

The Behavioral Risk Factor Surveillance System

Complex Sampling Weights and
Preparing 2017 BRFSS Module Data for Analysis

July 2018



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 non-institutionalized adults who reside in each of the states and selected 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 cellular telephones. 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 2017 BRFSS optional modules and/or combinations of module and core variables. The 2017 BRFSS data sets include data that respondents provided by landline telephone or cellular telephone. Data users should note that newer weighting procedures are likely to affect trend lines when comparing BRFSS data collected before and after 2011; 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. More information about the changes to the 2017 BRFSS is available in the [2017 Data Comparability Report](#).

Weighting BRFSS Core Data

BRFSS has grouped the 2017 data into four datasets:

- 2017 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)
- 2017 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 1 data (Landline Telephone and Cellular Telephone Version 1)
- 2017 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 2 data (Landline Telephone and Cellular Telephone Version 2)
- 2017 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 3 data (Landline Telephone and Cellular Telephone version 3)

As a first step to conducting research using the BRFSS data, users should understand that the data set they need is based on the location of the questions either in the core or in optional modules. Users should keep in mind that there are 4 possible data sets from which they will need to include data:

- 1) Researchers should use the Combined Landline Telephone and Cellular Telephone data if the questions are exclusively from the core section *or* if they are common modules asked on both the landline survey and cellular telephone survey.
- 2) Researchers should use the Version 1, Version 2, and/or Version 3 Combined Landline Telephone and Cellular Telephone data sets from states collecting multiple version questionnaires.

In all cases, the variable **_STSTR** should be used for stratification, and the variable **_PSU** should be used for clustering, in complex sampling analyses. 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	LLCP2017	_LLCPWT
Combined Landline Telephone and Cellular Telephone Version 1	LLCP17V1	_LCPWTV1
Combined Landline Telephone and Cellular Telephone Version 2	LLCP17V2	_LCPWTV2
Combined Landline Telephone and Cellular Telephone Version 3	LLCP17V3	_LCPWTV3

Researchers conducting analysis of variables from the *core-only* section should use the variable **_LLCPWT** for weighting. The variable **_STSTR** should be used for stratification, and the variable **_PSU** should be used as a primary sampling unit.

The example below uses the arthritis variable, HAVARTH3, and the physically healthy days variable, PHYSHLTH, in the core to demonstrate using the common weight for Louisiana. Examples are given in SAS, R, and SPSS.

Example SAS Code:

```
/* Set the location of the BRFSS data set */  
libname in '\BRFSS\';  
  
/* Calculate average number of physical healthy days */  
proc surveymeans
```

```

data = in.llcp2017 (where = (_state = 22));
/* Use common weight because question is from the core section */
weight _llcpwt;
strata _ststr;
cluster _psu;
var physhlth;
run;

/* Calculate percentage of population with arthritis */
proc surveyfreq
data = in.llcp2017 (where = (_state = 22));
/* Use common weight because question is from the core section */
weight _llcpwt;
strata _ststr;
cluster _psu;
table havarth3;
run;

```

Example R Code:

```

# 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
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, # Variable to analyze
  brfssdsgn,
  na.rm = TRUE)

# calculate percent in each arthritis category
svymean(~factor(havarth3),
  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 examples illustrate how to prepare module data for analysis when states have collected module data in a variety of ways.

Prior to conducting analyses, users should always review all relevant documents for the given data year (e.g., the [overview](#), [codebook](#), [module list](#), etc.). Users should also double-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 chosen to divide their samples and use different modules in the subsamples that were distinguished by the 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 cellular telephone.

Prior to using optional module data, users must identify the states that collected the data of interest and determine which questionnaire version the state used, if any. Check the 2017 document that lists [modules by state](#).

Example 1: Caregiver Module

The example below uses the caregiver module 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 cellular telephone or by splitting the sample and offering the module on one of the versions of the data collection. For ease in using the 2017 optional modules, we created a new table by combining names of modules, states, names of the data sets, names of the weight variables from the corresponding data sets, and descriptions of the data sets. The table is listed as module by state and state by module. If users want to analyze variables from certain modules, they need to search the list of modules by state.

- 1) Search Caregiver under the [Module-by-Category](#) list. 12 states collected data on this module.
 - a. Alaska (FIPS state code=2), Hawaii (FIPS state code=15), New Mexico (FIPS state code=35), Oregon (FIPS state code=41), and Rhode Island (FIPS state code=44) collected this data as a common module.
 - We need to extract Alaska, Hawaii, New Mexico, Oregon, and Rhode Island data from LLC2017.
 - b. Kansas (FIPS state code=20), Michigan (FIPS state code=26), and New Jersey (FIPS state code=34) used this module only with Questionnaire Version 1.
 - We therefore need to extract Kansas, Michigan, and New Jersey data from LLC17V1.
 - c. New York (FIPS state code=36), Oklahoma (FIPS state code=40), and Utah (FIPS state code=49) used this module only with Questionnaire Version 2.
 - We therefore need to extract New York, Oklahoma, and Utah data from LLC17V2.
 - d. Maryland (FIPS state code=24) used this module only with Questionnaire Version 3.
 - We therefore need to extract Maryland data from LLC17V3.
- 2) Obtain data sets:
 - To download the Combined Landline Telephone and Cellular Telephone data set (LLC2017):

- Go to the data [webpage](#), and choose the 2016 data set. From there, read all of the 2017 documents including the [overview](#), [codebook](#), and the [module list](#). The data set is available in ASCII and SAS transport formats. A record layout, format syntax, and SAS code are included to read in the raw data.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 data set:
 - Go to the [webpage](#) with Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data. Download the proper multiple-version telephone data. There are 3 separate SAS datasets corresponding to Questionnaire Version 1, Questionnaire Version 2, and Questionnaire Version 3. Choose Version 1: LLCP17V1.
- 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: LLCP17V2.
- 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: LLCP17V3.

Generate a new and uniform final weight variable from each of the data sets and combine data into a working data set. This step will require users to:

- a. keep all the states that collect the module data in each of the data sets,
- b. rename the corresponding weight variable to a consistent weight variable, and
- c. combine all the data sets into one data set that contains the consistent weight variable. Details are shown in the SAS syntax below.

Example SAS Code:

```
libname in '\BRFSS\';

*Extract state data from landline and cell phone combined data;

*Begin with the main dataset, where caregiver was used as a common module in the
landline and cellular telephone samples;
data llcp;

*Select states where the caregiver module was a common module;
set in.llcp2017 (where=( _state in (2, 15, 35, 41, 44)));
*FIPS state codes: 2 (Alaska), 15 (Hawaii), 35 (New Mexico), 41 (Oregon), 44 (Rhode
Island);

*Rename final weight variable to be consistent across new data sets;
rename _llcpwt=_finalwt;
run;

*Questionnaire Version 1: Extract states from Questionnaire V1, combined landline
and cell phone data;
data llcpv1;

*Select states where the caregiver module was only collected on questionnaire
version 1;
set in.llcp17v1 (where=( _state in (20, 26, 34)));
```

```

*FIPS state codes: 20 (Kansas), 26 (Michigan), 34 (New Jersey);

*Rename final weight variable to be consistent across new data sets;
rename _lcpwtv1=_finalwt;

run;

*Questionnaire Version 2: Extract states from Questionnaire V2, combined landline
and cell phone data;
data llcpv2;

*Select states where the caregiver module was only collected on questionnaire
version 2;
set in.llcp17v2 (where=( _state in (36, 40, 49)));
*FIPS state codes: 36 (New York), 40 (Oklahoma), 49 (Utah);

*Rename final weight variable to be consistent across new data sets;
rename _lcpwtv2=_finalwt;
run;

*Questionnaire Version 3: Extract states from Questionnaire V3, combined landline
and cell phone data;
data llcpv3;

*Select states where the caregiver module was only collected on questionnaire
version 3;
set in.llcp17v3 (where=( _state = 24));
*FIPS state code: 24 (Maryland);

*Rename final weight variable to be consistent across new data sets;
rename _lcpwtv3=_finalwt;
run;

*Combine 4 data sets together;
data caregiver;
set llcp llcpv1 llcpv2 llcpv3;
run;

```

Example R Code:

```

# Begin with the main data set for states that used the caregiver module
# as a common module.
llcp <- llcp2017[llcp2017$state %in% c(2, 15, 35, 41, 44), ]
# FIPS state codes: 2 (Alaska), 15 (Hawaii), 35 (New Mexico),
# 41 (Oregon), 44 (Rhode Island)

# Rename the final weight variable for the main data set
llcp$finalwt <- llcp$llcpwt

# Choose the states that used the caregiver module in version 1
llcpv1 <- llcp17v1[llcp17v1$state %in% c(20, 26, 34), ]
# FIPS state codes: 20 (Kansas), 26 (Michigan), 34 (New Jersey)

# Rename the final weight variable for the version 1 data set

```

```
llcpv1$finalwt <- llcpv1$lcpwtv1

# Choose the states that used the caregiver module in version 2
llcpv2 <- llcp17v2[llcp17v2$state %in% c(36, 40, 49), ]
# FIPS state codes: 36 (New York), 40 (Oklahoma), 49 (Utah)

# Rename the final weight variable for the version 2 data set
llcpv2$finalwt <- llcpv2$lcpwtv2

# Choose the states that used the caregiver module in version 3
llcpv3 <- llcp17v3[llcp17v3$state == 24, ]
# FIPS state codes: 24 (Maryland)

# Rename the final weight variable for the version 3 data set
llcpv3$finalwt <- llcpv3$lcpwtv3

# Combine the data sets
caregiver <- rbind(llcp, llcpv1, llcpv2, llcpv3)
```


Example 2: Social Determinants of Health Module

The example below uses the Social Determinants of Health Module to demonstrate how to combine and reweight data from multiple data sets. This module was selected for illustration because it covers multiple questionnaire versions.

- 1) Search Social Determinants of Health under the [Module-by-Category](#) list. 18 states used the Social Determinants of Health optional module in 2017.
 - 13 states used this module as a common version: Florida (FIPS state code=12), Georgia (FIPS state code=13), Iowa (FIPS state code=19), Kentucky (FIPS state code=21), Massachusetts (FIPS state code=25), Minnesota (FIPS state code=27), Mississippi (FIPS state code=28), New Hampshire (FIPS state code=33), Pennsylvania (FIPS state code=42), Utah (FIPS state code=49), West Virginia (FIPS state code=54), Wisconsin (FIPS state code=55), and Wyoming (FIPS state code=56).
 - We need to extract data from these 13 states from the data set LLCP2017.
 - Four states used this module with their questionnaire version 1: Colorado (FIPS state code=8), Maryland (FIPS state code=24), Ohio (FIPS state code=39), and Oklahoma (FIPS state code=40).
 - We need to extract Colorado, Maryland, Ohio, and Oklahoma data from LLCP17V1.
 - One state used this module with its questionnaire version 2: Maryland (FIPS state code=24). In addition, Maryland (FIPS state code=24) used this module with its questionnaire version 2. We need to extract Maryland's data from both LLCP17V1 and LLCP17V2.
- 2) If a state collects data on a module on *more than one—but not all—questionnaire versions*, there is an extra step to creating the appropriate analysis weight. For instance, Maryland collected the Social Determinants of Health module on Questionnaire Versions 1 and 2. In this case, we should run a frequency table to check the sample size in each of the multiple versions of the data set. Compare the sample size in these versions and decide how to calculate the final weight by multiplying by a proportion of the whole. Here is an example of SAS code and output showing Maryland's use of the Social Determinants of Health module; the state used more than one version and the sample sizes of each should be checked so that the weights can be adjusted proportionally. The SAS output shows that the frequencies of 4,412 in the first version and 4,343 in the second version of combined landline telephone and cellular telephone data are similar. Maryland's Version 1 analysis weight (_LCPWTV1) and Version 2 analysis weight (_LCPWTV2) can be divided by 2

Example SAS code:

```
proc freq data=in.llcp17v1;
tables _state;
where _state=24;
title "Maryland Sample Size, Questionnaire Version 1, Combined Landline and
Cell Phone";
run;

proc freq data=in.llcp17v2;
tables _state;
where _state=24;
```

```

title "Maryland Sample Size, Questionnaire Version 2, Combined Landline and
Cell Phone";
run;

```

SAS Output:

Maryland Sample Size, Questionnaire Version 1, Combined Landline and Cell Phone

The FREQ Procedure

FIPS STATE CODE				
_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
24	4412	100.0	4412	100.00

Maryland Sample Size, Questionnaire Version 2, Combined Landline and Cell Phone

The FREQ Procedure

FIPS STATE CODE				
_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
42	4343	100.00	4343	100.00

3) Obtain data sets:

- To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2017):
 - Go to the data [webpage](#), and choose the 2017 dataset. From there, read all of the 2017 documents including the [overview](#), [codebook](#), and the [module list](#). The data set is available in ASCII and SAS transport formats. A record layout, format syntax, and SAS code are included to read in the raw data.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 dataset:
 - Go to the webpage with Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data. Download the proper multiple-version telephone data. There are 3 separate SAS datasets corresponding to Questionnaire Version 1, Questionnaire Version 2, and Questionnaire Version 3. Choose Version 1: LLC17V1.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 dataset:
 - Follow the same step above for Version 1, but instead choose Version 2: LLC17V2.

Generate a new and uniform final weight variable from each of the data sets, and combine data into a working data set. This step will require users to:

- a. Keep all of the states that collect the module data in each of the data sets;
- b. Rename the corresponding weight variable to a consistent weight variable; and
- c. Combine all the data sets into one data set that contains the consistent weight variable.

Details are shown in the SAS syntax below.

Example SAS code:

```
libname in '\BRFSS\';

*Extract states from landline and cell phone combined data;

*Begin with the main dataset, where social determinants of health was used as a
common module in the landline and cellular telephone samples;
data llcp;

*Select states where the caregiver module was a common module;
set in.llcp2017 (where=(_state in (12, 13, 19, 21, 25, 27, 28, 33, 42,
49, 54, 55, 56)));
*FIPS state codes: 12 (Florida), 13 (Georgia), 19 (Iowa), 21 (Kentucky), 25
(Massachusetts), 27 (Minnesota), 28 (Mississippi), 33 (New Hampshire), 42
(Pennsylvania), 49 (Utah), 54 (West Virginia), 55 (Wisconsin), 56 (Wyoming);

*Rename final weight variable to be consistent across new data sets;
rename _llcpwt=_finalwt;
run;

*Questionnaire Version 1: Extract states from Questionnaire V1, combined landline
and cell phone data;
data llcpv1;

*Select states where the caregiver module was only collected on questionnaire
version 1;
set in.llcp17v1 (where=(_state in (8, 24, 39, 40)));
*FIPS state code: 8 (Colorado), 24 (Maryland), 39 (Ohio), 40 (Oklahoma);

*Rename final weight variable to be consistent across new data sets;
*Note: Maryland's version 1 weight is divided by 2 because Maryland collected
social determinants of health data on 2 questionnaire versions, and the sample size
in Version 1 is similar to Version 2;
if _state = 24 then _finalwt = _lcpwtv1/2;
else _finalwt = _lcpwtv1;
run;

*Questionnaire Version 2: Extract states from Questionnaire V2, combined landline
and cell phone data;
data llcpv2;

*Select states where the social determinants of health module was only collected on
questionnaire version 2;
set in.llcp17v2 (where=(_state = 8));
*FIPS state code: 8 (Maryland);

*Rename final weight variable to be consistent across new data sets.
*Note: Version 2 weight is divided by 2 because Maryland collected social
determinants of health data on 2 questionnaire versions, and the sample size in
Version 1 is similar to Version 2;
```

```

_finalwt=_lcpwtv2/2;
run;

*Combine all datasets together;
data sdhmodule;
set llcp llcpv1 llcpv2;
run;

```

Example R code:

```

# Determine number of observations in the version 1 data set for Maryland
nrow(llcp17v1[llcp17v1$state == 24, ])

# Determine number of observations in the version 1 data set for Maryland
nrow(llcp17v2[llcp17v2$state == 24, ])

# Begin with the main data set for states that used the social determinants
# of health module as a common module.
llcp <- llcp2017[llcp2017$state %in% c(12, 13, 19, 21, 25, 27, 28, 33, 42,
  49, 54, 55, 56), ]
# FIPS state codes: 12 (Florida), 13 (Georgia), 19 (Iowa), 21 (Kentucky),
# 25 (Massachusetts), 27(Minnesota), 28 (Mississippi), 33 (New Hampshire),
# 42 (Pennsylvania), 49 (Utah), 54 (West Virginia), 55 (Wisconsin),
# 56 (Wyoming)

# Rename the final weight variable for the main data set
llcp$finalwt <- llcp$llcpwt

# Choose the states that used the social determinants of health module in version 1
llcpv1 <- llcp17v1[llcp17v1$state %in% c(8, 24, 39, 40), ]
# FIPS state codes: 8 (Colorado), 24 (Maryland), 39 (Ohio), 40 (Oklahoma)

# Rename the final weight variable for the version 1 data set
# Maryland's version 1 weight is divided by 2 because Maryland collected
# social determinants of health module data in 2 questionnaire versions
# and the sample size in the two versions is roughly equal
llcpv1$finalwt <- ifelse(llcpv1$state == 24, llcpwtv1/2, llcpwtv1)

# Choose the states that used the social determinants of health module in version 2
# Maryland's version 2 weight is divided by 2 because Maryland collected
# social determinants of health module data in 2 questionnaire versions
# and the sample size in the two versions is roughly equal
llcpv2 <- llcp17v2[llcp17v2$state == 24, ]
# FIPS state codes: 24 (Maryland)

# Rename the final weight variable for the version 2 data set
llcpv2$finalwt <- llcpv2$llcpwtv2/2

# Combine the data sets
sdhmodule <- rbind(llcp, llcpv1, llcpv2)

```

Comparing Multiple Years of BRFSS Data

The example below uses the 2016 and 2017 BRFSS New York data to demonstrate how to combine and reweight data on the core section from multiple years.

In this case, we should run a frequency table to check the sample size in each of the data sets. Compare the sample size in these data sets and decide how to calculate the final weight by multiplying by a proportion of the whole. Here is an example of SAS code and output showing the combination of the 2016 and 2017 New York data sets; the sample sizes of each year should be checked so that the weights can be adjusted proportionally. The SAS output shows frequencies of 34,190 in the 2016 data set and 12,249 in the 2017 data set of combined landline telephone and cellular telephone data from New York. The 2016 observations make up about 75% of the total combined observations in 2016 and 2017, and the 2017 observations make up about 25% of the total combined observations in 2016 and 2017. The 2016 common weight (_LLCPWT) can be multiplied by 75% and the 2017 common weight (_LLCPWT) can be multiplied by 25%.

Obtain data sets:

- To download the 2016 Combined Landline Telephone and Cellular Telephone data set (LLCP2016):
 - Go to the data [webpage](#), and choose the 2016 dataset. From there, read all of the 2016 documents including the [overview](#), [codebook](#), and the [module list](#). The data set is available in ASCII and SAS transport formats. A record layout, format syntax, and SAS code are included to read in the raw data.
- To download the 2017 Combined Landline Telephone and Cellular Telephone Questionnaire data set:
 - Go to the data [webpage](#), and choose the 2017 dataset. From there, read all of the 2017 documents including the [overview](#), [codebook](#), and the [module list](#). The data set is available in ASCII and SAS transport formats. A record layout, format syntax, and SAS code are included to read in the raw data.

Generate a new and uniform final weight variable from each of the data sets, and combine data into a working data set. This step will require users to:

- a. Rename the corresponding weight variable to a consistent weight variable; and
- b. Combine all the data sets into one data set that contains the consistent weight variable.

Details are shown in the SAS syntax below.

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

Example SAS Code:

```
/* Check totals for each year */
proc sql;

title "Check Totals in 2016";
  select count(*)
         from in16.llcp2016
         where _state = 36;

title "Check Totals in 2017";
```

```

select count(*)
      from in17.llcp2017
      where _state = 36;

quit;

/* Create weighted dataset for multiple years */
/* Because the number of observations in 2016 account for about */
/* 75% of the total observations in 2016 and 2017, we multiply */
/* the 2016 weight by 3/4 and the 2017 weight by 1/4 */
data combineYears;
  set in16.llcp2016 (in = a)
      in17.llcp2017 (in = b);
  if a then finalwt = _llcpwt * (3/4);
  if b then finalwt = _llcpwt * (1/4);
  where _state = 36;

run;

```

Example R Code:

```

# Subset 2016 and 2017 data for New York
NY16 <- BRFSS16[BRFSS16$state == 36, ]
NY17 <- BRFSS17[BRFSS17$state == 36, ]

# Count number of observations in each year
nrow(NY16)
nrow(NY17)

# Create a final weight based on proportion of observations in each year
NY16$finalwt <- NY16$llcpwt * (3/4)
NY17$finalwt <- NY17$llcpwt * (1/4)

# Combine the data sets and reweight
NY1617 <- rbind(NY16, NY17)

```