

# Weighting the BRFSS Data

July 2020



## Weighting the BRFSS Data

When data are unweighted, each record counts the same as any other record. Unweighted data analyses are based on the assumption that each record has an equal probability of being selected and that noncoverage and nonresponse are equal among all segments of the population. When deviations from these assumptions are large enough to affect the results from a data set, weighting each record appropriately can help to adjust for assumption violations. In the BRFSS, such weighting serves as a blanket adjustment for noncoverage and nonresponse and forces the total number of cases to equal population estimates for each geographic region, which for the BRFSS sums to the state population. Regardless of state sample design, use of the final weight in analysis is necessary if users are to make generalizations from the sample to the population.

Following is a general description of the 2020 BRFSS weighting process. Where a factor does not apply, processors set its value to one for calculation. In order to reduce bias due to unequal probability of selection, design weighting is conducted. The BRFSS also uses iterative proportional fitting, or “raking,” to adjust for demographic differences between those persons who are sampled and the population that they represent. The weighting methodology is therefore comprised of two sections: design weight and raking.

Design weights are calculated using the weight of each geographic stratum (`_STRWT`), the number of landline phones within a household (`NUMPHON3`), and the number of adults who use those phones (`NUMADULT`). For cellphone respondents, both `NUMPHON3` and `NUMADULT` are set to 1. The formula for the design weight is:

$$\text{Design Weight} = \_STRWT * (1/NUMPHON3) * NUMADULT$$

In 2020, the inclusion of cellular telephone respondents who also have landline telephones in their residence and landline telephone respondents who also have a cellular telephone in their residence required an adjustment to the design weights to account for the overlapping sample frames. From each of the two sample frames, a compositing factor was calculated for the telephone dual sampling frame users. BRFSS multiplied the design weight by the compositing factor to generate a composite weight for the records in the overlapping sample frames as described in the section below. BRFSS then truncated the design weight within a geographic region using the mean plus/minus 1.96 times the standard deviation to calculate the truncation limits, which processors used as the raking input weight.

The stratum weight (`_STRWT`) accounts for differences in the probability of selection among strata (subsets of area code/prefix combinations). It is the inverse of the sampling fraction of each stratum.

There is rarely a complete correspondence between strata, defined by subsets of area code/prefix combinations, and regions, defined by the boundaries of government entities.

BRFSS calculates the stratum weight (**\_STRWT**) using the following items:

- Number of available records (NRECSTR) and the number of records users select (NRECSEL) within each geographic stratum and density stratum.
- Geographic strata (GEOSTR), which may be the entire state or a geographic subset such as counties, census tracts, etc.
- Density strata (**\_DENSTR**) indicating the density of the phone numbers for a given block of numbers as listed or not listed.

Within each **\_GEOSTR\*\_DENSTR** combination, BRFSS calculates the stratum weight (**\_STRWT**) from the average of the NRECSTR and the sum of all sample records used to produce the NRECSEL. The stratum weight is equal to  $NRECSTR / NRECSEL$ .

**1/ NUMPHON3** The inverse of the number of residential landline telephone numbers in the respondent's household.

**NUMADULT** The number of adults 18 years and older in the respondent's household.

**FINAL WEIGHT** BRFSS rakes the design weight to 8 margins (gender by age group, race/ethnicity, education, marital status, tenure, gender by race/ethnicity, age group by race/ethnicity, and phone ownership). If BRFSS includes geographic regions, it includes four additional margins (region, region by age group, region by gender, and region by race/ethnicity). If at least one county has 500 or more respondents, BRFSS includes four additional margins (county, county by age group, county by gender, and county by race/ethnicity).

**\_LLCPWT** The final weight assigned to each respondent.

BRFSS uses weight trimming to increase the value of extremely low weights and decrease the value of extremely high weights. The objective of weight trimming is to reduce errors in the outcome estimates caused by unusually high or low weights in some categories.

**2020 design weight correction for overlapping sample frame:**

The partial overlapping sample frames required an adjustment to address the respondent's probability of selection in both the landline sample frame and cell phone sample frame. The adjustment to the

design weights was made to records identified as available in both sample frames. Three possible telephone source contact categories were included for this adjustment:

- > 1 Land Line frame with a Cell Phone
- > 2 Cell Phone frame with a Land Line
- > 9 No Dual Phone Use

The adjustment to the design weight included the records identified as a landline sample record with a cell phone or cell phone sample record with a landline. A compositing factor was calculated for the overlapping sample frame users. The compositing factors were based on the effective sample size. The following descriptions apply to the overlapping sample frame telephone service categories compositing factor calculation:

$N_{\text{effective}} = N / DEFF$ , Where: N is the unweighted number of interviews, and

$DEFF = 1 + (\text{Standard deviation of design wt} / \text{Mean value of design wt})^2$ .

\_DUALUSE is the variable used to identify the Dual Phone use categories (\_DUALUSE = 1 Land Line with a Cell Phone, \_DUALUSE = 2 Cell Phone with a Land Line, \_DUALUSE = 9 No Dual Phone Use )

For the \_DUALUSE category 1 (Land Line with a Cell Phone) calculate the composite weight:

$\text{Composite\_wt} = \text{DESIGN\_WT} \times (\text{N effective value for category 1} / (\text{N effective value for category 1} + \text{N effective value for category 2}))$ .

For the \_DUALUSE category 2 (Cell Phone with a Landline) calculation:  $\text{Composite\_wt} =$

$\text{DESIGN\_WT} \times (\text{N effective value for category 2} / (\text{N effective value for category 1} + \text{N effective value for category 2}))$ .

The corresponding SAS code is similar to: If \_DUALUSE = 1 or \_DUALUSE = 2 then \_WT2RAKE\_C = \_WT2RAKE \* \_DUALCOR Else \_WT2RAKE\_C = \_WT2RAKE.

Where \_WT2RAKE is the design weight and \_DUALCOR is the composite factor calculated to adjust the design weight for the records collected from overlapping sample frames.

### **2020 design weight truncation:**

The design weight calculation is implemented separately for the landline sample (within \_GEOSTR) and the cell phone sample (within \_GEOSTR). In addition to the overlapping sample frame correction to the design weight, the combined landline and cell phone design weight has been truncated within \_REGION prior to raking. The primary purpose of the design weight truncation is to prevent any adults in a state from carrying extremely large weights into the raking. A secondary goal is to prevent any adults from having extremely small design weights (i.e. the responses should not completely disappear at this point).

The combined landline and cell phone samples within \_REGION are truncated using the mean plus/minus 1.96 times the standard deviation to calculate the truncation limits. \_LLCPWT2 holds the truncated design weight.

The child design weights have not been truncated prior to raking.

### **2020 integrated weight:**

The 2020 integrated weight includes the nine state level margins and allows up to eight additional margins to take advantage of additional adjustments to sub-state populations within the raking. There are four additional margins if a county has at least 500 interviews available. There are also four additional margins for \_REGION, if multiple regions have been defined for a state and each region has at least 500 interviews.

### **Order of margins:**

Ideally, convergence would be obtained quickly and all margins would achieve agreement with specified population control totals. In practice, however, given the complexity of this weighting system, this may not be feasible for all 16 margins. Thus, certain margins should match population control totals exactly (e.g., age\*gender and age\*race/ethnicity), and in the few difficult cases where the raking algorithm has not completely converged or has reached the point of diminishing returns, the algorithm may stop without matching a few of the margins exactly. The last margin will achieve exact agreement with the population control totals. Margins close to the last margin will almost always be very close to the population control totals.

For the 2020 integrated weight, the key state-level margin, age by gender, is last and other key state-level margins are included close to the last margin. The order is shown below.

*FIRST MARGIN:* sixteenth margin (county by sex)

Fifteenth margin (county by agecat7)

Fourteenth margin (county by race6cat)

Thirteenth margin (county, no collapsing)

Twelfth margin (region by race6cat)

Eleventh margin (region by sex)

Tenth margin (region by agecat7)

Ninth margin (region, no collapsing)

Eighth margin (telephone service)

Seventh margin (age3 by race6cat)

Sixth margin (sex by race6cat)

Fifth margin (own/rent)

Fourth margin (marital, no collapsing)

Third margin (education, no collapsing)

Second margin (race6cat)

*LAST MARGIN*: first margin (sex by agecat7, no collapsing)

### **Population estimates:**

The population estimates obtained for building the target totals are from similar sources used in previous years. Intercensal population estimates were purchased from Claritas, LLC, at the county-level for age, race/ethnicity, and gender. These population estimates are used as the population totals for a state across all margins. The five-year American Community Survey PUMS data set (2015–2019) was used to obtain estimates for margins 3, 4, and 5 (education, marital status, tenure). The non-institutionalized adults were weighted by the person-level weights to generate the population estimates. The percentages were then used in the raking margins. The telephone type estimates for margin 8 were taken from the state wireless estimate percentages produced by NCHS and released in [December, 2019](#)

### **Calculation of a Child Weight**

BRFSS calculates the design weight for child weighting from the stratum weight times the inverse of the number of telephones in the household and then multiplies by the number of children:

Child Design Weight =  $\_STRWT * (1/NUMPHON3) * CHILDREN$

CHIILDWT = BRFSS rakes the child design weight to 5 margins: age by gender, race/ethnicity, gender by race/ethnicity, age by race/ethnicity, and phone ownership.

$\_CLLCPWT$  is the weight assigned for each child interview.