

OVERVIEW: BRFSS 2011

1. BACKGROUND

The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project of the Centers for Disease Control and Prevention (CDC) and U.S. states and territories. The BRFSS, administered and supported by CDC's Behavioral Risk Factor Surveillance Branch, is an ongoing data collection program designed to measure behavioral risk factors for the adult population (18 years of age or older) living in households. 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, so that by 2001, 50 states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands were participating in the BRFSS. In this document, the term "state" is used to refer to all areas participating in the surveillance system, including the District of Columbia, Guam, the U.S. Virgin Islands, and the Commonwealth of Puerto Rico.

The BRFSS objective is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases that affect the adult population. Factors assessed by the BRFSS include tobacco use, health care coverage, HIV/AIDS knowledge and prevention, physical activity, and fruit-and-vegetable consumption. Data are collected from a random sample of adults (one per household) through a telephone survey.

BRFSS field operations are managed by state health departments that follow guidelines provided by the CDC. These health departments participate in developing the survey instrument and conduct the interviews either in-house or by using contractors. The data are transmitted to the CDC's Office of Surveillance, Epidemiology, and Laboratory Services, Behavioral Risk Factor Surveillance Branch for editing, processing, weighting, and analysis. An edited and weighted data file is provided to each participating health department for each year of data collection, and summary reports of state-specific data are prepared by CDC. Health departments use the data for a variety of purposes, including identifying demographic variations in health-related behaviors, targeting services, addressing emergent and critical health issues, proposing legislation for health initiatives, and measuring progress toward state and national health objectives.¹

The health characteristics estimated from the BRFSS pertain to the adult population, aged 18 years or older, who live in households. In 2011, additional questions were included as optional modules to provide a measure for several childhood indicators including asthma prevalence and influenza immunization for people aged 17 years or younger. As noted above, respondents are identified through telephone-based methods. Overall, an estimated 96.3% of U.S. households had telephone service in 2010.² The telephone coverage varies across states and subgroups. The increasing percentage of households that are abandoning their landline telephones for cell phones has significantly eroded the population coverage provided by landline-based surveys to pre-1970s levels. For the first half of 2011, the percentage of cell phone-only households was 31.6 percent.³ This is an increase of 1.9 percent over the preceding 6-month period. In households where both landline and wireless phone service is available, there is a trend toward increased use of wireless communication. In 2011, BRFSS respondents who received 100 percent of their calls on cell phones were eligible for participation in the cell phone survey. No direct method of compensating for non-telephone coverage is employed by the BRFSS; however, in 2011, a new weighting methodology, iterative proportional fitting (or "raking"), replaced the post stratification weighting method that had been used for the BRFSS for several years. Raking adjusts the data so that groups

which are underrepresented in the sample can be more accurately represented in the final dataset. Raking allows for the incorporation of cell phone survey data, permits the introduction of additional demographic characteristics and more accurately matches sample distributions to known demographic characteristics of populations. The use of raking reduces nonresponse bias and has been shown to reduce error within estimates. BRFSS raking 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/owner status, and age groups by race and ethnicity. In 2011, 50 states, the District of Columbia, Guam, and Puerto Rico collected samples of both landline and cell phone interviews while the Virgin Islands collected a sample of landline-only interviews.

2. DESIGN OF THE BRFSS

A. The BRFSS Questionnaire

The questionnaire has three parts: i. **Core component:** a standard set of questions asked by all states. It includes queries about current health-related perceptions, conditions, and behaviors (e.g., health status, health insurance, diabetes, tobacco use, disability, and HIV/AIDS risks), as well as demographic questions.

ii. **Optional CDC modules:** sets of questions on specific topics (e.g., cardiovascular disease, arthritis, women's health) that states elect to use on their questionnaires. In 2011, 34 optional modules were supported by CDC. The module questions are generally submitted by CDC programs and have been selected for inclusion in the editing and evaluation process by CDC. For more information, see <http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2011>.

iii. **State-added questions:** These are questions developed or acquired by participating states and added to their questionnaires. State-added questions are not edited or evaluated by CDC.

Each year, the states and CDC agree on the content of the core component and optional modules. 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 may have been tested and allows states to compare their data with those from other surveys. Any new questions proposed as additions to the BRFSS must go through cognitive testing and field testing prior to their inclusion on the survey. BRFSS guidelines 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 at the end of the questionnaire.

Although CDC supported 34 modules in 2011, it is not feasible for a state to use them all. States are selective about which modules and state-specific questions they add, to ensure the questionnaire is kept at a reasonable length--but there is wide variation across states in the total number of questions in a given year. New questionnaires are implemented in January and usually remain unchanged throughout the year. However, the flexibility of state-added questions does permit additions, changes, and deletions at any time during the year. The 2011 list of optional modules used on both the landline and cell phone surveys is available at <http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2011>.

B. Annual Questionnaire Development

The State BRFSS Coordinators Working Group meets three times a year with the Behavioral Risk Factor Surveillance Branch Management. One task of this group is to develop a 5-year, long-term plan for the BRFSS core instrument. The 2011 BRFSS questionnaire was the first year of a 5-year plan.

Before the beginning of the calendar year, CDC provides states with the text of the core component and the optional modules that will be supported for the coming year. States select their optional modules and choose any state-added questions. Each state then constructs its questionnaire. The order of the questioning is always the same: the core component is asked first, optional modules are asked next, and state-added questions last. This ordering ensures comparability across states and follows CDC guidelines. Generally, the only changes allowed are limited insertions of state-added questions on topics related to core questions. Such exceptions are to be agreed upon in consultation with CDC. However, despite this flexibility, not all states have adhered to the guidelines. Known deviations from the guidelines are noted in the Comparability of Data document.

Once the questionnaire content (core, modules, and state-added questions) is determined by a state, a hard-copy or electronic version of the instrument is constructed and sent to CDC. For states with Computer-Assisted Telephone Interview (CATI) systems, this document is used for CATI programming and general reference. The questionnaire is used without changes for one calendar year. The questionnaire is available at <http://www.cdc.gov/brfss/questionnaires/questionnaires.htm>. If a significant portion of the state population does not speak English, states have the option of translating the questionnaire into other languages. At the present time, CDC also provides a Spanish version of the core questionnaire and optional modules.

C. Sample Description

In a telephone survey, such as the BRFSS, a sample record is one telephone number in the list of all telephone numbers selected for dialing. To meet the BRFSS standard for the participating states' sample designs, sample records must be justifiable as a probability sample of all households with telephones in the state. All participating areas met this criterion in 2011. Fifty-one projects used a disproportionate stratified sample (DSS) design. Guam, Puerto Rico, and the U.S. Virgin Islands used a simple random sample design.

In the type of DSS design most commonly used in the BRFSS landline sampling, telephone numbers are divided 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 two digits of the suffix and all possible combinations of the last two digits. Numbers that come from hundred blocks with one or more listed household numbers ("1+ blocks," or "banks") are put in either the high-density stratum ("listed 1+ blocks") or medium-density stratum ("unlisted 1 + blocks"). The two strata are sampled to obtain a probability sample of all households with telephones.

Cell phone sampling frames are available from Marketing Systems Group (MSG)--Genesys Sampling Systems--and random samples of cell phone numbers can be called, provided that specific protocols are followed. Their sampling frame is based on the Telecordia database of telephone exchanges (e.g., 617-492-000 to 617-492-9999) and 1,000 banks (e.g., 617-492-0000 to 617-492-0999). MSG uses dedicated cellular 1,000 banks, sorted on the basis of area code and exchange within a state. An interval, K , is formed by dividing the population count of telephone numbers in the frame, N , by the desired sample size, n . The frame of telephone numbers is divided into n intervals of size K telephone numbers. From each interval, one 10-digit telephone number is drawn at random.

The target population for cell phone samples in 2011 consists of persons living in households who have a working cellular telephone, are aged 18 and older, and do not have a landline telephone.

In most cases, each state constitutes a single stratum. However, to provide adequate sample sizes for smaller geographically defined populations of interest, some states sample disproportionately from strata defined to correspond to substate regions. In 2011, the 48 states or territories with disproportionately sampled geographic strata were Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, the District of Columbia, 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, Virginia, the Virgin Islands, Washington, and Wisconsin.

Data for a state may be collected directly by the state health department or a contractor. In 2011, 11 state health departments collected their data in-house; 43 contracted data collection to university survey research centers or commercial firms. In 2011, the CDC's Behavioral Risk Factor Surveillance Branch provided samples purchased from MSG to all 54 states or territories.

3. DATA COLLECTION

Interviewing Procedures

In 2011, 54 states or territories used 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 CDC, 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 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 usually conducts the training using materials developed by CDC. These materials cover seven basic areas: overview of the BRFSS, role descriptions for staff involved in the interviewing process, the questionnaire, sampling, codes and dispositions (three-digit codes indicating the outcome of each call attempts), survey follow-up, and practice sessions. 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 BRFSS. Further specifics on interviewer training and procedures are available at <http://www.cdc.gov/brfss/training.htm>.

CDC expects interviewer performance to be monitored. In 2011, all BRFSS surveillance sites had the capability to monitor their interviewers. The system used for monitoring interviewers varied from listening to the interviewer only at an on-site location to listening to both the interviewer and respondent at a remote location. Verification call-backs were also used by some states in lieu of direct monitoring. Contractors typically conducted systematic monitoring by monitoring 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. All states were required to do verification callbacks for a sample of completed interviews as part of their quality-control practices.

Telephone interviewing was conducted during each calendar month, and calls were made seven days per week, during both daytime and evening hours. Standard procedures were followed for rotation of calls over days of the week and time of day. BRFSS procedural rules are described in the *BRFSS User's Guide* at <ftp://ftp.cdc.gov/pub/Data/Brfss/userguide.pdf>

Detailed information on interview response rates and item nonresponse rates are discussed at http://www.cdc.gov/brfss/technical_infodata/quality.htm.

4. DATA PROCESSING

A. Preparing for Data Collection and Data Processing

Data processing is an integral part of any survey. Because data are collected and sent to CDC during each month of the year, there are routine data processing tasks that need attention on an ongoing basis throughout the year. In addition, there are tasks that need to be conducted at different points in the annual BRFSS cycle. The preparation for the survey involves a number of steps that take place once the new questionnaire is finalized. This includes developing the edit specifications, programming portions of the Ci3 WinCATI software, programming the editing software, and producing telephone sample estimates for states that require them and ordering the sample from the contract vendor. A Ci3 WinCATI data entry module for each state that uses this software is produced. Skip patterns, together with some consistency edits, and response-code range checks are incorporated into the CATI system. These edits and skip patterns serve to reduce interviewer, data entry, and skip errors. Data conversion tables are then developed. These tables are used for reading the survey data from the entry module, calling information from the sample tracking module, and combining information into the final format specified for the data year. CDC also creates and distributes a Windows-based editing program that can perform data validations on properly formatted survey results files. This program is used to output lists of errors or warning conditions encountered 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, and continues processing data throughout the survey year. CDC receives and tracks monthly data submissions from the states. Once data are received from the state, editing programs and cumulative data quality checks are run against the data. Any problems in the file are noted, and a CDC programmer works with the state until the problems are resolved or agreement is reached that no resolution is possible. Response-rate data quality reports are produced and shared with the project officers and state coordinators, who review the reports and discuss any potential problems with the state. Once the entire year of data for a state has been received and validated, several year-end programs are run on the data. These programs perform some additional, limited data cleanup and fixes specific to the state and data year, and produce reports that identify potential analytic problems with the data set. Once these programs have been run, the data are ready for assigning weights and adding new variables.

Not all of the variables that appear on the public use data set are taken directly from the state files. CDC prepares a set of SAS programs that are used for end-of-year data processing. These programs prepare the data for analysis and add weighting and risk factor calculations as variables to the data file. The following variables are examples of results from this procedure and are created for the user's convenience: _RFSMOK3, _MRACE, _AGEG_, _TOTINDA. For more information, see the Calculated Variables and Risk Factors in Data Files document at http://www.cdc.gov/brfss/technical_infodata/surveydata/2011.htm. To create these variables, several variables from the data file are combined. The process of creating these variables varies in complexity; some are based only on combined codes, while others require sorting and combining of selected codes from multiple variables.

Almost every variable derived from the BRFSS interview has a code category labeled "refused" and generally given a value of "9," "99," or "999" value. Typically, the category consists of non interviews (a "non-interview" response results when an interview is terminated prior to this question and an interviewer

codes the remaining responses as “refused”) and persons for whom the question was not applicable because of a previous response or a personal characteristic (e.g., age). However, this code may capture some questions that were supposed to be answered, but for some reason were not, and appeared as a blank or other symbol. The combination of these types of responses into a single code requires vigilance on the part of data file users who wish to separate respondents who were skipped out of a question from those who were asked, but whose answer was unknown or who refused to answer a particular question.

B. Weighting the Data

When data are used without weights, each record counts the same as any other record. Implicit in such use are the assumptions that each record has an equal probability of being selected and that noncoverage and nonresponse are equal among all segments of the population. When deviations from these assumptions are large enough to affect the results obtained from a data set, then weighting each record appropriately can help to adjust for assumption violations. An additional, but conceptually unrelated, reason for weighting is to make the total number of cases equal to some desired number which, for state BRFSS data, is the number of people in the state who are aged 18 years and older. In the BRFSS, such raking serves as a blanket adjustment for noncoverage and nonresponse and forces the total number of cases to equal population estimates for each geographic region, which for the BRFSS is usually a state.

Following is a general description of the process that reflects factors taken into account in weighting the 2011 BRFSS data. Where a factor does not apply its value is set to one for calculation.

The Raking weighting methodology is comprised of two sections: Design weight and raking.

Design Weight = STRWT * (1/NUMPHON2) * NUMADULT

The 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 is rarely 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.

- The stratum weight (**STRWT**) is calculated using:
 - Number of available records (**NRECSTR**) and the number of records selected (**NRECSEL**) within each geographic strata and density strata.
 - Geographic strata (**GEOSTR**) which may be the entire state or a geographic subset such as counties, census tracts, etc.
 - 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 the stratum weight (**_STRWT**) is calculated from the average of the **NRECSTR** and the sum of all sample records used to produce the **NRECSEL**. The stratum weight is equal to **NRECSEL / NRECSTR**.

- **1/ NUMPHON2** is the inverse of the number of residential telephone numbers in the respondent’s household.
- **NUMADULT** is the number of adults 18 years and older in the respondent’s household.

FINAL WEIGHT = The design weight is raked to 8 margins (age group by gender, race/ethnicity, education, marital status, tenure, gender by race/ethnicity, age group by race/ethnicity, phone ownership). If geographic regions are included, four additional margins (region, region by age group, region by gender, region by race/ethnicity) are included.

_LLCPWT is the final weight assigned to each respondent.

Weight trimming is used to increase the value of extremely low weights and decrease the value of extremely high weights. The objective of weight trimming is to reduce errors in the outcome estimates caused by unusually high or low weights in some categories.

Calculation of a Child Weight

The design weight for the child weighting is calculated from the stratum weight times the inverse of the number of telephones in the household and then multiplied by the number of children

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

CHIILDWT = The child design weight is raked to 5 margins including age by gender, race/ethnicity, gender by race/ethnicity, age by race/ethnicity, and phone ownership.

_CLLCPWT is the weight assigned for each child interview.

REFERENCES

1. Remington PL, Smith MY, Williamson DF, Anda RF, et al. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–1987. *Public Health Reports* 1988;103(4):366–375.
2. Bureau of the Census, U.S. Department of Commerce. 2006-2010 American Community Survey 5-Year Estimates, Table B25043 – Tenure by Telephone Service Available by Age of Householder. (Universe: Occupied housing units, American FactFinder Database). Washington, DC: Bureau of the Census; 2010. (<http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>). (Accessed June 5, 2012)
3. Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2011. National Center for Health Statistics. December 2011. Available from <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201112.htm>