
Behavioral Risk Factor Surveillance System

OVERVIEW: BRFSS 2016

June 29, 2017





Background

The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project between all of the states in the United States (US) and participating US territories and the Centers for Disease Control and Prevention (CDC). The BRFSS is administered and supported by CDC's Population Health Surveillance Branch, under the Division of Population Health at the National Center for Chronic Disease Prevention and Health Promotion. The BRFSS is a system of ongoing health-related telephone surveys designed to collect data on health-related risk behaviors, chronic health conditions, and use of preventive services from the noninstitutionalized adult population ≥ 18 years residing in the United States.

The BRFSS was initiated in 1984, with 15 states collecting surveillance data on risk behaviors through monthly telephone interviews. Over time, the number of states participating in the survey increased; BRFSS now collects data in all 50 states as well as the District of Columbia and participating US territories. During 2016, all 50 states, the District of Columbia, Puerto Rico, Guam, and the US Virgin Islands collected BRFSS data. In this document, the term “state” is used to refer to all areas participating in the BRFSS, including the District of Columbia, Guam, the Commonwealth of Puerto Rico, and US Virgin Islands.

BRFSS’s objective is to collect uniform state-specific data on health risk behaviors, chronic diseases and conditions, access to health care, and use of preventive health services related to the leading causes of death and disability in the United States. Factors assessed by the BRFSS in 2016 included health status, healthy days/health-related quality of life, health care access, exercise, inadequate sleep, chronic health conditions, oral health, tobacco use, e-cigarettes, alcohol consumption, immunization, falls, seat belt use, drinking and driving, breast- and cervical cancer screening, prostate cancer screening, colorectal cancer screening, and HIV/AIDS knowledge. Since 2011, the BRFSS had been conducting both landline telephone- and cellular telephone-based surveys. All the responses were self-reported; proxy interviews are not conducted by the BRFSS. In conducting the landline telephone survey, interviewers collect data from a randomly selected adult in a household. In conducting the cellular telephone survey, interviewers collect data from adults answering the cellular telephones residing in a private residence or college housing. Beginning in 2014, all adults contacted through their cellular telephone were eligible regardless of their landline phone use (i.e., complete overlap).

The BRFSS field operations are managed by state health departments that follow protocols adopted by the states, with technical assistance provided by CDC. State health departments collaborate during survey

development and conduct the interviews themselves or use contractors. The data are transmitted to CDC for editing, processing, weighting, and analysis. An edited and weighted data file is provided to each participating state health department for each year of data collection, and summary reports of state-specific data are prepared by CDC. State health departments use the BRFSS data for a variety of purposes, including identifying demographic variations in health-related behaviors, designing, implementing, and evaluating public health programs, addressing emergent and critical health issues, proposing legislation for health initiatives, and measuring progress toward state health objectives.¹ For specific examples of how state officials use the finalized BRFSS data sets, please refer to the appropriate [state information](#) on the BRFSS website.

Health characteristics estimated from the BRFSS pertain to the noninstitutionalized adult population—aged 18 years or older—who reside in the United States. In 2016, an optional module was included to provide a measure for several childhood health and wellness indicators, including asthma prevalence for people aged 17 years or younger. BRFSS respondents are identified through telephone-based methods. According to the 2016 American Community Survey (ACS), 97.5% of all occupied housing units in the United States had telephone service available and telephone non-coverage ranged from 1.5% in Connecticut to 3.6% in Arkansas.² It is estimated that 4.5% of occupied households in Puerto Rico did not have telephone service.² The increasing percentage of households that are abandoning their landline telephones for cellular telephones has significantly eroded the population coverage provided by landline telephone-based surveys to pre-1970s levels. The preliminary results (July to December 2016) from the National Health Interview Survey (NHIS) indicate that 50.8% of American households had only wireless telephones.³ Using a dual-frame survey including landline and cellular telephones improved the validity, data quality, and representativeness of BRFSS data.

In 2011, a new weighting methodology called iterative proportional fitting (or “raking”)⁴ replaced the post-stratification method to weight BRFSS data. Raking allows incorporation of cellular telephone survey data and permits the introduction of additional demographic characteristics (e.g., education level, marital status, home renter/owner) in addition to age-race/ethnicity-gender that improves the degree and extent to which the BRFSS sample properly reflects the socio-demographic make-up of individual state. The 2016 BRFSS raking method includes categories of age by gender, detailed race and ethnicity groups, education levels, marital status, regions within states, gender by race and ethnicity, telephone source, renter or owner status, and age groups by race and ethnicity. In 2016, 50 states, the District of Columbia, Guam, Puerto Rico, and the US Virgin Islands collected samples of interviews conducted by landline and cellular telephone.

The BRFSS Design

The BRFSS Questionnaire

Each year, the states—represented by their BRFSS coordinators and CDC—agree on the content of the questionnaire. The BRFSS questionnaire consists of a core component, optional modules, and state-added questions. Many questions are taken from established national surveys, such as the National Health Interview

Survey or the National Health and Nutrition Examination Survey. This practice allows the BRFSS to take advantage of questions that have been tested and allows states to compare their data with those from other surveys. Any new questions that states, federal agencies, or other entities propose as additions to the BRFSS must go through cognitive testing and field testing before they can become part of the BRFSS questionnaire. In addition, a majority vote of all state representatives is required before questions are adopted. The BRFSS guidelines—agreed upon by the state representatives and CDC—specify that all states ask the core component questions without modification. They may choose to add any, all, or none of the optional modules and may add questions of their choosing as state-added questions.

The questionnaire has three parts:

- 1. Core component:** A standard set of questions that all states use. Core content includes queries about current health-related perceptions, conditions, and behaviors (e.g., health status, health care access, alcohol consumption, tobacco use, fruits and vegetable consumptions, HIV/AIDS risks), as well as demographic questions. The core component includes the annual core comprising questions asked each year and rotating core questions that are included in even- and odd-numbered years.
- 2. Optional BRFSS modules:** Sets of questions on specific topics (e.g., pre-diabetes, diabetes, sugar sweetened beverages, excess sun exposure, caregiving, shingles, cancer survivorship) that states elect to use on their questionnaires. Generally, CDC programs submit module questions and the states vote to adopt final questions that can be included as optional modules. For more information, please see the [questionnaire section](#) of the BRFSS website.
- 3. State-added questions:** Individual states develop or acquire these questions and add them to their BRFSS questionnaires. CDC does not edit, evaluate, or track the results of these questions.

The BRFSS supported 25 modules in 2016, but states limited modules and state-added questions to only the most useful for their state program purposes, in order to keep surveys at a reasonable length. Because different states have different needs, there is wide variation between states in terms of question totals each year. The BRFSS implements a new questionnaire in January and usually does not change it significantly for the rest of the year. The flexibility of state-added questions, however, does permit additions, changes, and deletions at any time during the year.

The 2016 list of [optional modules](#) used on both the landline telephone and cellular telephone surveys is available on the BRFSS website. In order to allow for a wider range of questions in optional modules, combined landline telephone and cellular telephone data for 2016 include up to three split versions of the questionnaire. A split version is when a subset of telephone numbers for data collection still followed the state sample design, and administrators used it as the state's BRFSS sample, but the optional modules and state-added questions may

have been different from other split-version questionnaires. For additional information on split version questionnaires, see the document 2016 module data appendix table.

Annual Questionnaire Development

The governance of the BRFSS includes a representative body of state health officials, elected by region. During the year, the State BRFSS Coordinators Working Group meets with CDC's BRFSS program management.

Before the beginning of the calendar year, CDC provides states with the text of the core component and the optional modules that the BRFSS will support in the coming year. States select their optional modules and ready any state-added questions they plan to use. Each state then constructs its own questionnaire. The order of the questioning is always the same—interviewers ask questions from the core component first, then they ask any questions from the optional modules, and the state-added questions. This content order ensures comparability across states and follows the BRFSS guidelines. Generally, the only changes that the standard protocol allows are limited insertions of state-added questions on topics related to core questions. CDC and state partners must agree to these exceptions. In some cases, however, states have not been able to follow all set guidelines. Users should refer to the yearly Comparability of Data document, which lists the known deviations.

Once each state finalizes its questionnaire content—consisting of the core questionnaire, optional modules, and state-added questions—the state prepares a hard copy or electronic version of the instrument and sends it to CDC. States use the [questionnaire](#) without changes for one calendar year, and CDC archives a copy on the BRFSS website. If a significant portion of any state's population does not speak English, states have the option of translating the questionnaire into other languages. Currently, CDC provides a Spanish version of the core questionnaire and optional modules. Specific wording of the Spanish version of the questionnaire may be adapted by the states to fit the needs of their Hispanic populations.

Sample Description

In a telephone survey such as the BRFSS, a sample record is one telephone number in the list of all telephone numbers the system randomly selects for dialing. To meet the BRFSS standard for the participating states' sample designs, one must be able to justify sample records as a probability sample of all households with telephones in the state. All participating areas met this criterion in 2016. Fifty-one projects used a disproportionate stratified sample (DSS) design for their landline samples. Guam, Puerto Rico, and the US Virgin Islands used a simple random-sample design.

In the type of DSS design that states most commonly used in the BRFSS landline telephone sampling, the BRFSS divides telephone numbers into two groups, or strata, which are sampled separately. The high-density and medium-density strata contain telephone numbers that are expected to belong mostly to households.

Whether a telephone number goes into the high-density or medium-density stratum is determined by the number of listed household numbers in its hundred block, or set of 100 telephone numbers with the same area code, prefix, and first 2 digits of the suffix and all possible combinations of the last 2 digits. BRFSS puts numbers from hundred blocks with 1 or more listed household numbers (1+ blocks, or banks) in either the high-density stratum (listed 1+ blocks) or medium-density stratum (unlisted 1 + blocks). The BRFSS samples the two strata to obtain a probability sample of all households with telephones.

Cellular telephone sampling frames are commercially available and the system can call random samples of cellular telephone numbers, but doing so requires specific protocols. The basis of the 2016 BRFSS sampling frame is the Telecordia database of telephone exchanges (e.g., 617-492-0000 to 617-492-9999) and 1,000 banks (e.g., 617-492-0000 to 617-492-0999). The vendor uses dedicated cellular 1,000 banks, sorted on the basis of area code and exchange within a state. The BRFSS forms an interval— K —by dividing the population count of telephone numbers in the frame— N —by the desired sample size— n . The BRFSS divides the frame of telephone numbers into n intervals of size K telephone numbers. From each interval, the BRFSS draws one 10-digit telephone number at random.

The target population (aged 18 years and older) for cellular telephone samples in 2016 consists of people residing in a private residence or college housing who have a working cellular telephone.

In the sample design, states begin with a single stratum. To provide adequate sample sizes for smaller geographically defined populations of interest, however, many states sample disproportionately from strata that correspond to sub-state regions. In 2016, the 48 states with geographic stratification were Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Puerto Rico, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, Wisconsin, and US Virgin Islands. As a precaution to protect the confidential responses provided by the respondent, specific variables (such as sub-state geographic identifiers, detailed race or ethnicity, and older than 80 years of age) in a given year are removed.

State health departments may directly collect data from their states or they may use a contractor. In 2016, 9 state health departments collected their data in-house and the remainder contracted with other data collectors. In 2016, the CDC provided samples purchased from Marketing Systems Group, Inc. (MSG) to all 54 states and territories.

Data Collection

Interviewing Procedures

In 2016, 54 states or territories used Computer-Assisted Telephone Interview (CATI) systems. CDC supports CATI programming using the Ci3 WinCATI software package. This support includes programming the core and module questions for data collectors, providing questionnaire scripting of state-added questions for states requiring such assistance, and contracting with a Ci3 consultant to assist states. Following guidelines provided by the BRFSS, state health personnel or contractors conduct interviews. The core portion of the questionnaire lasts an average of 18 minutes. Interview time for modules and state-added questions is dependent upon the number of questions used, but generally, they add 5 to 10 minutes to the interview.

Interviewer retention is very high among states that conduct the survey in-house. The state coordinator or interviewer supervisor conducts repeated training specific to the BRFSS. Contractors typically use interviewers who have experience conducting telephone surveys, but these interviewers are given additional training on the BRFSS questionnaire and procedures before they are approved to work on the BRFSS.

The BRFSS protocols require evaluation of interviewer performance. During 2016, all BRFSS surveillance sites had the capability to monitor their interviewers. Interviewer-monitoring systems vary from listening to the interviewer only at an on-site location to listening to both the interviewer and respondent at remote locations. Some states also use verification callbacks in addition to direct monitoring. Contractors typically conducted systematic monitoring of each interviewer a certain amount of time each month. All states had the capability to tabulate disposition code frequencies by interviewer. These data were the primary means for quantifying interviewer performance.

States conducted telephone interviews during each calendar month. They made calls 7 days per week, during both daytime and evening hours. They followed standard BRFSS procedures for rotation of calls over days of the week and time of day. Detailed information on interview response rates is available in the BRFSS 2016 Summary Data Quality Report.

Data Processing

Preparing for Data Collection and Data Processing

Data processing is an integral part of any survey. Because states collect and submit data to CDC each month, the BRFSS performs routine data processing tasks on an ongoing basis. Once the final version of the new questionnaire becomes available each year, CDC staff take steps to prepare for the next cycles of data collection. These steps include developing edit specifications, programming portions of the Ci3 WinCATI software, programming the editing software, and producing telephone sample estimates as requested by states

and ordering the sample from the contract vendor. CDC produces a Ci3 WinCATI data entry module for each state that requests it. CDC staff also must incorporate skip patterns, together with some consistency edits, and response-code range checks into the CATI system. These edits and skip patterns serve to reduce interviewer, data entry, and skip errors. Developers prepare data conversion tables that help processors read the survey data from the entry module, call information from the sample tracking module, and combine information into the final format for that data year. CDC also creates and distributes a Windows-based editing program that can perform data validations on files with proper survey result formats. This program helps users with output lists of errors or warns users about conditions of concern that may exist in the data.

CDC begins to process data for the survey year as soon as states (or their contractors) begin submitting data to the data management mailbox. Data processing continues throughout the survey year. CDC receives and tracks monthly data submissions from the states. Once data are received from a state, CDC staff run editing programs and cumulative data quality checks and note any problems in the files. A CDC programmer works with each state until any problems are optimally resolved. CDC staff generate data quality reports and share them with state coordinators, who review the reports and discuss any potential problems. Once CDC receives and validates the entire year of data for a state, processors run several year-end programs on the data. These programs perform some additional, limited data cleanup and fixes specific to each state and data year and produce reports that identify potential analytic problems with the data set. Once this step is completed, data are ready for assigning weights and adding calculated variables. Calculated variables are created for the benefit of users and can be noted in the data set by the leading underscore in the variable name. The following calculated variables are examples of results from this procedure:

- `_RFSMOK3`
- `_ECIGSTS`
- `_HCVU651.`
- `_AGE80`
- `_DRNKDRV`

For more information, see the [*Calculated Variables and Risk Factors in Data Files*](#) document. Several variables from the data file are used to create these variables in a process that varies in complexity. Some are based only on combined codes, while others require sorting and combining of particular codes from multiple variables.

Almost every variable derived from the BRFSS interview has a code category labeled refused and assigned values of 9, 99, or 999. These values may also be used to represent missing responses. Missing responses may be due to non-interviews (a non-interview response results when an interview ends prior to this question and an interviewer codes the remaining responses as refused) and missing responses due to skip patterns in the questionnaire. This code, however, may capture some questions that were supposed to have answers, but for

some reason do not have them, and appeared as a blank or another symbol. Combining these types of responses into a single code requires vigilance on the part of data file users who wish to separate (1) results of respondents who did not receive a particular question and (2) results from respondents who, after receiving the question, gave an unclear answer or refused to answer it.

Weighting the Data

The BRFSS is designed to obtain sample information on the population of interest i.e., the adult US population residing in different states. Data weighting helps make sample data more representative of the population from which the data were collected. BRFSS data weights incorporate the design of BRFSS survey and characteristics of the population. BRFSS weighting methodology comprises 1) design factors or design weight, and 2) some form of demographic adjustment of the population—by iterative proportional fitting, or raking.

The design weight takes into account the probability of selection, adjusts for nonresponse bias and non-coverage errors. Design weights are calculated using the weight of each geographic stratum (**_STRWT**), the number of phones within a household (**NUMPHON2**), and the number of adults aged 18 years and older in the respondent's household (**NUMADULT**). For cellphone respondents, both **NUMPHON2** and **NUMADULT** are set to 1. The formula for the design weight is

$$\text{Design Weight} = \text{_STRWT} * (1/\text{NUMPHON2}) * \text{NUMADULT}$$

The stratum weight (**_STRWT**) accounts for differences in the probability of selection among strata (subsets of area code or prefix combinations). It is the inverse of the sampling fraction of each stratum. There is rarely a complete correspondence between strata (which are defined by subsets of area code or prefix combinations) and regions—which are defined by the boundaries of government entities.

BRFSS calculates the stratum weight (**_STRWT**) using the following items:

- **Number of available records (NRECSTR)** and the number of records users select (**NRECSEL**) within each geographic strata and density strata.
- **Geographic strata (GEOSTR)**, which may be the entire state or a geographic subset (e.g., counties, census tracts).
- **Density strata (_DENSTR)** indicating the density of the phone numbers for a given block of numbers as listed or not listed.

Within each **_GEOSTR*_DENSTR** combination, BRFSS calculates the stratum weight (**_STRWT**) from the average of the **NRECSTR** and the sum of all sample records used to produce the **NRECSEL**. The stratum weight is equal to **NRECSTR/NRECSEL**.

The complete overlapping sample frames required an adjustment to address the respondent's probability of selection in both the landline and cell phone sample frame. A compositing factor was calculated for dual users in landline and cell phone sample frames. The design weight is adjusted by the compositing factor for the records in the overlapping sample frames and later truncated based on quartiles. The adjusted and truncated design weight was used as the raking input weight.

BRFSS uses *iterative proportional fitting, or raking*, to adjust for demographic differences between those persons who are sampled and the population that they represent. After combining landline and cellular telephone data, BRFSS performs raking by adjusting one or a combination of demographic categories at a time in an iterative process until a convergence of a set value is reached. The BRFSS rakes the design weight to 8 margins (gender by age group, race or ethnicity, education, marital status, tenure, gender by race or ethnicity, age group by race or ethnicity, phone ownership). If the state had geographic regions, it includes 4 additional margins (region, region by age group, region by gender, region by race or ethnicity). If the state had at least 1 county with 500 or more respondents, the BRFSS includes 4 additional margins (county, county by age group, county by gender, county by race or ethnicity). BRFSS, therefore, uses the adjusted and truncated design weight design weight for raking and produces **_LLCPWT**—the final weight assigned to each respondent.

The population estimates obtained for building the target totals for raking are from similar sources used in previous years. Intercensal population estimates were purchased from The Nielsen Company, LLC at the county-level for age, race or ethnicity, and gender. These population estimates are used as the population totals for a state across all margins. The 5-year year American Community Survey PUMS data set (2011–2015) was used to obtain estimates for margins 3, 4, and 5 (education, marital status, tenure). The noninstitutionalized adults were weighted by the person-level weights to generate the population estimates. The percentages were then used in the raking margins. The telephone ownership estimates for margin 8 were taken from the state wireless estimate percentages produced by the National Center for Health Statistics (NCHS) and released in August, 2016.

Calculation of a Child Weight

The BRFSS calculates the design weight for child weighting from the stratum weight times the inverse of the number of telephones in the household and then multiplies by the number of children:

Child Design Weight = _STRWT * (1/NUMPHON2) * CHILDREN

CHILDWT = BRFSS rakes the child design weight to 5 margins including age by gender, race or ethnicity, gender by race or ethnicity, age by race or ethnicity, and phone ownership.

_CLLCPWT is the weight assigned for each child interview.

References

1. Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hogelin GC. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981-87. Public health reports (Washington, D.C.: 1974). 1988;103(4):366-375.
2. Federal Communications Commission USA. Universal Service Monitoring Report. 2016; http://transition.fcc.gov/Daily_Releases/Daily_Business/2017/db0113/DOC-343025A1.pdf
3. Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2016. National Center for Health Statistics. May 2017. Available from: <https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201705.pdf> Accessed May 2017.
4. MP B. Improving Standard Poststratification Techniques for Random-Digit-Dialing Telephone Surveys. Surv Res Methods. 2008;2(1):9.