

# Complex Sampling Weights and Preparing 2021 BRFSS Module Data for Analysis

July 2022



## 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 2021 BRFSS optional modules and/or combinations of module and core variables. The 2021 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 2021 BRFSS data collection is available in the [2021 Data Comparability Report](#).

## Weighting BRFSS Core Data

BRFSS has grouped the 2021 data into four data sets:

- 2021 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)
 

*And three versions*

  - 2021 BRFSS Combined Landline Telephone and Cellular Telephone Version 1
  - 2021 BRFSS Combined Landline Telephone and Cellular Telephone Version 2
  - 2021 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	LLCP2021	_LLCPWT
Combined Landline Telephone and Cellular Telephone Version 1	LLCP21V1	_LCPWTV1
Combined Landline Telephone and Cellular Telephone Version 2	LLCP21V2	_LCPWTV2
Combined Landline Telephone and Cellular Telephone Version 3	LLCP21V3	_LCPWTV3

The examples below demonstrate how to use **\_LLCPWT**, **\_STSTR** & **\_PSU** when analyzing variable **HAVARTH5** (*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 2021 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 */

PROC SURVEYFREQ;
  DATA = IN.LLCP2021 (WHERE=(_STATE = 22));          /* Calculate percentage of population with arthritis */
  WEIGHT _LLCPWT;                                     /* Use _LLCPWT for weighted analysis */
  STRATA _STSTR;
  CLUSTER _PSU;
  TABLE HAVARTH5;
  RUN;

PROC SURVEymeans;
  DATA = IN.LLCP2021 (WHERE=(_STATE= 22));           /* Fips code for Louisiana is 22 */
  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
brfssdsgn <- svydesign(
  id=~1,
  strata = ~ststr,
  weights = ~llcpwt,
  data = BRFSS)

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

# calculate percent in each arthritis category
svymean(~factor(havarth5),
        brfssdsgn,
        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 by State](#) 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: Lung Cancer Screening

The example below uses the module **Lung Cancer Screening** 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 2021 optional modules, please refer to the two documents published with the annual data set: [Module weights for 2021](#) and [States by Module](#). In addition to showing which states used which modules, these tables list names and descriptions of the data sets, 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 2021

#### Modules by State

State	Description	Dataset	Data weight	Module(s)
<b>Maryland</b>	Combined Land Line and Cell Phone data	LLCP2021	_LLCPWT	Childhood Asthma Prevalence, Home/ Self-measured Blood Pressure, Industry and Occupation, Marijuana Use, Random Child Selection
	Combined Land Line and Cell Phone data, version 1	LLCP21V1	_LCPWTV1	Caregiver, Cognitive Decline, Tobacco Cessation
	Combined Land Line and Cell Phone data, version 2	LLCP21V2	_LCPWTV2	Lung Cancer Screening, Tobacco Cessation
	Combined Land Line and Cell Phone data, version 3	LLCP21V3	_LCPWTV3	Cancer Survivorship: Course of Treatment, Cancer Survivorship: Pain Management, Cancer Survivorship: Type of Cancer, Lung Cancer Screening

#### States by Module

Module	Description	Dataset	Data weight	State(s)
<b>Lung Cancer Screening</b>	Combined Land Line and Cell Phone data	LLCP2021	_LLCPWT	Maine, Michigan, New Jersey, Rhode Island
	Combined Land Line and Cell Phone data, version 1	LLCP21V1	_LCPWTV1	Oklahoma
	Combined Land Line and Cell Phone data, version 2	LLCP21V2	_LCPWTV2	Kansas, Maryland
	Combined Land Line and Cell Phone data, version 3	LLCP21V3	_LCPWTV3	Maryland

- 1) The module **Lung Cancer Screening** listed in the 2021 document [States by Module](#) shows that there are 7 states collected data on this module, 4 of them used the common version of the module, 1 state used version 1, 2 states used version 2 and 1 state used version 3 of the module.

See detailed information below:

- a. Four states collected this data as a common module; therefore, we need to extract data for these states from LLCP2021:

*Maine (23), Michigan (26), New Jersey (34), Rhode Island (44).*

- b. One state, *Oklahoma (40)*, used version 1 of the module. We therefore need to extract data from LLCP21V1 for Oklahoma.
- c. *Kansas (20) and Maryland (24)* used version 2 of the module, we will extract data from LLCP21V2 for Kansas and Maryland.
- d. *Maryland (24)* also used version 3 of the module, therefore, we need to extract data from LLCP21V3 for Maryland.

## 2) Obtain data sets:

- To download the Combined Landline Telephone and Cellular Telephone data set (LLCP2021):
  - Go to the [2021 data webpage](#), choose the 2021 data. From there, read all of the 2021 documents including the [overview](#), [codebook](#), and the [module list](#). The data set [LLCP2021](#) 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 [2021 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: [LLCP21V1](#).
- 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: [LLCP21V2](#).
- 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: [LLCP21V3](#)

## 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, 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 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.

Since Maryland used both version 2 and version 3 of the module, we need to manually adjust the Weight variable for each version of Maryland's data following these steps:

- 1) Find the sample size (n2) for Maryland in V2 data LLCP21V2, which is 5074
- 2) Find the sample size (n3) for Maryland in V3 data LLCP21V3, which is 5119
- 3) Combine sample sizes for LLCPV2 and LLCPV3 for Maryland,  
 $(n2 + n3) = (5074+5119) = 10193$
- 4) Use the sample size from each version divided by the combined sample size to get proportions, therefore,
  - the proportion for version 2 data is  $P2= n2/(n2+n3) = 5074/10193 \approx 0.50$
  - the proportion for version 3 data is  $P3= n3/(n2+n3) = 5119/10193 \approx 0.50$

As we can see, the sample sizes are similar (5074 vs 5119) for version 2 and version 3, 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\';

/* Extract data for states that used common module of Lung Cancer Screening:
   For Maine (23), Michigan (26), New Jersey (34), Rhode Island (44) */

DATA LLCP;
SET IN.LLCP2021 (WHERE=(_STATE IN (23,26,34,44)));
RENAME _LLCPWT=_FINALWT;
RUN;

DATA LLCPV1;           * Oklahoma(40) used version 1;
SET IN.LLCP21V1 (WHERE=(_STATE IN (40)));
RENAME _LCPWTV1=_FINALWT;
RUN;

/* Kansas (20) and Maryland (24) used v2 of the module, since Maryland used both v2 & v3, it's weight need to be adjusted */
DATA LLCPV2;
SET IN.LLCP21V2(WHERE=(_STATE IN (20,24)));
IF _STATE =24 THEN FINALWT=_LCPWTV2*.50;
                     *_LCPWTV2 times the calculated proportion or simply divided by 2;
ELSE FINALWT=_LCPWTV2;
DROP _LCPWTV2;
RUN;

DATA LLCPV3;
SET IN.LLCP21V3 (WHERE=(_STATE IN (24)));
FINALWT=_LCPWTV3*.50;      *_LCPWTV3 times the calculated proportion or simply divided by 2;
DROP _LCPWTV3;

```

**RUN;**

```
DATA LUNGCAANCER; *combine all versions of the dataset into one;
SET LLCP LLCPV1 LLCPV2 LLCPV3;
RUN;
```

The combined and reweighted dataset LUNGCAANCER is ready for analysis.

## Example in R:

```
# Begin with the main data set for states that used the Lung Cancer Screening as a common module.

llcp <- llcp2021[llcp2021$state %in% c(23,26,34,44), ]

# Rename the final weight variable for the main data set
llcp$finalwt <- llcp$llcpwt

# Subset data for states used version 1 of the module – Oklahoma(40)
llcpv1 <- llcp21V1[llcp21V1$state %in% c(40), ]

# Rename the final weight variable for the version 1 data set
llcpv1$finalwt <- llcpv1$lcpwtv1

# Subset data for states used version 2 of the module separately, Kansas (20)
llcpv2 <- llcp21V2[llcp21V2$state %in% c(20), ]

# Rename the final weight variable for Kansas version 2 data set
llcpv2$finalwt <- llcpv2$lcpwtv2

# subset version 2 of the module data for Maryland (24)
llcpv2md <- llcp21V2[llcp21V2$state %in% c(24), ]

# Adjust weight variable and rename the final weight variable for Maryland version 2 data set
llcpv2md$finalwt <- llcpv2md$lcpwtv2* (0.50)

# subset version 3 of the module data for Maryland
llcpv3md <- llcp21V3[llcp21V3$state %in% c(24), ]

# Adjust weight variable and rename the final weight variable for Maryland version 3 data set
llcpv3md$finalwt <- llcpv3md$lcpwtv3* (0.50)

# Combine v2 & v3 data sets for Maryland
mdlung <- rbind(llcpv2md, llcpv3md)

# Combine all data sets for analysis
lungcancer <- rbind(llcp, llcpv1, llcpv2, mdlung)
```

The combined and reweighted dataset lungcancer is ready for analysis.

## Example 2: Tobacco Cessation

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

- 1) The module **Tobacco Cessation** listed in the document [States by Module](#) shows that there are total of 23 states used it as optional module in 2021.

- 22 states used the common version of this module; therefore, we need to extract data for these 22 states from the dataset LLCP2021:

*Delaware(10), Georgia(13), Hawaii(15), Illinois(17), Indiana(18), Kentucky(21), Maine (23), Massachusetts(25), Minnesota(27), Mississippi(28), Missouri(29), Montana(30), Nebraska(31), Nevada(32), New Jersey(34), New York(36), Oregon(41), Rhode Island(44), Utah(49), Vermont, (50) Washington(53), West Virginia(54)*

- *Maryland (24)* state used version 1 of the module; therefore, we need to extract data from LLCP21V1 for Maryland.
- *Maryland (24)* also used version 2 of the module and therefore, we need to extract data from LLCP21V2 for Maryland.

Note: Maryland used both version 1 & version 2 of the Tobacco Cessation module for 2021. The weight variable will need to be adjusted 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 \(LLCP2021\):](#)
  - Go to the [2021data webpage](#), and choose the 2021 data set. From there, read all of the 2021 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 [2021data 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: [LLCP21V1](#).
- [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: [LLCP21V2](#).

## 3) Generate a working data set:

Create a new and uniform final weight variable from each of the data sets, combine data into a 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 Maryland used version 1 and version2 of the Tobacco Cessation module, here is how we should manually adjust the Weight for Maryland data:

- 1) Find the sample size (n1) for LLCPV1(Maryland), which is 5082
- 2) Find the sample size (n2) for LLCPV2(Maryland), which is 5074
- 3) Combine sample sizes for LLCPV1 and LLCPV2 for Maryland is  $(5082 + 5074) = 10156$
- 4) Use the sample size from each version divided by the combined sample size to get proportions, therefore,
 

the proportion for version 1 is  $P1 = n1/(n1+n2) = 5082/10156 \approx 0.50$

the proportion for version 2 is  $P2 = n2/(n1+n2) = 5074/10156 \approx 0.50$

Sample sizes from version 1 & version 2 are so similar for Maryland, 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.

#### Example code in SAS:

`LIBNAME IN "\BRFSS";`

*\*Begin with the main dataset, where **Tobacco Cessation** module was used as a common module in the combined landline and cellular telephone samples for 22 states.*

```

DATA LLCP;    * Extract data for 22 states used common version of the Questionnaire;
SET IN.LLCP2021 (WHERE=(_STATE
      in (10,13,15,17,18,21,23,25,27,28,29,30,31,32,34,36,41,44,49,50,53,54)));
RENAME _LLCPWT=FINALWT;      *Rename weight variable to be consistent across new data sets;
RUN;

DATA LLCPV1;          *Extract data for states that used Questionnaire V1;
SET IN.LLCP21V1 (WHERE=(_STATE IN (24));   *Maryland;
FINALWT= _LCPWT*0.5;           *Adjust weight variable;
RUN;

DATA LLCPV2;          *Extract data for states that used Questionnaire V2;
SET IN.LLCP21V2 (WHERE=(_STATE IN (24));   *Maryland;
FINALWT= _LCPWT*0.5;           *Adjust weight variable for Maryland;
RUN;

DATA Tobacco; *Combine all extracted datasets and ready for analysis;
SET LLCP LLCPV1 LLCPV2;
RUN;
```

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

### Example in R:

```
# Begin with the main data set for 22 states that used the common version of the Tobacco Cessation module
llcp <- llcp2021[llcp2021$state %in% c(10,13,15,17,18,21,23,25,27,28,29,30,31,32,34,36,41,44,49,50,53,54), ]

# Extract Questionnaire V1 data for Maryland
nrow(llcpv1md[llcp21v1$state %in% c(24), ])

# Adjust final weight variable for the version 1 MD data set
llcpv1md$finalwt <- llcpv1md$lcpwtv1 * (.50)

# Extract Questionnaire V2 data for Maryland
nrow(llcpv2md[llcp21v2$state %in% c(24), ])

# Adjust final weight variable for the version 2 MD data set
llcpv2md$finalwt <- llcpv2md$lcpwtv2 * (.50)

# Combine all data sets and it's ready for analysis
tobacco <- rbind(llcp, llcpv1md, llcpv2md)
```

## 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 2019 and 2021. In this case, we need to combine data from 2019 and 2021. The weight variable for 2019 and 2021 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 size for 2019 and 2021
- 2) Sum up the total sample size for 2019 and 2021
- 3) Find the proportion for 2019 (use the sample size for 2019 divided by the combined sample size for 2019 & 2021)
- 4) Find the proportion for 2021 (Use the sample size for 2021 divided by the combined sample size for 2019 & 2021)
- 5) The adjusted weight for 2019 is ( $\text{LLCPWT}(2019) * \text{proportion for 2019}$ )
- 6) The adjusted weight for 2021 is ( $\text{LLCPWT}(2021) * \text{proportion for 2021}$ )

The example below uses the 2019 and 2021 data from New York to demonstrate how to combine and reweight data on the core section from multiple years.

For 2019, the sample size for New York is 14,232. For 2021, the sample size is 39,089. The sample size for 2021 is much larger than that for 2019, therefore the weights need to be adjusted proportionally. Follow the steps listed above, we can find adjusted weights for New York for 2019 & 2021.

- 1) The sample size is 14,232 for 2019, and 39,089 for 2021
- 2) The combined sample size is 53,321

- 3) The proportion for 2019 is  $14,232/53,321 \approx .27$
- 4) The proportion for 2021 is  $39,089/53,321 \approx .73$ ;
- 5) The adjusted weight for 2019 is  $(\_LLCPWT * .27)$
- 6) The adjusted weight for 2021 is  $(\_LLCPWT * .73)$

### Obtain data sets:

- To download the 2019 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2019):
  - Go to the [2019 data Webpage](#), and choose 2019 data set. From there, read all of the 2019 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 2021 Combined Landline Telephone and Cellular Telephone Questionnaire data set (LLCP2021):
  - Go to the [2021 data Webpage](#), and choose 2021 data set. From there, read all of the 2021 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.

### Reweighting Example in SAS:

```

Libname D19 "BRFSS2019";
Libname D20 "BRFSS2021";

DATA NY19; /* Check sample size for each 2019 & 2021 */
  SET D19.LLCP2019(WHERE=(STATE=36));
  RUN; /* There were 14232 observations read from the data set D19.LLCP2019, WHERE _STATE=36*/

DATA NY21;
  SET D21.LLCP2021(WHERE=(STATE=36));
  RUN; /*There were 39089 observations read from the data set D21.LLCP2021, WHERE _STATE=36*/

DATA NY1921; /* Create a final dataset with the adjusted weight variable (FINALWT) */
  SET NY19 (IN=A)
        NY21 (IN=B);
  IF A THEN FINALWT = _LLCPWT * .27; * Use the calculated proportion to adjust the weight variable _LLCPWT;

```

```
IF B THEN FINALWT = _LLCPWT * .73;
RUN;
```

The combined and reweighted dataset NY1921 is ready for analysis.

### Reweighting Example in R:

```
# Subset 2019 and 2021 data for New York
NY19 <- llcp2019[llcp2019$state == 36, ]
NY21 <- llcp2021[llcp2021$state == 36, ]

# Count number of observations in each year
nrow(NY19)
nrow(NY21)

# Create a final weight based on proportion of observations in each year
NY19$finalwt <- NY19$llcpwt * (.27)
NY21$finalwt <- NY21$llcpwt * (.73)

# Combine the data sets and reweight
NY1921 <- rbind(NY19, NY21)
```

The combined and reweighted dataset NY1921 is ready for analysis.

~~~ **END**~~~