
Behavioral Risk Factor Surveillance System

Comparability of Data BRFSS 2014

(Version #1—Revised: September 2015)



Introduction

The Behavioral Risk Factor Surveillance System (BRFSS) is an ongoing, state-based, random-digit-dialed telephone survey of noninstitutionalized adults 18 years of age or older, residing in the United States.^{1, 2} For detailed descriptions of the BRFSS questionnaires, data, and reports, please see the BRFSS website. In 2014, all 50 states, the District of Columbia, the Commonwealth of Puerto Rico, and Guam conducted both household landline telephone and cellular telephone interviews for the BRFSS data collection.

The BRFSS data collection, structure, and weighting methodology changed in 2011 to allow the addition of data collection by cellular telephones. The BRFSS survey uses disproportionate stratified sample (DSS) design for landline telephone samples and random sample design for the cellular telephone survey. The BRFSS used iterative proportional fitting (IPF)—also known as raking—for weighting 2014 BRFSS data. Because of sample design and the multiple reporting areas, the BRFSS data showed some variation between states and territories for the 2014 data year. [The following sections identify important similarities and variations for the 2014 data year from previous years.](#)

A. 2014 Data Anomalies and Deviations from the Sampling Frame

The BRFSS intended to use a portion of sample records in all the states or territories. The intent was to use it for 1 month but it took more than 1 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 their cellular telephone sample quarterly for the first quarter only and later received their cellular telephone sample monthly.

Several states conducted fewer than 12 monthly data collections during the year. Texas and Minnesota did not collect the January data; Tennessee did not collect data in March; Virginia did not collect June and July data; North Carolina did not collect November and December data; and Mississippi did not collect June, July, August, and September data. The District of Columbia did not collect January, February, and October-to-December landline data and did not collect January and February cellular telephone data. In December, Louisiana did not collect landline data and Massachusetts did not collect cellular telephone data. Moreover, Florida did not collect January and February data for cellular telephone. California collected—beginning in July—with no data available for January through June. Twenty three states were unable to close out the December sample in 2014, and data collection continued into early 2015.

B. Protocol Changes from 2014 Data Collection

1. Cellular Telephone Data

Telephone coverage varies by state and also by subpopulation. According to the 2013 American Community Survey (ACS), 97.7% of all occupied housing units in the United States had telephone service available and telephone noncoverage ranged from 1.4% in Connecticut to 3.2% in Mississippi. It is estimated that 6.2% of occupied households in Puerto Rico did not have telephone service. (See data on telephone coverage in US households.) The percentage of households using only cellular telephones has been steadily increasing—43.1% of all adults lived in households with only cellular telephones in 2014.³ The increased use of cellular telephones required the BRFSS to begin to include the population of cellular telephone users in 2011. At that time, all adult cellular telephone respondents who had a landline telephone were not eligible for the survey. In 2012, the BRFSS changed the screening process. Cellular telephone respondents were eligible—even if they had landline phones—as long as they received at least 90% of all calls on their cell phones. In 2014, all adults contacted through their personal (nonbusiness) phone numbers were eligible regardless of their landline phone use (i.e., complete overlap).

2. Weighting Methodologies

Since 2011, the BRFSS has used the weighting methodology called iterative proportional fitting (IPF) or raking to weight data. Raking allows incorporation of cellular telephone survey data, and it permits the introduction of additional demographic characteristics that more accurately match sample distributions to known demographic characteristics of populations at the state level. (Refer to the BRFSS website for more information on [methodological changes](#)). Raking adjusts the estimates within each state using the margins (raking control variables). The raking method applies a proportional adjustment to the weights of the cases that belong to the same category of the margin. The iteration (up to 100 times) continues until a convergence to within a target percentage difference is achieved. In 2014, there were up to 16 raking margins used in the following order—county by gender, county by age, county by race or ethnicity, county, region by race or ethnicity, region by gender, region by age, region, telephone service (landline, cellular telephone or dual user), age by race or ethnicity, gender by race or ethnicity, tenure (rent or own), marital status, education, race or ethnicity, and gender by age.

In 2014, the inclusion of complete overlap in the cellular telephone survey 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 (landline and cellular sample) for dual users—individuals who had both cellular telephone and landline phone. The BRFSS multiplied the design weight by the compositing factor to generate a composite weight for the records in the overlapping sample frame. Later the design weight was truncated based on quartiles within geographic region, and the truncated weight was used as the input weight for the raking. At the last step of the raking process, weight trimming was 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 is based on two alternative methods, IGCV (Individual and Global Cap Value) and MCV (Margin Cap Value).

3. Other Issues

The data from an optional module were included if interviewers asked module questions 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 2014. If the module was not used for the entire data collection year, 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 2014. The data for the optional modules collected as common 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, CDC produced a separate data set with a raked weight for use with the landline survey data. The webpage with the landline data set and additional documentation is available from the Data Files portion of the 2014 BRFSS Survey Data and Documentation.

CDC has also provided limited technical support for the survey data collection of multiple (up to three in 2014) 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. States must use the 2014 core instrument without making any changes to it in any of their versions of the overall questionnaire. States can include an optional module on all versions or exclusively on a single version but, once a state chooses to use an optional module, the state must ask the module questions throughout the data collection year. The objective of the multiple-version questionnaire is to ask more questions, on additional topics, within a statewide sample. In 2014, 18 states conducted multiple-questionnaire-version surveys on both their landline telephone and cellular telephone surveys. Data users can find version-specific data sets and additional documentation regarding module data analysis in the 2014 BRFSS Survey Data and Documentation.

A 2012 change to the final disposition code assignment rules modified the requirements for a partial complete interview. If a participant terminated an interview during or after the demographics section, the BRFSS coded it as a partial complete. The coding of questions was discontinued at the point of interview termination. When determining which records to include in any analysis, data users should account for participants' missing and refused values.

More information about survey item nonresponse can be found in the 2014 BRFSS Summary Data Quality Report and in the respective states' Data Quality Reports.

C. Statistical and Analytic Issues

1. Analysis Procedures

To use the BRFSS data, the researcher needs to formulate a research question, review the existing data tabulations, develop an analytic plan, conduct the analyses, and use data for decision making.⁴ 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 93 SURVEYMEANS and SURVEYREG procedures, SUDAAN, and Epi Info's C-Sample are among those suitable for analyzing BRFSS data.^{5,6,7} SAS and SUDAAN can be used for tabular and regression analyses.^{4,5} Epi Info's C-sample can be used to calculate simple frequencies and two-way cross-tabulations.⁷ 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® 13.1 User's Guide⁵. For information about SUDAAN, see the SUDAAN Language Manual, Release 11⁵ and to find more about Epi Info, see *Epi Info, Version 7.0*.⁷

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 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 a 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 BRFSS replaced the confidence interval limitation with the relative standard error (RSE)—the standard error divided by the mean. The survey with the lower RSE has a more precise measurement. Because there is less variance around the mean. The BRFSS did not report percentage estimates where the RSE was greater than 30% or the denominator represented fewer than 50 respondents from an unweighted sample. Details of changes beginning with the 2011 BRFSS are available in the *Morbidity and Mortality Weekly Report (MMWR)*, which highlights weighting and coverage effects on trend lines.⁸ Because of the changes in the methodology, researchers are advised to avoid comparing data collected before the changes (up to 2010) with data collected from 2011 and onward.

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 interviewers may not be able to reach some US households by telephone. As mentioned earlier, approximately 97.7% 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.^{9,10,11} 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 younger than 25 years of age.¹² Raking adjustments for age, race, and sex, and more demographic variables, however, minimize the impact of differences to a greater extent in noncoverage, under-coverage, and nonresponse at the state level.

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

Despite the above limitations, the BRFSS data are reliable and valid.¹³ The 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).¹⁴ Please visit the BRFSS website for more information about [methodological studies](#).

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, which the BRFSS has created for the user's convenience, are examples of results from this procedure:

`_BMI5CAT, _RFSMOK3, _FLSHOT6, _TOTINDA, _RFHLTH, _AIDTST3`

The procedures for the variables vary in complexity. Some only combine codes, while others require sorting and combining selected codes from multiple variables. This 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 2014 BRFSS Survey Data and Documentation section of the BRFSS website.

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(RR-9):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, NY: Springer Publishing Company Inc; 2004:115-131.
3. Blumberg SJ, Luke JV. Wireless substitution: early release of estimates from the National Health Interview Survey, July–December 2014. National Center for Health Statistics. June 2015.
<http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201506.pdf> . Accessed 8 September 2015.
4. Frazier EL, Franks AL, Sanderson LM, Centers for Disease Control and Prevention. Behavioral risk factor data. In: *Using Chronic Disease Data: a Handbook for Public Health Practitioners*. Atlanta, GA: Centers for Disease Control and Prevention, US Dept. of Health and Human Resources; 1992.
5. SAS Institute Inc. 2013, *SAS/STAT*® 13.1 User's Guide. Cary, NC: SAS Institute, Inc.
6. Research Triangle Institute (2012). SUDAAN Language Manual, Volumes 1 and 2, Release 11.

7. Dean AG, Arner TG, Sunki GG, et al. Epi Info™, a database and statistics program for public health professionals. Atlanta, GA: Centers for Disease Control, US Dept of Health and Human Resources; 2011.
8. Pierannunzi C, Town M, Garvin W, et al. Methodologic changes in the Behavioral Risk Factor Surveillance System in 2011 and potential effects on prevalence estimates. *MMWR Morb Mortal Wkly Rep.* 2012;61(22):410-413.
www.cdc.gov/mmwr/preview/mmwrhtml/mm6122a3.htm Accessed 8 September 2015.
9. Groves RM, Kahn RL. *Surveys by Telephone: A National Comparison with Personal Interviews*, New York, NY: Academic Press; 1979.
10. 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.
11. 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, NY: John Wiley & Sons; 1988:25-49.
12. Massey JT, Botman SL. Weighting adjustments for random digit dialed surveys. In: Groves RM, et al, eds. *Telephone Survey Methodology*. New York, NY: John Wiley & Sons; 1988:143-160.
13. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol.* 2013;13:49.
14. 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;54(6):381-387.