

## 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 the Behavioral Surveillance Branch (BSB) of the CDC, is an on-going data collection program designed to measure behavioral risk factors in the adult population 18 years of age or over living in households. The BRFSS was initiated in 1984, with 15 states collecting surveillance data on risk behaviors through monthly telephone interviews. The number of states participating in the survey increased, so that by 1998, 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 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, physical activity, dietary practices, safety-belt use, and use of cancer screening services, among others. Data are collected from a random sample of adults (one per household) through a telephone survey.

Field operations for the BRFSS are managed by the health departments under guidelines provided by the BSB. These health departments participate in the development of the survey instrument and conduct the interviews either in-house or through use of contractors. The data are transmitted to the National Center for Chronic Disease Prevention and Health Promotion's Behavioral Surveillance Branch at CDC 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 the staff of the BSB. Health departments use the data for a variety of purposes, including identification of 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.

The health characteristics estimated from the BRFSS pertain only to the adult population age 18 years and older living in households. As noted above, respondents are identified through telephone-based methods. Although 95 percent of U.S. households have telephones, coverage ranges from 87-98 percent across states and varies for subgroups as well. For example, persons living in the South, minorities, and those in lower socioeconomic groups typically have lower telephone coverage ( see *Table 3\_98.* ). No direct method of compensating for non-telephone coverage is employed by the BRFSS; however, post-stratification weights are used, and 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 prevalences.

## 2. DESIGN OF THE BRFSS

### A. The BRFSS Questionnaire

The questionnaire has three parts: 1) the core component, consisting of the fixed, rotating, and emerging core; 2) optional modules; and 3) state-added questions.

**Core component.** The *fixed 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, selected cancer screening procedures, and HIV/AIDS risks) and questions on demographic characteristics. The *rotating core* consists of two distinct sets of questions, each asked in alternating years by all states, addressing different topics. In 1998, the rotating core items covered exercise, fruits and vegetables, and weight control. The *emerging core* is a set of up to five questions that are added to the fixed and rotating cores. Emerging core questions typically focus on issues of a “late breaking” nature and do not necessarily receive the same scrutiny that other questions receive prior to being added to the instrument. These questions are part of the core for one year and are evaluated during or soon after the year concludes to determine their potential value in future surveys. Emerging questions for 1998 focused on cigar smoking and use of weight loss pills.

**Optional CDC modules.** These are sets of questions on specific topics (e.g., smokeless tobacco, firearms) that states elect to use on their questionnaires. In 1998, 21 modules were supported by CDC, including 6 that were rotating core topics not used in the 1998 core component. (see *Table2\_98*. for a list of modules used by reporting areas.)

**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. For comparability, 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 prior to their inclusion on the survey. BRFSS protocol specifies that all states ask the core component questions without modification; they may choose to add any, all, or none of the optional modules; and states may add question(s) of their choosing at the end of the questionnaire.

Although CDC supported 21 modules in 1998, it is not feasible for a state to use them all. States are selective with their choices of modules and state-specific questions to keep the questionnaire at a reasonable length (though there is wide variation across states in the total number of questions for a given year, ranging from a low of about 90 to 150 or more). 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 1998 core and module questionnaire is in the file *QcoreMod\_98.rtf*.

## 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 question(s). Each state then constructs its questionnaire. 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 protocol. Generally, the only changes allowed are the limited insertion of state-added questions on topics related to core questions. Such exceptions are to be agreed upon in consultation with CDC. However, even with these exceptions, the policy has not been followed in every instance. Deviations from policy are noted in the *comparability of data* section of this document.

Once the content (core, modules, and state-added) of the questionnaire is determined by a state, a paper version of the instrument is constructed and a copy sent to CDC. For states with Computer Assisted Telephone Interview (CATI) systems, this copy is used for CATI programming and general reference. For states with manual systems, the paper copy is used as a prototype for a camera-ready master. The questionnaire is used without changes for one calendar year. The topics included on the 1998 questionnaire are shown in *Table1\_98*. 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 only provides a Spanish version of the core questionnaire and optional modules.

Each year, CDC provides Ci3 CATI programming, PC-EDITS programming, and state-specific annual data tables for questions on the core and selected modules.

### B. Sample description

The BRFSS standard for participating area sample designs is that sample records must be justifiable as a probability sample of all households with telephones in the state. All participating areas except Alaska, California, and Hawaii met this criterion in 1998.

The sample designs used by states in the 1998 BRFSS survey are shown in *Table4\_98*. Eleven states used the Mitofsky-Waksberg design. Thirty-seven states used a disproportionate stratified sample (DSS) design. Puerto Rico used a simple random sample design. The remaining three states--Alaska, California, and Hawaii--used a variety of other designs.

The Mitofsky-Waksberg sample design is a two-stage cluster design. In the first stage, telephone numbers are grouped into sets of 100, called primary sampling units (PSU's) or clusters. These sets consist of telephone numbers with identical area codes, prefixes, first two digits of the suffixes, and all possible combinations of the last two digits of the suffixes. Clusters are sampled randomly and, within each selected cluster, a single telephone number is randomly selected to be dialed. If this telephone number is a household, the entire PSU is selected for further sampling; if the telephone number is not a household the entire PSU is rejected. Telephone numbers from accepted PSU's are dialed until a target number (for the

BRFSS, the target is three) of completed interviews is obtained. The BRFSS implementation includes a third sampling stage, wherein an adult eighteen years of age or older is randomly selected from eligible households.

In a DSS design as most commonly practiced in the BRFSS, telephone numbers are divided into two groups, or strata, which are sampled separately. One group, the high-density stratum, contains telephone numbers which are expected to contain a large proportion of households. The other group, the low-density stratum, contains telephone numbers which are expected to contain a small proportion of households. Whether a telephone number goes into the high-density or low-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 high density stratum. Numbers that come from hundred blocks with no listed household numbers (0 blocks, or banks) are put in the low density stratum. Both strata are sampled to obtain a probability sample of all households with telephones. The high density stratum is sampled at a higher rate than the low density stratum (that is, disproportionately) to obtain a sample that contains a larger proportion of household numbers than would be the case if all numbers were sampled at the same rate.

Wisconsin and Michigan used variations of the DSS design described above. Wisconsin assigned telephone numbers to high and low density strata based on previous experience with their prefixes. Michigan defined its high density stratum as 2+ blocks and its low density stratum as 1- blocks.

The sample designs for Alaska, California, and Hawaii which, as noted above, do not conform to the BRFSS standard for state sample designs, are described in *Compare\_98.rtf*.

In most cases, each state constitutes a single stratum. In order to provide adequate sample sizes for smaller geographically defined populations of interest, however, sixteen states sampled disproportionately from strata defined to correspond to sub-state regions. The states with disproportionately sampled geographic strata are Alaska, Arizona, Connecticut, Delaware, Hawaii, Idaho, Iowa, Maryland, Massachusetts, Missouri, Nevada, North Carolina, Ohio, Utah, Virginia, and Wisconsin.

Data for a state may be collected directly by the state health department or through a contractor. Twenty-one state health departments collected their data in-house; thirty-one contracted out the data collection to a variety of governmental agencies, university survey research centers, and commercial firms in 1998.

In 1998, the Behavioral Surveillance Branch provided sample for twenty-nine states. Twenty-three states purchased their sample from a commercial sample provider: fourteen from Genesys Sampling Systems, two from Survey Sampling, Inc., two from the Wisconsin Survey Research Laboratory and five from a variety of other sources.

### **3. DATA COLLECTION**

## A. Interviewing Procedures

Interviews for 1998 were conducted through computer-assisted telephone interviewing (CATI) by 50 areas; paper questionnaires were used in the other two surveillance areas. CDC supports CATI programming using the Ci3 CATI software package. This support includes programming Ci3 for states and supporting a Ci3 consultant who is available to assist states. Following specifications provided by CDC, state health personnel or contractors conducted 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 covering seven basic areas: overview of the BRFSS, role descriptions for staff involved in the interviewing process, the questionnaire, sampling, codes and dispositions, 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 “certified” to work on BRFSS. Further specifics on interviewer training and procedures can be found in the *BRFSS Operations Manual, 1998* available at the internet web site [www.cdc.gov/nccdphp/brfss](http://www.cdc.gov/nccdphp/brfss).

Monitoring of interviewers is expected. In 1998, thirty of the 52 surveillance projects monitored interviews by “eavesdropping” on the interviewer without being able to hear the respondent. Thirty-seven states had systems connected to the telephones that enable monitoring of the respondent and interviewer. 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 a two-week period each month, and calls were made 7 days per week, during both day and evening hours. Standard procedures in the interviewing were followed for rotation of calls over days of the week and times of the day. BRFSS procedural rules are contained in the *BRFSS Operations Manual, 1998*.

Detailed information on interview response rates and item non-response rates are discussed in the *Summary Quality Control Report* available on the internet web site [www.cdc.gov/nccdphp/brfss](http://www.cdc.gov/nccdphp/brfss). The median response rate (Upper Bound) for 1998 was 73.4 %, but ranged from 45.4% to 95.4%.

## **4. DATA PROCESSING**

### **A. Preparing for collection and processing data collected**

Data processing is an integral part of any survey. Because data are collected and sent to BSB during each month of the year, there are routine data processing tasks that need attention at all times during 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 complete edit specifications, programming the Ci3 CATI software, programming the PC-EDITS software, and providing sample phone numbers for states that require them. 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. Incorporation of edits and skip patterns into the CATI instrument reduces interviewer errors, data entry errors, and skip errors. A program designed to read the data file from the survey data entry module and call information from the sample tracking module in Ci3 is written to combine information into the final format specified for the data year. For those state that do not use CATI software to perform the BRFSS interviews, CDC prepares and provides EPI-Info data entry formats. CDC also creates and distributes a DOS 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. A program which produces tables of response rates and other interviewing statistics is included with the edits program sent to the states.

CDC begins to process data for the data year beginning in February of that year, and continues processing data through the receipt of the final data for that year. CDC receives and tracks monthly data submissions from the states. Once data are received from the state, editing programs are run against the data. Any problems in the file are noted, and a BSB programmer works with the state until the problems are resolved, or agreement is reached that no resolution is possible. Response rate quality control reports are produced and shared with the Project Officers, who review the reports and discuss any potential problems with the state. After all of the data for a state are received and validated for the data year, three year-end programs are run on the data to consolidate the monthly transmissions into a single data file for the year, 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 are complete, the data are ready for assigning weights and adding new variables.

Not all of the variables that appear on the public use data tape are taken directly from the state files. CDC prepares a set of SAS programs that implement the end of year data processing. These programs not only prepare the data for analysis, but perform weighting and risk factor calculations which are added as variables to the data file. The following variables are examples of ones that result from this procedure, and are created for the user's convenience: \_RFSMOK2, RACE, \_AGEG, \_BMI, \_TOTINDX. To create these variables, several variables from the data file are combined. Creation of the variables vary in complexity; some only combine codes, while others require sorting and combining selected codes from multiple

variables.

Almost every variable derived from the BRFSS interview has a code category labeled “refused” and generally given a “9”, “99” or “999” value. Typically, the category consists of non-interviews and persons for whom the question was not applicable because of a previous response or a personal characteristics (i.e., age). However, this code may capture some responses 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 require 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 selection and that non-coverage and non-response 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 violations of the assumptions. An additional, but conceptually unrelated, reason for weighting is to make the total number of cases equal to some desired number. In the BRFSS, post-stratification serves as a blanket adjustment for non-coverage and non-response and forces the total number of cases to equal population estimates for each geographic stratum, which for the BRFSS is usually a state.

Following is a general formula that reflects all the factors taken into account in weighting the 1998 BRFSS data. Where a factor does not apply its value is set to one for calculation.

$$\text{FINALWT} = \text{GEOWT} * \text{DENWT} * 1 \text{ OVER NPH} * \text{NAD} * \text{CSA} * \text{POSTSTRAT}$$

**FINALWT** is the final weight assigned to each record.

**GEOWT** is the inverse of the ratio of the estimated sampling fraction of each area code/prefix combination subset to the area code/prefix combination subset with the largest estimated sampling fraction. It weights for the unequal probability of selection by area code/prefix combinations intended to cover specified geographic regions. Almost always, the regions are discrete subsets of counties and the boundaries of the area code/prefix combinations do not correspond exactly to the boundaries of the specified geographic regions.

**DENWT** is the inverse of the ratio of the sampling fraction of each subset of hundred blocks (sets of telephone numbers with identical first eight digits and all possible final two digits) sampled at a given rate to the hundred-block subset with the largest sampling fraction. It weights for the unequal probability of selection by presumed household density of hundred block. This is generally used in a design in which telephone numbers from hundred blocks with one or more listed residential numbers (one-plus blocks) are sampled at a higher rate than

telephone numbers from hundred blocks with no listed residential numbers (zero blocks).

**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.

**CSA** is the ratio of the expected cluster size to the actual cluster size.

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