

---

# Behavioral Risk Factor Surveillance System

Comparability of Data BRFSS 2012

July 15, 2013



## **Introduction**

The Behavioral Risk Factor Surveillance System (BRFSS) is an ongoing, state-based, random-digit-dialed telephone survey of non-institutionalized adults aged ≥18 years residing in the United States.<sup>1, 2</sup> Detailed descriptions of BRFSS questionnaires, data, and reports are available at [www.cdc.gov/brfss](http://www.cdc.gov/brfss). In 2012, all 50 states, the District of Columbia, the Commonwealth of Puerto Rico, and Guam participated in both landline telephone and cellular telephone interviews.

The BRFSS survey uses disproportionate stratified sample (DSS) design for landline telephone samples and random sample for cellular telephone survey. BRFSS data collection, structure, and weighting methodology changed in 2011 to allow data to be collected by cellular telephones, in addition to landline telephones. Iterative proportional fitting (IPF), also known as raking, was used for weighting 2012 BRFSS data. Due to sample design and the multiple reporting areas, BRFSS data showed some variation between states and territories. Sections A and B below identify important similarities and variations for the 2012 data year from previous years.

### **A. 2012 Data Anomalies and Deviations from Sampling Frame**

In all the states/territories, a portion of sample records intended for use during one month took more than one month to complete. In several instances, states used their monthly sample during a period of several months. This deviation will disproportionately affect analyses based on monthly (rather than annual), data. Michigan and California continued to receive their sample quarterly rather than monthly. Utah received cellular telephone sample quarterly rather than monthly.

Several states conducted fewer than 12 monthly data collections during the year. For landline telephone data collection, Georgia and North Carolina did not collect a sample in December and Utah did not collect a sample in January. For cellular telephone data collection, Puerto Rico did not collect a sample in January. Hawaii did not collect a sample in December. Wisconsin did not collect a sample in January or December. Rhode Island did not collect a sample in September, October, November, or December. Florida did not collect a sample in January through September. Thirty one (31) states were unable to close out the December sample in 2012, and data collection continued into early 2013.

### **B. Protocol Changes from 2011 Data**

#### **1. Addition of Cellular Telephone Data**

Telephone coverage varies by state and also by subpopulation. In 2011, the percentage of occupied housing units with telephone service averaged 97.4% for the United States, and non-coverage ranged from 1.5% in Connecticut to 5.1% in New Mexico. It is estimated that 6.2% of households in Puerto Rico were without

telephone service. Data on telephone coverage in U.S. households are available at <http://transition.fcc.gov/wcb/iatd/recent.html>. The percentage of cellular telephone-only households increased to 35.8% in 2012. Additionally, one of every six American homes received all or most telephone calls on a cellular telephone while a landline telephone was also available.<sup>3</sup> The increased use of cellular telephones required the BRFSS to begin to include the population of cellular telephone users in 2011. Now BRFSS is better able to reach the target population that was previously inaccessible, including those who had a cellular telephone but not a landline telephone. In 2012, BRFSS cellular telephone respondents who received 90 percent or more of their calls on cellular telephones were eligible for participation in the cellular telephone survey.

## 2. Weighting Methodologies

In 2011, a different weighting methodology, iterative proportional fitting (IPF), also known as raking, was applied to the BRFSS and was used to improve the extent to which a BRFSS sample can better reflect the socio-demographics at state level. (Refer to the [methodologic changes article](#) for more information). Raking adjusts the estimates within each state using the margins (raking control variables), including gender by age group, detailed race and ethnicity, education, marital status, tenure (rent or own), gender by race or ethnicity, age group by race or ethnicity, telephone sample source, region, region by age group, region by gender, and region by race or ethnicity. The raking method applies a proportional adjustment to the weights of the cases that belong to the same category of the margin. Raking can adjust a set of data to control totals on three or more variables. The iteration (up to 100 times) continues until a convergence of below 0.025 is achieved.

The new weighting methodology fits the dual frame survey design: landline telephone only, cellular telephone only, and combined landline telephone and cellular telephone. At the last step of the raking process, weight trimming takes place, which is used to increase the value of extremely low weights and decrease the value of extremely high weights, in order to reduce errors in the outcome estimates. Weight trimming has two alternative methods, IGCV (Individual and Global Cap Value) and MCV (Margin Cap Value). The raking method was used to weight the 2012 BRFSS data.

In 2012, the inclusion of cellular telephone respondents who received between 90 and 99 percent of their telephone calls on their cellular telephone required an adjustment to the design weights to account for the overlapping sample frames. A compositing factor was calculated from each of the two samples for “dual users”-- individuals who had both cellular telephones and landline phones but mostly used cellular telephones. The design weight was multiplied by the compositing factor to generate a composite weight, which is used as the raking input weight.

### 3. Other Issues

The data from an optional module was included if the module was given to all eligible respondents within a state for the entire data collection year. A state may have indicated the use of an optional module in 2012, but if the module was not given to the expected population, the data were moved into the state-added questions section. Several states collected data with optional modules by landline telephone and cellular telephone surveys in 2012. The data for the optional modules collected as common modules to both the landline telephone survey and cellular telephone survey are included in the available public release data file.

Several states used optional modules as part of their landline telephone survey only. In order to make these data available to users, a separate data set has been produced with a raked weight for use with the landline survey data. The Web page with the landline data set and additional documentation is available from the Data Files portion of the 2012 BRFSS Survey Data and Documentation.

The Division of Behavioral Surveillance has also provided limited support for the survey data collection of multiple (up to three in 2012) questionnaire versions. A state may ask a subset of its survey sample a different set of questions following the core, as long as the survey meets the minimum effective sample size (2,500 participants) for a given questionnaire version. The 2012 core instrument must be asked without any changes in any of the versions of the questionnaire. The optional modules can be included on all versions or exclusively on a single version but must be asked throughout the data collection year. The objective of the multiple-version questionnaire is to allow more questions--on additional topics--to be asked of a statewide sample. In 2012, 13 states conducted multiple questionnaire version surveys on both their landline telephone and cellular telephone surveys. The web page with the version-specific data sets and additional documentation regarding module data analysis is available from the 2012 BRFSS Survey Data and Documentation.

In 2012, 14 states conducted multiple questionnaire version surveys on their landline telephone surveys. The Web page with the version-specific data sets and additional documentation are available from the Data Files portion of the 2012 BRFSS Survey Data and Documentation.

A change in 2012 to the final disposition code assignment rules has modified the requirements for a partial complete interview. An interview terminated during or after the demographics section was coded as a partial complete. The coding of questions was discontinued at the point of interview termination. The missing and refused values should be taken into account when determining which records to include in any analysis.

Pennsylvania oversampled two counties in order to obtain local level estimates.

The proportion of the oversample in these two counties was disproportionate to the county's adult population and an adjustment factor was applied to the design weight prior to raking.

More information about survey item nonresponse can be found in the 2012 BRFSS Summary Data Quality Report and in the respective states' Data Quality Reports.

## **C. Statistical and Analytic Issues**

### **1. Analyses Procedures**

To use BRFSS data, one needs to formulate the research question, review the existing data tabulations, develop an analytic plan, conduct the analysis, and use data for decision making<sup>4</sup>. Unweighted data on the BRFSS represent the actual responses of each respondent before any adjustment is made for variation in the 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. Regardless of state sample design, use of the weight in analysis is necessary if generalizations are to be made from the sample to the population.

### **2. Statistical Issues**

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.<sup>5,6,7</sup> SAS and SUDAAN can be used for tabular and regression analyses.<sup>4,5</sup> SUDAAN also has these and additional options.<sup>6</sup> Epi Info's C-sample can be used to calculate simple frequencies and two-way cross-tabulations.<sup>7</sup> When using these software products, users must know the stratum, the primary sampling units, and the record weight—all of which are on the public use data file. For more information on calculating variance estimations using SAS, see the *SAS/STAT Users Guide, Version 9.2*.<sup>5</sup> For information about SUDAAN, see the *SUDAAN Users Manual, Release 7.5*.<sup>5</sup> For information about Epi Info, see *Epi Info, Version 6.0*.<sup>7</sup>

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. BRFSS previously followed a rule of not reporting or interpreting percentages based upon a denominator of fewer than 50 respondents (unweighted sample) or the half-width of 95% confidence interval greater than 10. For this reason, the FIPS County code is removed from the data file for any county with fewer than 50 respondents as well as counties with adult populations less than or equal to 10,000 residents. From 2011, the confidence interval limitation was replaced by the relative standard error (RSE): the standard error divided by the mean. The survey with the lower RSE has a more-precise measurement since there is less variance around the mean. BRFSS did not report percentage estimates where RSE was greater than 30% or the denominator represented fewer than 50 respondents from an unweighted sample.

### 3. Analytic Issues

#### a. Advantages and Disadvantages of Telephone Surveys

Compared with face-to-face interviewing techniques, telephone interviews are easy to conduct and monitor and are cost efficient. However, telephone interviews have limitations. Telephone surveys may have higher levels of noncoverage than face-to-face interviews because some U.S. households cannot be reached by telephone. As mentioned earlier, approximately 97.4% of households in the United States have telephones. A number of studies have shown that the telephone and non-telephone populations are different with respect to demographic, economic, and health characteristics.<sup>8,9,10</sup> Although the estimates of characteristics for the total population are unlikely to be substantially affected by the omission of the households without telephones, some of the subpopulation estimates could be biased. Telephone coverage is lower for population subgroups such as people with low incomes, people in rural areas, people with less than 12 years of education, people in poor health, and heads of households under 25 years of age.<sup>11</sup> Raking adjustments for age, race, and sex, and more demographic variables, however, minimize the impact of differences to a greater extent in noncoverage, undercoverage, and nonresponse at the state level than the poststratification adjustments used in previous years to weight the BRFSS.

Despite the above limitations, prevalence estimates from the BRFSS correspond well with findings from surveys based on face-to-face interviews, including the National Health Interview Survey (NHIS), and the National Health and Nutrition

Examination Survey (NHANES).<sup>12,13</sup> Refer to the BRFSS Web site for more information about [methodological studies](#).

Surveys based on self-reported information may be less accurate than those based on physical measurements. For example, respondents are known to underreport weight. This type of potential bias arises when conducting both telephone and face-to-face interviews, and the potential for underreporting should be taken into consideration when interpreting self-reported data.

#### b. New Calculated Variables and Risk Factors

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 processing. These programs prepare the data for analysis and add weighting, sample design, calculated variables, and risk factors to the data set. The following calculated variables and risk factors, created for the user's convenience, are examples of results from this procedure:

`_BMI5, _RFSMOK3, _TOTINDA, _RFHLTH, _RFBING5, _AIDTST3`

The procedures for the variables vary in complexity; some only combine codes, while others require sorting and combining selected codes from multiple variables, which may result in the calculation of an intermediate variable. For more information regarding the calculated variables and risk factors, refer to the document entitled *Calculated Variables in Data Files* in the 2012 BRFSS Survey Data and Documentation section of the BRFSS Web site.

## References

1. Mokdad AH, Stroup DF, Giles WH. Public health surveillance for behavioral risk factors in a changing environment: Recommendations from the Behavioral Risk Factor Surveillance Team. *MMWR Recomm Rep* 2003;52:1–12.
2. Holtzman D. The Behavioral Risk Factor Surveillance System. In: Blumenthal DS, DiClemente RJ, eds. *Community-based Health Research: Issues and Methods*. New York: Springer Publishing Company Inc.; 2004. 115–131.
3. Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2012. National Center for Health Statistics. December 2012. Available from <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201212.pdf>.
4. Frazier EL, Franks AL, Sanderson LM. Behavioral risk factor data. In: Using chronic disease data: a handbook for public health practitioners. Atlanta: US Department of Health and Human Services, Public Health Service, CDC; 1992. 4.1–4.17.
5. SAS Institute Inc. SAS/STAT user's guide, Version 9.2, 2nd ed. Cary (NC): SAS Institute, Inc. 2010.
6. Shah BV, Barnwell BG, Bieler GS. SUDAAN user's manual, Release 7.5. Research Triangle Park: Research Triangle Institute. 1997.
7. Dean AG, Dean JA, Coulombier D, et al. Epi Info, Version 6.0: A word processing, database, and statistics program for public health on IBM-compatible microcomputers. Atlanta: Centers for Disease Control and Prevention. 1995.
8. Groves RM, Kahn RL. Surveys by telephone: *A national comparison with personal interviews*, New York: Academic Press, Publishers; 1979.
9. Banks MJ. Comparing health and medical care estimates of the phone and nonphone populations. In: Proceedings of the Section on Survey Research Methods, American Statistical Association; 1983. 569–574.
10. Thornberry OT, Massey JT. Trends in United States telephone coverage across time and subgroups. In: Groves RM, et al, eds. *Telephone survey methodology*. New York: John Wiley & Sons; 1988. 25–49.
11. Massey JT, Botman SL. Weighting adjustments for random digit dialed surveys. In: Groves RM, et al, eds. *Telephone survey methodology*. New York: John Wiley & Sons; 1988. 143–160.
12. Li C, Balluz L, Ford ES, et al. A comparison of prevalence estimates for selected health indicators and chronic diseases or conditions from the

Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Health and Nutrition Examination Survey, 2007-2008. *Prev Med* 2012; Jun;54(6):381-7.

13. Groves RM. *Survey errors and survey costs*. New York: John Wiley & Sons; 1989. 265, 271–2.