

OVERVIEW: BRFSS 2007

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 Surveillance Branch, is an ongoing data collection program designed to measure behavioral risk factors in 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 objective of the BRFSS 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 in 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, who follow guidelines provided by the CDC. These health departments participate in developing the survey instrument and conduct the interviews either in-house or through use of contractors. The data are transmitted to the CDC's National Center for Chronic Disease Prevention and Health Promotion's Behavioral 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 and older, who live in households. In 2007 additional questions were included as optional modules to provide a measure for one health characteristic of the non-adult population aged 17 years and less. (These were for childhood asthma prevalence.) As noted above, respondents are identified through telephone-based methods. Although overall, approximately 95 percent of U.S. households have telephones, coverage ranges from 87 to 98 percent across states and varies for subgroups as well. For example, people living in the South, minorities, and those in lower socioeconomic groups typically have lower telephone coverage. No direct method of compensating for non-telephone coverage is employed by the BRFSS; however, post-stratification weights are used, which may partially correct for any bias caused by non-telephone coverage. These weights adjust for differences in probability of selection and nonresponse, as well as noncoverage, and must be used for deriving representative population-based estimates of risk behavior prevalence.

2. DESIGN OF THE BRFSS

A. The BRFSS Questionnaire

The questionnaire has three parts: 1) the core component; 2) optional modules; and 3) state-added questions.

Core component. The *core* is 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.

Optional CDC modules. These are sets of questions on specific topics (e.g., cardiovascular disease, arthritis, women's health) that states elect to use on their questionnaires. In 2007, 19 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 *2007 BRFSS Modules Used By States*, <http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2007>

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 26 modules in 2007, 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 (though there is wide variation across states in the total number of questions for a given year, ranging from 0 to approximately 200, in Massachusetts). 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 2007 core and module questionnaire is available at <http://apps.nccd.cdc.gov/BRFSSModules/ModByState.asp?Yr=2007>

Annual Questionnaire Development

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 section of this document.

Once the content (core, modules, and state-added questions) of the questionnaire 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 provides only a Spanish version of the core questionnaire and optional modules.

B. 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. In order 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 2007. 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, 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. A hundred block is a set of one hundred 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 the 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.

In most cases, each state constitutes a single stratum. However, in order to provide adequate sample sizes for smaller geographically defined populations of interest, some states sample disproportionately from strata defined to correspond to sub-state regions. In 2007, the 42 states with disproportionately sampled geographic strata are Alabama, Alaska, Arizona, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Puerto Rico, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Virgin Islands, Washington, and Wisconsin.

Data for a state may be collected directly by the state health department or through a contractor. In 2007, 12 state health departments collected their data in-house; 42 contracted data collection to university survey research centers or commercial firms.

In 2007, the Behavioral Surveillance Branch provided samples purchased from Genesys (Marketing Group Systems) to all 54 states or territories.

3. DATA COLLECTION

Interviewing Procedures

In 2007, 53 states used computer-assisted telephone interviewing (CATI). CDC supports CATI programming using the Ci3 CATI 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 who is available 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 10 minutes. Interview time for modules and state-added questions is dependent upon the number of questions used, but generally extend the interview period by an additional 5 to 10 minutes.

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, <http://www.cdc.gov/brfss/training.htm>.

CDC expects interviewer performance to be monitored. In 2007, 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*, <http://www.cdc.gov/brfss/pubrfdat.htm#users>.

Detailed information on interview response rates and item nonresponse rates are discussed in the *2007 Summary Data Quality Report*.

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 CATI software, programming the PC-EDITS software, and producing telephone sample estimates for states that require them and ordering the sample from the contract vendor. A Ci3-CATI 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/2007.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 age 18 and older. In the BRFSS, such poststratification 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 formula that reflects all the factors taken into account in weighting the 2007 BRFSS data. Where a factor does not apply its value is set to one for calculation.

$$\mathbf{FINALWT = STRWT \times 1 \text{ OVER } NPH \times NAD \times POSTSTRAT}$$

FINALWT is the final weight assigned to each respondent.

STRWT 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 seldom 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.

1/NPH is the inverse of the number of residential telephone numbers in the respondent's household.

NAD is the number of adults in the respondent's household.

POSTSTRAT is the number of people in an age-by-sex or age-by-race/ethnicity-by-sex category in the population of a region or a state divided by the sum of the preceding weights for the respondents in the same age-by-sex or age-by-race/ethnicity-by-sex category. It adjusts for noncoverage and nonresponse and forces the sum of the weighted frequencies to equal population estimates for the region or state.

$$\text{CHILDWT} = \text{STRWT} \times 1 \text{ OVER NPH} \times \text{CHILDREN} \times \text{POSTCH}$$

CHILDWT is the final weight assigned to each child.

STRWT 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 almost never 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.

1/NPH is the inverse of the number of residential telephone numbers in the respondent's household.

CHILDREN is the number of children (less than 18 years of age) in the respondent's household.

POSTCH is the number of children in an age-by-gender or age-by-race-by-gender category in the population of a region or a state divided by the sum of the preceding weights for the children in that same age-by-gender or age-by-race-by-gender category. It adjusts for non-coverage and non-response.

$$\text{HOUSEWT} = \text{STRWT} \times 1 \text{ OVER NPH} \times \text{POSTHH}$$

HOUSEWT is the weight assigned to each household.

STRWT 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 almost never 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.

1/NPH is the inverse of the number of residential telephone numbers in the household.

POSTHH is the number of households in the population of a region or a state divided by the sum of the products of the preceding weights for the households in that same category. It adjusts for non-coverage and non-response.

$$\text{ITSADJWT} = \text{ITSCF2} \times \text{ITSPOST}$$

ITSADJWT is the weight adjustment for non-telephone households or interruption in telephone service for 1 week or more during the past 12 months.

An estimate of the total number of adults in households with telephones and without telephones is obtained from the latest CPS data. The estimate of the total number of adults in households with telephone service interruptions is combined with the estimate of adults in non-telephone households. This total is used to determine a ratio to adjust the base weights of respondents with

telephone service interruptions. The estimate of the total number of adults in households with no telephone service interruptions is used to determine a ratio to adjust the base weights of respondents with no telephone service interruptions. The data is then post-stratified to the BRFSS population totals.

ITSCF2 is the interruption in telephone service correction factor (adjusted _WT2) using the variable TELSERV (indication of interruption in telephone service) response and the number of adults in households without telephones.

ITSPOST is the number of households in the population of a region or a state divided by the sum of the products of the preceding weights for the households in that same category. It adjusts for non-coverage and non-response.

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*. 1988; 103(4): 366–375.
2. Bureau of the Census. *Phoneless in America*. Statistical Brief 94–16. U. S. Department of Commerce, Bureau of the Census. July 1994.