

# **Behavioral Risk Factor Surveillance System**

**Module Data for Analysis: 2016 BRFSS**



## Overview

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey that includes information on a number of health outcomes, risk behaviors, use of preventive services, and chronic conditions for 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 new procedures have been used each year since 2011. The BRFSS includes a core set of questions and multiple optional modules that focus on specific health issues. Not all optional modules are collected by all of the states, and states may opt to include module data on portions of their surveys. Core questions are included on 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 2016 BRFSS optional modules and/or combinations of module and core variables. The 2016 BRFSS data sets include data that respondents provided by landline telephone or cellular telephone. 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, 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 2016 BRFSS is available in the [2016 Data Comparability Report](#).

## Using BRFSS Data from Multiple Data Sets

BRFSS has grouped the 2016 data into four data sets:

- 2016 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)
- 2016 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 1 data (Landline Telephone and Cellular Telephone Version 1)
- 2016 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 2 data (Landline Telephone and Cellular Telephone Version 2)
- 2016 BRFSS Combined Landline Telephone and Cellular Telephone Multiple Version Questionnaire—Version 3 data (Landline Telephone and Cellular Telephone Version 3),

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 2016 document that lists [modules by state](#).

As a first step to conducting research using the BRFSS module 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.

<b>Data description</b>	<b>Data set name</b>	<b>Final weight variable name</b>
Combined Landline Telephone and Cellular Telephone	LLCP2016	_LLCPWT
Combined Landline Telephone and Cellular Telephone Version 1	LLCP16V1	_LCPWTV1
Combined Landline Telephone and Cellular Telephone Version 2	LLCP16V2	_LCPWTV2
Combined Landline Telephone and Cellular Telephone Version 3	LLCP16V3	_LCPWTV3

## PREPARING 2016 BRFSS MODULE DATA FOR ANALYSIS: EXAMPLES

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 state FIPS codes in each data set to avoid duplication.

### 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 2016 optional modules, BRFSS 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. 20 states, the District of Columbia, and one territory collected data on this module.
  - a. Arkansas (state FIPS code=5), the District of Columbia (FIPS code=11), Georgia (state FIPS code=13), Minnesota (state FIPS code=27), Missouri (state FIPS code=29), Montana (state FIPS code=30), Nevada (state FIPS code=32), New Jersey (state FIPS code=34), North Dakota (state FIPS code=38), Oregon (state FIPS code=41), South Dakota (state FIPS code=46), Tennessee (state FIPS code=47), and Puerto Rico (FIPS code=72) collected this data as a common module.
    - We need to extract data from these 11 states, D.C., and Puerto Rico from LLCP2016.
  - b. Ohio (state FIPS code=39) and Utah (state FIPS code=49) used this module only with Questionnaire Version 1.
    - We therefore need to extract Ohio and Utah data from LLCP16V1.
  - c. Arizona (state FIPS code=4), California (state FIPS code=6), Connecticut (state FIPS code=9), and Texas (state FIPS code=48) used this module only with Questionnaire Version 2.
    - We therefore need to extract Arizona, California, Connecticut, and Texas data from LLCP16V2.
  - d. Colorado (state FIPS code=8) and New York (state FIPS code=36) used this module only with Questionnaire Version 3.
    - We therefore need to extract Colorado and New York data from LLCP16V3.
- 2) Obtain data sets:
  - To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2016):
    - Go to the data [webpage](#), and choose the 2016 data set. 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 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: LLCPI6V1.
- 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: LLCPI6V2.
- 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: LLCPI6V3.

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:

- keep all the states that collect the module data in each of the data sets,
- rename the corresponding weight variable to a consistent weight variable, and
- 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 data set, 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.llcp2016 (where=( _state in (5,11,13,27,29,30,32,34,38,41,46,47,72)));  
    *State FIPS codes: 5 (Arkansas), 11 (District of Columbia), 13  
    (Georgia), 27 (Minnesota), 29 (Missouri), 30 (Montana), 32 (Nevada),  
    34 (New Jersey), 38 (North Dakota), 41 (Oregon), 46 (South Dakota), 47  
    (Tennessee), 72 (Puerto Rico);
```

```
*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.llcp16v1 (where=( _state in (39, 49)));  
    *State FIPS codes: 39 (Ohio), 49 (Utah);
```

```

*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.llcp16v2 (where=(_state in (4,6,9,48)));
    *State FIPS codes: 4 (Arizona), 6 (California), 9 (Connecticut), 48
    (Texas);

*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 collected only on
questionnaire version 3;
set in.llcp16v3 (where=(_state in (8, 36)));
    *State FIPS code: 8 (Colorado), 36 (New York);

*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 2: Marijuana Module

The example below uses the marijuana 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 Marijuana under the [Module-by-Category](#) list. 12 states used the Marijuana optional module in 2016.
  - 10 states used this module as a common version: Alaska (state FIPS code=2), Colorado (state FIPS code=8), Florida (state FIPS code=12), Idaho (state FIPS code=16), Minnesota (state FIPS code=27), Mississippi (state FIPS code=28), Nebraska (state FIPS code=31), Ohio (state FIPS code=39), Tennessee (state FIPS code=47), and Wyoming (state FIPS code=56).
    - We need to extract data from these 10 states from the data set LLC2016.
  - One state used this module with its questionnaire version 1: Oklahoma (state FIPS code=40).
    - We need to extract Oklahoma's data from LLC16V1.
  - One state used this module with their questionnaire version 2: California (state FIPS code=6). In addition, California (state FIPS code=6) used this module with its questionnaire version 3. We need to extract California's data from both LLC16V2 and LLC16V3.
- 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, California collected the Marijuana module on Questionnaire Versions 2 and 3. 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 California's use of the Marijuana 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 3,415 in the second version and 3,444 in the third version of combined landline telephone and cellular telephone data are similar. California's Version 2 analysis weight (\_LCPWTV2) and Version 3 analysis weight (\_LCPWTV3) can be divided by 2

SAS code:

```
proc freq data=us16data.llcp16v2;
tables _state;
where _state=6;
title "California Sample Size, Questionnaire Version 2, Combined Landline
and Cell Phone";
run;
```

```
proc freq data=us16data.llcp16v3;
tables _state;
where _state=6;
title "California Sample Size, Questionnaire Version 3, Combined Landline
and Cell Phone";
run;
```

## SAS Output:

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

The FREQ Procedure

### STATE FIPS CODE

<b>_STATE</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>
<b>6</b>	3415	100.00	3415	100.00

California Sample Size, Questionnaire Version 3, Combined Landline and Cell Phone

The FREQ Procedure

### STATE FIPS CODE

<b>_STATE</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>
<b>6</b>	3444	100.00	3444	100.00

### 3) Obtain data sets:

- To download the 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 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 data sets corresponding to Questionnaire Version 1, Questionnaire Version 2, and Questionnaire Version 3. Choose Version 1: LLC16V1.
- 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: LLC16V2.
- 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: LLC16V3.



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 marijuana 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.llcp2016 (where=(_state in (2,8,12,16,27,28,31,39,47,56)));
    *State FIPS codes: 2 (Alaska), 8 (Colorado), 12 (Florida), 16 (Idaho),
    27 (Minnesota), 28 (Mississippi), 31 (Nebraska), 39 (Ohio), 47
    (Tennessee), 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.llcp16v1 (where=(_state =40));
    *State FIPS code: 40 (Oklahoma);

*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 cali2;

*Select states where the caregiver module was collected only on
questionnaire version 2;
set in.llcp16v2 (where=(_state =6));
    *State FIPS code: 6 (California);

*Rename final weight variable to be consistent across new data sets.
```

```
*Note: Version 2 weight is divided by 2 because California collected
marijuana data on 2 out of 3 questionnaire versions (i.e., California
collected marijuana data on Questionnaire Versions 2 and 3), and the sample
size in Version 2 is similar to Version 3;
_finalwt=_lcpwtv2/2;
run;
```

```
*Questionnaire Version 3: Extract states from Questionnaire V3, combined
landline and cell phone data;
data cali3;
set in.llcp16v3 (where=( _state =6));
    *State FIPS code: 6 (California);
```

```
*Rename final weight variable to be consistent across new data sets.
*Note: Version 3 weight is divided by 2 because California collected
marijuana data on 2 out of 3 questionnaire versions, and the sample size in
Version 3 is similar to Version 2;
finalwt=_lcpwtv3/2;
run;
```

```
*Combine all data sets together;
data marijuana;
set llcp llcpv1 cali2 cali3;
run;
```