



## Combining state data from the Behavioral Risk Factor Surveillance System (BRFSS) surveys

Ronaldo Iachan, ICF International

Kristie Healey, ICF International

## **Behavioral Risk Factor Surveillance System (BRFSS)**

- **The CDC coordinates state surveillance of behavioral risk factors through the Behavioral Risk Factor Surveillance System (BRFSS)**
- **The BRFSS surveys collect uniform state-specific data on preventive health practices and risk behaviors**
- **Target population: adults aged 18+ years living in households with phones**
- **Each state independently collects BRFSS data using a standardized instrument**

## Combining state BRFSS data

- **There is a natural demand for combining state level estimates into national estimates**
- **The aggregation 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 feasible for the following reasons:**
  - The surveys use a same sampling methodology across states
  - The surveys produce state-level weights with a same basic methodology
  - The surveys use the same core questionnaire across states

## National Weights

- **The increased uniformity of methods makes the aggregation more efficient than in previous investigations starting with Iachan et al. (1998). At that time, there was substantially more variation in the sampling and weighting methodologies used by different states**
- **This paper examines alternative approaches for generating national weights starting with the state-level weights now computed in the BRFSS system**

## Raking

- **The BRFSS weighting includes a raking process**
- **Raking is an iterative form of post-stratification**
- **Post-stratification and raking methods ensure that weights sum to known population totals for key demographics in each state**
- **Raking allows additional post-stratification dimensions and cross-classifications of variables**

## Comparisons of weighting methods

- **The baseline method for comparisons is a simple method that concatenates the data with the current state-level weights**
- **The variations involve raking at the national level**
  - Ensure the national distribution is met
- **Raking may include an additional layer for the states as a margin**
  - Ensure state-level distribution is preserved

## Assessment of weighting methods

- **Our assessment of the weights looks at estimated bias and variances for key health risk indicators**
- **Variances: Weight variability and variances of key survey estimates**
- **Bias: We compare the national weighted estimates with other benchmarks including data from the National Health Interview Survey (NHIS) for comparable health indicators**

## BRFSS Sampling Design

- **Dual frame landline and cell phone samples**
- **RDD stratified samples**
- **Density strata enhance sampling efficiency**
- **Some states add regional strata**



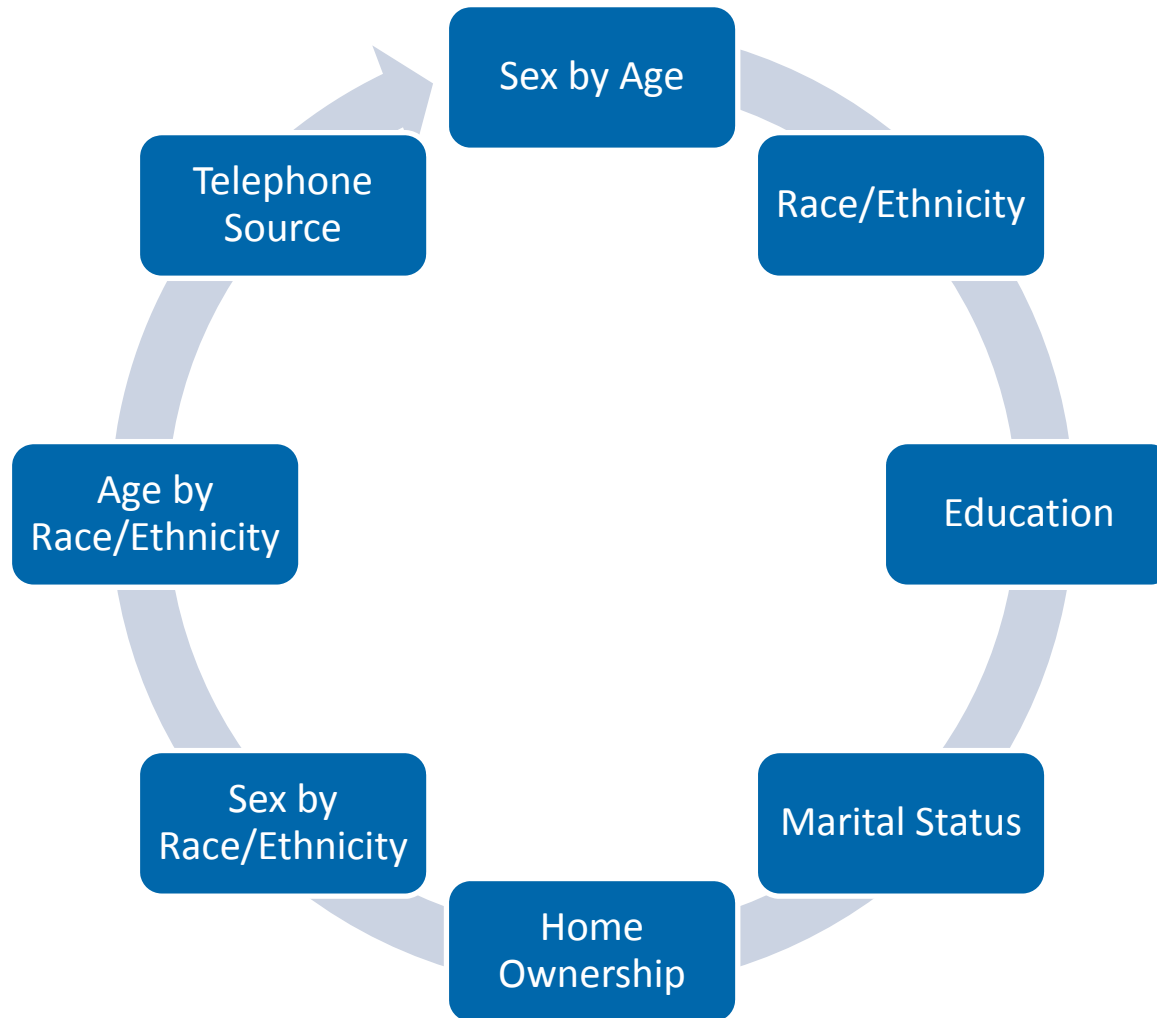
## State Weighting Process

- **Design weights**
  - Initial weights reflecting differential selection probabilities
- **Dual-use correction**
  - Accounts for the overlap in the landline and cell phone frames
- **Raking**

## Raking Categories

- **All states use eight raking margins:**
  - Sex by age
  - Race/Ethnicity
  - Education
  - Marital Status
  - Home Ownership
  - Sex by Race/Ethnicity
  - Age by Race/Ethnicity
  - Telephone Source

## Raking to Eight Margins



## Additional Raking Margins

- **States which include regions in the weighting use an additional four margins:**
  - Region (Margin 9)
  - Region by age (Margin 10)
  - Region by sex (Margin 11)
  - Region by race/ethnicity (Margin 12)

## Side investigation: limit weight variability

- **Ways to limit variability in the combined national weights**
- **Issue: very small states have sample sizes in the same ballpark as very large states leading to weight variability**
- **Possible approach: consider subsampling for states with relatively large sampling rates**
  - If warranted by large variability
- **Additional approach: trim national weights**
  - If warranted by large variability

## Limiting weight variability

- We developed optimal subsampling rates that maximize the effective sample size
- Comparison of approaches

### 1) Concatenated weights:

Total sample size= 6,713,925 and DEFF due to weighting= 3.96

Effective sample size of 1,694,766

2) Self-weighting subsample: 357,377

3) Optimal subsampling approach: Effective sample size = 474,062

Conclusion: subsampling reduces effective sample size, and not indicated

## Comparing national raking methods

- **Original BRFSS weights (concatenated)**
- **Reweight 1: National raking with 8 margins**
- **Reweight 2: National raking with 9 margins**
- **Reweight 3: National raking with 12 margins**
- **Reweight 4: National raking with 8 collapsed margins**
- **Reweight 5: National raking with 9 collapsed margins**
- **Reweight 6: National raking with 12 collapsed margins**

## National Raking Margins 1-8 (Similar to states)

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



## 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

## 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

## Collapsing Margins 6 and 7

- Collapsed Margin 7 (Race/Ethnicity by Age) to three age categories (18-34, 35-54, 55+) by four race/ethnicity categories
- Further collapsed Margins 6 (Sex by Race/Ethnicity) and 7 (Race/Ethnicity by Age) to achieve minimum sample size of 300 or minimum sample percentage of 5.0%

## 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+

## Collapsing Margins 10-12: National Raking by States

- Collapsed within region to achieve minimum sample size of 250 or minimum percentage of 5.0%
- Margin 10 (Age by State) collapsed 18-24 with 25-34 for 25 states
- Margin 11 (Sex by State) did not require any collapsing
- Margin 12 (Race/Ethnicity by State) collapsed all race categories for Maine, New Hampshire, and Vermont

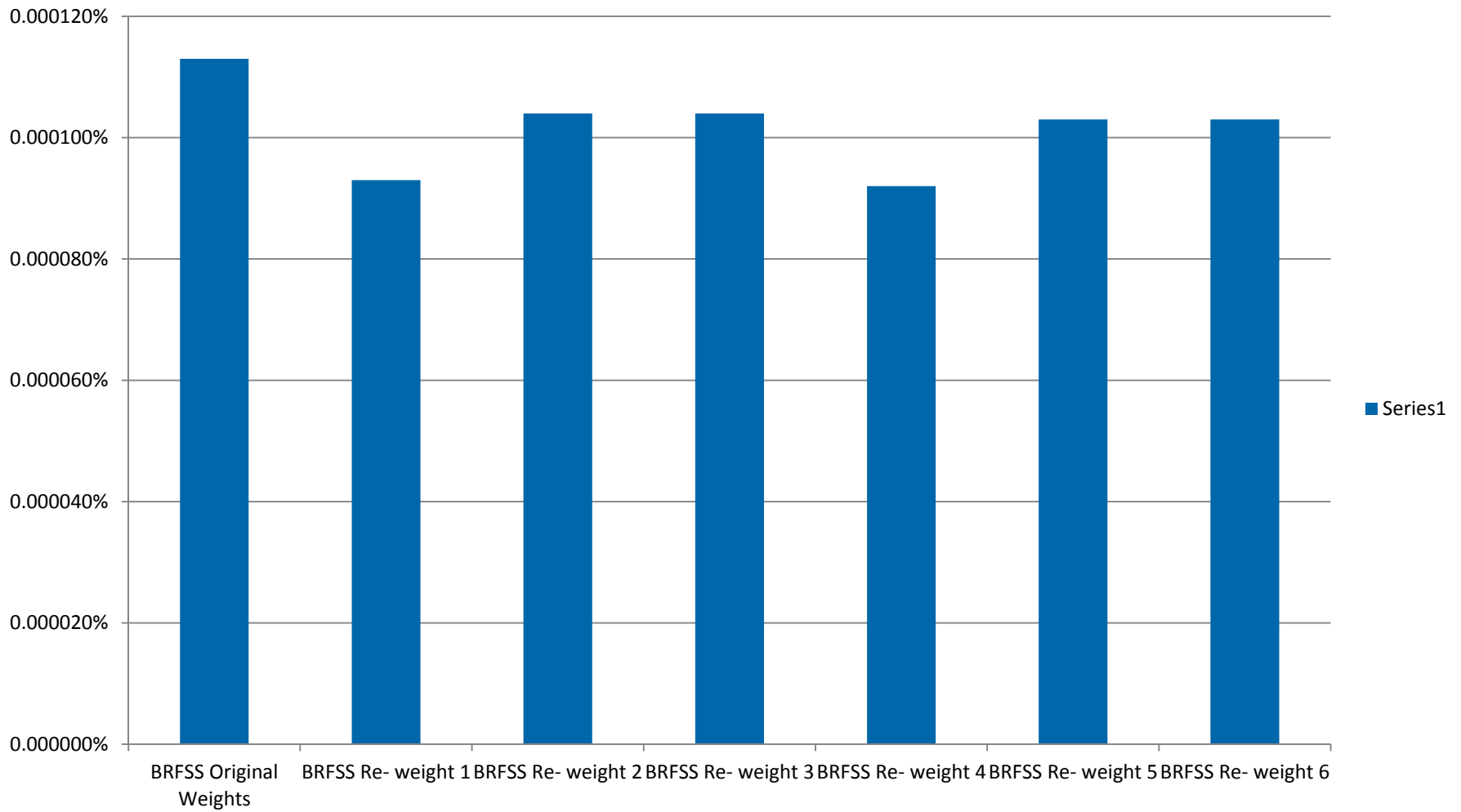
## Weight variability

Weight	CV	DEFF	Expected Margin of Error
Original BRFSS Weight	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%

## Bias Assessment: Comparisons with NHIS benchmark: Reweights #1 (8 margins), #2 (9 margins), #3 (12 margins)

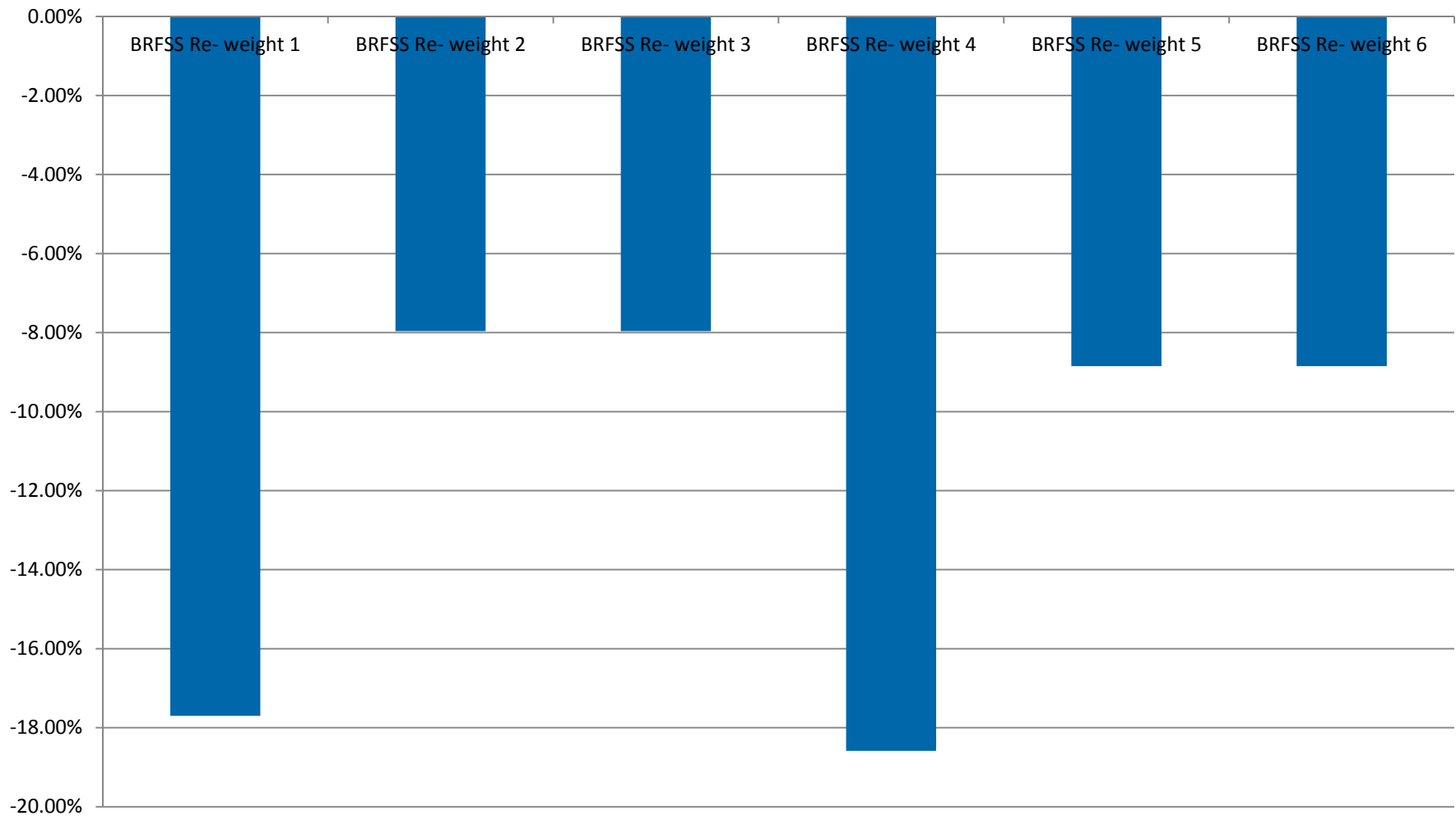
Indicator	NHIS Estimate	BRFSS Original Weights - NHIS	BRFSS Reweight 1 - NHIS	BRFSS Reweight 2 - NHIS	BRFSS Reweight 3 - NHIS	BRFSS Reweight 4 - NHIS	BRFSS Reweight 5 - NHIS	BRFSS Reweight 6 - NHIS
Current Smoker	18.06%	0.79%	1.15%	0.84%	0.89%	1.12%	0.80%	0.86%
Ever Told Had Diabetes	9.08%	1.04%	1.00%	0.96%	0.95%	1.00%	0.96%	0.95%
Ever Told Had Arthritis	22.08%	3.55%	3.61%	3.52%	3.49%	3.62%	3.53%	3.49%
Ever Told Had Asthma	12.63%	0.58%	0.48%	0.62%	0.63%	0.45%	0.59%	0.60%
Ever Told Had Heart Attack	3.26%	1.16%	1.19%	1.14%	1.14%	1.19%	1.14%	1.13%
Overweight or Obese	66.22%	0.20%	0.65%	0.40%	0.42%	0.66%	0.40%	0.43%

# Average Variance





# Average Change in Variance Relative to Stacked Weights



## Conclusions

- **Methods are nearly equivalent in terms of estimated bias**
- **Methods with 8 margins have better variance (lower variability in weights)**
- **Potential gains in bias reduction from deeper raking—involved in reweighting methods #2, #3, #5, #6—do not materialize, do not justify (or outweigh) added variance**
- **All methods seem effective for national BRFSS weights**

## Next steps

- **Look at the mean squared error (MSE)**
- **Investigate ways to estimate the variance of national weighted estimates**
- **Investigate influence of specific states**
- **Investigate weight trimming to limit overall variances of weighted national estimates**
- **Compare state-level estimates to published CDC estimates**

**Questions?**  
**Thank you!**  
**Contact information:**  
**Kristie.Healey@ICFI.Com**

