

# POVERTY TRACKER

## A Brief Guide to Using the Poverty Tracker (PT) Weights

*Last updated: May 17, 2023*

**ROBINHOOD**



This guide provides a brief overview of the weights for the Poverty Tracker (PT) data and how to use them in analyses. The purpose of this guide is to help data users understand the different sets of weights available in the public data files and how to use them to answer different types of questions. It is not intended to provide statistical advice. It simply describes the weights that are available to data users and other key information about their use. If you have any questions related to the methodology, or the study more generally, please reach out to us at [povertytracker@columbia.edu](mailto:povertytracker@columbia.edu).

I. Available Weights.....	1
II. Applying the Weights and Computing Variance.....	4

**Poverty Tracker Study Investigators:** Qin Gao, Irwin Garfinkel, Ronald Mincy, Kathryn Neckerman, Julien Teitler, Jane Waldfogel, Christopher Wimer

**Suggested citations:**

*For the Poverty Tracker Data*

Poverty Tracker Research Group at Columbia University (2023). *Poverty Tracker Data*. Center on Poverty and Social Policy at Columbia University and Columbia Population Research Center. <https://www.povertycenter.columbia.edu/poverty-tracker-data>.

*For the A Brief Guide to Using the Poverty Tracker (PT) Weights*

Yajun Jia, Sophie Collyer, Matthew Maury, Schuyler Ross, Christopher Wimer (2023). *Poverty Tracker Data User Guide*. Center on Poverty and Social Policy at Columbia University and Columbia Population Research Center. <https://www.povertycenter.columbia.edu/poverty-tracker-data>.

**Acknowledgments**

We are grateful for the support and partnership of Robin Hood. We thank Lily Bushman-Copp, Chantal Bannerman, Chloé Cargill, Lolita Colon, Katherine Friedman, Janira Gayle, Anastasia Koutavas, Xiaofang Liu, Kahlen Washington, Nicholas Wilkinson, Ho Yan Wong, and Christopher Yera for their contributions and assistance in the preparation of this guide.

**Disclaimer**

Note that all documentation files are working documents and are subject to change as the study evolves.

# I. Available Weights

## 1.1 PT Regular Weights (Person-level & Poverty-unit weights)

For each survey wave, users can apply weights to the data at two different levels: the person-level and the poverty-unit level.<sup>1</sup> More information about person-level and poverty-unit weights can be found in the [Poverty Tracker Technical Notes](#).

The person-level weights make the data representative of all adults (age 18 and over) in New York City in the year that the sample was recruited (i.e., 2012 for the first panel, 2015 for the second panel, and so on; see Table 1 for more details). The poverty-unit weights make the data representative of poverty units in New York City in the year in which the baseline data was collected (i.e., 2012 for the first panel, 2015 for the second panel, and so on; see Table 1 for more details). A poverty unit can be thought of as a family unit, comprised of individuals related by blood, marriage, or adoption; foster children; and unmarried partners and their relatives.

**Table 1. Poverty Tracker Cohorts**

Cohort	Initial recruitment year	Sample Size	Subsamples
1	2012	2,228	2,002 RDD 226 social service user oversample
2	2015	3,908	3,403 RDD (from <a href="#">CHS</a> ) 505 social service user oversample
3	2017	853	853 RDD
4	2020	1,912	1,491 RDD 421 Chinese-origin oversample
5	2022	1,548*	1,287 RDD* 261 Chinese-origin oversample*

For PT users, we recommend that the person-level weights be used when producing estimates related to individuals; the poverty-unit weights should be used when the unit of interest is the poverty unit or family. As an example, to weight the baseline data of panel 2 to be representative of adults in New York City, and answer the question “*What percent of adults experienced X in 2015?*”, you should apply the weight **qweight\_p**. If the question is instead “*What percent of families (poverty units) experienced X in 2015?*”, then you should apply the weight **qweight\_pu**.

When setting the weights in the statistical package, you will need to indicate the name of the weights you would like to use (person-level or poverty-unit weights). We have weights for every survey wave which, like the other variables, can be differentiated by their prefix (more details available in the [Poverty Tracker Data User Guide](#)). The variable names for these weights are listed in Table 2.

---

<sup>1</sup> We use the term “family” to describe poverty units, as defined by the Supplemental Poverty Measure.

**Table 2. PT Person-level & Poverty-Unit Weights**

Wave	Person-level Weights	Poverty-unit Weights
Baseline	qweight_p	qweight_pu
3-month	q1weight_p	q1weight_pu
6-month	q2weight_p	q2weight_pu
... and so on for each survey		

## 1.2 PT Replicate Weights

Replicate weights are currently available for PT data upon the user’s request.<sup>2</sup> In the PT data, there are 50 separate replicate weights at the person and poverty-unit levels that allow users to generate empirically derived standard error estimates.

In theory, the standard error of an estimate measures the variation of a statistic across multiple samples of a given population. A true standard error of estimates from a single sample can never be known with certainty. Replicate weights allow a single sample to simulate multiple samples, therefore generating more informed standard error estimates that mimic the “true” standard errors while retaining all information about the complex sample design.

For PT users, we recommend using replicate weights if you are interested in obtaining more precise confidence intervals and significance tests. Note that replicate weights usually increase standard errors -- however, the increase is generally not large enough to alter the significance level of coefficients (though marginally significant coefficients might become nonsignificant). The more notable effect of using replicate weights is that the width of the confidence intervals might change.

When setting the replicate weights in a statistical package, you will need to indicate the name of the set of replicate weights along with the regular person-level or poverty-unit weights you would like to use. The variable names for the replicate weights are listed in Table 3.

**Table 3. PT Person-level & Poverty-Unit Replicate Weights**

Wave	Person-level Replicate Weights	Poverty-unit Replicate Weights
Baseline	qweight_p_rep1	qweight_pu_rep1
	...	...
	qweight_p_rep50	qweight_pu_rep50
3-month	q1weight_p_rep1	q1weight_pu_rep1
	...	...
	q1weight_p_rep50	q1weight_pu_rep50

<sup>2</sup> Details of how to request data can be found in the [Poverty Tracker Data User Guide](#), “Accessing and Using the Data” section.

6-month	q2weight_p_rep1	q2weight_pu_rep1
	...	...
	q2weight_p_rep50	q2weight_pu_rep50
... and so on for each survey		

## II. Applying the Weights and Computing Variance

### 2.1 For Stata Users...

To use the weights in Stata, we recommend the code below. For more information on using weights in Stata, users should review the Stata manuals, particularly the “Survey Data” manual.<sup>3</sup>

To use PT weights in Stata, you must first **svyset** the data. The **svyset** command tells Stata about the design elements in the survey. Once this command has been issued, all you need to do for your analyses is use the **svy:** prefix before each command, which passes on the options you defined. Note that not all Stata commands can be run with the **svy:** prefix. Type **help svy\_estimation** to see a list of valid commands.

```
svyset [aweight = BASICWEIGHT], vce(bootstrap) bsrweight(REPLICATEWEIGHTS)
```

... where BASICWEIGHT and REPLICATEWEIGHTS are replaced with the relevant weights for your analyses.

For example, to run an analysis of baseline person-level data, the data user should first set the weights by entering...

```
svyset [aweight = qweight_p], vce(bootstrap) bsrweight(qweight_p_rep1 – qweight_p_rep50)
```

Or to run an analysis of baseline poverty-unit level data, the data user should first set the weights by entering...

```
svyset [aweight = qweight_pu], vce(bootstrap) bsrweight(qweight_pu_rep1 – qweight_pu_rep50)
```

Note that you can ignore the options after the comma if you are not using replicate weights...

### 2.2 For R Users...

To use the weights in R, we recommend the code below. For more information on using weights in R, users should review the R manuals, particularly the “**srvyr**” package reference manual.<sup>4</sup>

To use PT weights in R, you must first install the **srvyr** package. Next, you will need to create a survey object using the person or poverty-unit level weights and the replicate weights (if applicable). The **as\_survey** function has other arguments you can customize if needed. Any estimates you would like to make with replicate weights can be done with the object ‘svy’ instead of the object ‘data’.

---

<sup>3</sup> StataCorp. 2021. Stata: Release 17. Statistical Software. College Station, TX: StataCorp LLC. <https://www.stata.com/manuals/svy.pdf>

<sup>4</sup> srvyr. R package, version 1.2.0. <https://cran.r-project.org/web/packages/srvyr/srvyr.pdf>.

```
install.packages("srvyr")
```

```
library("srvyr")
```

```
svy <- as_survey(data, weight = BASEWEIGHT, repweights = matches("REPLICATEWEIGHTS"))
```

... where BASICWEIGHT and REPLICATEWEIGHTS are replaced with the relevant weights for your analyses.

For example, to run an analysis of baseline person-level data, the data user should first set the weights by entering...

```
svy <- as_survey(data, weight = qweight_p, repweights = matches("qweight_p_rep[1-50]+"))
```

Or to run an analysis of baseline poverty-unit level data, the data user should first set the weights by entering...

```
svy <- as_survey(data, weight = qweight_pu, repweights = matches("qweight_pu_rep[1-50]+"))
```

Note that you can ignore the options after the comma if you are not using replicate weights...