USER GUIDE

Medicare Current Beneficiary Survey 2015 Survey File, Public Use File

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. As a part of this goal, CMS is releasing the Medicare Current Beneficiary Survey (MCBS) 2015 Survey File Public Use File (PUF). The content of the MCBS 2015 Survey File 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, and other information regarding beneficiaries' knowledge of, attitudes toward, and satisfaction with their health care. This data release also contains demographic data and information on all types of health insurance coverage.

This user guide contains information about the MCBS 2015 Survey File PUF. Data users can obtain detailed information about the MCBS and specific background information to help data users understand and analyze the data. Readers should first familiarize themselves with the content of the 2015 Data User's Guide: General Information; the 2015 Data User's Guide: Survey File; and the 2015 MCBS Survey Methodology Report documents in order to obtain an overview of the survey, questionnaires, sample design, and other topics relevant to the MCBS in general. 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.

The MCBS 2015 Survey File PUF is not intended to replace the more detailed LDS files; rather, it provides a general-use publicly-available alternative that provides the very highest degree of protection to the Medicare beneficiaries' protected health information. The main benefits provided to potential researchers of the MCBS 2015 Survey File PUF are:

- 1. Increased data access through a free file download that is consistent with other HHS public-use survey files;
- 2. Increase 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;

- 3. Promote research to improve the Medicare program; and
- 4. Provide publically available data for key indicators of the Medicare population for benchmarking comparisons with other data sources.

2. Background of MCBS

2.1 Overview

Medicare is the nation's health insurance program for persons 65 years and older and for persons younger than 65 years who have a qualifying disability. The Medicare Current Beneficiary Survey (MCBS) contains data provided by a representative national sample of the Medicare population sponsored by the Centers for Medicare & Medicaid Services (CMS). 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 the population of Medicare beneficiaries—including both aged and disabled enrollees—in the United States, District of Columbia, and Puerto Rico. 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 MCBS 2015 Survey File PUF includes data for 12,311 community interview 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 PUF only and cannot be used for linking to other sources of data.

All variables on the PUF are numeric or integer. Formats and values for each variable are available in the PUF codebook that is available with the data download.

Variable groups are labeled with prefixes to help users identify these groups by topic area. Table 1 includes information about these variable prefixes and the associated questionnaire sections and original file locations (i.e. the 2015 LDS Survey File individual files, which are called data segments).

Table 1: MCBS 2015 Survey File PUF Variable Prefix

Prefix	Descriptions LDS Data Segments			
ADM_	Administrative Data	ADMNUTLS, HISUMRY, DEMO		
INS_	Insurance status, coverage, and type	HISUMRY, HITLINE		
INT_	Interview characteristics	INTERV		
IV_	Age, sex, and race groups	DEMO		
D_	Other demographic variables	DEMO		
ACC_	Access to, use, and	ACCESSCR, PMUSE,		
	satisfaction with health care	SATWCARE, USCPPIC		
HLT_	Health status/functioning	CHRNCOND, GENHLTH,		
	(Community)	NAGIDIS, PREVCARE,		
		VISHEAR		
PRV_	Health status/functioning	CHRNCOND, NAGIDIS,		
	(Prevention)	PREVCARE		
RSK_	Health behavior risk factors	NICOALCO		
FAL_	Falls	FALLS		
HOU_	Housing characteristics	HHCHAR		
MA_	HMO Supplement	MAPLANQX		
EE	Survey weights	EVRWGTS		

3.1 Data File Information

Detailed information about variables in the MCBS 2015 Survey File PUF can be found in the codebook. The codebook includes SAS variable names, labels, a note which indicates which respondents were eligible for the question, the question number for the question that was asked during the survey, and a label which summarizes the question text. For certain variables

some of the questionnaire categories do not match the categories provided in the public release file 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 based on combining two questions and their variable label indicates a recoded variable (e.g. HLT_ALZDEM). In cases where data was collected from both survey and administrative sources, especially for variables related to insurance coverage, administrative data was allowed to 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 2015 Survey File PUF are provided in the accompanying codebook file.

The MCBS 2015 Survey File PUF dataset is saved as a SAS export file. Directions and sample SAS code are given below to help users read the dataset into SAS.

Assume you have downloaded the MCBS 2015 Survey File PUF export (PUF2015.xpt) file in the folder "C:\MCBS\DOWNLOAD". You can then use the following SAS code to import it into SAS.

```
LIBNAME PUFLIB 'C:\MCBS\SASDATA';
FILENAME F "C:\MCBS\DOWNLOAD\PUF2015.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.

3.2 Comparison to LDS

The MCBS 2015 Survey File PUF differs from 2015 MCBS Survey LDS files because it has been evaluated for disclosure risk, and additional steps were taken to protect respondent confidentiality. The PUF contains data for 12,311 beneficiaries and 413 variables, which is a similar number of beneficiaries contained in the community-only 2015 MCBS LDS segments, but with fewer variables. Many LDS variables that were of narrow use or that posed disclosure risks were dropped or recoded to create the more general-use variable set for the PUF.

The PUF includes only those beneficiaries interviewed in the community and, thus, excludes all beneficiaries who were in a health care facility part or all of the year (n=1,288). Variables that were only created for facility residents were also not included. 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.

Additionally, the PUF contains no Medicare claims data in individual form. For users interested in this type of data, the MCBS 2015 Survey File LDS is available from CMS through a data use agreement. More information on the LDS process can be found here: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/.

A summary of the differences between the two data products appears in Table 2 below:

Table 2: Comparison Between the MCBS 2015 Survey File PUF and MCBS 2015 Survey File LDS

MCBS PUF	MCBS Survey File LDS			
Population – community dwelling	Population – community and facility			
Number of variables - 413	Number of variables – 4,041 variables across			
	33 data segments			
ID – new PUF_ID	ID – BASEID			
Randomly generated, can't be linked back to	Randomly generated, can't be linked back to			
BASEID	HIC number			
Date fields - NO	Date fields – YES			
Geographic identifiers – NO	Geographic identifiers – YES			
Cost/payment data – NO	Cost/payment data – YES			
Demographic data – YES	Demographic data – YES			
All variables are categorical	Continuous age variables available			
Insurance coverage – YES	Insurance coverage – YES			
Summarized to annual level	Monthly level			
Identifiable plan-related information for MA	Identifiable plan-related information for MA			
or Part D – NO	or Part D – YES			

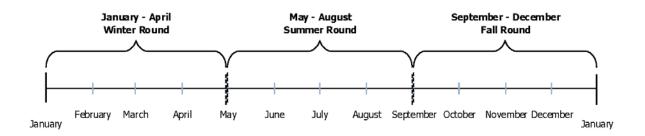
4. Survey sample information

4.1 Design of MCBS

Although participation in the survey is limited to four years, MCBS data collection is continual 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 by round configuration (rather than interviewing on an annual basis) is to have shorter periods of recall during the year in order to capture more complete health care costs and utilization from beneficiaries.

The 2015 MCBS data releases reflect data collected from January 2015 through early January 2016 (see Exhibit 4.1.1).1

Exhibit 4.1.1: Typical MCBS Data Collection Year

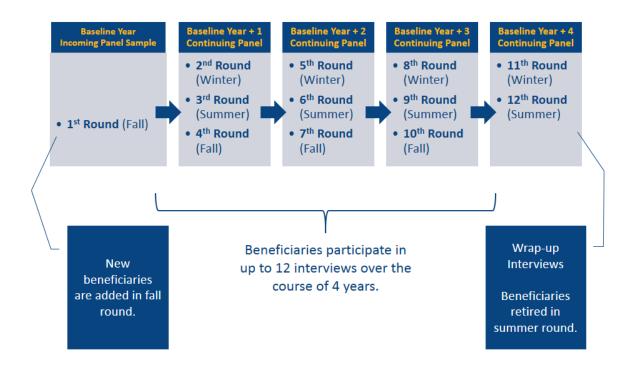


Initial interviews of newly-selected respondents take place in the fall round. Often the fall round begins early (i.e., late July or early August) to allow more time to conduct outreach and collect information from the new survey respondents who are selected to participate in the MCBS. That is, the early start of the fall round overlaps with the final weeks of data collection for the summer round. These small overlap periods as one round ends and another begins are acceptable design features of the survey. For example, the fall round usually extends into early January to allow for the completion of interviews that may have been postponed due to the holiday period.

Subsequent rounds, which occur every four months, involve the re-interviewing of the same respondent (or appropriate proxy respondents) until they have completed four years of participation (up to 12 interviews in total). Exhibit 4.1.2 depicts the timeline of participation for respondents selected to be in the MCBS sample.

¹ Due to the nature of some survey items, some data pulled is forward from a prior data collection year.

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.² Each sampled beneficiary is scientifically selected as part of a panel and is interviewed up to three times per year.³ 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.

² 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.

³ 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: 2011-2015 MCBS Rotating Panel Design

Data Collection Schedule		Panel						
Data Year	Season	Round#	2010	2011	2012	2013	2014	2015
2010	Winter	56						
	Summer	57]					
	Fall	58						
2011	Winter	59						
	Summer	60						
	Fall	61						
2012	Winter	62						
	Summer	63						
	Fall	64						
2013	Winter	65						
l	Summer	66						
	Fall	67						
2014	Winter	68						
	Summer	69						
	Fall	70						
2015	Winter/Summer	71/72						
	Fall	73	1					

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 2015 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 sample is designed to be representative of the Medicare population as a whole and by the following age groups: under 45, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 and over. Because of interest in their special health care needs, elderly beneficiaries (age 85 and over) and beneficiaries with disabilities (age 64 and under) are oversampled to permit more detailed analysis of these subpopulations. In 2013 and 2014, an additional oversample of

⁴ The MCBS sample frame includes all aged individuals, but only individuals 18 and over are interviewed and included in the LDS files.

beneficiaries in Accountable Care Organizations (ACOs) was conducted, and beginning in 2015, an oversample of Hispanic beneficiaries was implemented.⁵

The beneficiaries included in the 2015 MCBS LDS releases represent a random cross-section of all beneficiaries who were ever enrolled in either Part A or Part B of the Medicare program for any portion of 2015. A subset of these beneficiaries represent a random cross-section of all beneficiaries who were continuously enrolled from January 1, 2015 up to and including interviews conducted during Fall 2015. The Survey File and Cost Supplement File represent four separate MCBS panels identified by the year in which the panel was selected and first interviewed (i.e., for the MCBS 2015 PUF, the 2012, 2013, 2014 and 2015 panels). Exhibit 4.2.2 shows the distribution of each of the four panels included in the MCBS 2015 PUF.

Exhibit 4.2.2: 2015 MCBS Distribution of Panels in MCBS 2015 PUF						
Panel	Number of Beneficiaries Selected					
2012	1,992					
2013	2,276					
2014 ⁶	3,947					
2015	4,096					

4.3 Eligibility

4.3.1 Medicare Population Covered by the 2015 LDS

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 participation (i.e., Incoming or Continuing) and interview participation (i.e., Community or Facility), which is determined by residence

⁵ Note that once an oversample is implemented for a given fall round, those beneficiaries remain in the survey for four years.

⁶ Fall 2014 was the first round collected by NORC 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.

status. Although they appear in the 2015 MCBS LDS releases, beneficiaries residing in facilities were not included in the MCBS 2015 Survey File PUF.

4.4.1 Incoming and Continuing Cases

Every fall, 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 95 percent of the sample is interviewed in-person 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. However, over the course of a four year period, 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 (these cases are called Crossovers). In order to obtain an accurate representation of the experiences of all Medicare beneficiaries, the MCBS includes beneficiaries wherever they reside, even if they reside in and/or enter a facility for the duration of their four years with the study. The MCBS 2015 Survey File PUF does not contain data from Facility interviews.

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, clear, 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, 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 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

⁷ The fall round starts with a target of 230 field interviewers which, over the course of the year, is reduced due to staff turnover. Each summer, a cohort of new interviewers is hired for the MCBS.

Information Security Management Act of 2002. As stated in the MCBS OMB documentation, the information collected for MCBS is protected by NORC and by CMS. Respondent data is used only for research and statistical purposes. Identifiable information is not disclosed or released except those involved in research without the consent of the individual or the establishment except as required under the Privacy Act of 1974 (Public Law 93-579).

4.6 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 AAPOR⁸ and OMB. For the ever enrolled cross-sectional sample represented by the MCBS 2015 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 MCBS 2015 Survey File PUF, which excludes beneficiaries who only underwent Facility interviews.

5. Analysis

5.1 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 full sample weight is EEYRSWGT. EEYRSWGT is the "ever enrolled" cross- sectional weight and applies to both the continuing sample and to the supplemental sample that was fielded for the first time in Round 73. This weight is intended for use in cross-sectional statistics involving the total (combined) Round 73 sample. Each weight is greater than zero for all beneficiaries on

⁸ https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions2015 8theditionwithchanges April2015 logo.pdf

the file. EEYRSWGT should be used to make estimates of the levels of access to care for the Medicare population who were enrolled at any point in 2015 (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 2015 Survey File PUF can be used to calculate standard errors of the sample-based estimates as described below. 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 and STATA for estimating statistics follow below. Analysis of subgroups should utilize the domain functions within the statistical package of your choice, as shown below (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.

6. Sample Code

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;
```

Mean / Sum

proc surveymeans data=<**Analytic dataset>** plots=NONE sum VARMETHOD = brr (fay=.30);

```
var <Var name>;
      weight EEYRSWGT;
      repweight EEYRS001 - EEYRS100;
run;
Subgroup Analysis
proc surveymeans data=<Analytic dataset> plots=NONE sum VARMETHOD = brr (fay=.30);
      var <Var name>;
      weight EEYRSWGT;
      repweight EEYRS001 - EEYRS100;
      domain <Subgroup variable>;
run;
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 continuous variables, use:
```

svy brr, fay(.3) : mean <Var name>, over(<Var name>)

For subgroup analysis use:

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