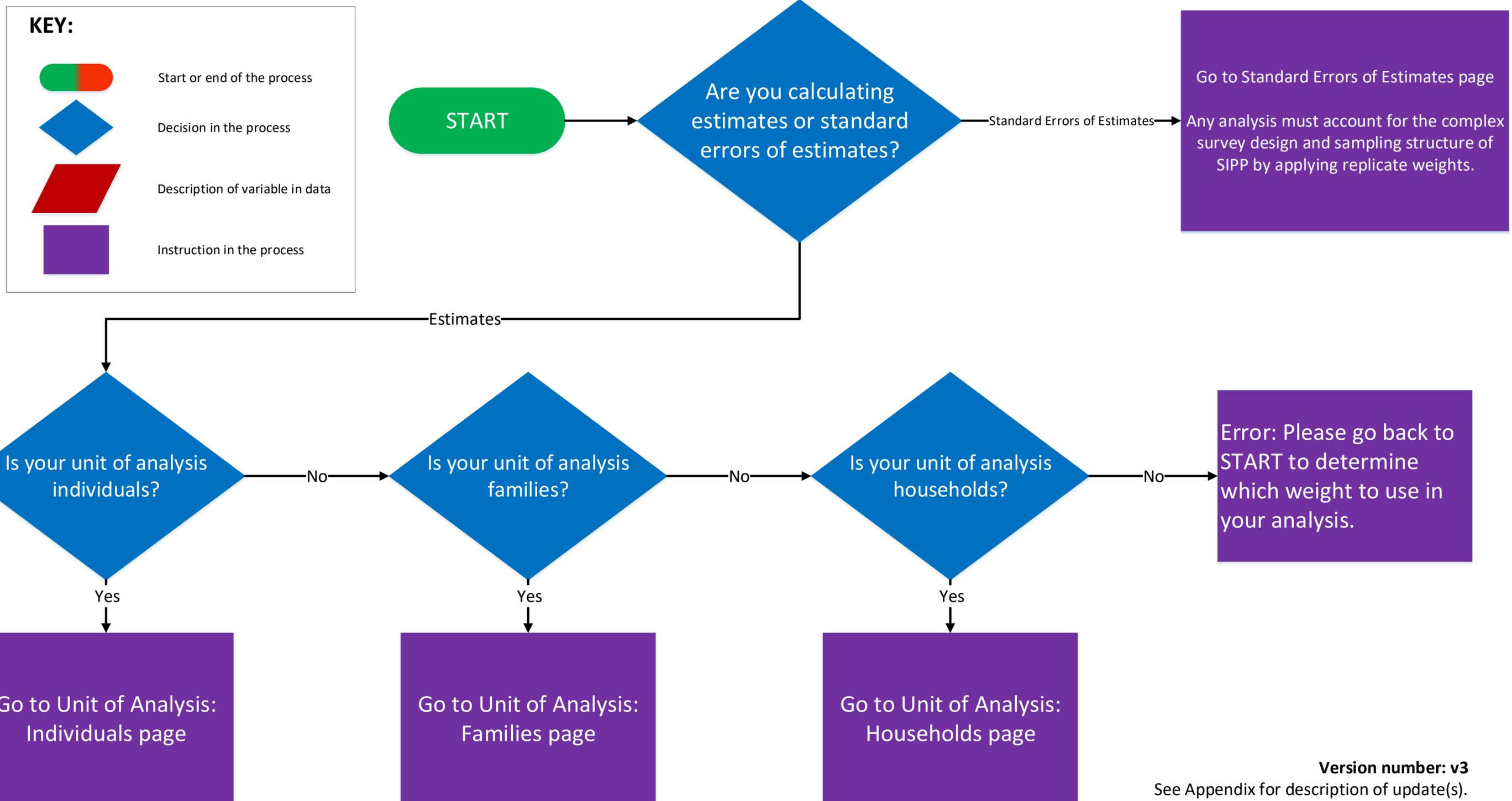


# Which Weight Should You Use for Your Analysis? Applying Weights in the 2014 SIPP Panel

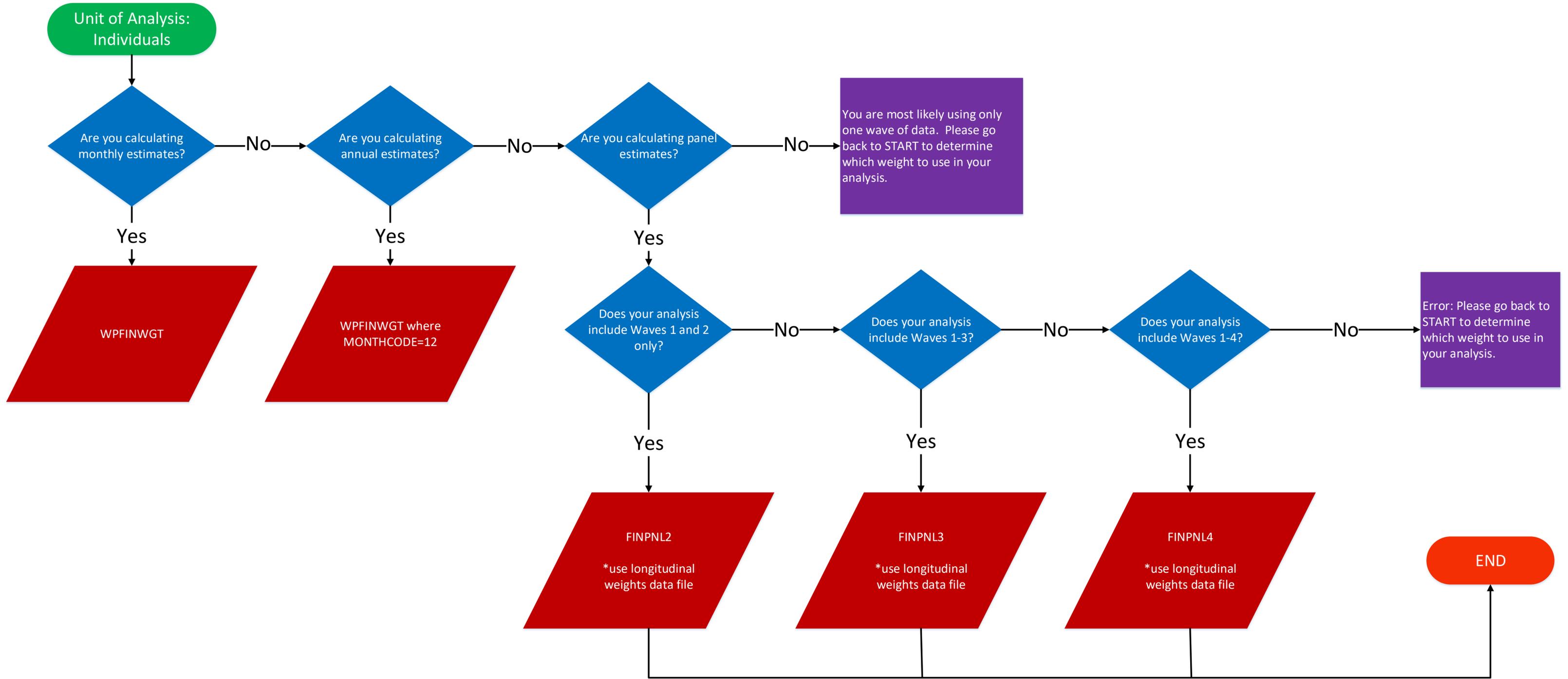
The decision of which weight to use for a given analysis depends on the time period (i.e., point in time, or time interval) and unit of interest (i.e., person, family, or household). Unweighted estimates show distributions for interviewed persons only. Using weights means that distributions are processed so as to be representative of the nation.

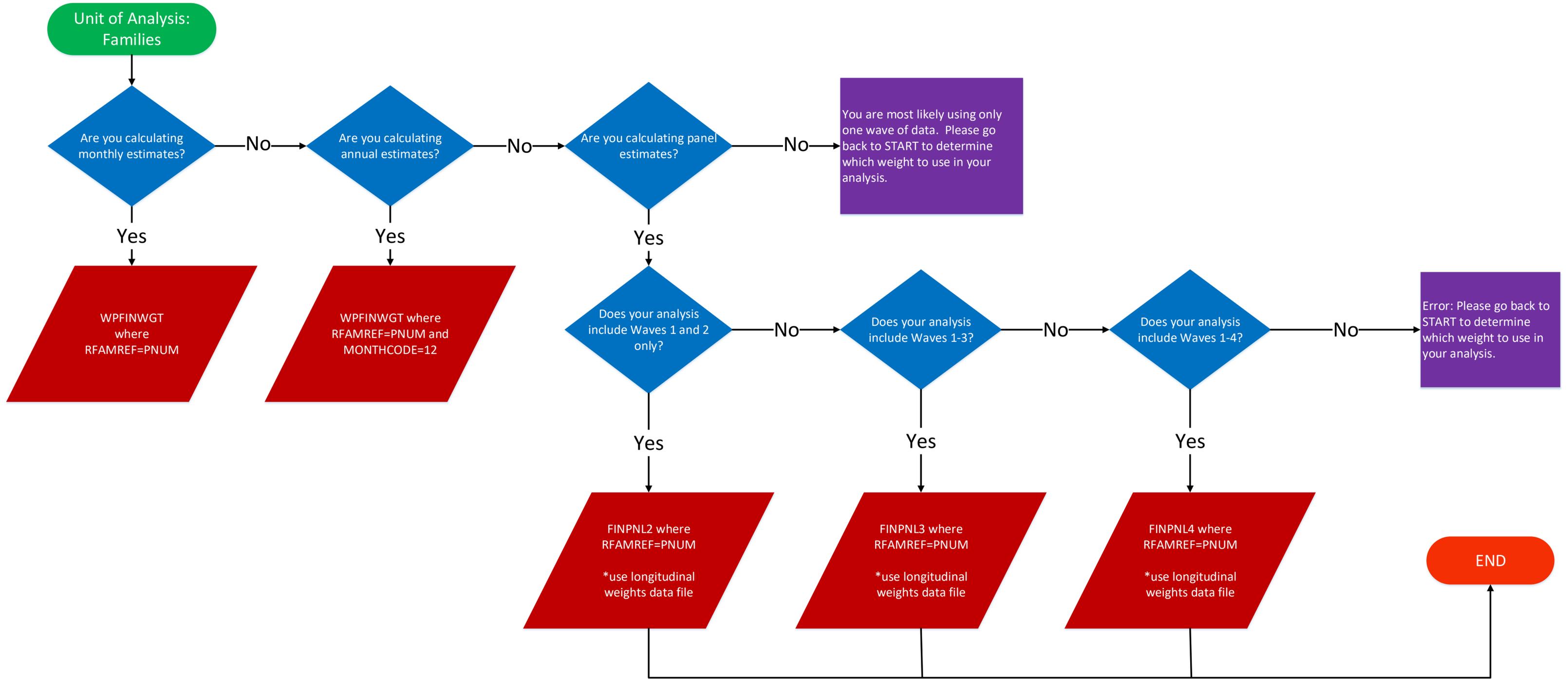
The figure below provides data users with a guide for selecting the appropriate weight for their analysis using 2014 SIPP Panel data. Data users must use either a monthly weight (i.e., cross-sectional person, family or household weight) or longitudinal weights (i.e., calendar or panel weights) for accurate population estimates. Replicate weights are used to generate standard errors and/or margins of error that are nationally representative and account for the SIPP sampling design. This guidance should not be considered exhaustive. Data users should be familiar with the 2014 SIPP Panel data and have a good understanding of which observations to include in their analyses. For more information about weighting, please see the SIPP users' guide for the 2014 SIPP Panel.

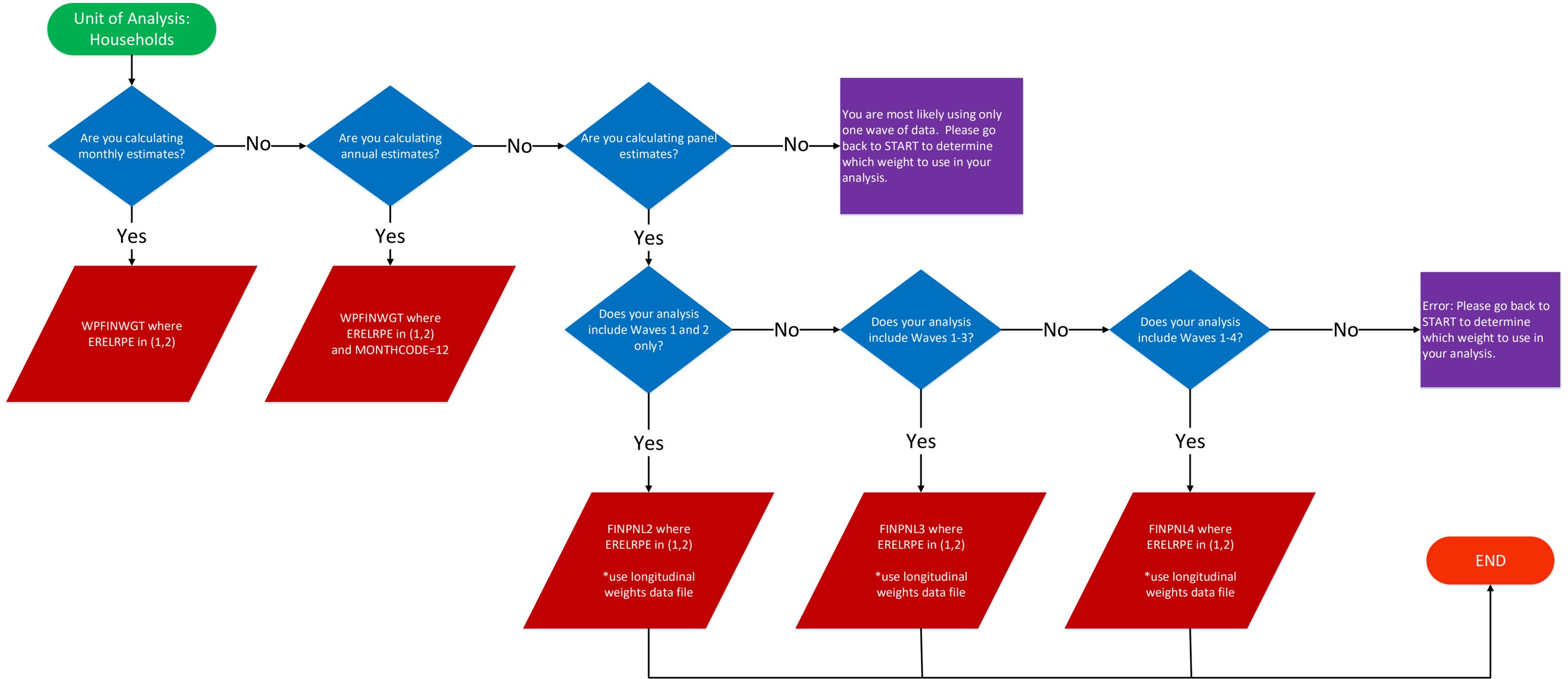


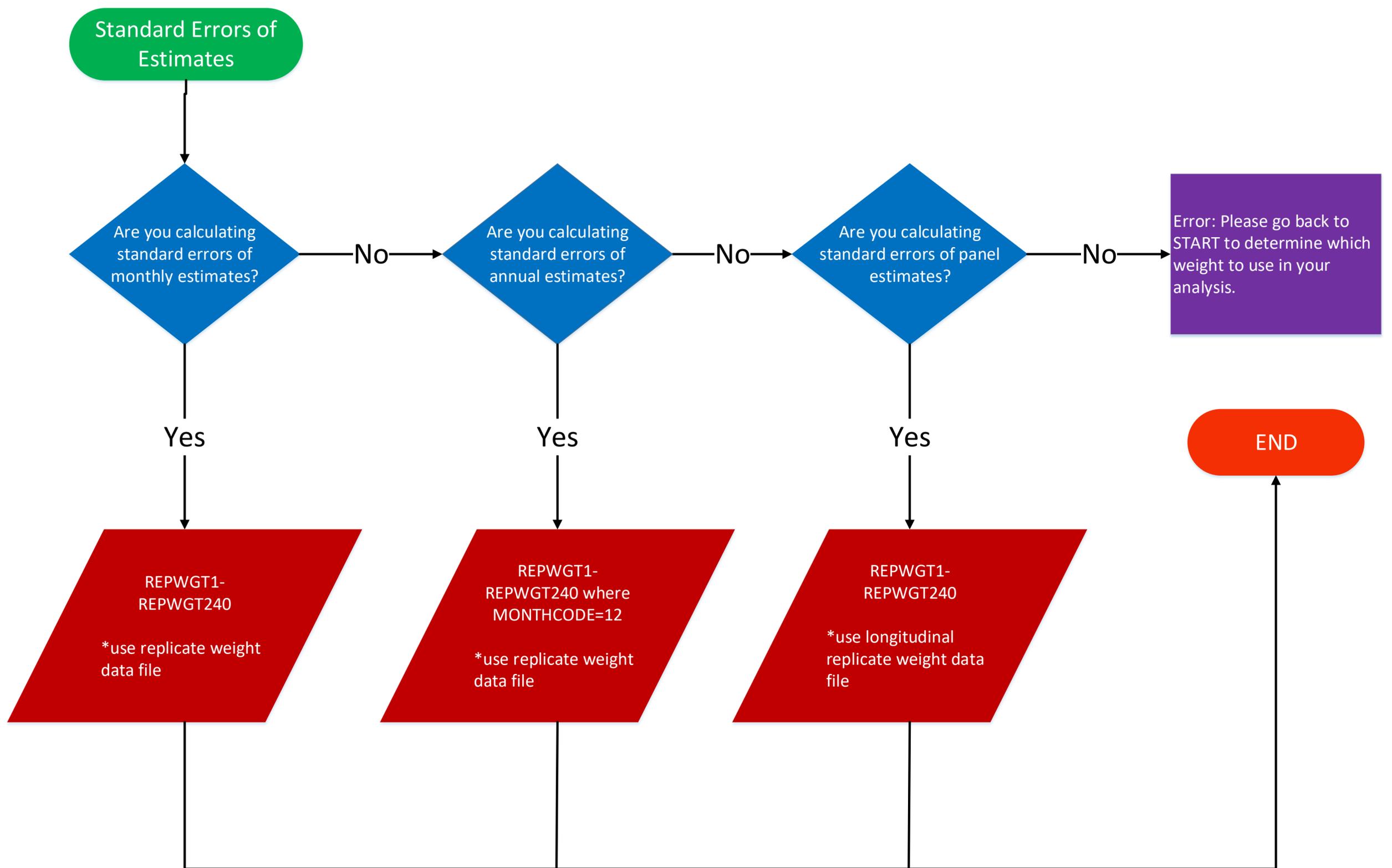
Version number: v3

See Appendix for description of update(s).









The tables below provide example SAS and Stata code that can produce unbiased estimates of personal income (TPTOTINC). Unit of analysis and time interval are taken into consideration for each example.

*Note:* Data users must first merge desired SIPP public-use dataset(s), longitudinal weights dataset, and/or replicate weights dataset(s) before calculating estimates and standard errors.

**Table 1. Unit of Analysis: Individual**

Time Interval	Example	SAS	Stata
Monthly	Personal income in September, Wave 1 (2013)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var TPTOTINC; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=9; run;</code>	<code>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean TPTOTINC if MONTHCODE==9</code>
Annual	Total personal income in Wave 1 (2013)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create sum variable of TPTOTINC for reference period and add here */; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=12; run;</code>	<code>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create sum variable of TPTOTINC for reference period and add here */ if MONTHCODE==12</code>
Waves 1 and 2	Total personal income in Waves 1 (2013) and 2 (2014)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */; weight FINPNL2; repweights repwgt1-repwgt240; run;</code>	<code>svyset [pw=FINPNL2], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */</code>
Waves 1-3	Total personal income in Waves 1 (2013), 2 (2014), and (2015)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-3 sum variable of TPTOTINC and add it here */; weight FINPNL3; repweights repwgt1-repwgt240; run;</code>	<code>svyset [pw=FINPNL3], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-3 sum variable of TPTOTINC and add it here */</code>
Waves 1-4	Total personal income in 2014 SIPP Panel (2013-2016)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-4 sum variable of TPTOTINC and add it here */; weight FINPN4; repweights repwgt1-repwgt240; run;</code>	<code>svyset [pw=FINPNL4], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-4 sum variable of TPTOTINC and add it here */</code>

*Note:* Naming convention of replicate weight variable names slightly changes throughout the 2014 SIPP Panel. For Wave 1, the replicate weight variable names are REPWGT1-REPWGT240. However, for Waves 2+, the replicate weight variable names are REPWT1-REPWT240, and the longitudinal replicate weight variables names are REPWGT1-REPWGT240. Note the replicate weight and longitudinal replicate weight are in separate data files for each wave in the 2014 SIPP Panel, so the naming convention of the replicate weight variables is unlikely to affect how data users manipulate the data (e.g., merging SIPP data with replicate weight data).

**Table 2. Unit of Analysis: Family**

Time Interval	Example	SAS	Stata
Monthly	Personal income for family reference person in September, Wave 1 (2013)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var TPTOTINC; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=9 &amp; RFAMREF=PNUM; run;</code>	<code>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean TPTOTINC if MONTHCODE==9 &amp; RFAMREF==PNUM</code>
Annual	Total personal income for family reference person in Wave 1 (2013)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create sum variable of TPTOTINC for reference period and add here */; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=12 &amp; RFAMREF=PNUM; run;</code>	<code>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create sum variable of TPTOTINC for reference period and add here */ if MONTHCODE==12 &amp; RFAMREF==PNUM</code>
Waves 1 and 2	Total personal income for family reference person in Waves 1 (2013) and 2 (2014)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */; weight FINPNL2; repweights repwgt1-repwgt240; where RFAMREF=PNUM; run;</code>	<code>svyset [pw=FINPNL2], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */ where RFAMREF==PNUM</code>
Waves 1-3	Total personal income for family reference person in Waves 1 (2013), 2 (2014), and (2015)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-3 sum variable of TPTOTINC and add it here */; weight FINPNL3; repweights repwgt1-repwgt240; where RFAMREF=PNUM; run;</code>	<code>svyset [pw=FINPNL3], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-3 sum variable of TPTOTINC and add it here */ where RFAMREF==PNUM</code>
Waves 1-4	Total personal income for family reference person in 2014 SIPP Panel (2013-2016)	<code>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-4 sum variable of TPTOTINC and add it here */; weight FINPN4; repweights repwgt1-repwgt240; where RFAMREF=PNUM; run;</code>	<code>svyset [pw=FINPNL4], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-4 sum variable of TPTOTINC and add it here */ where RFAMREF==PNUM</code>

**Note:** Naming convention of replicate weight variable names slightly changes throughout the 2014 SIPP Panel. For Wave 1, the replicate weight variable names are REPWGT1-REPWGT240. However, for Waves 2+, the replicate weight variable names are REPWT1-REPWT240, and the longitudinal replicate weight variables names are REPWGT1-REPWGT240. Note the replicate weight and longitudinal replicate weight are in separate data files for each wave in the 2014 SIPP Panel, so the naming convention of the replicate weight variables is unlikely to affect how data users manipulate the data (e.g., merging SIPP data with replicate weight data).

**Table 3. Unit of Analysis: Household**

Time Interval	Example	SAS	Stata
Monthly	Personal income for household reference person in September, Wave 1 (2013)	<pre>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var TPTOTINC; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=9 &amp; ERELRPE in (1:2); run;</pre>	<pre>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean TPTOTINC if MONTHCODE==9 &amp; (ERELRPE==1   ERELRPE==2)</pre>
Annual	Total personal income for household reference person in Wave 1 (2013)	<pre>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create sum variable of TPTOTINC for reference period and add here*/; weight WPFINWGT; repweights repwgt1-repwgt240; where MONTHCODE=12 &amp; ERELRPE in (1:2); run;</pre>	<pre>svyset [pw=wpfinwgt], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create sum variable of TPTOTINC for reference period and add here */ if MONTHCODE==12 &amp; (ERELRPE==1   ERELRPE==2)</pre>
Waves 1 and 2	Total personal income for household reference person in Waves 1 (2013) and 2 (2014)	<pre>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */; weight FINPNL2; repweights repwgt1-repwgt240; where ERELRPE in (1:2); run;</pre>	<pre>svyset [pw=FINPNL2], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1 and 2 sum variable of TPTOTINC and add it here */ where (ERELRPE==1   ERELRPE==2)</pre>
Waves 1-3	Total personal income for household reference person in Waves 1 (2013), 2 (2014), and 3 (2015)	<pre>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-3 sum variable of TPTOTINC and add it here */; weight FINPNL3; repweights repwgt1-repwgt240; where ERELRPE in (1:2); run;</pre>	<pre>svyset [pw=FINPNL3], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-3 sum variable of TPTOTINC and add it here */ where (ERELRPE==1   ERELRPE==2)</pre>
Waves 1-4	Total personal income for household reference person in 2014 SIPP Panel (2013-2016)	<pre>proc surveymeans data=&lt;PUT DATASET HERE&gt; varmethod = BRR (Fay=0.5) mean cv sum cvsum var T; var /* create total Waves 1-4 sum variable of TPTOTINC and add it here */; weight FINPNL4; repweights repwgt1-repwgt240; where ERELRPE in (1:2); run;</pre>	<pre>svyset [pw=FINPNL4], brrweight(repwgt1-repwgt240) fay(.5) vce(brr) mse  svy: mean /* create total Waves 1-4 sum variable of TPTOTINC and add it here */ where (ERELRPE==1   ERELRPE==2)</pre>

**Note:** Naming convention of replicate weight variable names slightly changes throughout the 2014 SIPP Panel. For Wave 1, the replicate weight variable names are REPWGT1-REPWGT240. However, for Waves 2+, the replicate weight variable names are REPWT1-REPWT240, and the longitudinal replicate weight variables names are REPWGT1-REPWGT240. Note the replicate weight and longitudinal replicate weight are in separate data files for each wave in the 2014 SIPP Panel, so the naming convention of the replicate weight variables is unlikely to affect how data users manipulate the data (e.g., merging SIPP data with replicate weight data).

# FAQ

[Question 1. Why do some people not have weights for all months in sample?](#)

[Question 2. When calculating annual estimates what population should be included?](#)

[Question 3. How do I identify people no longer living in a household?](#)

[Question 4. When calculating family-level estimates, why do I want to look at the family reference person \(RFAMREF=PNUM\)?](#)

[Question 5. When calculating household-level estimates, why do I want to look at the household reference person \(ERELRPE in \(1:2\)\)?](#)

[Question 6. How do I merge weights with the SIPP public-use data files?](#)

**Question 1. Why do some people not have weights for all months in sample?**

Answer: People do not have weights for months in which they are not part of the civilian non-institutionalized population. Reasons for this can include moves out of the country or into an institution.

**Question 2. When calculating annual estimates what population should be included?**

Answer: This depends on your research question. It may be of interest to include people in the SIPP sample who have positive weights for all 12 person months, if you are interested in looking at annual estimates for the population in the civilian non-institutionalized population for the entire calendar year. Another option is to include people in the SIPP sample who were in-sample and had a positive weight as of December of the reference period to capture people living in the United States at the end of calendar year.

**Question 3. How do I identify people no longer living in a household?**

Answer: THHLDSTATUS can identify people no longer living in the household. People who have a THHLDSTATUS value of “1” or “2” were living in the household at the time of the SIPP interview.

**Question 4. When calculating family-level estimates, why do I want to look at the family reference person (RFAMREF=PNUM)?**

Answer: There may be more than one family living in a household, and the identification variable, RFAMREF, identifies who is the family reference person. When calculating estimates, you should keep only one observation per family. The RFAMREF variable exists primarily for the purpose of calculating the family poverty measure and for the income recode variables. Recall family is defined as a set of people related by blood, marriage, or adoption. Alternatively, data users could average the final person weights for members in the family and use that average weight when calculating family-level estimates.

**Question 5. When calculating household-level estimates, why do I want to look at the household reference person (ERELRPE in (1:2))?**

Answer: The household reference person is the owner or renter of the housing unit, and the identification variable, ERELRPE, identifies who is the household reference person. When calculating estimates, you should keep only one observation per household. Recall a household is defined as a set of people living together. Alternatively, data users could average the final person weights for members in the household and use that average weight when calculating household-level estimates.

## Question 6. How do I merge weights with the SIPP public-use data files?

Answer: Below is SAS and Stata code that demonstrates how to merge the SIPP public-use data with the replicate weights file. You can also use this code to merge the SIPP public-use data with the longitudinal replicate weights file.

```
SAS libname sipp '<PUT PATHNAME HERE>'; /* location of where the data are stored */

/* Step 1: Read in 2014 SIPP Wave 1 public use dataset and create temporary
dataset named "public" */
data public ; set sipp.pu2014w1;
run;

/* Step 2: Sort "public" dataset by SSUID, PNUM, MONTHCODE */
proc sort data=public;
    BY SSUID PNUM MONTHCODE;
run;

/* Step 3: Read in 2014 SIPP Wave 1 replicate weight dataset and create temporary
dataset named "repwgt" */
data repwgt; set sipp.rw14w1;
run;

/* Step 4: Sort "repwgt" dataset by SSUID, PNUM, MONTHCODE */
proc sort data=repwgt;
    BY SSUID PNUM MONTHCODE;
run;

/* Step 5: Merge "public" and "repwgt" datasets by SSUID, PNUM, MONTHCODE and
create temporary dataset named "analysis" */
data analysis;
    merge public (in=a) repwgt (in=b);
    BY SSUID PNUM MONTHCODE;
    if a and b; /* include observations that are in both "public" and "repwgt"
datasets */
run;
```

```
Stata /* Step 1: Read in 2014 SIPP Wave 1 public use dataset and change variable names to lowercase */
use "<PUT PATH HERE>\pu2014w1.dta" /* use SIPP public use file */
rename *, lower /* change variable names to lowercase */

/* Step 2: Sort dataset by SSUID, PNUM, MONTHCODE */
sort ssuid pnum monthcode /* sort dataset by SSUID PNUM MONTHCODE */

/* Step 3: Save "public" dataset */
save "<PUT PATH HERE>\public.dta"

clear all /* clear dataset from memory */

/* Step 4: Read in 2014 SIPP Wave 1 replicate weight dataset */
use "<PUT PATH HERE>\rw14w1.dta", clear /* use replicate weight dataset */

destring pnum, replace /* Optional code: may need to destring pnum variable */

/* Step 5: Sort replicate weights dataset by SSUID, PNUM, MONTHCODE */
sort ssuid pnum monthcode

/* Step 6: Merge "public" and replicate weight dataset by SSUID, PNUM, MONTHCODE */
merge 1:1 ssuid pnum monthcode using "<PUT PATH HERE>\public.dta"

tab _merge /* keep observations that are in both "public" and the replicate weight datasets (i.e., the observations that match only) */
keep if _merge==3
```

## Glossary of Key Weight Terms

<b>Variable</b>	<b>Definition</b>
SSUID	Sample unit identifier
PNUM	Person number
WPFINWGT	Final person weight
MONTHCODE	Month during reference year
FINPNL2	Longitudinal weight for Waves 1 and 2
FINPNL3	Longitudinal weight for Waves 1, 2, and 3
FINPNL4	Longitudinal weight for Waves 1, 2, 3, and 4
RFAMREF	Family reference person
ERELRPE	Relationship to householder
REPWGT1- REPWGT240	Replicate weights used in calculating variance/standard errors

# Resources for Data Users

## *Data Resources*

- SIPP website: <http://www.census.gov/sipp>
- NBER SIPP page: <http://www.nber.org/data/sipp.html>

## *Technical Documentation*

- SIPP website (<http://www.census.gov/sipp>) is your best resource  
Currently available:
  - Users' Guide
  - Metadata
  - Release notes
  - User notes
  - Codebook
  - Crosswalks

## *Contact information*

- SIPP email: [Census.Sipp@census.gov](mailto:Census.Sipp@census.gov)
- SIPP phone number: 1-888-245-3076

# Appendix

<b>Document Revision History</b>			
<b>Version</b>	<b>Description</b>	<b>Date</b>	<b>Updated by</b>
v1	<ul style="list-style-type: none"><li>• Initial version</li></ul>	10/24/19	H. Fee
v2	<ul style="list-style-type: none"><li>• Answer for questions #2, #4, and #5 in FAQ section was revised.</li></ul>	2/28/20	H. Fee
v3	<ul style="list-style-type: none"><li>• Sample SAS and Stata code for annual estimates was revised.</li><li>• Link to Census FTP site removed from Resources for Data Users page.</li><li>• Reference to 2008-2014 and 2014-2008 crosswalks was removed from Resources for Data Users page.</li></ul>	1/22/21	H. Fee