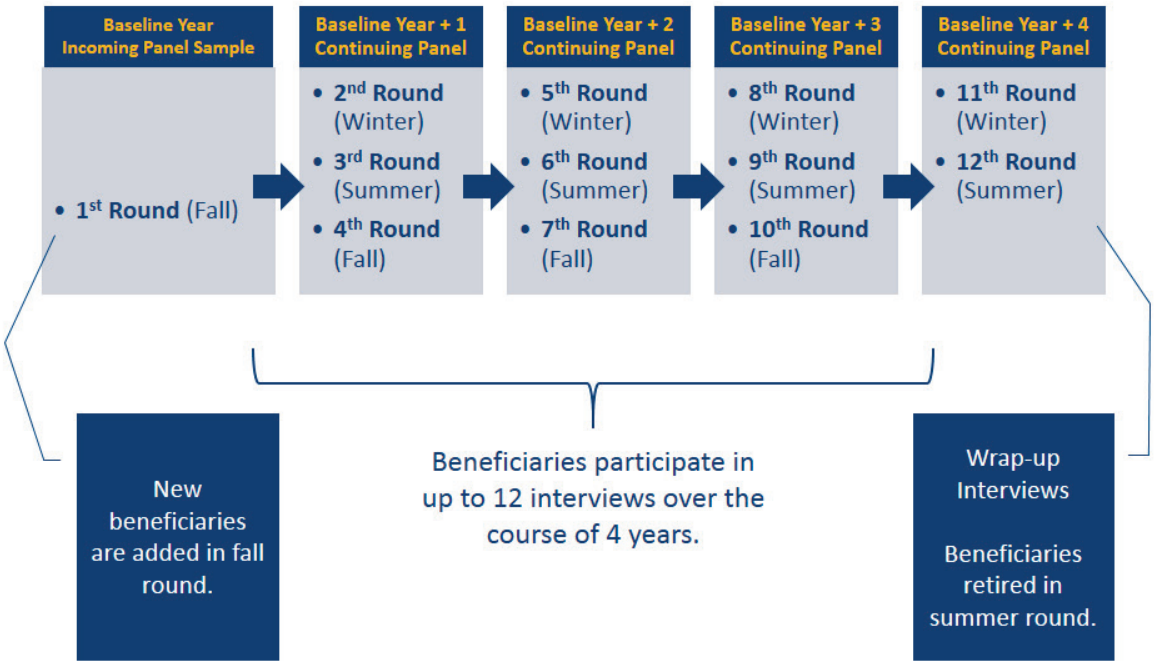


The initial interview of newly-selected respondents takes place in the fall round. Often the fall round begins early (i.e., late July or early August) to allow more time to locate and conduct outreach to these new survey respondents.

Subsequent rounds, which occur every four months, involve the re-interviewing of the same respondent (or appropriate proxy respondents) over a four year period (up to 12 interviews in total).[2] Exhibit 4.1.2 depicts the timeline of participation for respondents selected to be in the MCBS sample.

[2] Beginning in 2018, the number of rounds of participation was reduced from 12 to 11 and the last round of interviewing is the winter round, not the summer round.

Exhibit 4.1.2: MCBS Beneficiary Participation Timeline



4.2 Sample Design

The MCBS uses a rotating panel sample design, covering the population of Medicare beneficiaries residing in the continental U.S. (48 states and the District of Columbia) and Puerto Rico.[3] Each sampled beneficiary is scientifically selected as part of a panel and is interviewed up to three times per year.[4] One panel is retired during each summer round, and a new panel is selected to replace it each fall round (see Exhibit 4.2.1). The size of the new panel is designed to provide a stable number of respondents across all panels participating in the survey annually.

[3] Alaska and Hawaii are not included among the states from which the sample was selected due to the high cost of data collection in those areas; however, they are included in control totals for weighting purposes. Beginning in 2017, Puerto Rico data collection was discontinued.

[4] The three rounds per year are referred to seasonally. Respondents are interviewed in the winter round, the summer round, and the fall round each year.

Exhibit 4.2.1: 2011-2016 MCBS Rotating Panel Design

Data Collection Schedule			Panel					
Data Year	Season	Round#	2011	2012	2013	2014	2015	2016
2012	Winter	62						
	Summer	63						
	Fall	64						
2013	Winter	65						
	Summer	66						
	Fall	67						
2014	Winter	68						
	Summer	69						
	Fall	70						
2015	Winter/Summer	71/72*						
	Fall	73						
2016	Winter	74						
	Summer	75						
	Fall	76						

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

The MCBS employs a three-stage cluster sample design. Primary sampling units (PSUs) are made up of major geographic areas consisting of metropolitan areas or groups of rural counties. Secondary sampling units (SSUs) are made up of census tracts or groups of tracts within the selected PSUs. Medicare beneficiaries, the ultimate sampling units (USUs), are then selected from within the selected SSUs. The sample represented in the 2016 MCBS was drawn from 107 PSUs, which contained 1,250 SSUs. The MCBS sample is annually "supplemented" during the fall round to account for attrition (deaths, dis-enrollments, refusals) and newly enrolled persons. Each annual supplement is referred to as the Incoming Panel sample.

Respondents for the MCBS are sampled from the Medicare Administrative enrollment data. The beneficiaries included in the MCBS PUF represent a randomly selected cross-section of all beneficiaries who were ever enrolled in either Part A or Part B of the Medicare program for any portion of 2016.[5] The MCBS PUF represents four separate MCBS panels identified by the year in which the panel was selected and first interviewed (i.e., for the 2016 MCBS PUF, the 2013, 2014, 2015 and 2016 panels). Exhibit 4.2.2 shows the distribution of each of the four panels included in the 2016 MCBS PUF.

For more information on the sample design, please see Section 4.2 within the General Data User's Guide at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks-Items/2016SurveyFile.html.

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

Exhibit 4.2.2: 2016 MCBS Composition of Panels in the MCBS PUF

Data Year (Fall)	Number of Beneficiaries Selected
2013	1,790
2014[6]	2,737
2015	2,395
2016	5,930

4.3 Eligibility

4.3.1 Medicare Population Covered by the 2016 LDS and MCBS PUF

Beginning in 2015, beneficiaries who became eligible for Medicare Part A or B and enrolled anytime during the sampling year were eligible to be sampled as part of the annual panel. This is a substantial change in practice; prior to 2015, only beneficiaries enrolled in Medicare by January 1 of the sampling year were eligible to be sampled in an annual panel.

4.4 Case Types

MCBS respondents are classified by their phase of survey participation (i.e., Incoming or Continuing) and interview participation (i.e., Community or Facility), which is determined by residence status. Although they appear in the MCBS LDS releases, beneficiaries for whom only Facility-administered interviews were conducted during the data collection period are not included in the MCBS PUF. Researchers interested in the population of beneficiaries residing in facilities will need to use the MCBS LDS, as discussed in Section 3.3.

4.4.1 Incoming and Continuing Cases

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

4.4.2 Community Interviews

Approximately 90 percent of the interviews take place in the respondent or proxy's own residence or in a neutral interview location, such as a library or public venue. These interviews are called Community interviews; the remaining 10 percent of the interviews are from beneficiaries residing in a facility and these beneficiaries are not included in the MCBS PUF.

Over the course of a four year period, however, it is not uncommon for respondents to enter long-term care facilities (e.g., nursing homes) or to go back and forth between the community and a facility setting. To obtain an accurate representation of the experiences of all Medicare beneficiaries, the MCBS includes beneficiaries

[6] Fall 2014 was the first round collected by NORC at the University of Chicago after the contract transitioned from the prior incumbent of data collection following a transition between contractors. In September 2014, the final number of Summer 2014 completed interviews from the Continuing sample was provided to NORC. Because the completion rates for the Summer 2014 round were lower than anticipated, CMS and NORC agreed that the Incoming Panel sample size should be increased. Then in December 2014, CMS and NORC agreed to extend the Fall 2014 round to March 2015 so that final re-programming of all questionnaire sections could be completed for fielding of the next round. This decision resulted in an additional buffer of Incoming Panel sample released in January 2015 with a data collection period of about eight weeks.

wherever they reside, even if they enter or reside in a facility for the duration of their four years with the study. The MCBS PUF excludes those beneficiaries who were in a facility for each interview, due to disclosure concerns.

4.5 Interviewing and Training Procedures

4.5.1 Overview of Data Collection

CMS contracts with NORC at the University of Chicago (NORC) to conduct the MCBS. A national team of specially trained and certified NORC field interviewers conduct either face-to-face interviews with MCBS respondents or their designated proxies or they conduct face-to-face interviews with Facility administrators on behalf of respondents. The first interview conducted for an Incoming Panel respondent is relatively short as it does not collect health care utilization or cost data. Continuing respondent interviews are longer as field interviewers collect information about the respondent's health care utilization and associated costs. Telephone interviews are usually conducted for respondents who are in the 12th and final round of the MCBS as this interview is short and does not include questions on cost and utilization.

4.5.1.1 Overview of Recruitment of Beneficiaries and Scheduling Procedures

Medicare beneficiaries selected to participate in the MCBS receive a letter and brochure in the mail, introducing the study and explaining that an interviewer from NORC will contact them to schedule an appointment. For Incoming Panel respondents, initial contact is typically made in person; for Continuing respondents, outreach to set an appointment for the next interview is most often made by phone. If respondents are unable to answer questions or require language assistance, respondents 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 respondent if the respondent is incapacitated or unable to complete the interview. For Spanish speaking respondents, a Spanish version of the Community instrument is available and bilingual interviewers conduct the interview.

4.5.1.2 Computer-Assisted Personal Interviewing (CAPI)

Field interviewers complete MCBS interviews using a Computer-Assisted Personal Interviewing (CAPI) instrument loaded on a laptop computer. The CAPI program automatically guides the field interviewer through the questions, records the answers, and contains logic and skip flows that increase the output of timely and high quality data. The CAPI also contains follow-up questions where data were missing from the previous interview. When the interview is completed, the CAPI system allows the field interviewer to transmit the data electronically to the NORC central office.

4.5.2 Interviewer Training

Nationally, the MCBS employs an average of approximately 200 field interviewers[7], who participate in a combination of several targeted training initiatives and careful coaching and monitoring activities throughout data collection. Each training is customized to the level of experience of the interviewer (new to MCBS or MCBS-experienced), the type of interview (Community or Facility), the type of sample (Incoming Panel or Continuing), and the unique requirements of each round (changing questionnaire sections or data collection protocols). Field interviewers who are new to MCBS are always trained in-person; experienced field interviewers participate in a periodic in-person training program and receive continuous online refresher

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

training. Weekly field memos issued to all field managers and field interviewers cover important data collection tips, provide answers to interviewer questions, and reminders about how to handle complicated scenarios.

4.5.3 Privacy and Data Security

Field interviewer training stresses the importance of maintaining respondent privacy and project protocols are documented within the field interviewer manual. Field outreach and contacting procedures also maintain and ensure confidentiality. These procedures include the utilization of standard computer security protocol (dual authentication password protection for each interviewer laptop) and restrictions on submitting personally identifiable information (PII) through electronic mail. All MCBS survey staff directly involved in data collection and/or analysis activities are required to sign a Non-Disclosure Agreement and a confidentiality agreement.

NORC and CMS are committed to protecting respondent confidentiality and privacy, and both organizations diligently uphold provisions established under the Privacy Act of 1974, the NORC Institutional Review Board (IRB), the Office of Management and Budget (OMB), and the Federal Information Security Management Act of 2002. Respondent materials specify "The information you provide will be kept private to the extent permitted by law, as prescribed by the Privacy Act of 1974. The information you give will only be used for research and statistical purposes."

5. QUESTIONNAIRES

5.1 Overview

The MCBS Questionnaire structure features two components (Community and Facility), administered based on the beneficiary's residence status. Within each component, the flow and content of the questionnaire varies by interview type and data collection season (fall, winter, or summer). There are two types of interviews (Baseline, Continuing) containing two types of questionnaire sections (Core and Topical). See Exhibit 5.1 within the General Data User's Guide for a depiction of the MCBS Questionnaire structure: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks-Items/2016SurveyFile.html

- Community Component: Survey of beneficiaries residing in the community at the time of the interview (i.e., their residence or a household). Interview may be conducted with the beneficiary or a proxy.
- Facility Component: Survey of beneficiaries residing in facilities such as long-term care nursing homes or other institutions at the time of the interview. Interviewers do not conduct the Facility component with the beneficiary, but with staff members located at the facility (i.e., facility respondents). This is a key difference between the Community and Facility components.

As mentioned earlier, there are two types of interviews – an initial (Baseline) interview administered to new beneficiaries selected as part of the Incoming Panel, and an interview administered to repeat (Continuing) beneficiaries as they progress through the study.

- Baseline: The initial questionnaire administered to beneficiaries new to the study; administered in the fall round of the year they are selected into the sample (interview #1).
- Continuing: The questionnaire administered to beneficiaries as they progress through the study once they have completed a Baseline interview (interviews #2-12).

Depending on the interview type and data collection season (fall, winter, or summer), the MCBS Questionnaire includes Core and Topical sections:

- Core: These sections are of critical purpose and policy relevance to the MCBS, regardless of season of administration. Core sections collect information on beneficiaries' health insurance coverage, health care utilization and costs, and operational management data such as locating information.
- Topical: These sections collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

See Sections 5.2 and 5.3 within the General Data User's Guide for additional detail on the 2016 Core and Topical sections: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks-Items/2016SurveyFile.html

6. SAMPLING

6.1 Medicare Population Covered by the 2016 MCBS PUF

The MCBS data releases are a reflection of enrolled Medicare beneficiaries residing in the continental U.S. and Puerto Rico. Residents of foreign countries and U.S. possessions and territories other than Puerto Rico are excluded. The MCBS PUF further excludes Medicare beneficiaries who only provided facility-based interviews during the data year.

Exhibits 6.1.1 and 6.1.2 present estimates of the size of the ever enrolled, community-dwelling Medicare population by race, and age (as of December 31, 2016) by sex, in the 2016 MCBS PUF. Exhibit 6.1.3 presents the aggregated estimates of the size of the ever enrolled, community-dwelling Medicare population overall and by sex and race.

Exhibit 6.1.1: Estimated Male Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF

Race	Age as of 12/31/2016	Weighted Count
White non-Hispanic	Under 65 years	2,723,503
	65-74 years	9,536,572
	75-84 years	4,584,409
	85+ years	1,634,045
Black non-Hispanic	Under 65 years	629,581
	65-74 years	1,113,940
	75-84 years	451,476
	85+ years	98,243
Hispanic	Under 65 years	446,712
	65-74 years	889,310
	75+ years	587,607
Other*	Under 65 years	390,375
	65-74 years	978,603
	75+ years	435,587

SOURCE: 2016 MCBS PUF, weighted counts.

*The 'Other' race category includes other races, more than one race, and unknown race.

Exhibit 6.1.2: Estimated Female Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF

Race	Age as of 12/31/2016	Weighted Count
White non-Hispanic	Under 65 years	2,533,373
	65-74 years	10,957,946
	75-84 years	5,548,547
	85+ years	2,473,343
Black non-Hispanic	Under 65 years	768,103

Race	Age as of 12/31/2016	Weighted Count
	65-74 years	1,274,138
	75-84 years	632,903
	85+ years	272,067
Hispanic	Under 65 years	501,916
	65-74 years	1,264,783
	75+ years	827,943
Other*	Under 65 years	390,106
	65-74 years	980,387
	75+ years	617,967

SOURCE: 2016 MCBS PUF, weighted counts.

*The 'Other' race category includes other races, more than one race, and unknown race.

Exhibit 6.1.3: Estimated Community Medicare Beneficiaries by Race and Age, in the 2016 MCBS PUF

Group	Subgroup	Weighted Count
Overall Total		53,543,483
Sex	Male Total	24,499,962
	Female Total	29,043,521
Race	White non-Hispanic Total	39,991,738
	Black non-Hispanic Total	5,240,452
	Hispanic Total	4,518,270
	Other Total*	3,793,025

SOURCE: 2016 MCBS PUF, weighted counts.

*The 'Other' race category includes other races, more than one race, and unknown race.

6.2 Targeted Population and Sampling Strata

Historically, the targeted population for the MCBS consisted of persons enrolled in one or both parts of the Medicare program, that is, Part A or Part B, as of January 1 of the applicable sample-selection year, and whose address on the Medicare files is in one of the 48 contiguous states (excludes Alaska and Hawaii), the District of Columbia, or Puerto Rico. Beginning in 2015, the targeted population for the MCBS consisted of Part A and/or Part B enrollees as of December 31 of the sample-selection year. For example, for Fall Rounds 2013 and 2014 (the two rounds in which the 2013 and 2014 Panels included in the 2016 MCBS data were first selected), the targeted populations for the new panels included those individuals enrolled as of January 1 of 2013 and 2014 respectively. For Fall Rounds 2015 and 2016 (the two rounds in which the 2015 and 2016 Panels included in the 2016 MCBS data were selected), the targeted population for the new panels included those individuals enrolled as of December 31 of 2015 and 2016, respectively.

Beginning in 2015, the strata were expanded to separate U.S. Hispanic, U.S. non-Hispanic, and Puerto Rican beneficiaries by age group. Additionally, in the 2015 and 2016 Panels, beneficiaries residing within the U.S. who were Hispanic (based on a Hispanic ethnicity classification code in the Medicare enrollment data; see

Eicheldinger[8] for more details) were oversampled. For more information on the sampling strata, please see Section 6 within the General Data User's Guide: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks-Items/2016SurveyFile.html.

Exhibit 6.2.1 displays the beneficiaries included in the 2016 MCBS PUF, by age and ethnicity.

Exhibit 6.2.1: 2016 Panel of Selected Beneficiaries by U.S. Hispanic and U.S. Non-Hispanic Ethnicity Classification and Age Category

Age Category as of 12/31/2016	TOTAL Sample Size	TOTAL Weighted	U.S. Hispanic Sample Size	U.S. Hispanic Weighted	U.S. Non-Hispanic Sample Size	U.S. Non-Hispanic Weighted
Under 65 years	2,223	8,383,669	252	948,628	1,971	7,435,041
65-74 years	4,124	26,995,680	396	2,154,093	3,728	24,841,587
75+ years	6,505	18,164,136	579	1,415,549	5,926	16,748,587
Total	12,852	53,543,485	1,227	4,518,270	11,625	49,025,215

SOURCE: 2016 MCBS PUF.

6.3 Primary and Secondary Sampling Units

All of the panels in the 2016 data releases are distributed across the redesigned sample of 107 PSUs selected in 2001.[9] These PSUs are a representative, national sample of beneficiaries who are geographically dispersed throughout metropolitan areas and groups of non-metropolitan counties. Recall that SSUs are census tracts or groups of contiguous tracts.

6.4 Sample Selection

The MCBS sampling design provides nearly self-weighting (i.e., equal probabilities of selection) samples of beneficiaries within each of the 21 sampling strata. Within the selected PSUs and SSUs, a systematic sampling scheme with random starts is employed for selecting beneficiaries.[10] For each continuing beneficiary, the survey questions corresponding to the Survey File data release are administered in the fall of the data collection year. Similarly, for beneficiaries new to the MCBS, the survey questions are administered as part of the initial fall Baseline interview.

[8] Eicheldinger, C. "More Accurate Racial and Ethnic Codes for Medicare Administrative Data," *Health care financing review* 29, no. 3.

[9] An original set of 107 PSUs was selected at the start of the MCBS in 1991; the current PSUs were selected in 2001 with a focus on maximizing overlap with the original set of PSUs. With the rotating panel design, the PSU redesign is transparent to data users and no special processing is required. For more details on the PSU redesign, see Lo, A, A Chu, and R Apodaca. "Redesign of the Medicare Current Beneficiary Survey Sample," Proceedings of the Survey Research Section of the American Statistical Association 2002.

[10] The MCBS 2016 Panel was drawn by systematic random sampling with probability proportional to probabilities of selection with an independently selected random start within each PSU. For more information on this sampling method, please see the MCBS Methodology Report, available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks-Items/2015_MCBS_Survey_Methods_Report.html

7. TECHNICAL NOTES ON USING THE DATA

7.1 Weights and Variance Estimation

The sample design of MCBS includes stratification, clustering, multiple stages of selection, and disproportionate sampling. Furthermore, the MCBS sampling weights reflect adjustments for survey nonresponse. These survey design and estimation complexities require special consideration when analyzing MCBS data (i.e., it is not appropriate to assume simple random sampling).

To obtain accurate estimates from MCBS data, for either descriptive statistics or more sophisticated analyses based on multivariate models, the survey design complexities need to be taken into account by applying MCBS weights to produce estimates and using an appropriate technique to derive standard errors associated with the weighted estimates.

The MCBS PUF includes the ever enrolled, full sample cross sectional weight (EEYRSWGT) and does not include the continuously enrolled cross-sectional weight in order to protect the confidentiality of the respondents. The continuously enrolled cross-section weights are available, however, in the LDS.

The ever enrolled cross sectional weight applies to both the Continuing sample (beneficiaries sampled between 2013-2015) and to the Incoming Panel sample (beneficiaries sampled in 2016). This weight is intended for use in cross-sectional statistics involving the total (combined) Fall 2016 sample. Each weight is greater than zero for all beneficiaries on the file. EEYRSWGT should be used to make estimates of parameters for the Medicare population who were enrolled at any point in 2016 (i.e., the ever enrolled population).

To permit the calculation of random errors due to sampling, a series of replicate weights were computed. Unless the complex nature of the MCBS is taken into account, estimates of the variance of a survey statistic may be biased downward. The replicate weights included in the MCBS PUF can be used to calculate standard errors of the sample-based estimates. The replicate cross-sectional weights are labeled EEYRS001 through EEYRS100 corresponding to the ever enrolled weight EEYRSWGT.

Most commercial software packages today include techniques to accommodate the complex design, through replicate weight approaches. Among these are STATA®, SUDAAN®, R®, and the complex survey procedures in SAS®. When using the replicate weight approach to variance estimation, we recommend the variance estimation method of balanced repeated replication using Fay's adjustment of 0.3. Sample code in SAS, STATA and R for estimating statistics can be found in Appendix B. Analysis of subgroups should utilize the domain functions within the statistical package of your choice (e.g. the DOMAIN statement in SAS, or the OVER function in STATA); restricting the sample to the subgroup and then performing an analysis would lead to slightly biased point estimates and estimates of variance.

7.2 Item Non-Response

As with most data collection activities, some existing and new panel members will either be impossible to locate or will refuse to participate in an interview. The calculation of the study-wide response rates generally follow the guidelines specified in the American Association for Public Opinion Research (AAPOR) and OMB. For the ever enrolled cross-sectional sample represented by the MCBS 2016 Survey File, the calculated overall response rate was 65.2%. This rate includes non-response for persons in facilities, as the response rates are not calculated separated by questionnaire component. Therefore, this may not reflect exactly the response rate for the sample represented in the 2016 MCBS PUF, which excludes beneficiaries for whom only Facility-administered interviews were conducted during the data collection year.

7.3 Subgroup Analysis

When analyzing survey data, researchers are often interested in focusing their analyses on specific subgroups of the full population sample (e.g., Medicare beneficiaries age 65 and over, Hispanics, or females). A common pitfall when performing sub-group analysis of survey data when variance estimation methods such as Taylor-series is used, is to delete or exclude observations not relevant to the subgroup of interest. Standard errors for MCBS estimates are most accurate when the analytic file includes all beneficiaries. However, when replicate weights are used for variance estimation, deleting observations not relevant to the subgroup of interest prior to analyzing the subgroup will still produce accurate standard errors. Almost all statistical packages provide the capability to limit the analysis to a subgroup of the population.

8. REFERENCES

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

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

APPENDICES

9. APPENDICES

Appendix A: Glossary

Baseline interview: The initial questionnaire administered to new respondents to the study; administered in the fall round of the year they are selected into the sample (interview #1).

Beneficiary: An individual selected from MCBS' sample about whom the MCBS collects information. Beneficiary may also refer to a person receiving Medicare services who may or not be participating in the MCBS.

Community component: Survey of beneficiaries residing in the community at the time of the interview (i.e., not in a long-term care facility such as a nursing home).

Continuing interview: The questionnaire administered to repeat respondents as they progress through the study (interviews #2-12).

Continuously enrolled (aka always enrolled): A Medicare beneficiary who was enrolled in Medicare from the first day of the calendar year until the fall interview and did not die prior to the fall round. This population excludes beneficiaries who enrolled during the calendar year 2016, those who dis-enrolled or died prior to their fall interview, residents of foreign countries, and residents of U.S. possessions and territories other than Puerto Rico.

Core sections: These sections of the MCBS Questionnaire are of critical purpose and policy relevancy to the MCBS, regardless of season of administration.

Crossover: A respondents who enters a long-term care facility setting (e.g., nursing homes) or who alternates between a community and a facility setting.

Ever enrolled: A Medicare beneficiary who was enrolled at any time during the calendar year including those who dis-enrolled or died prior to their fall interview. Excluded from this population are residents of foreign countries and of U.S. possessions and territories other than Puerto Rico.

Facility component: Survey of beneficiaries residing in facilities, such as long-term care nursing homes or other institutions, at the time of the interview. Interviewers do not conduct the Facility component with the beneficiary, but rather, with a staff member located at the facility.

Field interviewer: The principal contact for collecting and securing respondent data.

Field manager: A supervisor who motivates and manages a group of field interviewers to meet the goals of high quality data collection on time and within budget limits.

Incoming Panel sample (formerly known as Supplemental Panel): A scientifically selected group of sampled beneficiaries that enter the MCBS in the fall round of a data collection year. One panel is retired during each summer round, and a new panel is selected to replace it each fall round. Panels are identified by the data collection year (e.g., 2016 panel) in which they were selected.

Internal Sample Control File: A data file that contains every beneficiary sampled back through the beginning of MCBS. The file contains sampling information, year of selection, primary sampling unit, secondary sampling unit, contact information, and other sampling demographic information as well as final disposition

codes to indicate completion status per round, component fielded per round, dates of death, and lost entitlement information.

Long-term care facility: A facility that provides rehabilitative, restorative, and/or ongoing skilled nursing care to patients or residents in need of assistance with activities of daily living.

Medicare: Medicare is the federal health insurance program for people who are 65 or over, certain younger people with disabilities, and people with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant, sometimes called ESRD). The different parts of Medicare help cover specific services:

- Hospital Insurance (Part A): covers inpatient hospital stays, care in a skilled nursing facility, hospice care, and some home health care.
- Medical Insurance (Part B): covers certain doctors' services, outpatient care, medical supplies, and preventive services.
- Medicare Advantage (Part C): an alternative to coverage under traditional Medicare (Parts A and B), a health plan option similar to a Health Maintenance Organization (HMO) or Preferred Provider Organization (PPO) administered by private companies.
- Prescription Drug Coverage (Part D): additional, optional coverage for prescription drugs administered by private companies.

For more information, please visit the Medicare.gov website at https://www.medicare.gov/sign-up-change-plans/decide-how-to-get-medicare/whats-medicare/what-is-medicare.html

Medicare Advantage (MA): Medicare Advantage Plans, sometimes called "Part C" or "MA Plans," are offered by private companies approved by Medicare. An MA provides, or arranges for the provision of, a comprehensive package of health care services to enrolled persons for a fixed capitation payment. The term "Medicare Advantage" includes all types of MAs that contract with Medicare, encompassing risk MAs, cost MAs, and health care prepayment plans (HCPPs).

Medicare beneficiary (aka, beneficiary): An individual who meets at least one of three criteria (is aged 65 years or over, is under age 65 with certain disabilities, or is of any age with End-Stage Renal Disease) and is entitled to health insurance benefits. (Source: https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html).

Panel: see Incoming Panel Sample

Primary Sampling Unit (PSU): Primary sampling unit refers to sampling units that are selected in the first (primary) stage of a multi-stage sample ultimately aimed at selecting individual elements (Medicare beneficiaries in the case of MCBS). PSUs are made up of major geographic areas consisting of metropolitan areas or groups of rural counties.

Race/ethnicity: Responses to race and ethnicity questions are self-reported by the respondent. Respondents who reported they were white and not of Hispanic origin were coded as white non-Hispanic; those who reported they were black/African-American and not of Hispanic origin were coded as black non-Hispanic; persons who reported they were Hispanic, Latino/Latina, or of Spanish origin, regardless of their race, were coded as Hispanic; persons who reported they were American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, two or more races, or other race and not of Hispanic origin were coded as other race/ethnicity.

Respondent: The person who answers questions about the beneficiary for the MCBS; this person can be the beneficiary themselves, a proxy, or a staff member located at a facility where the beneficiary resides.

Round: The MCBS data collection period. There are three distinct rounds each year; winter (January through April); summer (May through August); and fall (September through December).

Secondary Sampling Unit (SSU): SSUs are made up of census tracts or groups of tracts within the selected PSUs.

Topical sections: Sections of the MCBS Questionnaire that collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

Ultimate Sampling Unit (USU): USUs are Medicare beneficiaries selected from within the selected SSUs.

Appendix B: Technical Appendix – Sample Code and Output

SAS Analysis Statements

Cross-tabulations

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
        table <Var name>  / row  chisq lrchisq;
        weight EEYRSWGT;
        repweight EEYRS001 - EEYRS100;
run;
```

Subgroup Analysis

```
proc surveyfreq data=<Analytic dataset> VARMETHOD = brr (fay=.30);
        table <Var name> *  <Subgroup variable> / row  chisq lrchisq  ;
        weight EEYRSWGT;
        repweight EEYRS001 - EEYRS100;
run;
```

STATA Analysis Statements

Declare dataset as survey sample with replicate weights

```
svyset _n [pweight= EEYRSWGT ], brrweight(EEYRS001 - EEYRS100) fay(.3) vce(brr) singleunit(missing)
```

For categorical variables, use:

```
svy brr,  fay(.3) : tabulate <Var name> <Var name>
```

For subgroup analysis use:

```
svy brr, subpop(if <Subgroup>) fay(.3) : tabulate <Var name>, over(<Var name>)
```

R Analysis Statements

Declare MCBS survey design object with replicate weights

```
mcbs <- svrepdesign(
  weights = ~EEYRSWGT,
  repweights = "EEYRS[001-100]+",
  type = "Fay",
  rho = 0.3,
  data = <Source dataset>,
  combined.weights = TRUE
  )
```

For categorical variables, use:

```
svytable(~<Var name>,  design=mcbs)
```

For subgroup analysis use:

```
mcbs_subgrp <- subset(mcbs, <Subgroup criteria>)
svytable(~<Var name>,  design=mcbs_subgrp)
```