

Complex Sampling Weights  
and  
Preparing 2022 BRFSS Module Data for Analysis

July 2023



## Overview

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey that collects data on a number of health outcomes, health-related risk behaviors, use of preventive services, and chronic conditions from noninstitutionalized adults who reside in each of the states and participating US territories. In 2011, BRFSS changed its data collection procedures, structure, and weighting methodology so the traditional landline telephone-based data set could include data from participants using cell phones. The BRFSS uses a core set of questions and gives states the option to include multiple modules that focus on specific health issues. Not all optional modules are collected by all of the states, and states may opt to collect module data from just subsets of their survey participants. Core questions are used during all interviews. Researchers using BRFSS data should conduct analyses with complex sampling procedures; they also should appropriately stratify and weight the data in their work. Potential bias resulting from selection probabilities and noncoverage among segments of the population can be reduced through weighting.

Researchers conducting analysis of variables from the core-only section should use the variable `_LLCPWT` for weighting.

BRFSS has created this document to guide users analyzing variables from 2022 BRFSS optional modules and/or combinations of module and core variables. The 2022 BRFSS data sets include data that respondents provided by landline telephone or cell phone. Data users should note that newer weighting procedures are likely to affect trend lines when comparing BRFSS data collected up to 2010 with data from 2011 and afterward; because of these changes, users are advised NOT to make direct comparisons with pre-2011 data, and instead, begin new trend lines with that year.

Data users should become familiar with the information presented in this document prior to performing analyses. Information about the changes to 2022 BRFSS data collection is available in the [2022 Data Comparability Report](#).

## Weighting BRFSS Core Data

BRFSS has grouped the 2022 data into four data sets:

- 2022 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)

*And three versions*

- 2022 BRFSS Combined Landline Telephone and Cellular Telephone Version 1
- 2022 BRFSS Combined Landline Telephone and Cellular Telephone Version 2
- 2022 BRFSS Combined Landline Telephone and Cellular Telephone Version 3

Researchers should understand that BRFSS data were collected by asking core and/or module questions on the annual questionnaire. Users, therefore, may need to use data from up to 4 different data sets, described below.

- Use the Combined Landline Telephone and Cell phone data if questions are exclusively from the core section *or* if questions are in common modules that were asked on both the landline survey and cellphone survey (*Note: See the section below, Using BRFSS Data from Multiple Data Sets, for more about common modules.*)
- Use Version 1, Version 2, and/or Version 3 Combined Landline Telephone and Cell Phone data sets if states collected multiple version questionnaires.

In all cases, researchers conducting complex sampling analyses from the *core-only* section should use the variable `_LLCPWT` for weighting, `_STSTR` for stratification, and the variable `_PSU` for primary sampling unit, or clustering for short. The following table displays the description of the data, names of the data sets, and the variable names of the final weight.

Data description	Data set name	Final weight variable name
Combined Landline Telephone and Cellular Telephone	LLCP2022	_LLCPWT
Combined Landline Telephone and Cellular Telephone Version 1	LLCP22V1	_LCPWTV1
Combined Landline Telephone and Cellular Telephone Version 2	LLCP22V2	_LCPWTV2
Combined Landline Telephone and Cellular Telephone Version 3	LLCP22V3	_LCPWTV3

The examples below demonstrate how to use `_LLCPWT`, `_STSTR` & `_PSU` when analyzing variable **HAVARTH4** (*Ever told you had some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia*) and **PHYSHLTH** (*how many days during the past 30 days was your physical health not good?*) in the core section using the 2022 Louisiana data set. Examples are given in both SAS and R.

#### Example in SAS:

```
LIBNAME IN 'BRFSS';           /* Use LIBNAME to set the location of the BRFSS data set */
```

```
DATA LA22;                   /* Create a temporary data set for Louisiana */
SET IN.LLCP2022 (WHERE=( _STATE = 22));
RUN;
```

```
PROC SURVEYFREQ             /* Calculate percentage of population with arthritis */
DATA = LA22;
WEIGHT _LLCPWT;            /* Use _LLCPWT for weighted analysis*/
STRATA _STSTR;
CLUSTER _PSU;
TABLE HAVARTH4;
RUN;
```

```
PROC SURVEYMEANS           /* Calculate average number of physical healthy days */
DATA = LA22;
WEIGHT _LLCPWT;           /* Use _LLCPWT for weighted analysis */
```

```
STRATA _STSTR;
CLUSTER _PSU;
VAR PHYSHLTH;
RUN;
```

### Example in R:

```
# Install the survey package if it is not already installed
install.packages("survey")

# Call the library for the current R session
library(survey)

# Read in BRFSS data
load("\BRFSS\BRFSS.rdata")

# Subset the data for Louisiana(22)
BRFSS <- BRFSS[BRFSS$state == 22, ]

# Set options for allowing a single observation per stratum
options(survey.lonely.psu = "adjust")

# Create survey design
brfssdsn <- svydesign(
  id=~1,
  strata = ~ststr,
  weights = ~llcpwt,
  data = BRFSS)

# calculate average number of physical healthy days
svymean(~physhlth,
  brfssdsn,
  na.rm = TRUE)

# calculate percent in each arthritis category
svymean(~factor(havarth5),
  brfssdsn,
  na.rm = TRUE)
```

## Using BRFSS Data from Multiple Data Sets

Due to the complex nature of state-based data collection processes, users may have to create a data set that fits their research needs. The following two examples illustrate how to prepare module data for analysis when states have collected module data in a variety of ways.

Prior to conducting analyses using optional module data, users should always review all relevant documents for the given data year, such as the [Overview](#), [Codebook](#), [Comparability of Data](#), and [Modules Weights](#) in particular to identify states that collected data of interest and determine which questionnaire version states used. Remember to check FIPS state codes in each data set to avoid duplication.

Individual states may have chosen to use a number of optional modules, depending on each state's needs. Individual states may have also chosen to divide their samples and use different modules in the

subsamples that were distinguished by version of the surveys. Modules that appeared in each version of a state's questionnaire are called common modules. Each state that collected common modules did so on all its versions of combined landline telephone and cell phone.

## Example 1: Social Determinants and Health Equity

We will use the module **Social Determinants and Health Equity** to demonstrate how to combine and reweight data from multiple data sets for analysis. This module was selected because states collected the data either as a common module in the combined landline telephone and cell phone or by splitting the sample and offering the module on one or more versions of the data collection. For ease in using the 2022 optional modules, please refer to the two documents published with the annual data set: [Module weights for 2022](#) and [States by Module](#). In addition to showing which states used which modules, these tables list names and descriptions of the data sets, questionnaire versions, and weight variables from the corresponding data sets. If users want to analyze variables from certain modules, these two documents are a good place to start. The following information has been copied from the two module documents to show segments of the tables:

### Module weights for 2022

#### Modules by State

State	Description	Dataset	Data weight	Module(s)
Michigan	Combined Land Line and Cell Phone data	LLCP2022	_LLCPWT	Alcohol Screening & Brief Intervention (ASBI), Cancer Survivorship: Pain Management, Cancer Survivorship: Type of Cancer, Childhood Asthma Prevalence, Family Planning, Industry and Occupation, Pre-Diabetes, Random Child Selection, Sexual Orientation and Gender Identity (SOGI)
	Combined Land Line and Cell Phone data, version 1	LLCP22V1	_LCPWTV1	Cognitive Decline, Social Determinants and Health Equity
	Combined Land Line and Cell Phone data, version 2	LLCP22V2	_LCPWTV2	Reactions to Race, Social Determinants and Health Equity
	Combined Land Line and Cell Phone data, version 3	LLCP22V3	_LCPWTV3	Marijuana Use, Reactions to Race

#### States by Module

Module	Description	Dataset	Data weight	State(s)
<b>Social Determinants and Health Equity</b>	Combined Land Line and Cell Phone data	LLCP2022	_LLCPWT	Alabama, Alaska, Arizona, California, Connecticut, Delaware, District of Columbia, Florida, Georgia, Idaho, Indiana, Iowa, Kansas, Kentucky, Maine, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, Puerto Rico, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Virgin Islands, Washington, West Virginia, Wisconsin, Wyoming
	Combined Land Line and Cell Phone data, version 1	LLCP22V1	_LCPWTV1	Maryland, Michigan, Nebraska, Ohio

	Combined Land Line and Cell Phone data, version 2	LLCP22V2	_LCPWTV2	Maryland, Michigan, Oklahoma
--	---	----------	----------	------------------------------

1) The module **Social Determinants and Health Equity** listed in the 2022 document **States by Module** shows that 37 states used the common version of the module, 4 states used version 1, and 3 states used version 2 of the module. To prepare the data for analysis, we need to:

a. Extract data from LLCP2022 for the 37 states that collected the common version of the module:

*Alabama (1), Alaska (2), Arizona (4), California (6), Connecticut (9), Delaware (10), District of Columbia (11), Florida (12), Georgia (13), Idaho (16), Indiana (18), Iowa (19), Kansas (20), Kentucky (21), Maine (23), Massachusetts (25), Minnesota (27), Mississippi (28), Missouri (29), Montana (30), Nevada (32), New Hampshire (33), New Jersey (34), New Mexico (35), North Carolina (37), Rhode Island (44), South Carolina (45), Tennessee (47), Texas (48), Utah (49), Vermont (50), Washington (53), West Virginia (54), Wisconsin (55), Wyoming (56), Puerto Rico (72), Virgin Islands (78)*

b. Extract data from LLCP22V1 for the 4 states that collected version 1 of the module, *Maryland (24), Michigan (26), Nebraska (31) and Ohio (39)*

c. And extract data from LLCP22V2 for the 3 states that collected version 2 of the module, *Maryland (24), Michigan (26) and Oklahoma (40)*

## 2) Obtain data sets:

- To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2022):
  - Go to the [2022 data webpage](#), choose the 2022 data. From there, read all of the 2022 documents including the [overview](#), [codebook](#), and the [module list](#). The data set [LLCP2022](#) is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS datasets are also available in **SAS Resources** section of the page.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 data set:
  - Go to the [2022 data webpage](#), then find [The Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data](#). There are 3 separate SAS data sets corresponding to Questionnaire Version 1, 2, and 3. Choose Version 1: [LLCP22V1](#).

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 data set:

➤ Follow the same step above for Version 1, but instead, choose Version 2: [LLCP22V2](#).

### 3) Generate a working data set:

Create a new and uniform final weight variable from each version of the data sets downloaded in steps above, and combine data into one working data set. This step will require users to do the following:

- keep all the states that collected the module data in each version of the data sets,
- rename the corresponding weight variable to a new and consistent weight variable,
- combine all the data sets into one that contains the renamed weight variable.

Maryland and Michigan both used version 1 and version 2 of the module; we need to manually adjust the Weight variable for each version of their data following these steps:

#### Maryland:

- 1) Find the sample size (n1) for Maryland in V1 data LLCP22V1, which is 5328
- 2) Find the sample size (n2) for Maryland in V2 data LLCP22V2, which is 5028
- 3) Combine sample sizes for LLCPV1 and LLCPV2 for Maryland,  $(n1 + n2) = (5328+5308) = 10636$
- 4) Use the sample size from each version divided by the combined sample size to get proportions; therefore,
  - the proportion for version 1 data is  $P1 = n1/(n1+n2) = 5328/10636 \approx 0.50$
  - the proportion for version 2 data is  $P2 = n2/(n1+n2) = 5308/10636 \approx 0.50$

Michigan: follow the same steps:

- 1) Sample size for version 1,  $n1 = 3236$
- 2) Sample size for version 2,  $n2 = 3232$
- 3) Combined sample size for version 1 and version 2,  $(n1 + n2) = (3236 + 3232) = 6468$
- 4) the proportion for version 1 data is  $P1 = n1/(n1+n2) = 3236/6468 \approx 0.50$   
the proportion for version 2 data is  $P2 = n2/(n1+n2) = 3232/6468 \approx 0.50$

As we can see, the sample sizes are similar for version 1 and version 2, 5328 vs 5028, for Maryland, and 3236 vs. 3232 for Michigan, we can simply divide the \_FINALWT by 2. However, if the sample sizes are quite different, we should use the calculated proportions instead.

### Example in SAS:

```
LIBNAME IN 'BRFSS';
```

```
DATA LLCP; /* Extract data for states that used common version of the module*/
SET IN.LLCP2022 (WHERE=( _STATE IN
(1,2,4,6,9,10,11,12,13,16,18,19,20,21,23,25,27,28,29,30,32,33,34,35,37,44,45,47,48,49,50,53,54,55,56,72,78)));
```

```

RENAME _LLCPWT=_FINALWT;
RUN;

/* Extract data for states that used V1 of the module, adjust weights for states that used > 1 version of the module */
DATA LLCPV1;
SET IN.LLCP22V1 (WHERE=( _STATE IN (24 26 31 39)));
IF _STATE in (24 26) THEN FINALWT= (_LCPWTV1*.50); * _LCPWT1 times the calculated proportion;
ELSE FINALWT= _LCPWTV1;
DROP _LCPWTV1;
RUN;

/* Extract data for states that used V2 of the module, adjust weights for states that used > 1 version of the module */
DATA LLCPV2;
SET IN.LLCP21V2(WHERE=( _STATE IN (24 26 40)));
IF _STATE IN (24 26) THEN FINALWT= (_LCPWTV2*.50); * _LCPWTV2 times the calculated proportion;
ELSE FINALWT= _LCPWTV2;
DROP _LCPWTV2;
RUN;

DATA SDOH2022; *combine all versions of the dataset into one;
SET LLCP LLCPV1 LLCPV2;
RUN;

```

The combined and reweighted dataset SDOH2022 is ready for analysis.

## Example in R:

```

# Extract data from LLCP2022 for states that used the common version of the module

llcp <- llcp2022[llcp2022$state %in%
c(1,2,4,6,9,10,11,12,13,16,18,19,20,21,23,25,27,28,29,30,32,33,34,35,37,44,45,47,48,49,50,53,54,55,56,72,78), ]

llcp$finalwt <- llcp$llcpwt # Rename weight variable LLCPWT to FINALWT for the main data set

# Extract data for states used version 1 of the module, and adjust weight since they used both v1 & v2 of the module
llcpv1mdmi <- llcp22v1[llcp22v1$state %in% c(24,26), ]
llcpv1mdmi$finalwt <- llcpv1mdmi$lcpwtv1* (0.50)

# Extract data for states used only version 1 of the module and rename the LCPWTV1 to FINALWT
llcpv1neoh <- llcp22v1[llcp22v1$state %in% c(31,39), ]
llcpv1neoh$finalwt <- llcpv1neoh$lcpwtv1

#V1 - combine V1datasets
llcpv1<- rbind(llcpv1mdmi, llcpv1neoh)

# V2 - Extract data for states used version 2 of the module, adjust weight since they used both v1 & v2 of the module
llcpv2mdmi <- llcp22v2[llcp22v2$state %in% c(24,26), ]
llcpv2mdmi$finalwt <- llcpv2mdmi$lcpwtv2* (0.50)

# V2 – Extract data for state used v2 only and rename LCPWTV2 to FINALWT
llcpv2ok<- llcp22v2[llcp22v2$state %in% c(40), ]
llcpv2ok$finalwt <- llcpv2ok$lcpwtv2

# V2 – Combine V2 datasets
llcpv2<- rbind(llcpv2mdmi, llcpv2ok)

# Combine all data sets for analysis

```



```
sdoh2022 <- rbind(llcp, llcpv1, llcpv2)
```

The combined and reweighted dataset `sdoh2022` is ready for analysis.

## Example 2: Reactions to Race

The example below uses the Reactions to Race Module to demonstrate how to combine and reweight data from multiple data sets.

1) The module **Reactions to Race** listed in the document [States by Module](#) shows that there are total of 29 states used it as optional module in 2022.

- 25 states used the common version of this module; therefore, we need to extract data for these states from the dataset `LLCP2022`:

*California (6), District of Columbia (11), Georgia (13), Illinois (17), Indiana (18), Kansas (20), Kentucky (21), Louisiana (22), Maine (23), Maryland (24), Minnesota (27), Montana (30), Nevada (32), New Mexico (35), North Carolina (37), North Dakota (38), Ohio (39), Rhode Island (44), South Carolina (45), Tennessee (47), Vermont (50), Virginia (51), West Virginia (54), Wisconsin (55), Wyoming (56)*

- *Nebraska (31) and New York (36)* used version 1 of the module; therefore, we need to extract data from `LLCP22V1` for these 2 states.
- *Michigan (26), New York (36) and Oklahoma (40)* used version 2 of the module and therefore, we need to extract data from `LLCP22V2` for these 3 states.
- *Michigan (26)* also used version 3 of the module and we need to extract data from `LLCP22V3` for Michigan.

Note: New York used both version 1 & version 2 of the Reaction to Race module, and Michigan used version 2 and version 3 of the module. The weight variable will need to be adjusted accordingly for New York and Michigan before we combine and analyze the data. We shall follow the same steps for weight adjustment as we did in example 1.

## 2) Obtain data sets:

- To download the Combined Landline Telephone and Cellular Telephone data set (`LLCP2022`):
  - Go to the [2022 data webpage](#), and choose the 2022 data set. From there, read all of the 2022 documents including the [Overview](#), [Codebook](#), and the [States by Module](#). The data set is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS datasets are also available in **SAS Resources** section of the page.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 dataset:
  - Go to the [2022 data webpage](#), select [BRFSS the Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data](#). There are 3 separate SAS data sets corresponding to Questionnaire Version 1, 2, and 3. Choose Version 1: [LLCP22V1](#).
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 data set:
  - Follow the same step above for Version 1, but instead, choose Version 2: [LLCP22V2](#).
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 3 data set:
  - Follow the same step above for Version 1, but instead, choose Version 3: [LLCP22V3](#).

### 3) Generate a working data set:

Create a new and uniform final weight variable from each of the data sets, combine data into one working data set. This step will require users to do the following:

- a. Keep all states that collected the module data in each of the data sets
- b. Rename the corresponding weight variable to a consistent weight variable and adjust the weight variable if any state uses more than 1 version of the module
- c. Combine all the data sets into one that contains the renamed weight variable. Details on how to adjust weight are shown below.

### 4) Weight adjustment:

Since New York used version 1 and version 2 of the Reactions to Race module, Michigan used version 2 and version 3 of the module, we should manually adjust the Weight for New York and Michigan data:

#### **New York:**

- 1) Find the sample size (n1) for New York from LLCP22V1, which is 5722
- 2) Find the sample size (n2) for New York from LLCP22V2, which is 5645
- 3) Combine sample sizes for version 1 and version 2 is  $(5722 + 5645) = 11367$
- 4) Use the sample size from each version divided by the combined sample size to get proportions, therefore,
  - the proportion for version 1 is  $P1 = n1/(n1 + n2) = 5722/11367 \approx 0.50$
  - the proportion for version 2 is  $P2 = n2/(n1 + n2) = 5645/11367 \approx 0.50$

Sample sizes from version 1 & version 2 are so similar for New York, we can simply adjust the FINALWT by dividing it by 2. However, if the sample sizes are very different, calculated proportions should be used for weight adjustment.

**Michigan** – follow the same steps:

- 1) Sample size for version 2,  $n_2 = 3232$
- 2) Sample size for version 3,  $n_3 = 3235$
- 3) Combined sample size for version 2 and version 3,  $(n_2 + n_3) = (3232+3235) = 6467$
- 4) the proportion for version 2 data is  $P_2 = n_2/(n_2+n_3) = 3232/6467 \approx 0.50$   
the proportion for version 3 data is  $P_3 = n_3/(n_2+n_3) = 3235/6467 \approx 0.50$

Sample sizes from version 2 & version 3 are so similar for Michigan, we can simply adjust the FINALWT by dividing it by 2.

### Example code in SAS:

```
LIBNAME IN 'BRFSS';
```

*\*Begin with the main dataset, LLC2022, where Reaction to Race module was used as a common module;*

```
DATA RR; *Extract data for 25 states that used common version of the module;
```

```
SET IN.LLCP2022 (WHERE=( STATE
```

```
IN (6,11,13,17,18,20,21,22,23,24,27,30,32,35,37,38,39,44,45,47,50,51,54,55,56)));
```

```
RENAME _LLCPWT=FINALWT; *Rename weight variable to be consistent across new data sets;
```

```
RUN;
```

```
DATA RRV1; *Extract V1 data ;
```

```
SET IN.LLCP22V1 (WHERE=( STATE IN (31,36)));
```

```
IF _STATE = 36 THEN FINALWT= _LCPWTV1 * 0.5; *Adjust _LCPWT since NY used > 1 version of the module;
```

```
ELSE FINALWT= _LCPWTV1;
```

```
RUN;
```

```
DATA RRV2; *Extract V2 data;
```

```
SET IN.LLCP22V2 (WHERE=( STATE IN (26,36,40)));
```

```
IF _STATE IN (26, 36) THEN FINALWT = _LCPWTV2 * 0.5; *Adjust weight variable for MI & NY;
```

```
RUN;
```

```
DATA RRV3; *Extract V3 data for MI;
```

```
SET IN.LLCP22V3 (WHERE=( STATE IN (26)));
```

```
FINALWT = _LCPWTV3 * 0.5; *Adjust weight variable for MI;
```

```
RUN;
```

```
DATA ReactionToRace; *Create combined dataset for analysis;
```

```
SET RR RRV1 RRV2 RRV3;
```

```
RUN;
```

The combined and reweighted SAS dataset, ReactionToRace, is ready to be analyzed.

### Example code in R:

*# Begin with the main data set LLC2022 for 25 states that used the common version of the Reaction To Race module*

```
llcp <- llcp2022[llcp2022$state %in% c(6,11,13,17,18,20,21,22,23,24,27,30,32,35,37,38,39,44,45,47,50,51,54,55,56), ]
```

*# Extract V1 data for NE & NY*

```
nrow(nelcpv1[llcp22v1$state %in% c(31), ])
```

```

nrow(nyllep1[llep2v1$state %in% c(36),])

# Rename LCPWTV1 to FINALWT for NE
nlep1$finalwt <- nlep1$lcpwtv1

# Adjust weight for NY version 1 data set
nylep1$finalwt <- nylep1$lcpwtv1 * (.50)

# Combine NE & NY version 1 data set
rrv1 <- rbind(nlep1, nylep1)

# Extract V2 data for MI, NY
nrow(minylep2[llep2v2$state %in% c(26,36),])

# Adjust final weight variable for version 2 data sets MI & NY
minylep2$finalwt <- minylep2$lcpwtv2 * (.50)

#Extract V2 data for OK
nrow(oklep2[llep2v2$state %in% c(40),])

# Rename LCPWTV2 to FINALWT for OK
oklep2$finalwt <- oklep2$lcpwtv2

# Combine V2 data sets
rrv2 <- rbind(minylep2, oklep2)

# Extract V3 data & adjust weight for MI
nrow(rrv3[llep2v3$state %in% c(26),])
rrv3$finalwt <- rrv3$lcpwtv3 * (.50)

# Combine all data sets and it's ready for analysis
reactiontrace <- rbind(llcp, rrv1, rrv2, rrv3)

```

## Combining Multiple Years of BRFSS Data

There are times we need to analyze data from multiple years, such as calculating the average prevalence of a variable in core sections for 2020 and 2022. In this case, we need to combine data from 2020 and 2022. The weight variable for 2020 and 2022 needs to be adjusted proportionally for the combined data set. There are two ways to adjust the weight variable. The sample size for each year helps to determine how we should adjust them. If the sample sizes are very similar, we can simply divide the original weight variable by 2. If they are different, we need to adjust the weight variable proportionally by following these steps:

- 1) Find out the sample sizes for 2020 and 2022
- 2) Sum up the total sample sizes for 2020 and 2022
- 3) Find the proportion for 2020 (use the sample size for 2020 divided by the combined sample size for 2020 & 2022)
- 4) Find the proportion for 2022 (Use the sample size for 2022 divided by the combined sample size for 2020 & 2022)
- 5) The adjusted weight for 2020 is  $(\_LLCPWT(2020) * \text{proportion for 2020})$
- 6) The adjusted weight for 2022 is  $(\_LLCPWT(2022) * \text{proportion for 2022})$

The example below uses the 2020 and 2022 data from New York to demonstrate how to combine and reweight data from multiple years.

The sample size for 2020 New York data is 14,232 and 17,800 for 2022. Since sample sizes are somewhat different, the weights need to be adjusted proportionally. Follow the steps listed above to get the adjusted weights for New York for 2020 & 2022.

- 1) The sample size is 14,769 for 2020, and 17,800 for 2022
- 2) The combined sample size is 32,569
- 3) The proportion for 2020 is  $14,769 / 32,569 \approx .45$
- 4) The proportion for 2022 is  $17,800 / 32,569 \approx .55$ ;
- 5) The adjusted weight for 2020 is (`_LLCPWT * .45`)
- 6) The adjusted weight for 2022 is (`_LLCPWT * .55`)

### Obtain data sets:

- To download the 2020 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2020):
  - Go to the [2020 data Webpage](#), and choose 2020 data set. From there, read all of the 2020 documents including the [Overview](#), [Codebook](#), and the [Modules By State](#). The data set is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS datasets are also available in **SAS Resources** section of the page.
- To download the 2022 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2022):
  - Go to the [2022 data Webpage](#), and choose 2022 data set. From there, read all of the 2022 documents including the [Overview](#), [Codebook](#), and the [Modules By State](#). The data set is available in ASCII and SAS transport formats. Record layout, formats, and SAS code are also available in **SAS Resources** section of the page.
- Generate a new and uniform final weight variable (the adjusted weight variables) from each of the data sets and combine the data into one working data set. This step will require users to:
  - Rename the corresponding weight variable to a consistent weight variable; and
  - Combine all data sets into one with the consistent weight variable.

Again, users should note that new weighting procedures are likely to affect trend lines when comparing BRFSS data collected before and after 2011; because of the changes, users are advised NOT to make direct comparisons with pre-2011 data, and instead, to begin new trend lines with that year.

### Example code in SAS:

```
Libname D20 "\BRFSS2020";
Libname D22 "\BRFSS2022";
```

```

DATA NY20;  /* Check sample size for each 2020 & 2022 */
  SET D20.LLCP2020(WHERE=(STATE=36));
  RUN;  /* There were 14769 observations read from the data set D20.LLCP2020, WHERE STATE=36*/

```

```

DATA NY22;
  SET D22.LLCP2022(WHERE=(STATE=36));
  RUN;  /*There were 17800 observations read from the data set D22.LLCP2022, WHERE STATE=36*/

```

```

DATA NY2YR;  /* Create a final dataset with the adjusted weight variable (FINALWT) */
  SET NY20 (IN=A)
      NY22 (IN=B);
  IF A THEN FINALWT = LLCPWT * .45;  * Use the calculated proportion to adjust LLCPWT;
  IF B THEN FINALWT = LLCPWT * .55;
  RUN;

```

The combined and reweighted dataset NY2YR is ready for analysis.

### Example code in R:

```

# Subset 2019 and 2022 data for New York
NY20 <- llcp2020[llcp2020$state == 36, ]
NY22 <- llcp2022[llcp2022$state == 36, ]

# Count number of observations in each year
nrow(NY20)
nrow(NY22)

# Create a final weight based on proportion of observations in each year
NY20$finalwt <- NY20$llcpwt * (.45)
NY22$finalwt <- NY22$llcpwt * (.55)

# Combine the data sets and reweight
NY2YR <- rbind(NY20, NY22)

```

The combined and reweighted dataset NY2YR is ready for analysis.

~~~ *END* ~~~