BRFSS National Weighted Estimates

Ronaldo Iachan[1], Kristie Healey[2], Pedro Saavedra[1]

[1]ICF International, 530 Gaither Road, Suite 500, Rockville, MD 20850

[2]ICF International, 126 College Street, Burlington, VT 05401

Abstract

The Behavioral Risk Factor Surveillance System (BRFSS) is the world's largest on-going telephone health survey system. Conducting interviews in 50 states and the District of Columbia, the BRFSS surveys monitor state-level prevalence of the major behavioral risks among adults associated with premature morbidity and mortality. The weighting methodology was revised in 2011 to use raking as a means of incorporating more demographic variables, including telephone ownership, into the weighting. Designed as a state level survey, there has been increasing interest in using the data to provide estimates at the national level. One challenge is to develop design variables for use in accurate variance estimation. We present the development of weights that support combining state-level data to produce statistical valid national estimates. This work updates research conducted by Iachan et al (1998). The paper presents methods, a summary of the resulting weights, and compares resulting estimates with those obtained from other national estimation systems.

1. Overview

The CDC coordinates state surveillance of behavioral risk factors through the Behavioral Risk Factor Surveillance System (BRFSS). The primary purpose of the BRFSS is to collect uniform state-specific a data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries and preventable infectious diseases. The BRFSS is a state-based survey of adults aged 18+ years living in households with phones. Each state independently collects BRFSS data using a standardized instrument.

There is a natural demand for national BRFSS estimates as researchers rely on the BRFSS as the only or primary source for a number of health indicators. The aggregation of state level BRFSS data requires the development of national weights as well as a methodology for computing the associated variance estimates.

Combining the state level survey data into a national data set is a valid initiative for the following reasons:

- The surveys use the same sampling methodology across states
- The surveys yield state-level weights with a same basic methodology
- The surveys use the same core questionnaire across states

The increased uniformity of methods makes the aggregation more efficient than in earlier investigations, starting with Iachan et al. (1999), conducted at a time when there was substantially more variation in the sampling and weighting methodologies used by different states.

This paper examines alternative approaches for generating national weights. These approaches all begin with the state-level weights now computed in the BRFSS system. The baseline method for our comparisons is a simple method that concatenates the data with the current state-level weights. The BRFSS weighting includes a raking process, an iterative form of post-stratification that ensures that weights sum to known population totals for key demographics in each state. Most of the variations involve an additional layer for the raking that adds the state as a margin.

An assessment of the weights considers estimated bias and variances, as well as the mean squared error (MSE) for key health risk indicators. We compare the national estimates with a benchmark provided by the National Health Interview Survey (NHIS) data for comparable health indicators.

We developed a range of weighting methods that can potentially improve on the simple method that stacks the state data with the state level weights to form a national data set. Potential improvements can take place along both bias and variance dimensions. Section 2 outlines the methods used in state level weighting. Section 3 suggests some approaches for reducing the variance. Section 4 introduces a family of raking methods for reducing or controlling the potential bias. Section 5 provides an assessment of the alternative weighting methods, and a conclusion.

2. State weights

This section provides an overview of the methods used by CDC in computing state level weights. The weights start from design weights—base weights or sampling weights—computed as the reciprocal of the probabilities of selection. They include a correction for dual frame, or dual use of landline and cell phones. Finally, the weights are raked, or iteratively fitted to population distributions, along a range of socio-demographic variables shown in Table 1.

Table 1: State level Raking and National Raking Margins 1-8

Margin	*Categories*
1: Sex by Age	Male and Female by Age categories: 18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+
2: Race/Ethnicity	Non-Hispanic White; Non-Hispanic Black; Hispanic; Other
3: Education	Less than HS; HS Grad; Some College; College Grad

4: Marital Status	Married; Never married/member of unmarried couple; Divorced/widowed/separated.
5: Home Ownership	Own; Rent/Other
6: Sex by Race/Ethnicity	Male; Female by Non-Hispanic White; Non-Hispanic Black; Hispanic; Other
7: Race/Ethnicity by Age	Non-Hispanic White; Non-Hispanic Black; Hispanic; Other by 18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+
8: Phone Usage	Cell Only; Landline Only; Dual Usage

3. Variances

The variability in state-level weights is very large due to the variability in base weights (design weights or sampling weights) that reflect the highly unequal sampling rates adopted across states. Recall that the base weights are computed as the reciprocal of sampling probabilities, and that for a stratified random sampling design, the probabilities are in essence the sampling rates in different strata and overall. Because sample sizes are reasonably constant across states, the sampling rates are much larger in the smaller states (i.e., states with small populations like ID, DE, SD, MT and so on) than in the larger states (i.e., states with very large populations such as CA, FL, NY, TX and so on).

Table 2 presents the design effect (DEFF) due to weighting at the state level, the component of the DEFF due to unequal weighting effects. It gauges the impact of the weight variability on sampling error under two scenarios: a) under simple random sampling, and b) by allowing for the impact of unequal weighting effects. The measure of sampling error shown in this table is the margin of error, i.e., the half-width of a 95% confidence interval.

Table 2: Design effect (DEFF) due to unequal weighting for each state

State	*N*	*DEFF*	*Margin of Error (SRS)*	*Effective Sample Size*	*Expected Margin of Error*
Nationwide	467333	4.50862	0.14%	103653.27	0.30%
Alabama	9026	2.23185	1.03%	4044.17	1.54%
Alaska	4345	2.01437	1.49%	2157.00	2.11%
Arizona	7306	2.78326	1.15%	2624.98	1.91%
Arkansas	5187	1.93093	1.36%	2686.28	1.89%
California	14574	2.13646	0.81%	6821.56	1.19%
Colorado	12255	1.83758	0.89%	6669.10	1.20%
Connecticut	8781	2.03684	1.05%	4311.09	1.49%

Delaware	5174	1.91926	1.36%	2695.83	1.89%
District of Columbia	3827	2.71521	1.58%	1409.47	2.61%
Florida	7624	2.45716	1.12%	3102.77	1.76%
Georgia	6100	2.02842	1.25%	3007.26	1.79%
Hawaii	7582	2.36436	1.13%	3206.78	1.73%
Idaho	5896	3.12673	1.28%	1885.68	2.26%
Illinois	5579	2.12142	1.31%	2629.85	1.91%
Indiana	8645	1.75329	1.05%	4930.72	1.40%
Iowa	7166	1.59735	1.16%	4486.17	1.46%
Kansas	11801	1.84120	0.90%	6409.41	1.22%
Kentucky	11223	2.45243	0.93%	4576.28	1.45%
Louisiana	9068	2.70278	1.03%	3355.06	1.69%
Maine	9921	1.59249	0.98%	6229.85	1.24%
Maryland	12812	2.96643	0.87%	4318.99	1.49%
Massachusetts	21723	2.43757	0.66%	8911.75	1.04%
Michigan	10499	1.91571	0.96%	5480.48	1.32%
Minnesota	12246	1.83205	0.89%	6684.31	1.20%
Mississippi	7788	2.15611	1.11%	3612.06	1.63%
Missouri	6754	2.16672	1.19%	3117.16	1.76%
Montana	8679	1.76679	1.05%	4912.31	1.40%
Nebraska	19173	2.08117	0.71%	9212.62	1.02%
Nevada	4846	2.19393	1.41%	2208.82	2.09%
New Hampshire	7530	2.14577	1.13%	3509.23	1.65%
New Jersey	15761	2.21211	0.78%	7124.86	1.16%
New Mexico	8776	1.64383	1.05%	5338.74	1.34%
New York	6060	2.04353	1.26%	2965.45	1.80%
North Carolina	11898	1.75513	0.90%	6778.99	1.19%
North Dakota	4879	1.83068	1.40%	2665.13	1.90%
Ohio	13026	1.98417	0.86%	6564.97	1.21%
Oklahoma	8015	1.77471	1.09%	4516.24	1.46%
Oregon	5302	1.82039	1.35%	2912.56	1.82%
Pennsylvania	19958	2.42702	0.69%	8223.24	1.08%
Rhode Island	5480	1.97042	1.32%	2781.13	1.86%
South Carolina	12795	2.25215	0.87%	5681.25	1.30%

South Dakota	7878	2.57234	1.10%	3062.58	1.77%
Tennessee	7056	2.05803	1.17%	3428.52	1.67%
Texas	9129	2.06138	1.03%	4428.58	1.47%
Utah	12436	1.86681	0.88%	6661.64	1.20%
Vermont	6056	1.75788	1.26%	3445.06	1.67%
Virginia	7398	1.76563	1.14%	4190.01	1.51%
Washington	15319	1.95472	0.79%	7836.92	1.11%
West Virginia	5409	1.45052	1.33%	3729.01	1.60%
Wisconsin	5299	2.24324	1.35%	2362.20	2.02%
Wyoming	6273	2.79994	1.24%	2240.41	2.07%

We investigated an approach for reducing the variance of national estimates by balancing more closely the sampling weights. This approach aims at balancing the sampling rates used across states by selecting subsamples from those states with the largest rates (largest weights). The idea is to reduce the effective sample size at the national level, defined as the aggregate sample size divided by the design effect (DEFF) due to unequal weighting.

Without any subsampling, the effective sample size of selected phone numbers at the national level can be calculated as follows:

Effective-n = (Total sample size) / (DEFF due to unequal weighting)

= 6,617,262/ 3.97

= 1,667,160

With about 7.1 percent of selected telephone numbers resulting in a completed interview, this results in an effective sample size of approximately 117,740 interviews.

With optimal subsampling, the effective “n” is 468,677selected records and 33,099 interviews. For comparison purposes, a self-weighting subsample would have an effective “n” of 353,458 selected records and 24,962 interviews.

The design effect due to weighting can be due to differential selection of records, differential response rates in different strata and weight adjustments to correct for the above. The sampling process began with the selection of records from strata, leading to initial weights equal to the number in the strata divided by the number selected. A total of 6,617,262 records were selected. A design effect due to this initial weighting process was calculated to be 3.97 – not that different from the 4.51 attributed to the final weight.

4. Bias and raking

The use of the stacked state-level weights may lead to biases at the national level to the extent that for key demographics, the national weighted distribution does not match the national population distribution. Note also that bias in each key survey outcome may still occur to the extent that the survey health outcome is not strongly associated to those variables used in the raking.

To control for this potential bias, the national weights should be raked at the national level using as many of the raking dimensions—used at the state level—as possible for convergence and stability. In addition, it is desirable for the national raking to use states as marginal layers to preserve the state totals.

This motivates the different weighting (or reweighting) methods that use a range of raking marginal defined in Table 3 in addition to Table 1. Table 1 above describes the variables and categories used in the raking for the states but now used at the national level. These variable/categories constitute eight marginal classes inherited for the national raking. The national raking also includes the state by demographic variable classifications shown in Table 3. Three initial reweighting methods were identified. Each started with the original design weights and readjusted the raking process at the national level. The first reweight used the original raking margins as described in Table 1. The second reweight used the original eight raking margins as well as state (Margin 9). The third reweight included all twelve raking margins defined in Tables 1 and 3.

Table 3: National Raking Margins 9-12 (new for National)

Margin	*Categories*
9: State	State FIPS code
10: Age by State	18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+ by State FIPS Code
11: Sex by State	Male; Female by State FIPS Code
12: Race/Ethnicity by State	Non-Hispanic White; Other by State FIPS Code

An additional three reweighting methods were tested in an effort to reduce the overall variability of the weights. These three methods used the same overall raking margins as the first three methods, but collapsed some demographics into larger categories. Table 4 describes the preliminary collapsing of categories recommended for some of the margins (#6 and #7) at the national level. Some additional collapsing of margins 6, 7, 10, and 12 was performed on individual cells to ensure that all cells obtained a minimum sample percentage of 5.0%.

Table 5 summarizes the six alternative weighting methods, or reweights, investigated in this research. The first three methods use 8, 9 and 12 margins, while the last three methods use the same margins but with the collapsing described in Table 4.

Table 4: Collapsed Margins 6 and 7

Margin	*Categories*
6: Sex by Race/Ethnicity	Male non-Hispanic White; Male Other; Female non-Hispanic White; Female non-Hispanic Black; Female Other
7: Race/Ethnicity by Age	Non-Hispanic White 18-34; Non-Hispanic White 35-54; Non-Hispanic White 55+; Other 18-34; Other 35-54; Other 55+

Table 5 Additional reweights

National Reweight#1: 8 Margins	National Reweight#4: 8 Margins, Collapsed Categories
National Reweight#2: 9 Margins	National Reweight#5: 9 Margins, Collapsed Categories
National Reweight#3: 12 Margins	National Reweight#6: 12 Margins, Collapsed Categories

5. Assessment of alternative national weighting methods

This section compares the six modified weighting methods with the simple method that stacks the state-level weights. The methods are compared in terms of the estimated variance and bias of resulting weighted survey estimates. The estimated variances are gauged in two ways. First, in terms of the variability in the weights, we assess a pure contribution of unequal weighting to the design effects and survey variances. Second, in a more empirical way, we look at the estimated variances for a number of key survey estimates, i.e., for six key health indicators. The indicators are for current smoking, diabetes, arthritis, asthma, heart attacks and heart disease, each estimated via the prevalence of ever having the condition.

Table 6 presents the variability in the weights as measured by the design effect due to unequal weighting for each method. It also shows the margin of error (half-width for the 95% confidence interval) for each method. The table suggests a slight superiority for the two methods using 8 marginal classes.

Table 6: Weight Variability

Weight	*CV*	*DEFF*	*Expected Margin of Error*
Stacked state weights	187.313	4.50862	0.30%

National 8 Margins	166.082	3.75832	0.28%
National 9 Margins	178.950	4.20230	0.29%
National 12 Margins	177.140	4.13784	0.29%
National 8 Collapsed Margins	165.590	3.74200	0.28%
National 9 Collapsed Margins	178.380	4.18192	0.29%
National 12 Collapsed Margins	176.533	4.11637	0.29%

Table 7 provides a more comprehensive and empirical picture by looking at potential biases as well as variances estimated for the six key indicators described above. Moreover, these estimates also permit the computation of a Mean Squared Error (MSE), estimated as the variance plus the square of the bias. The bias is estimated as the difference between the weighted estimate and the NHIS estimate used as a benchmark.

Table 7 presents a summary, average MSE used to compare the six methods and the canonical, simple method stacking the state-level weights. While no method is superior across all health indicators in terms of MSE, the average MSE values suggest the superiority of (re)weighting methods #2, #3 and #6 based on nine margins and on twelve margins.

Table 7: Standard Error and Mean Squared Error (MSE)

Weight	*Current Smoker*	*Ever Told Had Diabetes*	*Ever Told Had Arthritis*	*Ever Told Had Asthma*	*Ever Told Had Heart Attack*	*Ever Told Had Heart Disease*	*Average MSE*
BRFSS Stacked Weights	Pct: 18.85% SE: 0.12% MSE: 0.006%	Pct: 10.12% SE: 0.09% MSE: 0.011%	Pct: 25.63% SE: 0.12% MSE: 0.126%	Pct: 13.22% SE: 0.11% MSE: 0.004%	Pct: 4.42% SE: 0.06% MSE: 0.014%	Pct: 4.50% SE: 0.06% MSE: 0.059%	0.037%
BRFSS Re-weight 1	Pct: 19.22% SE: 0.12% MSE:	Pct: 10.08% SE: 0.08% MSE:	Pct: 25.69% SE: 0.11% MSE:	Pct: 13.12% SE: 0.10% MSE:	Pct: 4.45% SE: 0.05% MSE:	Pct: 4.52% SE: 0.05% MSE:	0.038%

	0.013%	0.010%	0.130%	0.002%	0.014%	0.058%	
BRFSS Re-weight 2	Pct: 18.91% SE: 0.12% MSE: 0.007%	Pct: 10.04% SE: 0.08% MSE: 0.009%	Pct: 25.60% SE: 0.11% MSE: 0.124%	Pct: 13.25% SE: 0.10% MSE: 0.004%	Pct: 4.40% SE: 0.05% MSE: 0.013%	Pct: 4.51% SE: 0.05% MSE: 0.059%	0.036%
BRFSS Re-weight 3	Pct: 18.95% SE: 0.12% MSE: 0.008%	Pct: 10.03% SE: 0.08% MSE: 0.009%	Pct: 25.57% SE: 0.12% MSE: 0.122%	Pct: 13.26% SE: 0.10% MSE: 0.004%	Pct: 4.40% SE: 0.05% MSE: 0.013%	Pct: 4.50% SE: 0.05% MSE: 0.059%	0.036%
BRFSS Re-weight 4	Pct: 19.18% SE: 0.12% MSE: 0.013%	Pct: 10.08% SE: 0.08% MSE: 0.010%	Pct: 25.71% SE: 0.11% MSE: 0.131%	Pct: 13.08% SE: 0.10% MSE: 0.002%	Pct: 4.45% SE: 0.05% MSE: 0.014%	Pct: 4.52% SE: 0.05% MSE: 0.058%	0.038%
BRFSS Re-weight 5	Pct: 18.87% SE: 0.12% MSE: 0.007%	Pct: 10.04% SE: 0.08% MSE: 0.009%	Pct: 25.61% SE: 0.11% MSE: 0.125%	Pct: 13.22% SE: 0.10% MSE: 0.004%	Pct: 4.40% SE: 0.05% MSE: 0.013%	Pct: 4.51% SE: 0.05% MSE: 0.059%	0.036%
BRFSS Re-weight 6	Pct: 18.92% SE: 0.12% MSE: 0.007%	Pct: 10.03% SE: 0.08% MSE: 0.009%	Pct: 25.57% SE: 0.12% MSE: 0.122%	Pct: 13.24% SE: 0.10% MSE: 0.004%	Pct: 4.39% SE: 0.05% MSE: 0.013%	Pct: 4.50% SE: 0.05% MSE: 0.059%	0.036%
NHIS	Pct: 18.06% SE: 0.30%	Pct: 9.08% SE: 0.19%	Pct: 22.08% SE: 0.30%	Pct: 12.63% SE: 0.25%	Pct: 3.26% SE: 0.12%	Pct: 6.93% SE: 0.18%	

In conclusion, the investigation suggests the advantages in using a national raking method that also ensures state totals and state distributions are preserved by adding states as layers in the raking. It also suggests that methods with additional margins lead to additional gains.

Reference

Iachan R, Schulman J, Collins S, Powell-Griner E, Nelson D. Methods for combining state BRFSS survey data for national estimation. Proceedings of the Section on Survey Research Methods, American Statistical Association; 1999. p. 711-4.