

Behavioral Risk Factor Surveillance System

Module Data for Analysis for 2015 BRFSS

July 2016





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. Since 2011, BRFSS has 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 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 weighted 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

2015 BRFSS modules and/or combinations of module and core variables. The 2015BRFSS 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 2015 BRFSS is available in the [2015 data comparability report](#).

Using BRFSS Data from Multiple Datasets

BRFSS has grouped the 2015 data into four datasets:

- 2015 BRFSS questionnaire data (combined landline telephone and cellular telephone),
- 2015 BRFSS combined landline telephone and cellular telephone multiple version questionnaire version 1 data (landline telephone and cellular telephone version 1),
- 2015 BRFSS combined landline telephone and cellular telephone multiple version questionnaire version 2 data (landline telephone and cellular telephone version 2), and
- 2015 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 used 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 2015 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 use the combined landline telephone and cellular telephone data if the questions are exclusively from the core section or common modules asked on both the landline survey and cellular telephone survey.
- 2) Researchers use the combined landline telephone and cellular telephone data sets from states collecting multiple version questionnaires (up to 3).

In all cases, researchers should use the variable `_STSTR` for stratification and `_PSU` should be used for cluster 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	LLCP2015	<code>_LLCPWT</code>
Combined landline telephone and cellular telephone Version 1	LLCP15V1	<code>_LCPWTV1</code>
Combined landline telephone and cellular telephone Version 2	LLCP15V2	<code>_LCPWTV2</code>
Combined landline telephone and cellular telephone Version 3	LLCP15V3	<code>_LCPWTV3</code>

Examples of Preparing 2015 BRFSS Module Data for Analysis

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 the analysis, users should always go through all documents relative to the given data year before conducting analysis and double-check state FIPS codes in each data set to avoid duplication.

Example 1: Sodium or Salt-related Behavior Module

The example below uses the sodium or salt-related behavior 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 module tables, in 2015, we created a new table by combining names of modules, states, names of the data set, names of the weight variable from the corresponding data sets, and the description of the data sets. The table was 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 Sodium or Salt-related Behavior under the [Module-by-Category](#) list. 7 states and one territory used this module as a common version: Alabama (State FIPS code=1),

Indiana (State FIPS code=18), Iowa (State FIPS code=19), Kentucky (State FIPS code=21), Maine (State FIPS code=23), North Carolina (State FIPS code=37), Tennessee (State FIPS code=47), and Puerto Rico (FIPS code=72). We need to extract the data from LLC2015. Nebraska (State FIPS code=31) used this module only with a questionnaire version 1. Oregon (State FIPS code=41) used this module only with a questionnaire version 2. We therefore need to extract the Nebraska data from the data set LLC15V1 and Oregon data from the dataset LLC15V2.

2. Obtain data sets:

- To download the landline telephone and cellular telephone data set (LLC2015): Go to the data [webpage](#) and choose the 2015 data set. From there, read all the 2015 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 the raw data.
- To download the combined landline telephone and cellular telephone data set with questionnaire version 1:
 - Go to the [webpage](#) with data for the combined landline and cellular telephone multiple version questionnaire module. Download the proper multiple-version landline telephone data. There are 3 separate SAS datasets corresponding to version 1, version 2, and version 3.
Choose Version 1: LLC15V1.
- To download the combined landline telephone and cellular telephone data set for questionnaire version 2:
 - Follow the same step above for Version 1, but then choose Version 2: LLC15V2.

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 on the next page.

Example SAS code:

```
/*Example SAS code*/
*extract states from landline and cell phone combined data;
data llcp; *observation number=56,195;
*originally downloaded data selecting states that collected data using
sodium or salt-related behavior module from combined landline and cell phone
data for common module;
set in.llcp2015 (where=(_state in (1,18,19,21,23,37,47,72)));
*state FIPS code:
1(Alabama),18(Indiana),19(Iowa),21(Kentucky),23(Maine),
37(North Carolina),47(Tennessee),72(Puerto Rico);
*rename final weight variable to be consistent across new data sets;
rename _llcpwt=_finalwt;
run;

*extract states from combined landline and cell phone data, multiple versions;
*multiple version 1;
data llcpv1; *observation number=8,774;
*originally downloaded data selecting states that collected data for the
sodium or salt-related behavior module from combined landline and cell phone data,
multiple versions;
set in.llcp15v1 (where=(_state in (31))); *state FIPS code: 31(Nebraska);
*rename final weight variable to be consistent across new data sets;
rename _lcpwtv1=_finalwt;
run;

*multiple version 2;
data llcpv2; *observation number=2,536;
*originally downloaded data selecting states that collected data for the
sodium or salt-related behavior module from combined landline and cell phone data,
multiple versions;
set in.llcp15v2 (where=(_state in (41))); *state FIPS code: 41(Oregon);
*rename final weight variable to be consistent across new data sets;
rename _lcpwtv2=_finalwt;
run;

*combine 3 data sets together;
data sodium; *observation number=67,505;
set llcp llcpv1 llcpv2;
run;
```

Example 2: Arthritis Management Module

The example below uses the arthritis management 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 Arthritis Management under the [Module-by-Category](#) list. 7 states used this module as a common version: Kentucky (State FIPS code=21), Minnesota (State FIPS code=27), Missouri (State FIPS code=29), Montana (State FIPS code=30), Pennsylvania (State FIPS code=42), Rhode Island (State FIPS code=44), and South Carolina (State FIPS code=45). We need to extract the data of the 7 states from the data set LLCP2015. Three states used this module with their questionnaire version 1. They are California (State FIPS code=6), Oregon (State FIPS code=41), and Utah (State FIPS code=49). Two states used this module with their questionnaire version 2. They are Kansas (State FIPS code=20) and New York (State FIPS code=36). We therefore need to extract the data of California, Oregon, and Utah from LLCP15V1, and the data of Kansas and New York from LLCP15V2.
2. In some of the modules (e.g., Arthritis Management) where the states have multiple versions of combined landline telephone and cellular telephone data, 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 Michigan's use of the Arthritis Management 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 995 in the second version and 998 in the third version of combined landline telephone and cellular telephone data are similar.

SAS code:

```
proc freq data=in.llcp15v2;
where ARTHWGT ne . and _state=26;
*ARTHWGT is one of the variables in the Arthritis Management Module';
tables _state;
title1 "Combined landline and cell phone multiple Version 2:";
title2 "Michigan sample size for Arthritis Management Module";
run;
proc freq data=in.llcp15v3;
where ARTHWGT ne . and _state=26;
*ARTHWGT is one of the variables in the Arthritis Management Module';
tables _state;
title1 "Combined landline and cell phone multiple Version 3:";
title2 "Michigan sample size for Arthritis Management Module";
run;
```

SAS output:

```
Combined landline and cell phone multiple Version 2:
Michigan sample size for Arthritis Management Module
```

The FREQ Procedure

STATE FIPS CODE				
_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
26	995	100.00	995	100.00

Combined landline and cell phone multiple Version 3:
Michigan sample size for Arthritis Management Module

The FREQ Procedure

STATE FIPS CODE				
_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
26	998	100.00	998	100.00

3. Obtain data sets:

- To download the landline telephone and cellular telephone data set (LLCP2015):
Go to the data [webpage](#) and choose the 2015 data set. From there, read all the 2015 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 the raw data.
- To download the combined landline telephone and cellular telephone dataset with questionnaire versions 1:
 - Go to the combined landline and cellular telephone multiple version questionnaire module data webpage to download multiple-version landline telephone data. There are 3 separate SAS datasets corresponding to version 1, version 2, and version 3. Choose Version 1: LLCP15V1.
- To download the combined landline telephone and cellular telephone dataset with questionnaire versions 2:
 - Follow the same step above, but then choose Version 2: LLCP15V2.
- To download the combined landline telephone and cellular telephone dataset with questionnaire versions 3:
 - Follow the same step above, but then choose Version 3: LLCP15V3.

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:

```
/*Example SAS code*/
*extract states from landline and cell phone combined data;
data llcp; *observation number=62,478;
*originally downloaded data selecting states that collected data using
arthritis management module from combined landline and cell phone data
as a common version;
set in.llcp2015 (where=(state in (21,27,29,30,42,44,45)));
*state FIPS code:
21(Kentucky),27(Minnesota),29(Missouri),30(Montana),42(Pennsylvania),
44(Rhode Island),45(South Carolina);
*rename final weight variable to be consistent across new data sets;
rename _llcpwt=_finalwt;
run;

*extract states from combined landline and cell phone data, multiple versions;
*multiple version 1;
data llcpv1; *observation number=12,010;
*originally downloaded data selecting states that collected data for the
Arthritis Management Module from combined landline and cell phone data,
multiple versions;
set in.llcp15v1 (where=(state in (6,41,49)));
*state FIPS code: 6(California),41(Oregon),49(Utah);
*rename final weight variable to be consistent across new data sets;
rename _lcpwtv1=_finalwt;
run;

*extract states from combined landline and cell phone data, multiple versions;
*multiple version 2;
data llcpv2; *observation number=17,471;
*originally downloaded data selecting states that collected data for the
Arthritis Management Module from combined landline and cell phone data,
multiple versions;
set in.llcp15v2 (where=(state in (20,36)));
*state FIPS code: 20(Kansas),36(New York);
*rename final weight variable to be consistent across new data sets;
rename _lcpwtv2=_finalwt;
run;

*extract Michigan data because it collected this module from
2 multiple versions of combined landline and cellphone data;
data MI2; *observation number=2,900;
*original downloaded data, select Michigan that collected data from
combined landline and cell phone, Version 2;
set in.llcp15v2 (where=(state=26));
*rename final weight variable to be consistent across new data sets;
```

```

_finalwt = _lcpwtv2/2;
*Note: Version 2 weight is divided by 2 because Michigan collected
2 versions and the sample size is similar as Version 3 weight;
drop _lcpwtv2;
run;

data MI3; *observation number=2,973;
*original downloaded data, select Michigan that collected data from combined
landline and cell phone, Version 3;
set in.llcp15v3 (where=(state=26));
*rename final weight variable to be consistent across new data sets;
_finalwt = _lcpwtv3/2;
*Note: Version 3 weight is divided by 2 because Michigan collected 2
versions and the sample size is similar as Version 2 weight;
drop _lcpwtv3;
run;

*combine all data sets together;
data arthritis; *observation number=97,832;
set llcp llcpv1 llcpv2 MI2 MI3;
run;

```