# Complex Sampling Weights
# and
# Preparing 2025 BRFSS Module Data for Analysis

August 2026

## 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 dataset 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 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 2025 BRFSS optional modules and/or combinations of module and core variables. The 2025 BRFSS datasets 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 2025 BRFSS data collection is available in the 2025 Data Comparability Report.

## Weighting BRFSS Core Data

BRFSS has grouped the 2025 data into four datasets:

- 2025 BRFSS Questionnaire data (Combined Landline Telephone and Cellular Telephone)

  *And three versions*

- 2025 BRFSS Combined Landline Telephone and Cellular Telephone Version 1
- 2025 BRFSS Combined Landline Telephone and Cellular Telephone Version 2
- 2025 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 datasets, 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 Datasets****, for more about common modules.)*

- Use Version 1, Version 2, and/or Version 3 Combined Landline Telephone and Cell Phone datasets 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 datasets, and the variable names of the final weight.

| Data description | Dataset name | Final weight variable name |
|---|---|---|
| Combined Landline Telephone and Cellular Telephone | LLCP2025 | _LLCPWT |
| Combined Landline Telephone and Cellular Telephone Version 1 | LLCP25V1 | _LCPWTV1 |
| Combined Landline Telephone and Cellular Telephone Version 2 | LLCP25V2 | _LCPWTV2 |
| Combined Landline Telephone and Cellular Telephone Version 3 | LLCP25V3 | _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 2025 Louisiana dataset. 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 LA25;                         /* Clean and create a temporary data set for Louisiana */
 SET IN.LLCP2025 (WHERE=(_STATE = 22));
 IF HAVARTH4 IN (7,9) THEN HAVARTH4=.;
 IF PHYSHLTH = 88 THEN PHYSHLTH =0;
  ELSE IF PHYSHLTH IN (77,99) THEN PHYSHLTH=.;

 PROC FREQ DATA=LA25;       * Check to make sure the variables are re-coded correctly;
 TABLE HAVARTH4 PHYSHLTH;
RUN;

PROC SURVEYFREQ DATA = LA25;          /* HAVARTH4, Calculate percentage of population with arthritis */
 WEIGHT _LLCPWT;                      /* Use _LLCPWT for weighted analysis*/
 STRATA _STSTR;
 CLUSTER _PSU;
 TABLE HAVARTH4;
RUN;

PROC SURVEYMEANS DATA =  LA25;        /* PHYSHLTH, Calculate average number of physical healthy days */
```

```
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(c("survey", "haven", "dplyr", "tidyverse"))

# Call the library for the current R session
library(survey)
library(haven)
library(dplyr)
library(tidyverse)

# Read in BRFSS data
LLCP2025 <- read_sas("LLCP2025.sas7bdat")

# Subset the data for Louisiana(22)
LA25 <- LLCP2025 %>%
 filter(STATE == 22) %>%
 mutate(
  HAVARTH4 = if_else(HAVARTH4 %in% c(7, 9), NA_real_, HAVARTH4),
  PHYSHLTH = case_when(
   PHYSHLTH == 88 ~ 0,
   PHYSHLTH %in% c(77, 99) ~ NA_real_,
   TRUE ~ PHYSHLTH
  )
 )

# 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
    nest = TRUE)

# calculate percent in each arthritis category
svymean(~factor(HAVARTH4), design = brfssdsgn, na.rm = TRUE)

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

## Using BRFSS Data from Multiple Datasets

Due to the complex nature of state-based data collection processes, users may have to create a dataset 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 Module 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 dataset 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 appear 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: HPV - Vaccination

We will use the module **HPV - Vaccination** to demonstrate how to combine and reweight data from multiple datasets 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 2025 optional modules, please refer to the two documents published with the annual dataset: BRFSS Modules Used by State and BRFSS Modules Used by Category. In addition to showing which states used which modules, these tables list names and descriptions of the datasets, questionnaire versions, and weight variables from the corresponding datasets. 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 2025

### States by Module

| Module | Description | Dataset | Data Weight | State(s) |
|---|---|---|---|---|
| **HPV - Vaccination** | Combined Land Line and Cell Phone data | LLCP2025 | _LLCPWT | Connecticut(9), Georgia(13), Hawaii(15), New Jersey(34), Rhode Island(44) & Virginia(51) |
| | Combined Land Line and Cell Phone data, version 1 | LLCP25V1 | _LCPWTV1 | Maryland(24), Massachusetts(25), New York(36) |
| | Combined Land Line and Cell Phone data, version 2 | LLCP25V2 | _LCPWTV2 | Maryland(24) |

### Modules by State

| State | Description | Dataset | Data weight | Module(s) |
|---|---|---|---|---|
| **Maryland** | Combined Land Line and Cell Phone data | LLCP2025 | _LLCPWT | Marijuana Use, Home/ Self-measured Blood Pressure, Sexual Orientation |
| | Combined Land Line and Cell Phone data, version 1 | LLCP25V1 | _LCPWTV1 | Firearm Safety, Cancer Survivorship: Type of Cancer, Cancer Survivorship: Course of Treatment, Cancer Survivorship: Pain Management, **HPV – Vaccination** |
| | Combined Land Line and Cell Phone data, version 2 | LLCP25V2 | _LCPWTV2 | Family Planning, Caregiver, Cognitive Decline, **HPV - Vaccination**, Social Determinants |

| | | | | |
|---|---|---|---|---|
| | Combined Land Line and Cell Phone data, version 3 | LLCP25V3 | _LCPWTV3 | Family Planning, Social Determinants |

1) The **HPV - Vaccination** module listed in the 2025 document **Modules by Category** shows: Common version of the HPV module is used by Connecticut (9), Georgia (13), Hawaii (15), New Jersey (34), Rhode Island (44) & Virginia (51), Version 1 of the HPV module is used by Maryland (24), Massachusetts (25) & New York (36) and Version 2 of the module is used by Maryland (24) only.

   *Note: numbers in parentheses are the FIPS codes for states.*

   We will use 2025 data to explain how to reweight data for HPV -Vaccination module analysis.

   a. Extract data from LLCP2025 for the 6 states that collected the common version of the HPV module, Connecticut (9), Hawaii (15), New Jersey (34), Rhode Island (44) & Virginia (51).

   b. Extract data from LLCP25V1 for Maryland (24), Massachusetts (25) & New York (36) that collected version 1 of the module

   c. Extract data from LLCP25V2 for Maryland (24) that used version 2 of the HPV module

2) **Obtain datasets**:

- To download the Combined Landline Telephone and Cellular Telephone dataset (LLCP2025):

  ➢ Go to the 2025 data webpage, choose 2025 data and review the 2025 documents including the overview, codebook, and the modules by category list provided here. The dataset LLCP2025 is available in ASCII and SAS transport formats. Record layout, formats, and SAS code that converts ASCII or SAS transportable file to SAS datasets also are available in the **SAS Resources** section of the page.

- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 dataset:

  ➢ Go to the 2025 data webpage, find The Combined Landline and Cellular Telephone Multiple Questionnaire Version Data including Optional Modules There are 3 sections, one for each version of the 2025 questionnaire datasets. Choose Version 1: LLCP25V1.

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

  ➢ Follow the same steps above and choose Version 2 data: LLCP25V2.

- Since version 3 of the HPV module was not used by any state in 2025, it is not necessary to download it. (Had you needed these data, you would have downloaded the Version 3 dataset of the Combined Landline Telephone and Cellular Telephone Questionnaire by following the same steps above, then choosing Version 3 data: LLCP25V3 .)

**3) Generate a working dataset**:

Create a new and uniform final weight variable from each version of the datasets downloaded in the steps above and combine data into one working dataset. This will require users to follow these steps:

a. Keep all states that collected the module data in each version of the datasets.
b. Rename the corresponding weight variable to a new and consistent weight variable.
c. Combine all the datasets into one that contains the renamed weight variable.

Since Maryland used versions 1 and 2 of the HPV module, we need to adjust the Weight variable manually for each version of its data, by following these steps:

***Maryland***:

1) Find sample size for version 1, n1 = 4311
2) Find sample size for version 2, n2 = 4216
3) Combined sample size for version 1 and version 2, (n1 + n2) = (4311 + 4216) = 8527
4) The proportion for version 1 data is **P1**= n1/(n1+n2) = 4311/8527 ≈ **0.51**
   and the proportion for version 2 data is **P2**= n2/(n1+n2) = 4216/8527 ≈ **0.49**

As we can see, the sample sizes for version 1 and version 2 of the 2025 HPV -Vaccination module for Maryland, are 4311 and 4216, we can use the proportion we calculated above .51 vs .49 or simply divide the FINALWT by 2 since the two sample sizes are very similar. But if the sample sizes were quite different, we would have to use the calculated proportions instead.

## Example in SAS:

```
LIBNAME IN  '\BRFSS\';

DATA HPVC;    /* Extract data for states that used common version of the module*/
SET IN.LLCP2025 (WHERE=(_STATE IN (9 13 15 34 44 51)));
RENAME _LLCPWT=FINALWT;
RUN;

/* Extract data for states that used V1 of the module,  Adjust weights for MD that used V1 & V2 of the module */
DATA HPV1;
  SET IN.LLCP25V1 (WHERE=(_STATE IN (24 25 36)));
  IF _STATE = 24 THEN FINALWT= (_LCPWTV1*.51); * LCPWT1 times the calculated proportion;
  ELSE FINALWT=_LCPWTV1;
  DROP _LCPWTV1;
RUN;

/* Extract V2 data for MD, Rename LCPWTV2 to FINALWT, weight adjustment needed */
DATA HPV2;
```

```
    SET IN.LLCP25V2(WHERE=(_STATE IN ( 24)));
    FINALWT=_LCPWTV2*.49;
    DROP _LCPWTV2;
RUN;

DATA HPV;  *combine all versions of the HOV datasets into one;
    SET   HPVC  HPV1  HPV2;
RUN;
```

The combined and reweighted dataset HPV is ready for analysis.

## Example in R:

```
    # Extract data from LLCP2025 for states that used the common version of the module
    # Note - R does not allow variable names to begin with underscores, Such as _LLCPWT, _LCPWTV1,
    _LCPWTV2 and _LCPWTV3, remove underscores before analysis

#read SAS dataset LLCP2025 into R
library(haven)
library(dplyr)

LLCP2025 <- read sas("LLCP2025.sas7bdat")

#Extract data for states that used common version of the module
HPVC <- LLCP2025 %>%
 filter(STATE %in% c(9, 13, 15, 34, 44, 51)) %>%
 rename(FINALWT = LLCPWT)

# HPV1: Extract data for states that used version 1 of the module
# FINALWT=LCPWT1 * the calculated proportion for MD (24), drop LCPWT1

HPV1 <- LLCP25V1 %>%
 filter(STATE %in% c(24, 25, 36)) %>%
mutate( FINALWT = if_else(STATE == 24, LCPWTV1 * 0.51, LCPWTV1) %>%
select(-LCPWTV1)

# HPV2: Extract V2 data, FINALWT= LCPWTV2* the calculated proportion for adjusted weight
HPV2 <- LLCP25V2 %>%
filter(_STATE == 24) %>%
mutate(FINALWT = LCPWT * 0.49) %>%
select(-LCPWTV2)

# No state used V3 of HPV module, no need to extract V3 data
# HPV: combine all versions of the HPV datasets into one
 HPV <- bind_rows(HPVC, HPV1, HPV2)
```

The combined and reweighted dataset HPV is ready for analysis.

## Example 2: Childhood Asthma Prevalences

The example below uses the **Childhood Asthma Prevalence** Module to demonstrate how to combine and reweight data from multiple datasets.

**1)** The module **Childhood Asthma Prevalence** listed in the document Modules by Category shows that a total of 23 states used it as optional module in 2025.

23 states used the common version of the module; therefore, we need to extract data for these states from the dataset LLCP2025:

Arizona (4) Connecticut (9) Delaware (10) District of Columbia (11) Georgia (13) Hawaii (15) Illinois (17) Indiana (18) Kansas (20) Louisiana (22) Minnesota (27) Montana (30) Nebraska (31) New Jersey (34) New Mexico (35) Pennsylvania (42) Rhode Island (44) Texas (48) Vermont (50) Virginia (51) West Virginia (54) Wisconsin (55) Puerto Rico (72)

- Ohio (39) used version 1 of the module, so data should be extracted from LLCP25V1 for Ohio.
- Massachusetts (25) used version 2 of the module; therefore, data should be extracted from LLCP25V2 for Massachusetts.
- None of the states used Version 3 of the module, so no data will be extracted from LLCP25V3.

**2) Obtain datasets**:

- To download the Combined Landline Telephone and Cellular Telephone dataset (LLCP2025):
  - Go to the 2025 data webpage, find the 2025 Data files section of the webpage. 2025 documentation--including Overview, Codebook, and Module Weights --are all provided. The 2025 dataset is available in ASCII and SAS transport formats. Record layout, formats, and SAS codes that convert ASCII or SAS transportable file to SAS datasets are also available in the **SAS Resources** section of the page.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 1 dataset:
  - Go to the 2025 data webpage, find The Combined Landline and Cellular Telephone Survey Multiple Questionnaire Version Data Including Optional Modules. There are 3 sections, one for each version of the questionnaire datasets. Choose Version 1: LLCP25V1.
- To download the Combined Landline Telephone and Cellular Telephone Questionnaire Version 2 dataset:
  - Follow the same step above for version 1, but choose Version 2: LLCP25V2.
  - Since version 3 of the Childhood Asthma Prevalence module was not used by any state in 2025, it is not necessary to download it. (You would have downloaded it by following the same step above for Version 1, but choosing Version 3: LLCP25V3.)

3) **Generate a working dataset**:
Create a new and uniform final weight variable from each of the datasets, then combine data into one working dataset. This step will require users to follow these steps:

a. Keep all states that collected the module data in each version of the questionnaire.
b. Rename the corresponding weight variable from each version of the dataset to a consistent weight variable and adjust the weight variable if any state used more than 1 version of the module.
c. Combine all the datasets into one that contains the renamed weight variable.

4) **Weight adjustment**:
Since none of the states used more than one version of the Childhood Asthma Prevalence module, it is not necessary to adjust weight for any state. But we will use weight variables specific to children instead of the usual weight variable we use for adults. These child-specific weight variables are:

_CLLCPWT, Final Child Weight: Landline and Cellphone Data (LLCP2025)
_CLCWTV1, Version 1 Child Weight: Landline and Cellphone Data (LLCP25V1)
_CLCWTV2, Version 2 Child Weight: Landline and Cellphone Data (LLCP25V2)
_CLCWTV3, Version 3 Child Weight: Landline and Cellphone Data (LLCP25V3)

Note: these child-specific weight variables need to be renamed to a common weight variable before datasets are combined for analysis.

## Example code in SAS:

```
LIBNAME IN '\BRFSS\'; *specify the libname where BRFSS datasets are downloaded and stored;

DATA CHASP;          * Extract data for  states that used common version of the module;
SET IN.LLCP2025 (WHERE= (_STATE IN (4 9 10 11 13 15 17 18 20 22 27 30 31 34 35 42 44 48 50 51 54 55 72 )));
FINALWT =  _CLLCPWT;     *FINALWT, new weight variable for new data sets;
RENAME  _CLLCPWT = FINALWT;
RUN;

DATA CHASP1;       *Extract V1 data for OH;
SET IN.LLCP25V1 (WHERE=(_STATE IN (39)));
RENAME  _CLCWTV1 = FINALWT;
RUN;

DATA CHASP2;       *Extract V2 data for MA;;
SET IN.LLCP25V2 (WHERE=(_STATE IN (25)));
RENAME _CLCWTV2 = FINALWT ;
RUN;

DATA ASTHMA;   *Create combined dataset for analysis;
SET   CHASP CHASP1 CHASP2;
RUN;
```

The combined Childhood Asthma Prevalence dataset is ready for analysis.

**Example code in R:**

```
# Begin with the main data set LLCP2025 for states that used the common version of the module
# Since R does not allow variable names to begin with underscore, such as _LLCPWT, _LCPWTV1, _LCPWTV2 and _LCPWTV3
# Remove underscores from all variables in the dataset before analysis
#Create common weight variable FINALWT, drop original weight variables
#Only one state used each version of the V1, V2 module data, weight adjustment is not needed

library(dplyr)

# List of states that used the common version of the module
states_common_module <- c(4, 9, 10, 11, 13, 15, 17, 18, 20, 22, 27, 30, 31, 34, 35, 42, 44, 48, 50, 51, 54, 55, 72)

# Extract data for states that used the common version of the module
CHASP <- LLCP2025 %>%
 filter(STATE %in% states_common_module) %>%
 mutate(FINALWT = CLLCPWT)
 select(-CLLCPWT)

# Extract V1 data for OH
CHASP1 <- LLCP25V1 %>%
 filter(STATE %in% c(39)) %>%
 mutate(FINALWT = CLCWTV1) %>%
 select(-CLCWTV1)

# Extract V2 data for MA
CHASP2 <- LLCP25V2 %>%
 filter(STATE %in% c(25)) %>%
 mutate(FINALWT = CLCWTV2) %>%
 select(-CLCWTV2)

# Create combined dataset for analysis
ASTHMA <- bind_rows(CHASP, CHASP1, CHASP2)
```

The combined Childhood Asthma Prevalence dataset is ready for analysis.

## Combining Multiple Years of BRFSS Data

Sometimes we need to analyze data from multiple years, such as calculating the average prevalence of a variable in core sections for 2023 and 2025. In this case, we need to combine data from 2023 and 2025. The weight variable for 2023 and 2025 needs to be adjusted proportionally for the combined dataset. There are two ways to adjust the weight variable. The sample size from each year helps to determine how we should adjust them. Again, if the sample sizes are very similar, we can simply divide the original weight variable by 2. Otherwise, we need to adjust the weight variable proportionally by following these steps:

1) Find out the sample sizes for 2023 and 2025
2) Sum up the total sample size for 2023 and 2025
3) Find the proportion for 2023 (Use the sample size for 2023 divided by the combined sample size for 2023 & 2025)
4) Find the proportion for 2025 (Use the sample size for 2025 divided by the combined sample size for 2023 & 2025)
5) The adjusted weight for 2023 is (_LLCPWT(2023) * proportion for 2023)
6) The adjusted weight for 2025 is (_LLCPWT(2025) * proportion for 2025)

The example below uses the 2023 and 2025 data from New York to demonstrate how to combine and reweight data from multiple years.

The sample size for 2023 New York data is 17,349 and 14,876 for 2025 respectively. Since sample sizes are different, the weights need to be adjusted proportionally. Follow the steps listed above to adjust weights for New York for 2023 & 2025.

1) The sample size is 17,349 for 2023, and 14,876 for 2025 respectively
2) The combined sample size is 32,225
3) The proportion for 2023 is 17,349 / 32,225≈ .54
4) The proportion for 2025 is 14,876 / 32,225≈ .46;
5) The adjusted weight for 2023 is (_LLCPWT * **.54**)
6) The adjusted weight for 2025 is (_LLCPWT * **.46**)

**Obtain datasets**:

- To download the 2023 Combined Landline Telephone and Cellular Telephone Questionnaire dataset (LLCP2023):
  - Go to the 2023 data Webpage, and choose 2023 dataset. From there, review all 2023 documents including the Overview, Codebook, and the Modules by Category. The dataset is available in ASCII and SAS transport formats. A 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 2025 Combined Landline Telephone and Cellular Telephone Questionnaire dataset (LLCP2025):
  - Go to the 2025 data Webpage, and choose 2025 dataset. From there, read all of the 2025 documents including the Overview, Codebook, and the Modules by Category. The dataset 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 datasets and combine the data into one working dataset. This step will require users to do these steps:

    - Rename the corresponding weight variable to a consistent weight variable; and
    - Combine all datasets into one with the consistent weight variable.

    Again, users should be aware 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 D23 "\BRFSS2023";
Libname D25 "\BRFSS2025";

DATA NY23; /* Check sample size for each YEAR, 2023 & 2025 */
  SET D23.LLCP2023(WHERE=(_STATE=36));
  RUN;    /* 17349 observations extracted from LLCP2023 for NY (36) */

DATA NY25;
  SET D25.LLCP2025(WHERE=(_STATE=36));
  RUN;    /* 14876 observations extracted fromLLCP2025 for NY (36) */

DATA NY2YR;         /* Create a final dataset with the adjusted weight variable (FINALWT) */
  SET NY23 (IN=A)
      NY25 (IN=B);
  IF A THEN FINALWT = _LLCPWT * .54;       * Use the calculated proportion to adjust _LLCPWT;
  IF B THEN FINALWT = _LLCPWT * .46;
  RUN;
```

The combined and weight adjusted dataset NY2YR is ready for analysis.

**Example code in R:**

```
# Subset 2023 and 2025 data for New York
# Since R does not allow variable names to begin with underscore, such as _LLCPWT, _LCPWTV1, _LCPWTV2
and #_LCPWTV3, remove underscores from the dataset before analysis
# install.packages(c("readr", "dplyr"))

 library(dplyr)

# Check sample sizes for 2023 & 2025 NY common version dataset
# 17349 observations extracted from LLCP2023 for NY (36)

NY23 <- LLCP2023 %>%
  filter(STATE == 36)
nrow(NY23)

# 14876 observations extracted from LLCP2025 for NY (36)
NY25 <- LLCP2025 %>%
```

```
 filter(STATE == 36)
nrow(NY25)

# The calculated proportions are .54 & .46 for 2023 & 2025 respectively
# Create a final dataset with the adjusted weight variable (FINALWT) using the calculated proportion

NY23 <- NY23 %>%
 mutate(FINALWT = LLCPWT * .54)

NY25 <- NY25 %>%
 mutate(FINALWT = LLCPWT * .46)

NY2YR <- bind_rows(NY23, NY25)
```

The combined and reweighted dataset NY2YR is ready for analysis.

***~~~ END~~~***