

# The BRFSS Data User Guide

August 15, 2013



---

## **Table of Contents**

Introduction .....	3
The BRFSS Process .....	3
Survey Protocol.....	5
Sampling Design .....	6
Weighting.....	7
Using BRFSS Data Sets .....	9
Calculated Variables .....	10

## Introduction

In 1984, the Centers for Disease Control and Prevention (CDC) initiated the state-based Behavioral Risk Factor Surveillance System (BRFSS)--a cross-sectional telephone survey that state health departments conduct monthly over landline telephones and cellular telephones with a standardized questionnaire and technical and methodologic assistance from CDC. BRFSS is used to collect prevalence data among adult U.S. residents regarding their risk behaviors and preventive health practices that can affect their health status. Respondent data are forwarded to CDC to be aggregated for each state, returned with standard tabulations, and published at year's end by each state. In 2011, more than 500,000 interviews were conducted in the states, the District of Columbia, and participating U.S. territories and other geographic areas.

This document is intended to provide a brief overview of BRFSS to data users. Specific information regarding data quality, response and/or cooperation rates, or calling outcome can be found in the Summary Data Quality Report produced each year in conjunction with the annual data release. Data users needing more information about comparability across years should refer to the annual Comparability of Data document, particularly before using the data to conduct trend analyses. (For an example, please see the [2011 Comparability of Data document](#)).

## The BRFSS Process

The BRFSS questionnaire was developed in collaboration between CDC and public health departments in each of the states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Data derived from the questionnaire provide health departments, public health officials and policymakers with behavioral information that, when combined with mortality and morbidity statistics, inform public health officials as they establish health-related policies and priorities as well as address and assess strategies to promote good health.

### *Annual questionnaire construction*

The BRFSS questionnaire is comprised of an annual standard core, a biannual rotating core, optional modules, and state-added questions.

**Standard Core Questions:** The portion of the questionnaire that is included each year and must be asked by all states. Each year, the core includes questions about emerging or “late-breaking” health issues. After one year, these questions are either discontinued or incorporated into the fixed core, rotating core, or optional modules.

**Rotating Core Questions:** The portion of the questionnaire asked by all states on an every-other year basis.

**Optional Modules:** Sets of standardized questions on various topics that each state may select and include in its questionnaire. Once selected, a module must be used in its entirety and asked of all eligible respondents. If an optional module is modified in any way (e.g., if a question is omitted), then the questions will be treated as state-added questions (see below).

In order to achieve a wide range of data, states may choose to “split” samples that include only selected modules. Some modules, therefore, may appear only on “versions” of questionnaires. For example, if the questionnaire adopted by a state is too long to ensure respondent cooperation, different modules may be separated among respondents in order to include more modules. States are required to conduct at least 2,500 interviews for each of the versions of the questionnaire in order to have enough responses for weighting purposes. Data users who are analyzing data from different versions of modules should read technical documents for the years included in the analyses in order to understand different weighting variables included in the datasets. For an example of these documents, refer to the [module support document for 2011](#) data. Refer to the [list of the modules](#) to see which ones were used by each state in each year and find a comprehensive listing of questions and topics by year.

**State-added Questions:** States are encouraged to gather data on additional topics related to their specific health priorities through the use of extra questions they choose to add to their questionnaire. All questions included in the BRFSS are cognitively tested prior to inclusion in the questionnaire.

The exact wording of the questions in any part of the BRFSS is determined at the annual BRFSS meeting in March, where BRFSS State Coordinators vote to adopt questions submitted by CDC programs. A governing group of state BRFSS coordinators, known as the BRFSS Working Group, may add questions on emerging issues (such as the H1N1 flu questions added in 2009). After the BRFSS meeting, CDC then designs core components and optional modules as well as produces data processing layouts, while taking state priorities, potential funding, and other practical aspects into consideration. The new BRFSS materials for the next surveillance year are then sent to the states, which then may add their own questions that they have designed or acquired.

### *Data Collection*

States may opt to contract with a private company or university to conduct interviews or conduct interviews internally, but regardless of who conducts data collection, it is done according to BRFSS protocols. States generally follow a suggested BRFSS interviewing schedule and complete all calls for a given survey month within the same sample month. Up to 15 calling attempts may be made for each phone number in the sample, depending on state regulations for calling and outcomes of previous calling attempts. Rules on calling attempts are set by the CDC and are provided annually in the Summary Data Quality Reports. In general, surveys are conducted using the following calling occasions:

- Conduct 20% of the interviews on weekdays
- Conduct 80% on weeknights and weekends

- Change schedules to accommodate holidays and special events
- Make weekday calls just after the dinner hour
- Make appointment callbacks during hours that are not scheduled for other interviews, generally on weekdays

Whether data are collected in-house or by contracted organizations, states must develop and maintain procedures to ensure respondents' confidentiality, assure and document the quality of the interviewing process, and supervise and monitor the interviewers.

### *Disposition Codes*

Each telephone number in the sample must be assigned a final disposition code to indicate a particular result of calling the number:

- > A completed interview or
- > A determination that:
  - \*A household was eligible to be included but an interview was not completed or
  - \*A telephone number was ineligible or could not have its eligibility determined.

The final disposition codes are then used to calculate response rates, cooperation rates and refusal rates. The distribution of individual disposition codes and the rates of cooperation, refusal and response are published annually in the Summary Data Quality Reports. The BRFSS uses standards set by the American Association of Public Opinion Research (AAPOR) to determine disposition codes and response rates.

## **Survey Protocol**

In order to maintain consistency across states, the BRFSS sets standard protocols for data collection. These standards allow for state-to-state data comparison in data. The following steps are included in the BRFSS survey protocol:

1. All states ask the core questions without modification. States may choose to add any, all, or none of the optional modules and state-added questions after the core component.
2. Systematic, unobtrusive electronic monitoring is a routine and integral part of monthly survey procedures for all interviewers. States may also use callback verification procedures to ensure data quality.
3. An eligible household is defined as a housing unit that has a separate entrance, where occupants eat separately from other persons on the property, and that is occupied by its members as their principal or secondary place of residence. The following are non-eligible households: vacation homes not occupied by household members for more than 30 days per year, group homes, institutions, and (in the landline sample) households in states other than the one conducting the particular BRFSS questionnaire. Since 2011, adult students living in college housing were included as eligible respondents.

4. Eligible household members include all related adults (aged 18 years or older), unrelated adults, boarders/roomers, and domestic workers who consider the household their home, even though they may not be home at the time of the call. Persons living in college housing are treated as single adult households. Household members do not include adult family members who are currently living elsewhere.
5. Proxy interviews are not conducted within the BRFSS. For persons interviewed on landline telephones, individual respondents are randomly selected from all adults, aged 18 years and older, living in a household and are interviewed in accordance with BRFSS protocol. Cellular telephone interviews are conducted with respondents who answer the number called and are treated as one-person households.
6. An interview is considered complete if data are collected for age, race, and sex (approximately half-way through the core BRFSS questionnaire). If values on age or race are not entered, imputed values will be generated and used only to assign weights.
7. Unless electronic monitoring of interviewers is being routinely conducted, a 5% random sample of each month's interviews must be called back to verify selected responses for quality assurance.
8. With the exception of verbally abusive respondents, eligible persons who initially refuse to be interviewed will be contacted at least one additional time and given the opportunity to be interviewed. Preferably, this second contact will be made by a supervisor or a different interviewer.
9. States are required to give a final disposition for every number in the sample, usually within a single month.

## Sampling Design

In some instances, states design samples within boundaries of substate geographic regions. States may determine that they would like to sample by county, public health district or other sub-state geography in order to make comparisons of geographic areas with their states. In order to conduct the BRFSS, states obtain samples of telephone numbers from CDC. States then review their sampling methodology with a state statistician and CDC to make sure data collection procedures are in place to follow the methodology. If any change in sampling methodology is considered, states consult with CDC before making changes. The BRFSS uses two samples: one for landline telephone respondents and one for cellular telephone respondents. Since landline telephones are often shared among persons living within a residence, household sampling is used in the landline sample. Household sampling requires interviewers to collect information on the number of adults living within a residence and then select randomly from all eligible adults. Cellular telephone respondents are weighted as single adult households.

### *The Landline Sample*

Disproportionate stratified sampling (DSS) has been used for the landline sample since 2003. DDS draws telephone numbers from two strata (lists) that are based on the presumed density of

known telephone household numbers. In this design, telephone numbers are classified into strata that are either high density (**listed 1+ block** telephone numbers) or medium density (**not listed 1+ block** telephone numbers) to yield residential telephone numbers. Telephone numbers in the high density stratum are sampled at the highest rate. The rate at which each stratum is sampled is called the *sampling rate*. The ratio of the sampling rate of one stratum to sampling rate of a reference stratum is called the *sampling ratio*. For BRFSS the landline sampling ratio of high to medium density is 1:1.5.

The DSS design attempts to find a way of differentiating, before sampling begins, between a set of telephone numbers that contains a large proportion of target numbers (the high-density block) and a set that contains a smaller proportion of target numbers (the medium-density block). It is possible to create more than two groups, but for BRFSS, only two groups are used. DSS sampling telephone numbers is more efficient than simple random sampling.

### *The Cellular Telephone Sample*

The cellular telephone sample is randomly generated from a sampling frame of confirmed cellular area code and prefix combinations. Cellular telephone respondents are randomly selected with each having equal probability of selection. States complete approximately 20% of their completed interviews with respondents on cell phones. CDC provides a separate cellular telephone sample to each state, according to the total number of completes that the state is targeting for that year. Some of the numbers in the cell phone sample will reach respondents who have moved into other states. Each state completes the core BRFSS interview with respondents from other states. Data from out-of-state interviews are transferred to the appropriate states at the end of each data-collection period.

### *Sample Size*

Sample size refers to the number of telephone numbers that must be called within a given period of time. The BRFSS goal is to support at least 4,000 interviews per state each year. Factors influencing sample size include the cost involved in data collection for a larger sample and the states' need for obtaining estimates for subpopulations within states.

### *Geographic Stratification*

The BRFSS samples landline telephone numbers based on substate geographic regions. Regional sampling is used to target data collection to geographic subpopulations (such as residents within a public health district). Large numbers of geographic strata increase costs but can provide information regarding smaller areas. In effect, the data collection structure for each stratum is the same as that for an entire state. For a state with seven strata, the time, cost and effort involved is the same as for seven states, each with one stratum. In 2011 and 2012, cellular telephone samples were not stratified by substate geographies and were stratified only by state. Beginning in 2013, cellular telephone stratification is conducted by the BRFSS, although geographic specificity is less reliable for cellular telephone numbers than for landline numbers.

## **Weighting**

Data weighting is an important statistical process that attempts to remove bias in the sample. The BRFSS weighting process includes two steps: design weighting and iterative proportional fitting

(also known as “raking” weighting). Because raking does not require demographic information for small geographic areas, it allows for the introduction of more demographic variables than were used by the BRFSS in the past. Since 2011, telephone ownership, education level, marital status, and home ownership were added to age, sex, race, ethnicity and region, which were the variables used in prior years. Moreover, since state level demographic characteristics of cellular telephone-only households are not available, weighting with the previous method of post stratification was no longer feasible.

Design Weighting takes into account the number of phones and the number of adults in each household. It also takes into account the number of available records (NRECSTR) and the number of records selected (NRECSEL) within each geographic strata (\_GEOSTR) and density strata (\_DENSTR). The first step is to calculate the stratum weight (\_STRWT) from the number of records in the strata and the number of records selected. The design weight is calculated in the following way within each \_GEOSTR\*\_DENSTR combination:

$$\_STRWT = \frac{NRECSTR}{NRECSEL}$$

Once the stratum weight is calculated, the number of adults within the household and the number of phones are used to calculate the design weight:

$$Design\ Weight = STRWT \left( \frac{1}{Number\ of\ phones} \right) * Number\ of\ adults\ in\ the\ household$$

Questions on the number of adults and the number of telephones in each household are asked during the screening process of each landline telephone interview. For cellular telephone respondents, the number of adults and the number of telephones in the household are set to 1, and cellular telephone respondents are treated as one adult/one phone households in the design weighting process.

Since 2011, BRFSS’s new weighting protocols have ensured that data are representative of the population on a number of demographic characteristics including sex, age, race, education, marital status, home ownership, phone ownership (landline telephone, cellular telephone or both) and sub-state region. Because raking considers each of the weighting variables separately, there is less likelihood that categories of age and/or race would be collapsed than under previous weighting methods (see the [Fact Sheet on Weighting Changes](#)). Design weights continue to be used with the new weighting protocols, with the exception that for cellular telephone respondents, the number of telephones and the number of adults in the household are set to 1. The final weight is based on the following formula:

$$Design\ Weight * Raking\ Adjustment.$$

Raking weighting incorporates the known characteristics of the population into the sample. If the sample is disproportionately female, raking will adjust the responses of females in the sample to accurately represent the proportion of females in the population. This is done in an iterative



process, with each demographic factor introduced in a sequence. The sequence of factors may be multiple times before the sample is found to accurately represent the population on all factors under consideration. BRFSS raking variables include race and ethnicity in detailed categories, sex, age, home ownership, education, marital status, phone ownership and region.

The final BRFSS data set includes several variables that are used in the weighting process. The following table explains data weighting variables:

Weighting Variables Included in the BRFSS Dataset	
Variable Name	Description
_LLCPWT	The final weight assigned to each respondent for the landline telephone and cellular telephone combined data. This weight should be used when landline telephone and cellular telephone data are included in analyses.
_LANDWT	The weight assigned to each respondent for the landline telephone sample only. This weight should be used in analyses of landline telephone data when cellular telephone data are not included.
_STRWT	Stratum weight accounts for differences in the basic probability of selection among strata (subsets of area code/prefix combinations). It is the inverse of the sampling fraction of each stratum. There almost never is a complete correspondence between strata, which are defined by subsets of area code/prefix combinations, and regions, which are defined by the boundaries of government entities.
NUMPHON2	The number of residential telephone numbers in the respondent's house.
NUMADULT	The number of adults in the respondent household.
NRECSEL	Number of records selected to be included in the sample.
NRECSTR	Number of records available in the strata.

## Using BRFSS Data Sets

The BRFSS Web site includes a variety of data sets for analysis. All BRFSS data should be analyzed using complex sampling procedures. The statistical procedures or modules of statistical software (e.g., SUDAAN, SAS and SPSS) include statements for stratification, clustering, and sample weight to account for complex sampling design of survey data. The following are examples for specifying strata, cluster and sample weights *for the combined landline telephone and cellular telephone data* when analyzing the BRFSS data:

SAS or SPSS:  
 Strata: \_STSTR;  
 Cluster: \_PSU;

Weight\_LLCPWT;

SUDAAN:  
NEST\_STSTR\_PSU;  
WEIGHT\_LLCPWT;

If users are analyzing data from landline telephone interviews only, the weighting variable should be \_LANDWT. Users should refer to the technical documentation available on the BRFSS Web site for the annual data used in analyses. *Lack of attention to the complex survey design specifications in analyses will yield biased estimates of population prevalence and of associated confidence intervals.* Users should refer to the appropriate SUDAAN, SAS, SPSS, or other software user manuals for detailed discussion concerning incorporation of complex survey designs before conducting data analyses.

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

## Calculated Variables

The BRFSS includes a number of variables that have been calculated from responses provided during the interviews. For example, respondents are asked about their height and weight from which their Body Mass Index (BMI) is calculated. Calculated variables are identifiable by a preceding underscore before alpha characters (i.e. \_BMI). If a downloaded SAS dataset has been

read into SPSS, an @ will appear before the underscore (i.e. @\_BMI). The codebooks for each year of BRFSS data indicate which variables have been calculated as well as the methods used to create the variables. In addition, an annual document entitled Calculated Variables in Data Files are available on the BRFSS Web site to illustrate the SAS code used to create each of the calculated variables.