Preparing 2011 BRFSS Module Data for Analysis

February 2013

Division of Behavioral Surveillance

Public Health Surveillance and Informatics Program Office

Office of Surveillance, Epidemiology, and Laboratory Services

Centers for Disease Control and Prevention

Contents

Overview and Purpose of this Document

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey that includes information on a number of health outcomes, risk behaviors, and chronic conditions for persons residing in each of the states and selected U.S. territories. The BRFSS data collection, structure, and weighting methodology changed in 2011 to allow data to be collected by cell phones, in addition to landline telephones. The BRFSS includes a core set of questions and multiple modules that focus on specific health issues. Not all 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. Analysis using BRFSS data should be conducted using complex sampling analyses. Data should be appropriately stratified and weighted in the analyses. Weighting can help selection probabilities and noncoverage among segments of the population.

If users intend to analyze the variables from the core section only, _LLCPWT is the appropriate weight for analysis. This document has been created to guide users who are analyzing the variables from 2011 BRFSS data using the module or module and core sections. Please note that the dataset naming convention and weighting variables have changed from previous years in order to illustrate that weighting variables are different from those used in the past. Data users should become familiar with the information presented in this document prior to performing analyses. More information about the changes to the 2011 BRFSS is available at http://www.cdc.gov/surveillancepractice/reports/brfss/brfss.html.

Using BRFSS Data from Multiple Datasets

For 2011 BRFSS data, there are 5 datasets:

- 2011 BRFSS data (combined landline and cell phone),
- 2011 BRFSS landline questionnaire data (landline only),
- 2011 BRFSS landline multiple version questionnaire version 1 data (landline version 1),
- 2011 BRFSS landline multiple version questionnaire version 2 data (landline version 2), and
- 2011 BRFSS landline multiple version questionnaire version 3 data (landline version 3).

Individual states may have chosen to use a number of optional modules, depending on a state's needs. Although core questions are always collected on both landline and cell phones, module data may have been collected by landline and/or cell phone. In addition, states collecting module data by landline only may have chosen to split their modules in order to achieve a wider range of data. By splitting the modules, the states divided their samples and used different modules in the subsamples that were distinguished by the version of the surveys. Some modules may appear only on versions of questionnaires given to landline respondents. There are no split versions in cell phone only data. Modules that appear in every version of a state's questionnaire are called "common" modules. Common modules may be collected by cell phone and landline or by landline only.

Prior to using data that have been collected in optional modules, users must identify the states that collected the data of interest and determine which questionnaire version the state used, if any. A listing of modules by state and by category is available at http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2011. As a first step to conduct research using the BRFSS module data, users should understand that the dataset they need is based on the location of the questions in the core, in optional modules collected by landline, or in optional modules collected by landline and cell phone. Keep in mind that:

1) The combined landline and cell phone data are used if the questions are exclusively from the core section or the questions also come from the module data but are in both landline and cell phone surveys.
2) The landline-only data are used if the questions are only asked in the common version of the landline survey.
3) The combined landline and cell phone data, the landline-only data, and up to 3 datasets of landline multiple versions are used if the questions include module data with different questionnaire versions.

In all cases, the variable _STSTR should be used for stratification and _PSU should be used for cluster in complex sampling analyses. The description of the data, name of the datasets, and the variable name of the final weight are illustrated in the following table. If the user wants to study the core question only, combined landline and cell phone data are recommended.

Table: Data Description, Dataset Names, and Variable Name of the Final Weight

Data description	Dataset name	Final weight variable name
Combined landline and cell phone	LLCP2011	_LLCPWT
Landline only	LAND2011	_LANDWT
Landline only Version 1	LAND11V1	_LNDWTV1
Landline only Version 2	LAND11V2	_LNDWTV2
Landline only Version 3	LAND11V3	_LNDWTV3

Examples of Preparing 2011 BRFSS Module Data for Analysis

Due to the complex nature of state-based data collection processes, users may have to create a dataset 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, always go through all documents on the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011.htm before conducting analysis and double-check state FIPS codes in each dataset to avoid overlap.

Example 1: Cognitive Impairment Module

The example below uses the Cognitive Impairment Module to demonstrate how to combine and reweight data from multiple datasets that can be used for analysis. This module was selected because states collected the data either as a common module (cell phone and landline or landline only), or by splitting the sample and offering the module on one of the versions of the survey (landline only).

1. In the 2011 landline and cell phone data, search for states that have chosen to use this module: Go to the Web page at http://apps.nccd.cdc.gov/BRFSSModules/ModByCat.asp?Yr=2011 to find states that have data for the cognitive impairment module recorded in the combined landline-cell phone dataset. These states are Hawaii (15*), Illinois (17), New Hampshire (33), South Carolina (45), Tennessee (47), West Virginia (54), and Wisconsin (55). (*Note: Numbers in parentheses are the State FIPS codes. The state FIPS codes can be found in http://www.bls.gov/lau/lausfips.htm.)

2. Search for states that collect data for this module among landline-only users, common module: Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline.htm and click on the .rtf icon for the document named "Use of BRFSS Landline Questionnaire Data"; find the table of "2011 Landline Survey States and Modules," which begins on page 3. Information in this table is listed by state. Search for "cognitive impairment" in this table to find the states used landline. They are Arkansas (5), Florida (12), Hawaii (15), Illinois (17), Iowa (19), Louisiana (22), New Hampshire (33), North Carolina (37), South Carolina (45), Tennessee (47), West Virginia (54), and Wisconsin (55). Some of these states may also use the Cognitive Impairment Module to collect data from participants using cell phones; therefore, these states need to be compared with those that used both landline and cell phone combined data to obtain the states for landline only, common module. These are Arkansas (5), Florida (12), Iowa (19), Louisiana (22), and North Carolina (37).

3. Search for states that collect data for this module using multiple versions among landline-only users: Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline_multiple.htm and click on the .rtf icon for the document named "Use of BRFSS Landline Multiple Version Questionnaire Data" and find the table of "2011 Multiple-Version Questionnaire States and Modules." The table is listed by state. Search for "cognitive impairment" in this table to find the states that used multiple versions of the landline survey. They are California (6) for Version 1, Maryland (24) for Version 2, Michigan (26) for Version 1, Nebraska (31) for Version 3, New York (36) for Version 2, Oklahoma (40) for Version 2, Texas (48) for Version 2, Utah (49) for Version 3, and Washington (53) for Version 1.

4. Obtain datasets:

- To download the landline and cell phone dataset (LLCP2011):
  Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011.htm. From there, please read all the documents including the overview, codebook, and the module list. The dataset comes in ASCII, SAS transport formats, and SAS files with record layout, format syntax to read the raw data. The Web site also provides the link to access 2011 BRFSS landline data or click on the Web link as follows in the next step.

- To download the landline-only dataset:
  Go to the Web page at
  http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline.htm. The Web site also provides the link to access 2011 landline multiple version questionnaires or click on the Web site below, in the next bullet item.

5. To download multiple-version landline data:
   Go to the Web page at
   http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline_multiple.htm to download multiple-version landline data. There are 3 separate SAS datasets corresponding to version 1, version 2, and version 3.

6. Generate a new and uniform final weight variable from each of the datasets and combine data into a working dataset. This will require users to:
   a. keep all the states that collect the module in each of the datasets,
   b. rename the corresponding weight variable to a consistent weight variable,
   c. combine all the datasets into one dataset that contains the consistent weight variable.

Details in SAS Syntax

```
/*Example SAS code*/
*extract states from landline and cell phone combined data;
data llcp; *observation number = 48,889;
   *originally downloaded data selecting states that collected data using the
cognitive impairment module from combined landline and cell phone data;
    set libname.llcp2011 (where=(_state in (15, 17, 33, 45, 47, 54, 55)));
   *State FIPS code: 15 (Hawaii), 17 (Illinois), 33 (New Hampshire), 45 (South
Carolina), 47 (Tennessee), 54 (West Virginia), 55 (Wisconsin);
   *rename final weight variable to be consistent across new datasets;
   _finalwt = _llcpwt;
   drop _llcpwt;
run;

*extract states from landline-only data, common module;
data land; *observation number = 41,360;
```

```
  *original downloaded data selecting states that collected data for the cognitive
impairment module from landline only data;
  set libname.land2011 (where=(_state in (5, 12, 19, 22, 37)));
  *State FIPS code: 5 (Arkansas), 12 (Florida), 19 (Iowa), 22 (Louisiana), 37
(North Carolina);
  *rename final weight variable to be consistent across new datasets;
  _finalwt = _landwt;
  drop _landwt;
run;

*extract states from landline data, multiple Version 1 only;
data landv1; *observation number = 17,871;
  *original downloaded data selecting states that collected data for the cognitive
impairment module from landline only data, Version 1;
   set libname.land11v1 (where=(_state in (6, 26, 53)));
  *State FIPS code: 6 (California), 26 (Michigan), 53 (Washington);
  *rename final weight variable to be consistent across new datasets;
   _finalwt = _lndwtv1;
  drop _lndwtv1;
run;

*extract states from landline data, multiple Version 2 only;
data landv2; *observation number = 18,082;
  *originally downloaded data selecting states that collected data for the
cognitive impairment module from landline only data, Version 2;
  set libname.land11v2 (where=(_state in (24, 36, 40, 48)));
  *State FIPS code: 24 (Maryland), 36 (New York), 40 (Oklahoma), 48 (Texas);
  *rename final weight variable to be consistent across new datasets;
  _finalwt = _lndwtv2;
  drop _lndwtv2;
run;

*extract states from landline data, multiple Version 3 only;
data landv3; *observation number = 12,672;
  *originally downloaded data selecting states that collected cognitive impairment
   module from landline only data, Version 3;
  set libname.land11v3 (where=(_state in (31, 49)));
  *State FIPS code: 31 (Nebraska), 49 (Utah);
  *rename final weight variable to be consistent across new datasets;
  _finalwt = _lndwtv3;
  drop _lndwtv3;
run;

*combine all datasets;
data cognitive; *observation number = 138,874;
  *final dataset containing uniform weight variable and all states that collected
   cognitive impairment module data from 5 datasets;
  set llcp land landv1 landv2 landv3;
run;
```

Example 2: Chronic Obstructive Pulmonary Disease (COPD) Module

The example below uses the Chronic Obstructive Pulmonary Disease (COPD) Module to demonstrate how to combine and reweight data from multiple datasets. This module was selected because states collected the data either as a common module (cell phone and landline or landline only) or by splitting the sample and offering the module on one or more versions of the survey (landline only).

1. Search for states that use this module to collect data by landline and cell phone:
   Go to the Web page at http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2011 to find states that have the COPD module in the combined landline and cell phone dataset. They are Arizona (4), Connecticut (9), Illinois (17), Michigan (26), Minnesota (27), Montana (30), North Carolina (37), Tennessee (47), and West Virginia (54).

2. Search for states that collect this module for landline only, common module:
   Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline.htm and click on the .rtf file icon for the document titled, "Use of BRFSS Landline Questionnaire Data" and find the table of "2011 Landline Survey States and Modules." This table is listed by state. Search for "COPD" in this table to find the states that used this module and contacted participants through landline; they are Arizona (4), Connecticut (9), District of Columbia (11), Illinois (17), Iowa (19), Kentucky (21), Massachusetts (25), Michigan (26), Minnesota (27), Montana (30), Nevada (32), North Carolina (37), Oregon (41), Tennessee (47), West Virginia (54), and Puerto Rico (72). Some of these states may also collect cell phone data on the COPD module; therefore, these states need to be compared with those that are landline and cell phone combined to obtain the states for landline only, common module. These are District of Columbia (11), Iowa (19), Kentucky (21), Massachusetts (25), Nevada (32), Oregon (41), and Puerto Rico (72).

3. Search for states that used this module for landline-only participants, multiple versions:
   Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline_multiple.htm and click on the .rtf file icon for the document titled, "Use of BRFSS Landline Multiple Version Questionnaire Data" and find the table of "2011 Multiple-Version Questionnaire States and Modules." The table is listed by state. Search for "COPD" in this table to find the states that used multiple versions of landline. They are California (6) for Versions 1 and 2, Kansas (20) for Version 1, Maine (23) for Version 2, Nebraska (31) for Versions 2 and 3, New Jersey (34) for Version 1, Ohio (39) for Version 1, and Utah (49) for Version 2.

4. Obtain datasets:

- To download the landline and cell phone dataset:
  Go to the Web page at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011.htm. From there, please read all the documents including the overview, codebook, and the module list. The dataset comes in ASCII, SAS transport formats, and SAS files with record layout, format syntax

to read the raw data. The Web site also provides the link to access 2011 BRFSS landline data or click on the Web link in the next bulleted item, below.

- To download the landline-only dataset:
  Go to the Web page at
  http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline.htm. The Web site also provides the link to access 2011 landline multiple version questionnaires or click on the Web link in the next bulleted item, below.

- To download multiple-version landline data:
  Go to the Web page at
  http://www.cdc.gov/brfss/technical_infodata/surveydata/2011/2011_landline_multiple.htm.
  There are 3 separate SAS datasets corresponding to Version 1, Version 2, and Version 3.

5. To generate a new and uniform final weight variable from each of the datasets and combine these datasets into one new dataset for analysis, follow the SAS syntax, below. In some of the modules (e.g., COPD) where the states have multiple versions of landline data, run a frequency table to check the sample size in each of the multiple versions of landline data. 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 COPD module. The SAS output shows that the frequencies of 344 in the first Version and 362 in the second version of landline are similar.

SAS Code and Outputs of Example 2

```
  proc freq data=libname.land11v1;
  where copdtest not = . and _state=6; *copdtest is one of the COPD module
variables;
  tables _state;
  titles "Landline only Version 1: California sample size for COPD module";
run;
proc freq data= libname.land11v2;
  where copdtest not = . and _state=6; *copdtest is one of the COPD module
variables;
  tables _state;
  titles "Landline only Version 2: California sample size for COPD module";
run;
```

SAS Output:

Landline only Version 1: California sample size for COPD module

STATE FIPS CODE

_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
6	344	100.00	344	100.00

Landline only Version 2: California sample size for COPD module

STATE FIPS CODE

_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
6	362	100.00	362	100.00

SAS Code and Output: Nebraska COPD

Here is an example of SAS code and output showing Nebraska's use of the COPD module. The SAS output shows that the sample size of 337 in the second Version is about one-third of the sum of the sample sizes in the second and third Version of landline (n = 337+678 = 1,015).

```
  proc freq data=libname.land11v2;
  where copdtest not = . and _state=31; *copdtest is one of the COPD module
variables;
  tables _state;
  titles "Landline only Version 2: Nebraska sample size for COPD module";
run;
proc freq data= libname.land11v3;
  where copdtest not = . and _state=31; *copdtest is one of the COPD module
variables;
  tables _state;
  titles "Landline only Version 3: Nebraska sample size for COPD module";
run;
```

SAS Output:

Landline only Version 2: Nebraska sample size for COPD module

STATE FIPS CODE

_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
31	337	100.00	337	100.00

Landline only Version 3: Nebraska sample size for COPD module

STATE FIPS CODE

_STATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
31	678	100.00	678	100.00

/*Example SAS code*/

```
*extract states from landline and cell phone combined data;
data llcp; *observation number = 78,254;
    *originally downloaded data selecting states that collected COPD from combined
     Landline and cell phone data;
    set libname.llcp2011 (where=(_state in (4, 9, 17, 26, 27, 30, 37, 47, 54)));
    *State FIPS code: 4 (Arizona), 9 (Connecticut), 17 (Illinois), 26 (Michigan),
27 (Minnesota), 30 (Montana), 37 (North Carolina), 47 (Tennessee), 54 (West
Virginia);
    *rename final weight variable to be consistent across new datasets;
    _finalwt = _llcpwt;
    drop _llcpwt;
run;

*extract states from landline only data, common module;
data land; *observation number = 54,099;
  *originally downloaded data selecting states that collected COPD from landline
   only data;
  set libname.land2011 (where=(_state in (11, 19, 21, 25, 32, 41, 72)));
  *rename final weight variable to be consistent across new datasets;
  _finalwt = _landwt;
  drop _landwt;
run;

*extract states from landline data, multiple Version 1 only;
data landv1; *observation number = 18,928;
  *originally downloaded data selecting states that collected COPD from landline
   only data, Version 1;
  set libname.land11v1 (where=(_state in (20, 34, 39)));
  *State FIPS code: 20 (Kansas), 34 (New Jersey), 39 (Ohio);
  *rename final weight variable to be consistent across new datasets;
    _finalwt = _lndwtv1;
   drop _lndwtv1;
run;

*extract states from landline data, multiple Version 2 only;
data landv2; *observation number = 6,629;
   *originally downloaded data selecting states that collected COPD from landline
   only data, Version 2;
  set libname.land11v2 (where=(_state in (23, 49)));
  *State FIPS code: 23 (Maine), 49 (Utah);
  *rename final weight variable to be consistent across new datasets;
  _finalwt = _lndwtv2;
   drop _lndwtv2;
run;

*Note: for the COPD module, there is no state collecting landline data, Version 3
only. Instead, California collected both Version 1 and Version 2, Nebraska
collected both Version 2 and Version 3. The datasets for these 2 states need to be
created separately;

*extract California because it collected COPD module data from 2 multiple versions
of landline data;

data CA1; *observation number = 5,439;
  *originally downloaded data, landline only Version 1, select CA only;
  set libname.land11v1 (where=(_state=6));
```

```
  *rename final weight variable to be consistent across new datasets;
    _finalwt = _lndwtv1*(1/2);
 *Note: Version 1 weight is divided by 2 because California collected 2 versions
and the sample size is same as Version 2;
   drop _lndwtv1;
run;

data CA2; *observation number = 5,661;
  *originally downloaded data, landline only Version 2, select CA only;
  set libname.land11v2 (where=(_state=6));
  *rename the final weight variable to be consistent across new datasets;
  _finalwt = _lndwtv2*(1/2);
*Note: Version 2 weight is divided by 2 because California collected 2 versions and
the sample size in Version 2 (n = 362) is similar as Version 1 (n = 344);
  drop _lndwtv1;
run;

*extract Nebraska because it collected COPD module data from 2 multiple versions of
landline data;

data NE2; *observation number = 5,138;
  *originally downloaded data, landline only Version 2, select NE only;
  set libname.land11v2 (where=(_state=31));
  *rename final weight variable to be consistent across new datasets;
   _finalwt = _lndwtv2*(1/3);
*Note: Version 2 weight is divided by 3 because Nebraska collected 2 versions and
the sample size in Version 2 is one-third of the sum of the sample sizes in Version
2 and Version 3;
   drop _lndwtv2;
run;

data NE3; *observation number = 10,089;
  *originally downloaded data, landline only Version 3, select NE only;
  set libname.land11v3 (where=(_state=31));
  *rename the final weight variable to be consistent across new datasets;
  _finalwt = _lndwtv3*(2/3);
*Note: Version 3 weight should be multiplied by 2/3 because Nebraska collected 2
versions and the sample size in Version 3 (n = 678) is two-thirds of the sum of the
sample sizes in Version 2 and Version 3 (678+337 = 1,015);
  drop _lndwtv3;
run;

*combine all datasets together;

data COPD; *observation number = 184,237;
  set llcp land landv1 landv2 CA1 CA2 NE2 NE3;
run;
```