

# New York Behavioral Risk Factor Surveillance Survey

## Overview

Center for Community Health  
Division of Chronic Disease and Injury Prevention  
Bureau of Chronic Disease, Epidemiology and Surveillance

**OPEN Data NY**

## Behavioral Risk Factor Surveillance Survey

### General description

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual statewide telephone surveillance system designed by the Centers for Disease Control and Prevention, and conducted by the NYSDOH, Division of Chronic Disease and Injury Prevention, Bureau of Chronic Disease Epidemiology and Surveillance through a cooperative agreement. Interviews are collected on an ongoing basis in both English and Spanish. 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. New York State's BRFSS sample represents the non-institutionalized adult household population, aged 18 years and older. The BRFSS is designed to provide estimates at the state level. All data collected are self-report. Survey results are analyzed and disseminated as reports to advance disease prevention and control in community and statewide settings.

The questions on the BRFSS are not the same every year, although there is a set of core required questions asked in all states every year, or on a regular rotating basis, such as every other year. CDC also offers states the option of including questions from a list of optional CDC questions. In addition, states can add additional questions to serve their own specific state needs.

### 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. The 2009 NYS BRFSS questionnaire included the following optional modules: pre-diabetes, diabetes, inadequate sleep, actions to control high blood pressure, heart attack and stroke, random child selection, and childhood asthma prevalence.

**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.

The 2009 NYS BRFSS questionnaire included the following state-added modules: adult TV viewing, childhood TV environment, food security/social context, access to fruits and vegetables, sweetened beverage consumption, disability limitations, caregiver and sexual behavior STD/female condom.

Interviews were conducted monthly, from January through December of 2009. Calls were made seven days a week, during daytime and evening hours. Standard calling procedures were followed for rotation of calls during days of the week and time of day throughout the month. Detailed information on interview response rate and item non-response rates are discussed in the quality control reports produced by the CDC, available at [http://www.cdc.gov/brfss/technical\\_infodata/quality.htm](http://www.cdc.gov/brfss/technical_infodata/quality.htm).

### **Data collection and processing methodology**

A BRFSS 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 represent a probability sample of all households with telephones in the state. In the type of DSS (Disproportionate Stratified Sampling) design most commonly used in the BRFSS (including NYS 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. In most cases, each state constitutes a single stratum. 2009 data for NYS was collected as a single stratum.

CDC supports computer-assisted telephone interviewing (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 to assist states. Following guidelines provided by CDC, New York State health personnel contract with Clearwater Research Inc. to conduct interviews. The core portion of the questionnaire lasts an average of 15 minutes. Interview time for modules and state-added questions is dependent upon the number of questions used; additional questions extend the interview period by an additional 5 to 10 minutes.

The NYS BRFSS data set is a combined file using information from CDC and the state for landline households. The CDC end-of-year data file sent to the state is processed to include state-added questions. The data set contains the CDC risk factor calculations as variables to the data files. More information about the calculated variables and risk factors in BRFSS data files is available at [http://www.cdc.gov/brfss/technical\\_infodata/surveydata.htm](http://www.cdc.gov/brfss/technical_infodata/surveydata.htm).

### **Weighting the data**

The end of year data file supplied to the states by CDC contains weights to be used for analyses. 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 violations in these assumptions. 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 post-stratification serves as a blanket

adjustment for noncoverage and nonresponse and forces the total number of cases to equal population estimates for the state.

Following is a general formula that reflects all the factors taken into account in weighting the 2009 BRFSS data. Where a factor does not apply its value is set to one for calculation.

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

**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.

$$\mathbf{CHILDWT = STRWT \times 1/NPH \times CHILDREN \times POSTCH}$$

**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 noncoverage and nonresponse.

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

**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 noncoverage and nonresponse.

### **Statistical and analytic issues**

Unweighted data on the BRFSS represent the actual responses of each respondent, before any adjustment is made for variation in respondents' probability of selection, disproportionate selection of population subgroups relative to the state's population distribution, or nonresponse. Weighted BRFSS data represent results that have been adjusted to compensate for these issues. Irrespective of state sample design, use of the final weight in analysis is necessary if generalizations are to be made from the sample to the population.

The procedures for estimating variances described in most statistical texts and used in most statistical software packages are based on the assumption of simple random sampling (SRS). However, the data collected in the BRFSS are obtained through a complex sample design; therefore, the direct application of standard statistical analysis methods for variance estimation and hypothesis testing may yield misleading results. There are computer programs available that take such complex sample designs into account. SAS Version 8 SURVEYMEANS and SURVEYREG procedures, SUDAAN, and Epi Info's C-Sample are among those suitable for analyzing BRFSS data. SAS and SUDAAN can be used for tabular and regression analyses. SUDAAN also has these and additional options. Epi Info's C-sample can be used to calculate simple frequencies and two-way cross-tabulations. When using these software products, users must know the stratum, the primary sampling units, and the record weight—all of which are on the data file. For more information on calculating variance estimations using SAS, see the *SAS/STAT Users Guide, Version 8*. For information about SUDAAN, see the SUDAAN Users Manual, Release 7.5. For information about Epi Info, see *Epi Info, Version 6.0*

### **Limitations of data use**

Although the overall number of respondents in the BRFSS is more than sufficiently large for statistical inference purposes, subgroup analyses can lead to estimates that are unreliable. Consequently, users need to pay particular attention to the subgroup sample when analyzing subgroup data, especially within a single data year or geographic area. Small sample sizes may produce unstable estimates. Reliability of an estimate depends on the actual unweighted number of respondents in a category, not on the weighted number. Interpreting and reporting weighted numbers that are based on a small, unweighted number of respondents can mislead the reader into believing that a given finding is much more precise than it actually is. The BRFSS follows a rule of not reporting or interpreting percentages based upon a denominator of fewer than 50 respondents (unweighted sample).

Telephone coverage by landline varies by state and also by subpopulation. In 2008 telephone coverage averaged 94.1% for the United States as a whole. Data on telephone coverage in U.S. households are available at <http://www.fcc.gov/wcb/iatd/trends.html>. The report shows telephone coverage by state as a combination of landline and cellular telephone. The percentage of cellular phone service only households increased to 22.7% in 2009.<sup>1</sup> Cellular telephones were not included as part of the regular BRFSS sample in 2009. However the BRFSS is making adjustments to include this segment of the population.

**Other issues**

*Analyzing Subgroups*

Before conducting subgroup analyses, it is necessary to determine whether the total number of respondents will yield the precision needed, which depends upon the intended use of the estimate. For example, greater precision would be required to justify implementing expensive programs than that needed for general information only. The table below presents sample size requirements for each of several levels of precision. Precision is indicated by the width of the 95% confidence interval around the prevalence estimate.

<u>Precision Desired</u>	<u>Sample Size Needed</u>
2%	3600
4%	900
6%	400
8%	225
10%	144
15%	64
20%	36

*Multiple Questionnaires*

The New York State BRFSS included multiple questionnaires (Survey 1 and Survey 2). The analysis of the multiple questionnaire data requires some careful consideration of which records to use with the appropriate weight. For the core questions, optional modules and state-added questions asked on the “Common” questionnaire the `_FINALWT` variable should be used to produce estimates. For optional modules and state added questions used only on “Survey 1” the `_FINALQ1` variable should be used to produce estimates for records with variable `QSTVER = 1`. For optional and state added modules used only on “Survey 2” the `_FINALQ2` variable should be used to produce estimates for records with variable `QSTVER = 2`. Below are the common and survey specific optional modules and state-added questions.

**Common:** Random Child Selection, Childhood Asthma Prevalence

**Survey 1:** Heart Attack and Stroke, Actions to Control High Blood Pressure, Inadequate Sleep, Sexual Behavior related to STD (sexually transmitted disease) and Female Condom use/Awareness

**Survey 2:** Diabetes, Pre-Diabetes, Adult and Child TV Viewing, Food Security/Social Context, Access to Fruits and Vegetables, Sweetened Beverage Consumption, Disability Limitations, and Caregiver.

References:

<sup>1</sup> Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2009. National Center for Health Statistics. December 2009.