NYC POVERTY TRACKER SYMPOSIUM



May 18, 2023

General Overview of the Project

- Goals
- Core measures and additional topics covered
- Cohorts, timeline, and evolution
- Survey schedule
- Pooling Poverty Tracker cohort data

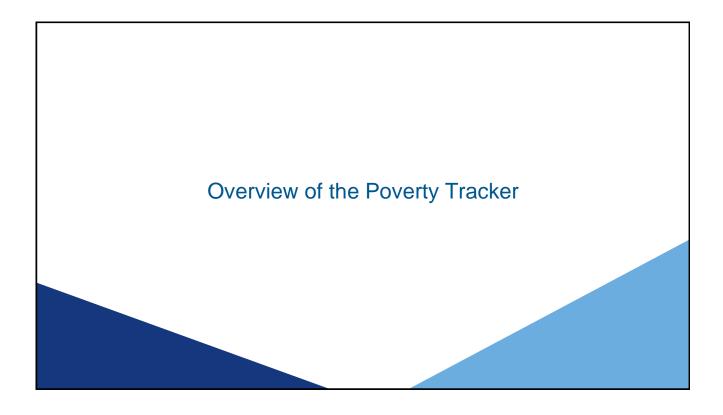
Imputation and Weighting

- Imputation methods
- Weighting methods
- Types of weights and how to use them

Accessing Poverty Tracker Data and Documentation







Original goals of the project

Get a better measure of poverty in New York City.





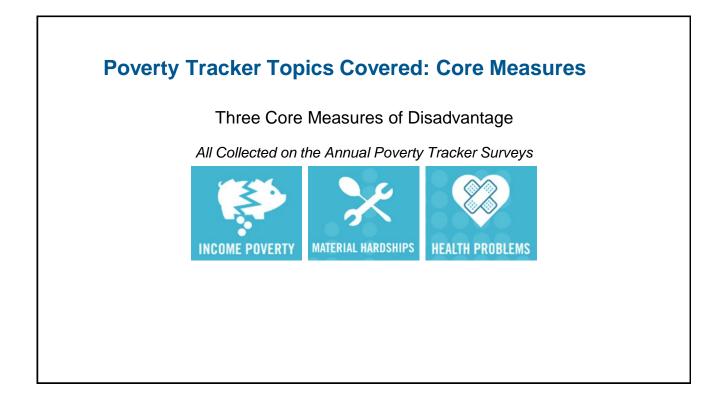
Provide a more comprehensive understanding of disadvantage beyond poverty.

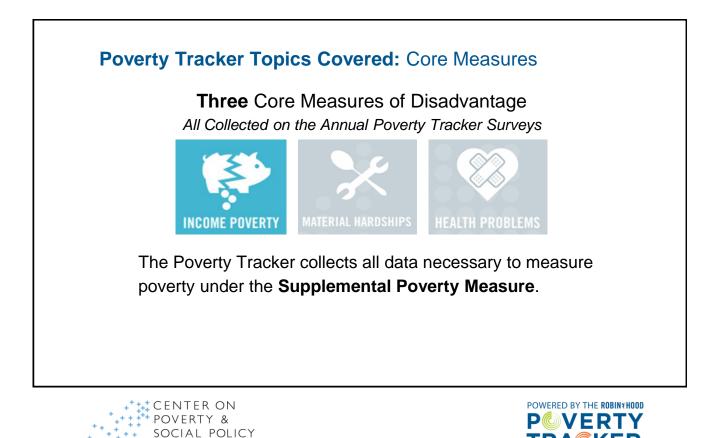


Understand the dynamics of poverty and disadvantage how these experiences respond to policy interventions and other life events and circumstances.



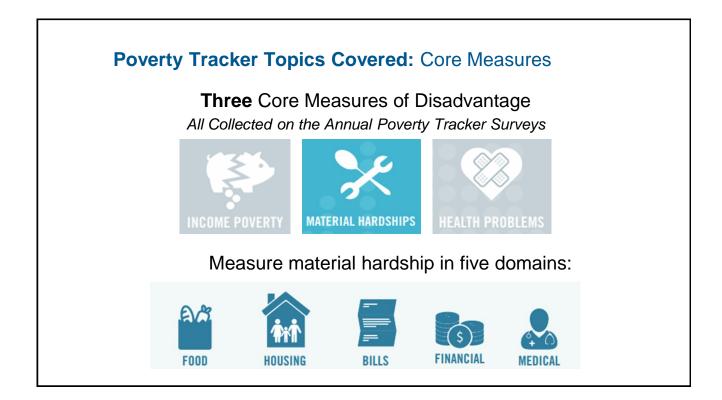






at Columbia University

TRACKFR

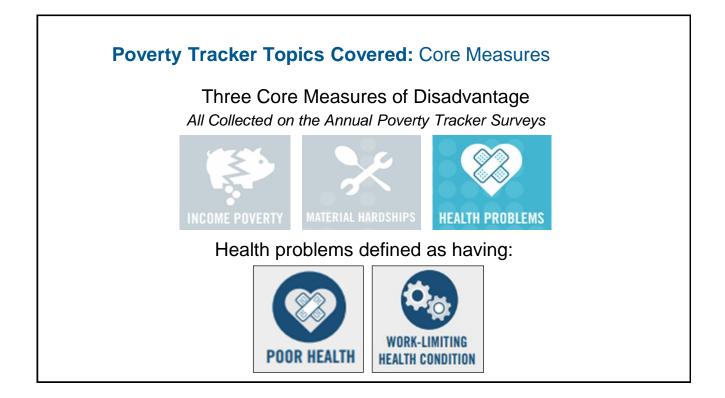


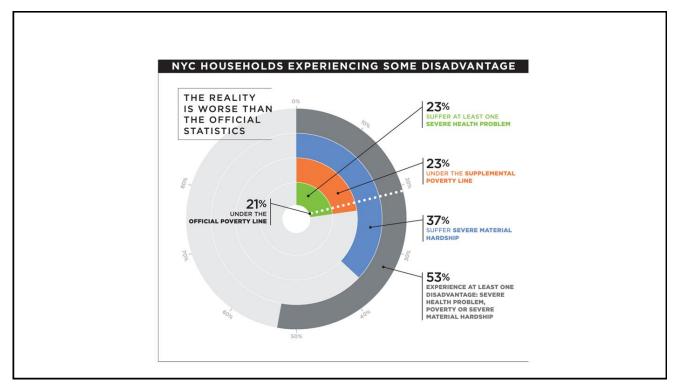


TRACKER

SOCIAL POLICY

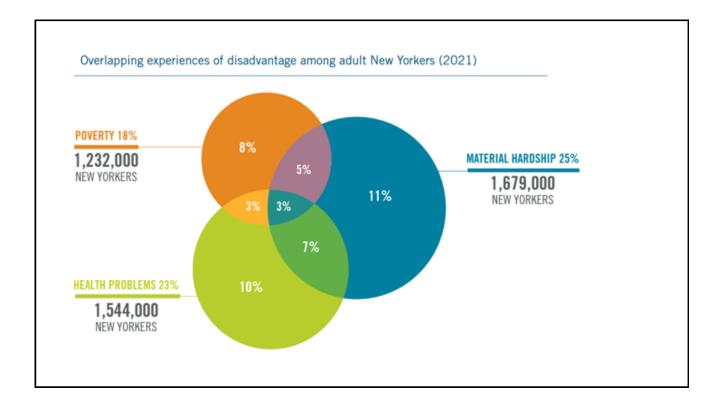
at Columbia University







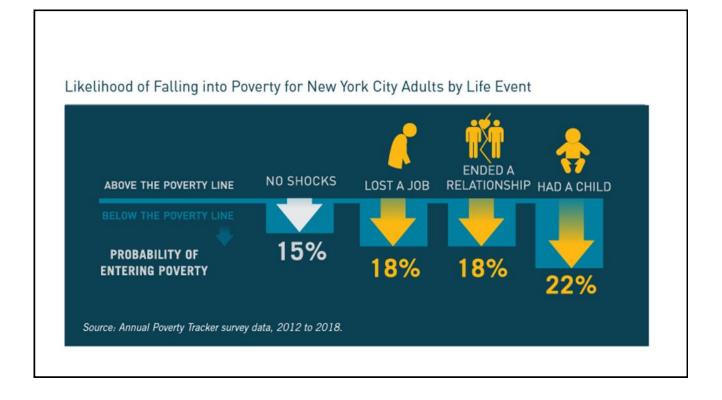


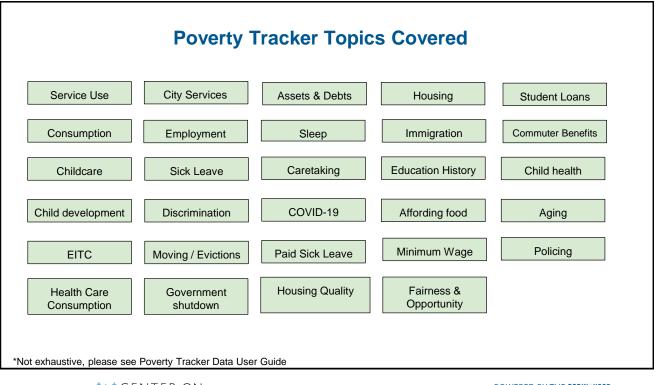


Poverty ⁻	Tracker Topics	Covered: "Sh	ocks"
	e measures, the Poverty nocks that a respondent		•
Starting or losing a job	Having a child	Having an accident or injury	Being a victim of a crime
Beginning to receive benefits or losing benefits	Having a major increase or decrease in income	Being arrested or being stopped by police	Lost/broke expensive belongings or having a major unanticipated expense
Changes in household composition and moving	Starting or ending a relationship	Changes in childcare	Having to take time off work to talk with your child's teacher or guidanc counselor













Cohorts, Timeline, and Evolution

The study surveys a **representative sample of adult New Yorkers** multiple times a year for a period of up to **six years.** (Originally planned to just survey for two years but later extended to six.)

Since initial launch, the study recruited **five distinct samples** of New Yorkers: **Cohorts 1 through 5** (more details to come!)

The majority of these samples were recruited through a **Random Digit Dial (RDD) methodology**. Periodically includes **supplementary samples**, including:

- · Respondents recruited from social service agencies (in the first two cohorts)
- New Yorkers of Chinese-origin (in the fourth and fifth cohorts)

Cohorts, Timeline, and Evolution

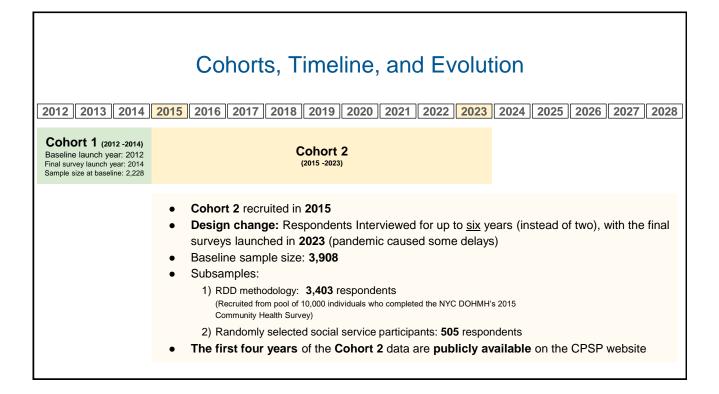
2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028

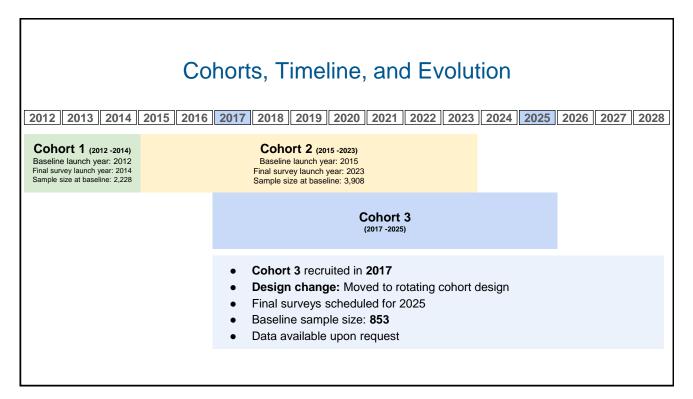
Cohort 1 (2012 -2014)

- Cohort 1 recruited in 2012
- Respondents interviewed for up to two years, with the final surveys launching in 2014
- Baseline sample size: 2,228
- Subsamples:
 - 1) RDD methodology: 2,002 respondents
 - (Incl. oversampling of New Yorkers in high-poverty zip codes)
 - Randomly selected social service participants: 226 respondents (Oversample of social service users that may be disproportionately low-income)
- All of the Cohort 1 data is publicly available on the CPSP website.



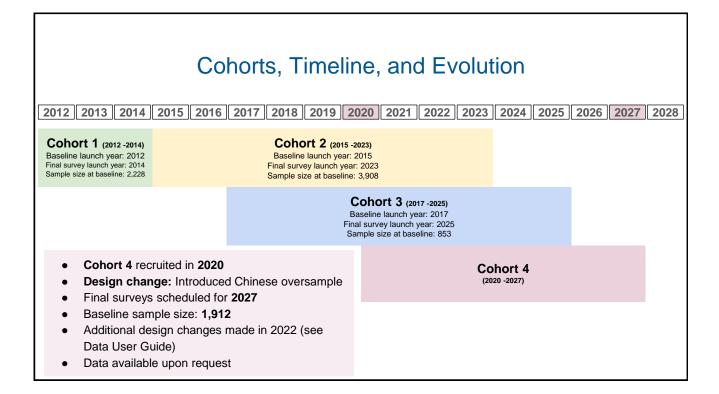


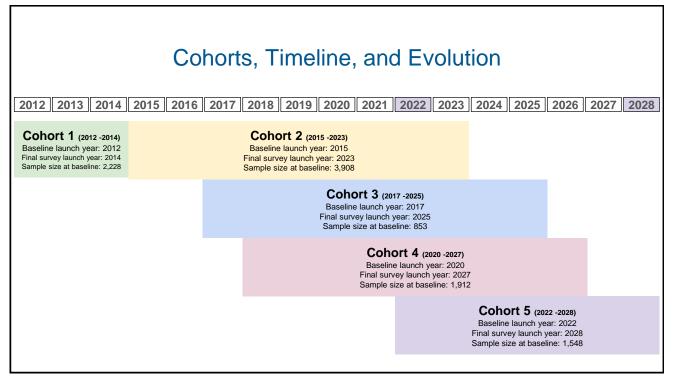






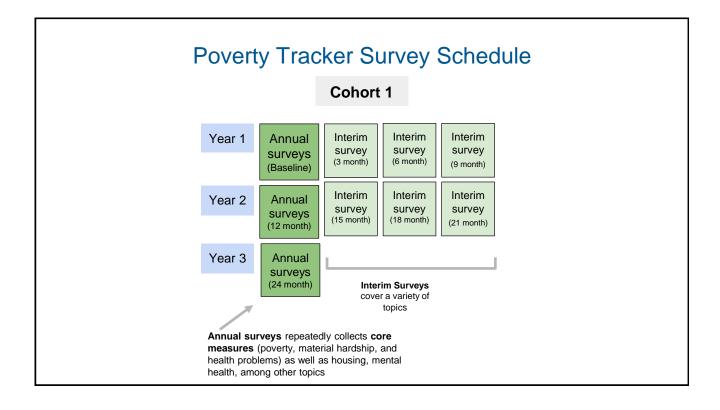


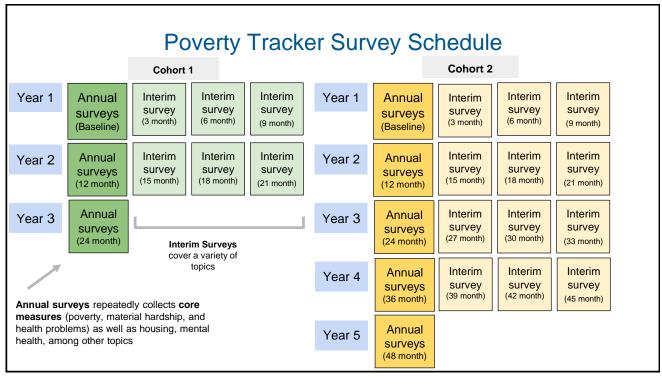






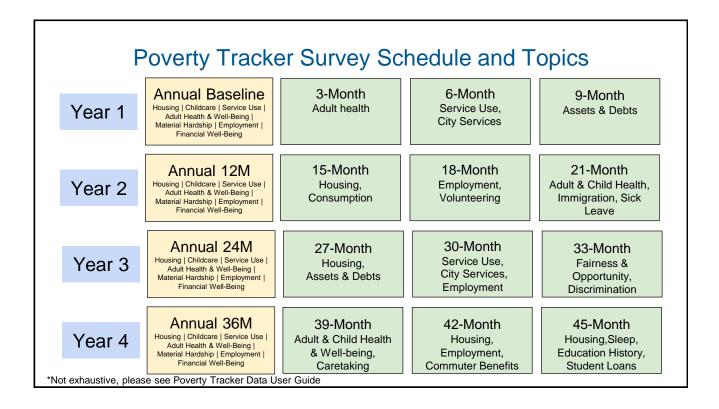


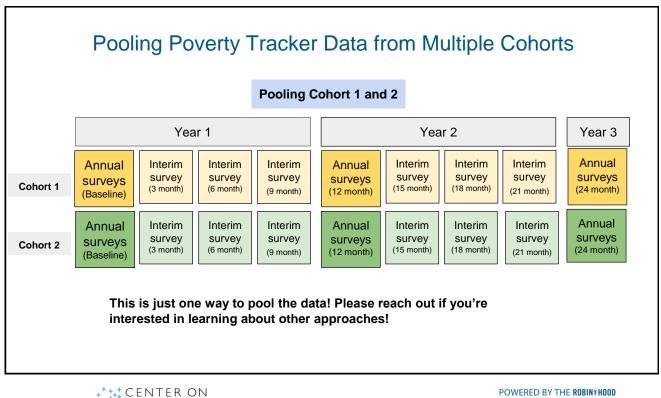












POVERTY &

at Columbia University

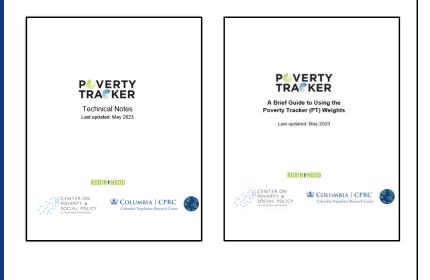
SOCIAL POLICY





Documents

- Technical Notes
- Guide for Using Weights







Dealing with Missingness

Missing Data

- **Common in survey research** respondents might by skeptical about data protection, uncomfortable disclosing personal information, or experience survey fatigue...
- **Going back** to conduct the survey **may not be the best solution** because you'd have to spend twice as much time and money.
- Most commonly, folks impute due to issues of power associated with reduced sample size due to missingness.

Imputation

- Substituting missing information with an estimated value depending on other available information/data.
- Imputation **preserves all cases** so that we would be able to analyze "complete" data once all values have been imputed.

How the Poverty Tracker Deals with Missingness

Missing at Random (MAR)

- It's an **underlying assumption** (missing at random but not completely at random) for Poverty Tracker imputations where we impute missing values using other information/characteristics in the data.
- MAR means there might be systematic differences between the missing and observed values, but these can be entirely explained by other observed variables.

Multiple Imputation (MI)

- Simple imputation assume perfect estimation of imputed values and ignore between-imputation variability.
- In contrast, MI generates replacement values ("imputations") for missing data and repeating this procedure many times, resulting in many data sets with replaced missing information.
- Incorporates uncertainty into the standard errors of imputed values by accounting for variability between imputed solutions.

Acock, 2005; Graham, 2009; Hibel, Farkas, & Morgan, 2010; Schafer, 1999





Variables Imputed in Poverty Tracker Data

Demographics [Baseline]

• Demographic variables (e.g. education, immigration status)

Core Measures [Baseline & Post-Baseline]

- Health (e.g. limiting health condition, mental health)
- Material hardship (e.g. housing, bill, medical, financial, food hardships)
- Income and expenses
 - Earnings (e.g. respondent/spouse months worked, earnings)
 - Income from other sources (e.g. retirement, disability, welfare, unemployment, SNAP, WIC)
 - Two-step imputation
 - · Binary value indicating whether or not received type of income
 - Amount of income
 - Income from other family members
 - Expenses (e.g. childcare, work, medical-related expenses)
- Housing (e.g. mortgage, rent)

Variables Imputed in Poverty Tracker Data

Impute Variables for Weighting

- Phone type (necessary for producing design weights; e.g. cellphone/landline)
- Service use frequency

Impute for Other Measures

Assets and debts





Weighting What is a Survey Weight? A value assigned to each case in the data, normally used to adjust statistical parameters/estimates so that inferences from the data are more representative of the target population. In this case, survey weight allows us to use Poverty Tracker sample to make statements about the city's population. Example: statements about what share of adult New Yorkers or families are in poverty instead of statements just about the share of the sample.

Weights in the Poverty Tracker

Survey weights are used to ensure that the PT sample is representative of adults and poverty units (families) in New York City.

The weighting approach adjusts for **oversampling**, **random over- or under-representation**, **non-response**, and **attrition**.

For example, the value of the weight indicates how much each case will count in a statistical procedure. A weight of 2000 in Poverty Tracker data indicates that this weighted case is representative of 2000 New Yorkers.

To achieve this, Poverty Tracker weights are produced by...

- Step 1: Adjust for sampling design
- Step 2: Adjust for non-response post-stratify to "population"
- Step 3: Adjust post-baseline surveys for sample attrition





Step 1: Adjust for Sampling Design

How Design Weights Adjust for Sampling Design?

- Design weights compensate for over- or under-sampling of specific cases or for disproportionate stratification.
- Example: oversample Chinese New Yorkers 4 times more than White New Yorkers...





Design Weights in Poverty Tracker

- Household size
- Phone availability
- Oversample households in high-poverty neighborhoods (cohort 1)
- Service use frequency

Gelman, & Little, 1998; Lohr, 2009

Step 2: Adjust for Non-response - Post-stratify to "Population"

Post-stratification Weights Adjust for Non-response

- Post-stratification weights compensate for the fact that persons with certain characteristics are not as likely to respond to the survey (e.g. more females than males). Corrects for departures from known population totals.
- **Requires** the use of **auxiliary information about the population** and may take a number of different variables into account.
- · Commonly adjusted respondent characteristics: age, gender, education, race/ethnicity, etc.

Post-stratification Weights in Poverty Tracker

- "Population" data: data are weighted to three-year American Community Survey (ACS) dataset.
- **Post-stratification variables** include gender, age, education, race/ethnicity, number of children, number of seniors, number of working-aged people, income-to-needs measure, and interactions between many of the demographics and the income-to-needs measure to account for dependencies between these factors.





Step 3: Adjust Post-Baseline Surveys for Attrition

Longitudinal Weights in Poverty Tracker

Longitudinal weights are computed from two components:

- The weight from the baseline survey
- The weight calculated to adjust for attrition between the waves
 - Logistic regression with response to the wave as outcome variable (0=no, 1=yes)
 - Predict probability of responding
 - attrition weight = Inverse the probability of responding

Using Poverty Tracker 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 wave				

Person-level weights: generalize to characteristics of adults in NYC...

"What percent of adults experienced X in 2015?" Apply qweight_p

Poverty-unit weights: generalize to characteristics of poverty units/families in NYC...

"What percent of families (poverty units) experienced X in 2015?" Apply **qweight_pu**





Replicate Weights

Replicate weights allow a single sample to simulate multiple samples, thus generating more informed standard error estimates that mimic the theoretical basis of **standard errors** while retaining all information about the complex sample design. These standard errors can then be used to obtain more precise confidence intervals and significance test. [Guide for Using Poverty Tracker Weights]

Wave	Person-level Replicate Weights	Poverty-unit Replicate Weights qweight_pu_rep1 qweight_pu_rep50		
Baseline	qweight_p_rep1 qweight_p_rep50			
3-Month	q1weight_p_rep1 q1weight_p_rep50	q1weight_pu_rep1 q1weight_pu_rep50		
and so on for each wave				



POVERTY & SOCIAL POLICY

at Columbia University



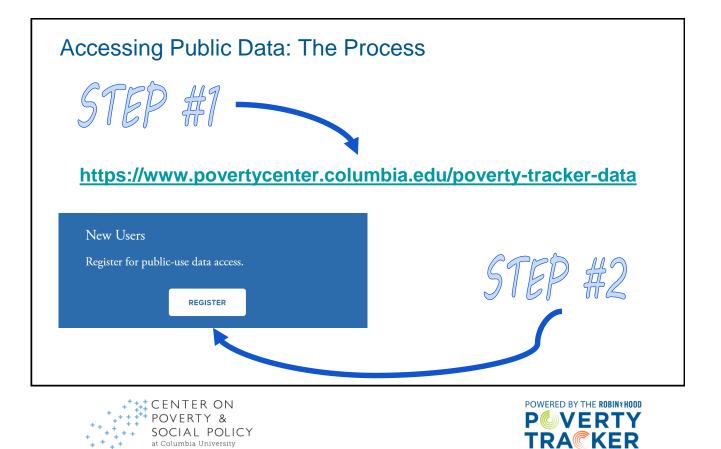
Public Data

Much of the data we've collected over the years is publicly available on the CPSP website.

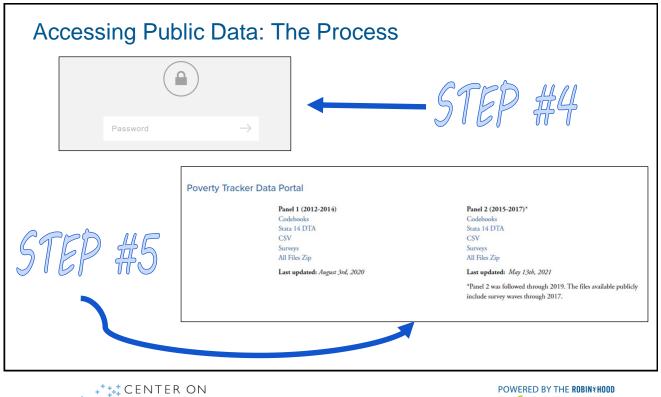
As of today, you can download the full two years of data from our first cohort (2012 - 2014) and the first four years of data from our second cohort (2015 - 2019). Data collection with the second cohort is ongoing.

Similar resources are available for both cohorts:

- Data files in both Stata 14 and CSV format. (Stata file includes labels.)
- Complete copies of all **survey questionnaires** administered to respondents.
- **Codebooks** for all survey waves that list the variables, full questions, answer options, and labels.



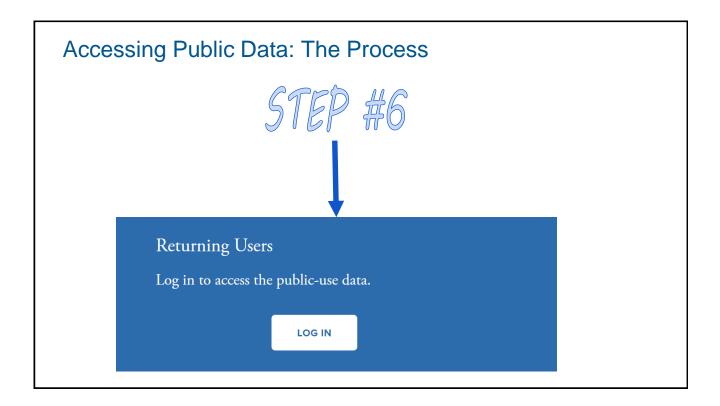
Accessing Pub	lic Data: Th	e Process	
	REGISTRATION - POVERTY TI	RACKER DATA	
registration or	New Users For any questions regarding the registration or data, please email povertytracker@columbia.edu	Name *	
		First Name Last Name Email Addross *	
		Affiliated Institution *	
		List your university, school, organization, or business; if none write "not applicable." Research Interests *	
		Describe what you would like to learn.	
		Other Notes Let us know any additional information pertinent to your research.	
		Privacy Agreement *	
		SUBMIT	

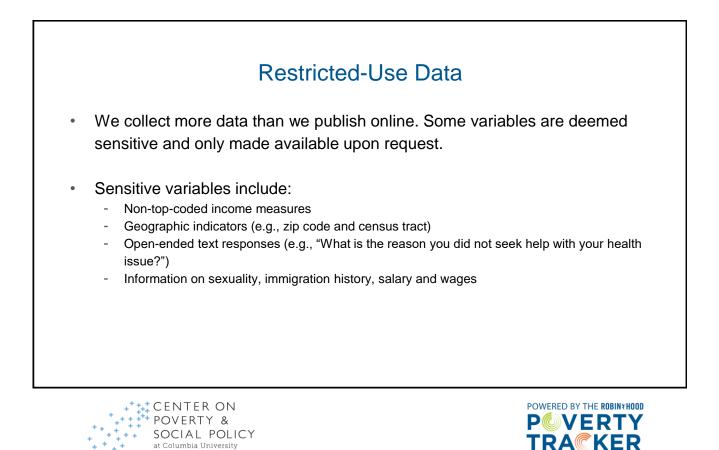


POVERTY & SOCIAL POLICY

at Columbia University

POWERED BY THE RUBIN RUUD POVERTY TRACKER





Accessing Restricted-Use Data: The Process

- 1. Let us know! Put your request in the public data registration form or shoot us an email at <u>povertytracker@columbia.edu</u>. Include a brief overview of your research project and variables-of-interest.
- 2. Set up a meeting with a member of the data team to discuss details, timeline, and next steps.
- 3. Obtain approval from Poverty Tracker project directors.
- 4. Submit a Data Usage Agreement (DUA) packet.

So what's a "DUA"?

- Extended Research Abstract
- Application for Obtaining Restricted Use Data (Cover Page)
- Restricted Use Data License Application
- Supplemental Agreement with Research Staff
- Restricted Use Data Protection Plan
- Copies of CVs for all Research Staff
- Proof of IRB Approval from Researcher's Institution (Proposal + Data Protection Plan)

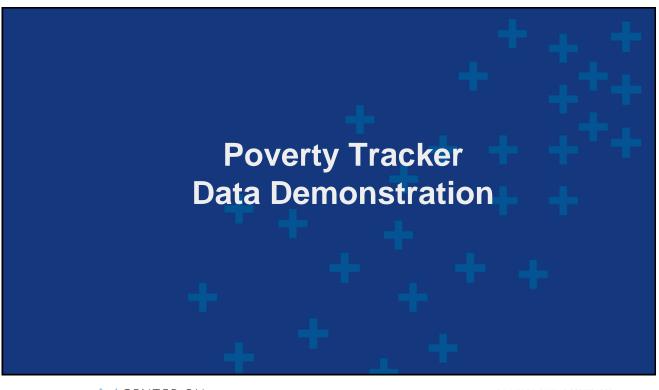
Don't worry!! We're here to help!





New Public Resources

- 2-Page Overview of the Poverty Tracker
- Comprehensive Guide to the Poverty Tracker
- Supplemental Guide to Using the Weights
- Technical Appendix on Imputation and Weighting
- Searchable Question Library Spreadsheet
 Includes all questions/variables in the public datasets
 - Organized by survey wave and topic







Poverty Tracker Data Demonstration

- Downloading the Poverty Tracker Data
- Pointers for using the Poverty Tracker data
- Producing summary statistics across survey waves

Example analysis:

What shocks are correlated with entrances and exits from poverty and hardship?

Downloading the Poverty Tracker Data

→ Go to the website: povertycenter.columbia.edu/poverty-tracker-data

Panel 1 (2012-2014 | Baseline through 24-month) Codebooks Stata 14 DTA CSV Surveys All Files Zip Last updated: August 3rd, 2020 Panel 2 (2015-2019 | Baseline through 48-month)* Codebooks Stata 14 DTA CSV Surveys All Files Zip Last updated: *May 17th, 2023* *The files available publicly include survey waves through 2019.

Supplementary Materials and Documentation

Poverty Tracker Two Page Overview Data User Guide Guide to Using the Poverty Tracker Weights Technical Notes on Imputation and Weighting





Pointers for using the Poverty Tracker data

Variable Naming Convention: Prefixes

Prefix indicating the survey on which the data stored in the variable was collected. Begin with a "*q*" for questionnaire, followed by a number corresponding to a particular survey.

Survey	Baseline	3-month	6-month	9-month	12-month	15-month	18-month	21-month
Prefix	q	q1	q2	q3	q4	q5	q6	q7
Survey	24-month	27-month	30-month	33-month	36-month	39-month	42-month	45-month
Prefix	q8	q9	q10	q11	q12	q13	q14	q15
Survey	48-month	60-month	72-month					
Prefix	q16	q17	q18					

Note: This table provides a guide to understanding the variable naming structure in the Poverty Tracker data. A "q" denotes each questionnaire, followed by the associated survey number.

Pointers for using the Poverty Tracker data

Variable Naming Convention: Suffixes

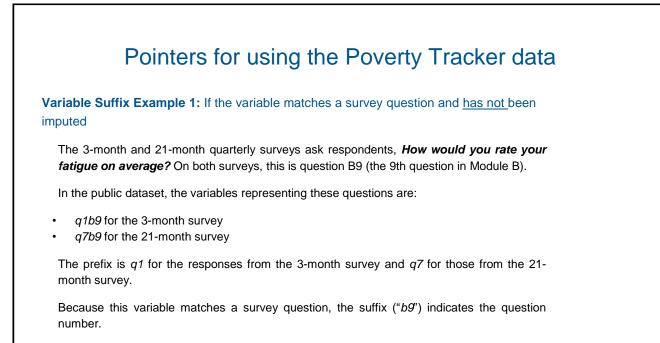
The variables also have a **suffix** that indicates one of three things:

- 1. if the variable matches a survey question and has not been imputed;
- 2. if the variable matches a survey question but was imputed; or
- 3. if the variable does not match a survey question, but was constructed based on responses to survey questions.

*You can also use the suffix to identify if it is a top-coded variable because you will see the suffix "_*tc*" at the end of the variable name.







Warning: Suffix does not always match question number - please consult codebooks and questionnaires!

Pointers for using the Poverty Tracker data

Variable Suffix Example 2: if the variable matches a survey question but was imputed

Imputes variables concerning income (all forms), hardship, health, assets, debts, or demographics.

Imputed variable names are not associated with their survey question number. They are descriptive.

In the public dataset, the variables representing disability income are:

- *imp_qincdis_tc* for the baseline annual survey
- imp_q4incdis_tc for the 12-month annual survey
- *imp_q8incdis_tc* for the 24-month annual survey ... and so on for each annual survey

You can see the complete list of imputed variables in **Poverty Tracker Technical Notes**.





Pointers for using the Poverty Tracker data

Variable Suffix Example 3: if the variable does not match a survey question, but was constructed based on responses to survey questions.

The Poverty Tracker data includes several variables *constructed* using responses to the individual survey questions (e.g. poverty status)

All of the constructed variables also have descriptive names.

For example the variables representing poverty status (measured using the SPM) are:

- *qspmpov* for the responses from the baseline survey
- *q4spmpov* for the responses from the 12-month survey
- *q8spmpov* for the responses from the 24-month survey... and so on for each annual survey

Questions about this research?

Contact:

General Inquiries, <u>povertytracker@columbia.edu</u> Yajun Jia, Survey Methodologist, <u>yj2648@columbia.edu</u> Schuyler Ross, Data Manager, <u>sr3820@columbia.edu</u> Sophie Collyer, Research Director, <u>smc2246@columbia.edu</u>

Acknowledgments

We thank Robin Hood for funding the Poverty Tracker, which provided data for this study. We are grateful to all members of the Poverty Tracker team based at Columbia, past and present! povertycenter.columbia.edu cpsp@columbia.edu twitter: @cpsppoverty

Join our emailing list for policy analysis updates and information about upcoming policy seminars.



