Complex Sampling Weights
and
Preparing 2023 BRFSS Module Data for Analysis

July 2024

Overview

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey that collects data on a number of health outcomes, health-related risk behaviors, use of preventive services, and chronic conditions from noninstitutionalized adults who reside in each of the states and participating US territories. In 2011, BRFSS changed its data collection procedures, structure, and weighting methodology so the traditional landline telephone-based data set could include data from participants using cell phones. The BRFSS uses a core set of questions and gives states the option to include multiple modules that focus on specific health issues. Not all optional modules are collected by all of the states, and states may opt to collect module data from just subsets of their survey participants. Core questions are used during all interviews. Researchers using BRFSS data should conduct analyses with complex sampling procedures; they also should appropriately stratify and weight the data in their work. Potential bias resulting from selection probabilities and noncoverage among segments of the population can be reduced through weighting.

Researchers conducting analysis of variables from the core-only section should use the variable _LLCPWT for weighting.

BRFSS has created this document to guide users analyzing variables from 2023 BRFSS optional modules and/or combinations of module and core variables. The 2023 BRFSS data sets include data that respondents provided by landline telephone or cell phone. Data users should note that newer weighting procedures are likely to affect trend lines when comparing BRFSS data collected up to 2010 with data from 2011 and afterward; because of these changes, users are advised NOT to make direct comparisons with pre-2011 data, and instead, begin new trend lines with that year.

Data users should become familiar with the information presented in this document prior to performing analyses. Information about the changes to 2023 BRFSS data collection is available in the 2023 Data Comparability Report.

Weighting BRFSS Core Data

BRFSS has grouped the 2023 data into four data sets:

- 2023 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)

  And three versions

- 2023 BRFSS Combined Landline Telephone and Cellular Telephone Version 1
- 2023 BRFSS Combined Landline Telephone and Cellular Telephone Version 2
- 2023 BRFSS Combined Landline Telephone and Cellular Telephone Version 3

Researchers should understand that BRFSS data were collected by asking core and/or module questions on the annual questionnaire. Users, therefore, may need to use data from up to 4 different data sets, described below.

- Use the Combined Landline Telephone and Cell phone data if questions are exclusively from the core section ***or*** if questions are in common modules that were asked on both the landline survey and cellphone survey *(Note: See the section below,* ***Using BRFSS Data from Multiple Data Sets****, for more about common modules.)*

- Use Version 1, Version 2, and/or Version 3 Combined Landline Telephone and Cell Phone data sets if states collected multiple version questionnaires.

In all cases, researchers conducting complex sampling analyses from the *core-only* section should use the variable **_LLCPWT** for weighting, **_STSTR** for stratification, and the variable **_PSU** for primary sampling unit, or clustering for short. The following table displays the description of the data, names of the data sets, and the variable names of the final weight.

Data description	Data set name	Final weight variable name
Combined Landline Telephone and Cellular Telephone	LLCP2023	_LLCPWT
Combined Landline Telephone and Cellular Telephone Version 1	LLCP23V1	_LCPWTV1
Combined Landline Telephone and Cellular Telephone Version 2	LLCP23V2	_LCPWTV2
Combined Landline Telephone and Cellular Telephone Version 3	LLCP23V3	_LCPWTV3

The examples below demonstrate how to use _LLCPWT, _STSTR & _PSU when analyzing variable **HAVARTH4** (*Ever told you had some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia*) and **PHYSHLTH** (*how many days during the past 30 days was your physical health not good?*) in the core section using the 2023 Louisiana data set. Examples are given in both SAS and R.

Example in SAS:

```
LIBNAME IN '\BRFSS\';                /* Use LIBNAME to set the location of the BRFSS data set */

DATA LA23;                  /* Create a temporary data set for Louisiana */
 SET IN.LLCP2023 (WHERE=(_STATE = 22));
 RUN;

PROC SURVEYFREQ             /* Calculate percentage of population with arthritis */
  DATA = LA23;
  WEIGHT _LLCPWT;                    /* Use _LLCPWT for weighted analysis*/
  STRATA _STSTR;
  CLUSTER _PSU;
  TABLE HAVARTH4;
RUN;


PROC SURVEYMEANS            /* Calculate average number of physical healthy days */
DATA =   LA23;
WEIGHT _LLCPWT;             /* Use _LLCPWT for weighted analysis */
```

```
STRATA _STSTR;
CLUSTER _PSU;
VAR PHYSHLTH;
RUN;
```

Example in R:

```
# Install the survey package if it is not already installed
install.packages("survey")

# Call the library for the current R session
library(survey)

# Read in BRFSS data
load("\\BRFSS\\BRFSS.rdata")

# Subset the data for Louisiana(22)
BRFSS <- BRFSS[BRFSS$state == 22, ]

# Set options for allowing a single observation per stratum
options(survey.lonely.psu = "adjust")

# Create survey design
brfssdsgn <- svydesign(
   id=~1,
   strata = ~ststr,
   weights = ~llcpwt,
   data = BRFSS)

# calculate average number of physical healthy days
svymean(~physhlth,
      brfssdsgn,
      na.rm = TRUE)

# calculate percent in each arthritis category
svymean(~factor(havarth5),
      brfssdsgn,
      na.rm = TRUE)
```

Using BRFSS Data from Multiple Data Sets

Due to the complex nature of state-based data collection processes, users may have to create a data set that fits their research needs. The following two examples illustrate how to prepare module data for analysis when states have collected module data in a variety of ways.

Prior to conducting analyses using optional module data, users should always review all relevant documents for the given data year, such as the Overview, Codebook, Comparability of Data, and Module by State in particular to identify states that collected data of interest and determine which questionnaire version states used. Remember to check FIPS state codes in each data set to avoid duplication.

Individual states may have chosen to use a number of optional modules, depending on each state's needs. Individual states may have also chosen to divide their samples and use different modules in the

subsamples that were distinguished by version of the surveys. Modules that appeared in each version of a state's questionnaire are called common modules. Each state that collected common modules did so on all its versions of combined landline telephone and cell phone.

Example 1: Other Tobacco Use

We will use the module **Other Tobacco Use** to demonstrate how to combine and reweight data from multiple data sets for analysis. This module was selected because states collected the data either as a common module in the combined landline telephone and cell phone or by splitting the sample and offering the module on one or more versions of the data collection. For ease in using the 2023 optional modules, please refer to the two documents published with the annual data set: BRFSS 2023 Modules Used by State and BRFSS 2023 Modules Used by Category. In addition to showing which states used which modules, these tables list names and descriptions of the data sets, questionnaire versions, and weight variables from the corresponding data sets. If users want to analyze variables from certain modules, these two documents are a good place to start. The following information has been copied from the two module documents to show segments of the tables:

Module weights for 2023

Modules by State

State	Description	Dataset	Data weight	Module(s)
Maryland	Combined Land Line and Cell Phone data	LLCP2023	_LLCPWT	Sexual Orientation and Gender Identity (SOGI)
	Combined Land Line and Cell Phone data, version 1	LLCP23V1	_LCPWTV1	HPV Vaccination, Industry and Occupation, Reactions to Race, Social Determinants and Health Equity
	Combined Land Line and Cell Phone data, version 2	LLCP23V2	_LCPWTV2	COVID Vaccination, Cancer Survivorship: Course of Treatment, Cancer Survivorship: Pain Management, Cancer Survivorship: Type of Cancer, Caregiver, Cognitive Decline, Excess Sun Exposure, Indoor Tanning, Industry and Occupation, Lung Cancer Screening, Other Tobacco Use
	Combined Land Line and Cell Phone data, version 3	LLCP23V3	_LCPWTV3	Adverse Childhood Experiences, COVID Vaccination, Lung Cancer Screening, Marijuana Use, Other Tobacco Use, Social Determinants and Health Equity

States by Module

Module	Description	Dataset	Data weight	State(s)
Other Tobacco Use	Combined Land Line and Cell Phone data	LLCP2023	_LLCPWT	Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Guam, Hawaii, Illinois, Indiana, Iowa, Maine, Minnesota, Mississippi, Missouri, Montana, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, North Dakota, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, West Virginia, Wisconsin, Wyoming
	Combined Land Line and Cell Phone data, version 1	LLCP23V1	_LCPWTV1	Kansas, Michigan
	Combined Land Line and Cell Phone data, version 2	LLCP23V2	_LCPWTV2	Maryland, Nebraska, Ohio

	Combined Land Line and Cell Phone data, version 3	LLCP23V3	_LCPWTV3	Maryland, Michigan

1) The module **Other Tobacco Use** listed in the 2023 document **Modules (2023) by Category** shows that 35 states used the common version of the module, 2 states used version 1, 3 states used version 2, and 2 states used version 3 of the module. To prepare the data for analysis, we need to:

    a. Extract data from LLCP2023 for the 35 states that collected the common version of the module (numbers in parentheses are the values for the variable _STATE):

    Arizona (4), Arkansas (5), California (6), Colorado (8), Connecticut (9), Delaware (10), Florida (12), Georgia (13), Hawaii (15), Illinois (17), Indiana (18), Iowa (19), Maine (23), Minnesota (27), Mississippi (28), Missouri (29), Montana (30), Nevada (32), New Hampshire (33), New Jersey (34), New Mexico (35), North Carolina (37), North Dakota (38), Oregon (41), Rhode Island (44), South Carolina (45), South Dakota (46), Tennessee (47), Texas (48), Utah (49), Vermont (50), West Virginia (54), Wisconsin (55), Wyoming (56), Guam (66)

    b. Extract data from LLCP23V1 for the 2 states that collected version 1 of the module, *Kansas (20) and Michigan (26)*
    c. Extract data from LLCP23V2 for the 3 states that collected version 2 of the module, *Maryland (24), Nebraska (31) and Ohio (39)*
    d. Extract data from LLCP23V3 for the 2 states that collected version 3 of the module, *Maryland (24) and Michigan (26)*

2) **Obtain data sets**:

- To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2023):

    - Go to the 2023 data webpage, choose the 2023 data. From there, read all of the 2023 documents including the overview, codebook, and the modules by category list. The data set LLCP2023 is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS data sets are also available in the **SAS Resources** section of the page.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 data set:

    - Go to the 2023 data webpage, then find The Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data. There are 3 separate SAS data sets corresponding to Questionnaire Versions 1, 2, and 3. Choose Version 1: LLCP23V1.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 data set:

- Follow the same step above for Version 1, but instead, choose Version 2: LLCP23V2.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 3 data set:

  - Follow the same step above for Version 1, but instead, choose Version 3: LLCP23V3

3) Generate a working data set:

Create a new and uniform final weight variable from each version of the data sets downloaded in steps above and combine data into one working data set. This step will require users to do the following steps:

a. Keep all the states that collected the module data in each version of the data sets,
b. Rename the corresponding weight variable to a new and consistent weight variable,
c. Combine all the data sets into one that contains the renamed weight variable.

Maryland used versions 2 and 3 of the module, whereas Michigan used versions 1 and 2 of the module; we need to manually adjust the Weight variable for each version of their data following these steps:

Maryland:

1) Find the sample size (n2) for Maryland in V2 data LLCP23V2, which is 5198
2) Find the sample size (n3) for Maryland in V3 data LLCP23V3, which is 5260
3) Combine sample sizes for LLCPV2 and LLCPV3 for Maryland, (n2 + n3) = (5198+5260) = 10458
4) Use the sample size from each version divided by the combined sample size to get proportions; therefore,
   the proportion for version 2 data is **P2**= n2/(n2+n3) = 5198/10458≈ **0.50**
   the proportion for version 3 data is **P3**= n3/(n2+n3) = 5260/10458≈ **0.50**

Michigan: follow the same steps:

1) Sample size for version 1, n1 = 3012
2) Sample size for version 3, n3 = 3078
3) Combined sample size for version 1 and version 2, (n1 + n3) = (3012 + 3078) = 6090
4) the proportion for version 1 data is **P1**= n1/(n1+n3) = 3012/6090 ≈ **0.50**
   the proportion for version 2 data is **P3**= n2/(n1+n3) = 3078/6090 ≈ **0.50**

As we can see, the sample sizes are similar for version 2 and version, 3, 5198 vs. 5260, for Maryland, and version 1 and version 3, 3012 vs. 3078, for Michigan; we can simply divide the _FINALWT by 2. However, if the sample sizes are quite different, we should use the calculated proportions instead.

Example in SAS:

```
LIBNAME IN  '\BRFSS\';

DATA LLCP;  /* Extract data for states that used common version of the module*/
```

```
SET IN.LLCP2023 (WHERE=(_STATE IN (4 5 6 8 9 10 12 13 15 17 18 19 23 27 28 29 30 32 33 34 35 37 38 41
44 45 46 47 48 49 50 54 55 56 66)));

RENAME _LLCPWT=FINALWT;
RUN;

/* Extract data for states that used V1 of the module, adjust weights for states that used > 1 version of the module */
DATA LLCPV1;
  SET IN.LLCP23V1 (WHERE=(_STATE IN (20 26)));
  IF _STATE = 26 THEN FINALWT= (_LCPWTV1*.50); * _LCPWT1 times the calculated proportion;
  ELSE FINALWT=_LCPWTV1;
  DROP _LCPWTV1;
RUN;

/* Extract data for states that used V2 of the module, adjust weights for states that used > 1 version of the module */
DATA LLCPV2;
  SET IN.LLCP23V2(WHERE=(_STATE IN (24 31 39)));
  IF _STATE = 24 THEN FINALWT= (_LCPWTV2*.50); * _LCPWTV2 times the calculated proportion;
  ELSE FINALWT=_LCPWTV2;
  DROP _LCPWTV2;
RUN;

/* Extract data for states that used V3 of the module, adjust weights for states that used > 1 version of the module */
DATA LLCPV3;
  SET IN.LLCP23V3(WHERE=(_STATE IN (24 26)));
  IF _STATE IN (24 26) THEN FINALWT= (_LCPWTV3*.50); * _LCPWTV3 times the calculated proportion;
  ELSE FINALWT= _LCPWTV3;
  DROP _LCPWTV3;
RUN;


DATA OTUSE2023;    *combine all versions of the dataset into one;
  SET   LLCP   LLCPV1   LLCPV2 LLCPV3;
RUN;
```

The combined and reweighted dataset OTUSE2023 is ready for analysis.

Example in R:

```
# Extract data from LLCP2023 for states that used the common version of the module
# Note - R does not allow variable names to begin with underscores, Such as _LLCPWT, _LCPWTV1, _LCPWTV2
# and _LCPWTV3, remove underscores from variable names before analysis using R

llcp <- llcp2023[llcp2023$state %in%
c(4,5,6,8,9,10,12,13,15,17,18,19,23,27,28,29,30,32,33,34,35,37,38,41,44,45,46,47,48,49,50,54,55,56,66), ]

llcp$finalwt <- llcp$llcpwt # Rename weight variable LLCPWT to FINALWT for the main data set

# Extract data for states that used version 1 of the module only
llcpv1ks <- llcp23v1[llcp23v1$state %in% c(20), ]
llcpv1ks$finalwt <- llcpv1ks$lcpwtv1

# Extract data for states that used > 1 version of the module & Adjust weight
llcpv1mi <- llcp23v1[llcp23v1$state %in% c(26), ]
llcpv1mi$finalwt <- llcpv1mi$lcpwtv1* (0.50)

#V1 - combine V1 datasets
```

```
llcpv1<- rbind(llcpv1ks, llcpv1mi)

# V2 - Extract data for states that used only version 2 of the module, rename lcpwtv2 to FINALWT
llcpv2oroh <- llcp23v2[llcp23v2$state %in% c(31,39), ]
llcpv2oroh$finalwt <- llcpv2oroh$lcpwtv2

# V2 – Extract data for that states used > 1 version  of the module, & Adjust weight and rename to FINALWT
llcpv2md <- llcp23v2[llcp23v2$state %in% c(24), ]
llcpv2md$finalwt <- llcpv2md$lcpwtv2* (0.50)

# V2 – Combine V2 datasets
llcpv2<- rbind(llcpv2oroh, llcpv2md)

# V3 – Extract Version3 data for states that used > 1 version of the module, adjust weight LCPWTV3 and rename to
FINALWT
llcpv3<- llcp23v3[llcp23v3$state %in% c(24,26), ]
llcpv3$finalwt <- llcpv3mdmi$lcpwtv3* (0.50)

# Combine all data sets for analysis
otuse2023 <- rbind(llcp, llcpv1, llcpv2, llcpv3)
```

The combined and reweighted dataset **otuse2023** is ready for analysis.

Example 2: Marijuana Use

The example below uses the **Marijuana Use** Module to demonstrate how to combine and reweight data from multiple data sets.

1) The module **Marijuana Use** listed in the document Modules (2023) by Category shows that a total of 19 states used it as optional module in 2023.

- 16 states used the common version of this module; therefore, we need to extract data for these states from the dataset LLCP2023:

  Connecticut (9), Delaware (10), Guam (66), Illinois (17), Indiana (18), Maine (23), Mississippi (28), Montana (30), Nebraska (31), Nevada (32), New Mexico (35), Oregon (41), Vermont (50), Virgin Islands (78), Virginia (51), Wyoming (56)

- *Oklahoma (40)* used version 1 of the module; therefore, we need to extract data from LLCP23V1 for Oklahoma.

- *Ohio (39)* used version 2 of the module and therefore, we need to extract data from LLCP23V2 for Ohio.

- *Maryland (24)* used version 3 of the module and we need to extract data from LLCP23V3 for Maryland.

  Note: Since all of states that used **Marijuana Use** module used only one version of the module, there is no need to adjust the weight variable when we combine and analyze the module data.

2) **Obtain data sets**:

- To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2023):

➢ Go to the 2023 data webpage, and choose the 2023 data set. From there, read all of the 2023 documents including the Overview, Codebook, and Module by State. The data set is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS data sets are also available in the **SAS Resources** section of the page.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 data set:

➢ Go to the 2023 data webpage, select BRFSS the Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data. There are 3 separate SAS data sets corresponding to Questionnaire Versions 1, 2, and 3. Choose Version 1: LLCP23V1.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 data set:

➢ Follow the same step above for Version 1, but instead, choose Version 2: LLCP23V2.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 3 data set:

➢ Follow the same step above for Version 1, but instead, choose Version 3: LLCP23V3.

3) **Generate a working data set**:
Create a new and uniform final weight variable from each of the data sets, combine data into one working data set. This step will require users to do the following steps:

a. Keep all states that collected the module data in each of the data sets
b. Rename the corresponding weight variable to a consistent weight variable and adjust the weight variable if any state uses more than 1 version of the module
c. Combine all the data sets into one that contains the renamed weight variable. Details on how to adjust weight are shown below.

4) **Weight adjustment**:

Since none of the states used more than one version of the Marijuana Use module, there is no need to adjust weight for the module analysis; however, if any state used more than one version of the module, the weight variable will need to be adjusted as we did in example 1, with the Other Tobacco Use module.

Example code in SAS:

```
LIBNAME IN '\BRFSS\';

*Begin with the main dataset, LLCP2023, where Marijuana Use module was used as a common module;

DATA MJUSE;   * Extract data for 16 states that used common version of the module;
SET IN.LLCP2023 (WHERE=(_STATE
          IN (9, 10, 17, 18, 23, 28, 30, 31, 32, 35, 41, 50, 51, 56, 66, 78)));
FINALWT=_LLCPWT;      *FINALWT, new weight variable for new data sets;
RUN;

DATA MJV1;    *Extract V1 data ;
SET IN.LLCP23V1 (WHERE=(_STATE IN (40)));
FINALWT=_LCPWTV1;
RUN;

DATA MJV2;    *Extract V2 data;
SET IN.LLCP23V2 (WHERE=(_STATE IN (39)));
FINALWT = _LCPWTV2;
RUN;

DATA MJV3;    *Extract V3 data;
SET IN.LLCP23V3 (WHERE=(_STATE IN (24)));
FINALWT = _LCPWTV3;
RUN;

DATA MARJUSE;    *Create combined dataset for analysis;
SET   MJUSE   MJV1   MJV2   MJV3;
RUN;
```

The combined SAS data set, MARJUSE, is ready to be analyzed.

Example code in R:

```
# Begin with the main data set LLCP2023 for 16 states that used the common version of the Marijuana Use module
# Since R does not allow variable names to begin with underscore, such as _LLCPWT, _LCPWTV1, _LCPWTV2 and _LCPWTV3
# Remove underscores from the dataset before analysis

llcp <- llcp2023[llcp2023$state %in% c(9, 10, 17, 18, 23, 28, 30, 31, 32, 35, 41, 50, 51, 56, 66, 78), ]
llcp$finalwt <- llcp$llcpwt

# Extract V1 data for OK
nrow(llcpv1[llcp23v1$state %in%  c(40), ])

# Rename LCPWTV1 to FINALWT for OK
llcpv1$finalwt <- llcpv1$lcpwtv1

# Extract V2 data for OH
nrow(llcpv2[llcp23v2$state %in%  c(39), ])

# Rename LCPWTV2 to FINALWT for OH
llcpv2$finalwt <- llcpv2$lcpwtv2

# Extract V3 data for MD & rename LCPWTV3 to FINALWT for MD
nrow(llcpv3[llcp23v3$state %in%  c(24), ])
llcpv3$finalwt <- llcpv3$lcpwtv3
```

```
# Combine all data sets and it's ready for analysis
marjuse <- rbind(llcp, llcp1, llcpv2, llcpv3)
```

Combining Multiple Years of BRFSS Data

There are times we need to analyze data from multiple years, such as calculating the average prevalence of a variable in core sections for 2021 and 2023. In this case, we need to combine data from 2021 and 2023. The weight variable for 2021 and 2023 needs to be adjusted proportionally for the combined data set. There are two ways to adjust the weight variable. The sample size for each year helps to determine how we should adjust them. If the sample sizes are very similar, we can simply divide the original weight variable by 2. If they are different, we need to adjust the weight variable proportionally by following these steps:

1) Find out the sample sizes for 2021 and 2023
2) Sum up the total sample size for 2021 and 2023
3) Find the proportion for 2021 (Use the sample size for 2021 divided by the combined sample size for 2021 & 2023)
4) Find the proportion for 2023 (Use the sample size for 2023 divided by the combined sample size for 2021 & 2023)
5) The adjusted weight for 2021 is (_LLCPWT(2021) * proportion for 2021)
6) The adjusted weight for 2023 is (_LLCPWT(2023) * proportion for 2023)

The example below uses the 2021 and 2023 data from New York to demonstrate how to combine and reweight data from multiple years.

The sample size for 2021 New York data is 39,095 and 17,336 for 2023. Since sample sizes are quite different, the weights need to be adjusted proportionally. Follow the steps listed above to get the adjusted weights for New York for 2021 & 2023.

1) The sample size is 39,095 for 2021, and 17,336 for 2023
2) The combined sample size is 56,431
3) The proportion for 2021 is 39,095 / 56,431≈ .69
4) The proportion for 2023 is 17,336 / 56,431≈ .31;
5) The adjusted weight for 2021 is (_LLCPWT * .69)
6) The adjusted weight for 2023 is (_LLCPWT * .31)

Obtain data sets:

- To download the 2021 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2021):
  - Go to the 2021 data Webpage, and choose 2021 data set. From there, read all of the 2021 documents including the Overview, Codebook, and the Modules by Category. The data set is available in ASCII and SAS transport formats. A record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS data sets are also available in **SAS Resources** section of the page.

- To download the 2023 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2023):
  - Go to the 2023 data Webpage, and choose 2023 data set. From there, read all of the 2023 documents including the Overview, Codebook, and the Modules by Category. The data set is available in ASCII and SAS transport formats. Record layout, formats, and SAS code are also available in **SAS Resources** section of the page.

- Generate a new and uniform final weight variable (the adjusted weight variables) from each of the data sets and combine the data into one working data set. This step will require users to do these steps:
  - Rename the corresponding weight variable to a consistent weight variable; and
  - Combine all data sets into one with the consistent weight variable.

    Again, users should note that new weighting procedures are likely to affect trend lines when comparing BRFSS data collected before and after 2011; because of the changes, users are advised NOT to make direct comparisons with pre-2011 data, and instead, to begin new trend lines with that year.

Example code in SAS:

```
Libname D21 "\BRFSS2021";
Libname D23 "\BRFSS2023";

DATA NY21;    /* Check sample size for each 2021 & 2023 */
  SET D21.LLCP2021(WHERE=(_STATE=36));
  RUN;   /* There were 39095 observations read from the data set D21.LLCP2021 for NY (36) */

DATA NY23;
  SET D23.LLCP2023(WHERE=(_STATE=36));
  RUN;   /*There were 17336 observations read from the data set D23.LLCP2023 for NY (36) */

DATA NY2YR;      /* Create a final dataset with the adjusted weight variable (FINALWT) */
  SET NY21 (IN=A)
      NY23 (IN=B);
  IF A THEN FINALWT = _LLCPWT * .69;      * Use the calculated proportion to adjust _LLCPWT;
  IF B THEN FINALWT = _LLCPWT * .31;
  RUN;
```

The combined and reweighted dataset NY2YR is ready for analysis.

Example code in R:

```
# Subset 2021 and 2023 data for New York
# Since R does not allow variable names to begin with underscore, such as _LLCPWT, _LCPWTV1, _LCPWTV2 and
_LCPWTV3
# Remove underscores from the variable names in the dataset before analysis using R

NY21 <- llcp2021[llcp2021$state == 36, ]
NY23 <- llcp2023[llcp2023$state == 36, ]

# Count number of observations in each year
nrow(NY21)
nrow(NY23)
```

```
# Create a final weight based on proportion of observations in each year
NY21$finalwt <- NY21$llcpwt * (.69)
NY23$finalwt <- NY23$llcpwt * (.31)

# Combine the data sets and reweight
NY2YR <- rbind(NY21, NY23)
```

The combined and reweighted data set NY2YR is ready for analysis. ***~~~ END~~~***