COMPARABILITY OF DATA: BRFSS 2011

The Behavioral Risk Factor Surveillance System (BRFSS) is a cross-sectional surveillance survey currently involving 54 reporting areas.[1, 2] BRFSS questionnaires, data, and reports are available on the BRFSS Web site. It is important to note that any survey will have natural variations across sample sites; therefore, some variation between states is to be expected. The complex sample design and the multiple reporting areas complicate the analysis of the BRFSS. In 2011, there were some major changes including collecting cell phone data and using new weighting methodologies. The following Sections A and B identify other known variations for the 2011 data year.

A. 2011 Data Anomalies and Deviations from Sampling Frame

In 31 states, 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.

Several states conducted fewer than 12 monthly data collections during the year. For landline data, the District of Columbia, New Hampshire, and Utah did not collect data in January. Nevada did not collect data in January and February. Illinois did not collect data in April and May. North Carolina did not collect data in June and July. For cell phone data, the District of Columbia, New Hampshire and Rhode Island did not collect data in January. Hawaii did not collect data in January, February, and December. Nevada did not collect data in January and February. South Carolina did not collect data from January to March. Puerto Rico did not collect data in January and from September to December. Tennessee did not collect data from January to October. Wisconsin did not collect data in November or December.

Ten states and one U.S. territory were unable to close out the December sample in 2011, and data collection continued into early 2012: Colorado, Illinois, Kentucky, Louisiana, Nebraska, New Mexico, North Carolina, Oregon, Utah, Wisconsin, and Puerto Rico.

B. Protocol Changes in 2011 Data

1. Addition of Cell Phone Data

Telephone coverage varies by state and also by subpopulation. In 2008, telephone coverage averaged 98.2% for the United States as a whole, but non-coverage ranged from 0.7% in Utah to 4.3% in New Mexico. It is estimated that 8.1% of households in Puerto Rico are without telephone service. Click here to access data on telephone coverage in U.S. households. The report shows telephone coverage by state as a combination of landline and cell phone use. The percentage of cell-phone-service-only households increased to 31.6% in 2011. Additionally, one of every six American homes received all or most of their phone calls on a cell phone while a landline was also available. [3] The increased use of cell phones has required the BRFSS to include the population of cell phone users in the survey. By doing so, BRFSS is able to reach the target population that was previously inaccessible, including those who had a cell phone but not a landline phone.

2. New Weighting Methodologies

Previously, BRFSS processed weighting using post-stratification to estimate how non-respondents may have answered survey questions. The final weight is achieved by multiplying design weight (stratum weight * 1 / number of phone lines * number of adults) by post-stratification adjustment. Post-stratification forces the sum of weighted frequencies to equal the population estimates for the region or state by age, race, and gender groups. One big limitation is that it was not possible to control for the full distribution of other important demographic information such as marital status, race or ethnicity, education, etc.

In 2011, a different weighting methodology, raking, was applied to the BRFSS, which aimed to improve the extent to which a BRFSS sample can better reflect the socio-demographics at state levels. Raking adjusts estimates within each state using the margins (raking control variables), including age group by gender, 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. Raking, also known as iterative proportional fitting, usually proceeds one variable at a time, applying a proportional adjustment to the weights of the cases that belong to the same category of the margins. Unlike post-stratification, 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 weight methodology fits in a dual frame survey design: landline only, cell phone only, and combined landline and cell phone. 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).

3. Other Issues

The data from an optional module is included if asked of all eligible respondents within a state for the entire data collection year. A state may have indicated the use of an optional module in 2011, but the data set may have been moved into the state-added questions if it does not represent all eligible respondents. The states that included optional modules on both the landline and cell phone surveys include Alaska (2), Arizona (10), Connecticut (10), Hawaii (10), Illinois (4), Michigan (4), Minnesota (8), Nebraska (1), New Hampshire (5), New Jersey (4), New Mexico (5), North Carolina (4), Oregon (3), Pennsylvania (6), Puerto Rico (3), South Carolina (8), Tennessee (14), Texas (2), Utah (2), Vermont (3), Washington (4), West Virginia (7), Wisconsin (10), and Wyoming (4).

Several states collected optional modules on the landline-only survey in 2011. 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 version-specific data sets and additional documentation are available from the Data Files portion of the 2011 BRFSS Survey Data and Documentation. The Division of Behavioral Surveillance has also provided limited support for the landline survey data collection of multiple (up to three) questionnaire versions in 2011. 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 (2500) for a given questionnaire version. The 2011 core instrument must be asked without any changes in all 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 is to allow more questions to be asked of a statewide sample for additional topics. There were 18 states that conducted multiple questionnaire version surveys in 2011. The web page with the version-specific data sets and additional documentation are available from the Landline Data Files portion of the 2011 BRFSS Survey Data and Documentation.

A change in 2002 to the final disposition codes has continued to present some inconsistencies in closing-out the questionnaire. Prior to 2002, interviews that were terminated during or after the demographics section were coded as complete interviews, and any remaining unanswered questions were coded as refused by the interviewer. In 2002, a revised procedure was implemented for handling partial completes: to stop coding questions at the point of interview termination to assign the appropriate disposition code. The missing and refused values should be taken into account when determining which records to include in an analysis. Records with a termination in the questionnaire followed by coded refusals for the remainder of the eligible responses have been dispositioned as 120 Partial Completes.

Another issue regarding partial completes is the inappropriate coding of the remaining questions.

For example, some questions allow responses of 01-76, 77, 88, and 99 (with 99 as the refusal code). A problem occurs when an interviewer incorrectly codes the remaining questions as refused by entering '9' instead of a '99' for these questions; where nine (9) is a valid response for these questions and should not have been used to indicate refusal. When reviewing responses to a partial complete interview, data users should be aware that a core section or module that follows the demographics section may contain questions incorrectly coded as refused (9 filled). Records are evaluated upon submission to identify core sections following the demographics section or optional modules with more than half of the responses coded as unknown, refused, or missing (7, 77, 777, 9, 99, 999, or blank). Records meeting the criteria have their disposition code changed to reflect a partial complete (120) and the data for the corresponding variables set to missing to remove some of the confusion of miscoded refusals.

Several states continue to ask the diabetes module questions directly after the diabetes question in the core of the survey. In addition, several states ask the adult asthma module questions after the asthma questions in the core. Some states have also asked the childhood asthma module questions in the demographics section of the core survey after question 12.7, (CHILDREN) — number of children younger than age 18 in the household.

More information about survey item nonresponse can be found in the 2011 BRFSS Summary Data Quality Report and in the respective states' Data Quality Reports.

C. STATISTICAL AND ANALYTIC ISSUES

1. Estimation Procedures

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 final 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.[4,5,6] SAS and SUDAAN can be used for tabular and regression analyses.[4,5] SUDAAN also has these and additional options.[5] Epi Info's C-sample can be used to calculate simple frequencies and two-way cross-tabulations.[6] 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*.[4] For information about SUDAAN, see the SUDAAN User's Manual, Release 7.5.[5] For information about Epi Info, see *Epi Info, Version 6.0* [6]

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. In 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 98.2% of households in the United States have telephones. A number of studies have shown that the telephone and nontelephone populations are different with respect to demographic, economic, and health characteristics.[7,8,9] Although the estmates 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.[10] However, raking adjustments for age, race, and sex, and more demographic variables 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 studies conducted by the National Institute on Alcohol Abuse and Alcoholism, CDC's National Center for Health Statistics, and the American Heart Association.[11, 12] A summary of methodological studies of BRFSS is provided in the publication section of the BRFSS Web site.

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. However, when measuring change over time, this type of bias is likely to be constant, and is therefore not a factor in trend analysis.

b) Aggregating Data Over Time

When data from one time period are insufficient for estimating the prevalence of a risk factor, data from multiple periods can be combined as long as the prevalence of the risk factor of interest did not substantially change during one of the periods. One method that can be used to assess the stability of the prevalence estimates is as follows[11]:

i. Compute the prevalence for the risk factor for each period.

ii. Rank the estimates from low to high.

iii. Identify a statistical test appropriate for comparing the lowest and the highest estimates at the 5% level of significance. For example, depending on the type of data, a t-test, or the sign test might be appropriate.

iv. Test the hypothesis that prevalence is not changing by using a two-sided test in which the null hypothesis is that the prevalences are equal.

v. Determine whether the resulting difference could be expected to occur by chance alone

less than 5% of the time (i.e., test at the 95% confidence level).

c) Analyzing Subgroups

Provided that the prevalence of risk factors did not change rapidly over time, data combined for two or more years may provide a sufficient number of respondents for additional prevalence estimates for population groups (such as age/sex/race subgroups or county populations). Before combining data for subgroups, 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 shows the sample size required for each of several levels of precision, based on a calculation in which the estimated risk factor prevalence is 50% and the design effect is 1.5.

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

Precision is indicated by the width of the 95% confidence interval around the prevalence estimate. For example, precision of 2% indicates that the 95% confidence interval is plus (+) or minus (-) 2% of 50%, or 48% to 52%. As shown in the table, to yield this high a level of precision, the sample size required is about 3,600 persons. When a lower level of precision is acceptable, the sample size can be considerably smaller.

The design effect is a measure of the complexity of the sampling design that indicates how the design differs from simple random sampling. It is defined as the variance for the actual sampling design divided by the variance for a simple random sample of the same size.[11, 13] For most risk factors in most states, the design effect is less than 1.5. If it is more than 1.5, however, sample sizes may need to be larger than those shown in the table above.

The standard error of a percentage is largest at 50% and decreases as a percentage approaches 0% or 100%. From this perspective, the required sample sizes listed in the table above are conservative estimates. They should be reasonably valid for percentages between 20% and 80% but may significantly overstate the required sample sizes for smaller or larger percentages.

d) Creating Synthetic Estimates

Even after combining data for several years, sample sizes may still be inadequate for risk factor estimates for some geographic areas (e.g., counties) or subpopulations (e.g., people with diabetes). In such situations, the analyst may wish to derive synthetic estimates by extrapolating from BRFSS data collected at the state level.

Synthetic estimates can be calculated using the population estimates for the subgroup of interest

and the statewide BRFSS risk factor prevalences for that subgroup. This approach assumes that the risk factor prevalences for specific subgroups in each area are the same as the statewide risk factor prevalences for the same subgroups. For example, it assumes that the risk factor prevalences for black women in every county of a state are the same as those for black women in the entire state. The accuracy of the estimate depends on the validity of this assumption, which is often impossible to judge. However, a "ballpark" estimate may be sufficient for establishing broad goals and objectives for prevention strategies. For a discussion of the precision of such estimates, see Levy and Lemeshow, 1991.[14]

An example for estimating the number of people with hypertension in a hypothetical county, as well as the overall prevalence of hypertension in that county, is shown below. The sex and race distribution of the county's population differs from the statewide population, and these differences need to be taken into account. By developing a table like the one below, a synthetic estimate for the overall county prevalence of hypertension can be made.

Synthetic Estimates of Prevalence of Hypertension in a Hypothetical County, 2010

Subgroup	Statewide Prevalence* 2010	County Population 2010	County Population With Hypertension 2010
Men			
White	15.6	10,000	1,560
Black	27.0	25,000	6,750
Women			
White	19.5	12,000	2,340
Black	26.5	28,000	7,420
Total		75,000	18,070
***Per 100 persons**			

The statewide prevalence values, given as rates per 100 persons are computed from the BRFSS data. The estimated number of persons with hypertension for each race-sex group in the county was obtained by multiplying the statewide prevalence for that group by the county population for the group. To determine the total county prevalence, the number of people with hypertension in each race-sex group in the county was summed and this sum (18,070) was divided by the county's total population (75,000) to yield an overall prevalence of 24.1 per 100 persons.

e) 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:

MAXVO2_, ACTINT2_, _PACAT, _RFHLTH, _RFCHOL, _RFHYPE5

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 (e.g., MAXVO2_, ACTINT1_, _PACAT). For further details regarding the calculated variables and risk factors, refer to the document entitled *Calculated Variables and Risk Factors for the 2011 Behavioral Risk Factor Surveillance System*, on the 2011 Annual Data page 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, 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, January–June 2011. National Center for Health Statistics. December 2011.

4. SAS Institute Inc., SAS/STAT User's Guide, Version 9.2, Second edition. Cary, NC: SAS Institute, Inc.; 2010.

5. Shah BV, Barnwell BG, Bieler GS. SUDAAN User's Manual, Release 7.5. Research Triangle Park, NC: Research Triangle Institute; 1997.

6. 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, GA: Centers for Disease Control and Prevention; 1995.

7. Groves RM, Kahn RL. Surveys by Telephone: *A National Comparison with Personal Interviews*, New York, NY: Academic Press, Publishers; 1979.

8. Banks MJ. Comparing health and medical care estimates of the phone and nonphone populations. Proceedings of the Section on Survey Research Methods, American Statistical Association; 1983;569–574.

9. 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.

10. 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.

11. 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. Available on this page.

12. Nelson DE, Powell-Griner E, Town M, et al. A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *Am J Public Health* 2003;93(8);1335–41.

13. Groves RM. *Survey Errors and Survey Costs*. New York, NY: John Wiley & Sons; 1989:265, 271–2.

14. Levy PS, Lemeshow S. *Sampling of Populations: Methods and Applications*. New York, NY: John Wiley & Sons; 1991;347–50.