Data User Guide

Last updated: *July 27, 2023*

Poverty Tracker Data User Guide

Table of Contents

Introduction

There is no place quite like New York City. It's a large, diverse, and ever-changing city. That's why Robin Hood and Columbia University teamed up to rigorously study poverty, disadvantage, and the impact of anti-poverty strategies among those who call New York City home. This partnership gave rise to the Poverty Tracker, a unique dataset that captures the various dimensions and dynamics of economic insecurity in New York City. It allows researchers to explore how experiences of economic disadvantage and insecurity interact with housing, health, employment, discrimination, neighborhood characteristics, debt, consumption, and many other aspects of everyday life. The Poverty Tracker also allows data users to examine whether New Yorkers experience economic mobility and to identify possible barriers to their economic security. Notably, the Poverty Tracker samples are representative of the majority of the city's population, meaning that researchers can use the data to show how issues affect *all* New Yorkers, not just those living below or near the poverty line.

Conceptualized in 2010 and launched in 2012, the Poverty Tracker has come a long way. It has emerged as one of the nation's most comprehensive longitudinal studies of poverty and associated hardships. The frequency and flexibility of survey administration also allow for timely analysis of contemporary issues facing New Yorkers, such as COVID-19, systemic racism, and housing instability.

This document aims to provide an overview of the Poverty Tracker and its major components for all who are interested in the project and data. In turn, it discusses the Poverty Tracker cohorts and sampling, surveys, core measures, and data access/use. The **Poverty Tracker Technical Notes** provide more detailed information on our imputations and weighting methodology.

Cohorts and Sampling

The Poverty Tracker study surveys a representative sample of adult New Yorkers multiple times a year for up to six years. Before 2022, Poverty Tracker respondents were surveyed **quarterly.** In 2022, the study underwent a redesign whereby respondents are now surveyed **tri-annually.** Since initially launching, the study recruited five distinct samples of New Yorkers, referred to as study Cohorts 1 through 5 (See Table 1 for details about the study cohorts.) For all cohorts, the study recruited a majority of the sample through a Random Digit Dial (RDD) methodology. To ensure that the data is representative of the diverse experiences of all New Yorkers, the Poverty Tracker periodically includes supplementary samples of subgroups of New Yorkers in addition to the larger RDD sample. These subsamples of New Yorkers include those using social service agencies (Cohorts 1 and 2) and New Yorkers of Chinese origin (Cohorts 4 and 5). Below, we provide additional information about the various Poverty Tracker cohorts. See **Appendix A** for details about cohort retention.

Note on the timing of cohort recruitment and length of study participation

Table 1 lists the time of recruitment by study cohort, and Figure 1 depicts the length of cohort participation. Note that surveys are administered on a rolling basis as respondents are recruited. As such, a respondent may be recruited several months after initial cohort recruitment begins. For this reason, the length of time that a cohort is surveyed exceeds the designated time of study participation. Some surveys with Cohorts 2, 3, and 4 were also delayed due to the COVID-19 pandemic. For example, the study intended to survey Cohort 2 for up to 6 years; initial recruitment began in 2015, but it ended up surveying Cohort 2 through 2023 because of COVID-related delays and because some Cohort 2 respondents joined the study later in 2015 or in early 2016. **Survey administration dates for all respondents across all surveys are included in the public-use data files and can be used to identify the timing of survey administration**.

Table 1. Poverty Tracker Cohorts

Cohort	Initial recruitment year	Sample size at baseline	Subsamples	Notes
1	2012	2,228	2,002 RDD 226 social service user oversample	2-year design. Oversample of high-poverty zip codes.
2	2015	3,908	3,403 RDD (from CHS) 505 social service user oversample	6-year design introduced and carried forward for subsequent cohorts. Surveyed quarterly during the first 4 years and annually in the final 2 years.
3	2017	853	853 RDD	First rotating "refresher" cohort under rotating cohort design. Surveyed quarterly during the first 4 years and annually in the final 2 years.
4	2020	1,912	1,491 RDD 421 Chinese-origin oversample	Chinese-origin subsample introduced. Redesign changes made **after** 1st year of study participation: • Survey schedules of active cohorts are harmonized (cohorts 4 and 5).* • Respondents participate in three surveys per year instead of four.
5	2022	1,548	1,287 RDD 261 Chinese-origin oversample	Survey schedules of active cohorts are harmonized (Cohorts 4 and 5)* Respondents participate in three surveys per year instead of four.

Note: Table 1 summarizes each cohort's recruitment year, sample size, and relevant subsamples. It also includes additional details regarding these cohorts in the Notes column. **See Appendix A for additional sample size and retention information across all cohorts and survey waves.**

*See Tables 2 and 3 for an overview of the survey schedules before and after 2022.

Figure 1. Poverty Tracker Timeline

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

Cohort 1
Baseline launch year: 2012
Final survey launch year: 2014
Sample Size at Baseline: 2,228

Cohort 2
Baseline launch year: 2015 | Final survey launch year: 2023
Sample Size at Baseline: 3,908

Cohort 3
Baseline launch year: 2017 | Final survey launch year: 2025
Sample Size at Baseline: 853

Cohort 4
Baseline launch year: 2020 | Final survey launch year: 2027
Sample Size at Baseline: 1,912

Cohort 5
Baseline launch year: 2022 | Final survey launch year: 2028
Sample Size at Baseline: 1,548

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

Note: Figure 1 depicts the timing and length of participation in the Poverty Tracker study by cohort. See **Note on the timing of cohort recruitment and length of study participation** for additional details about the length of study participation. Survey administration dates for all respondents across all surveys are included in the public-use data files and can be used to identify the timing of survey administration.
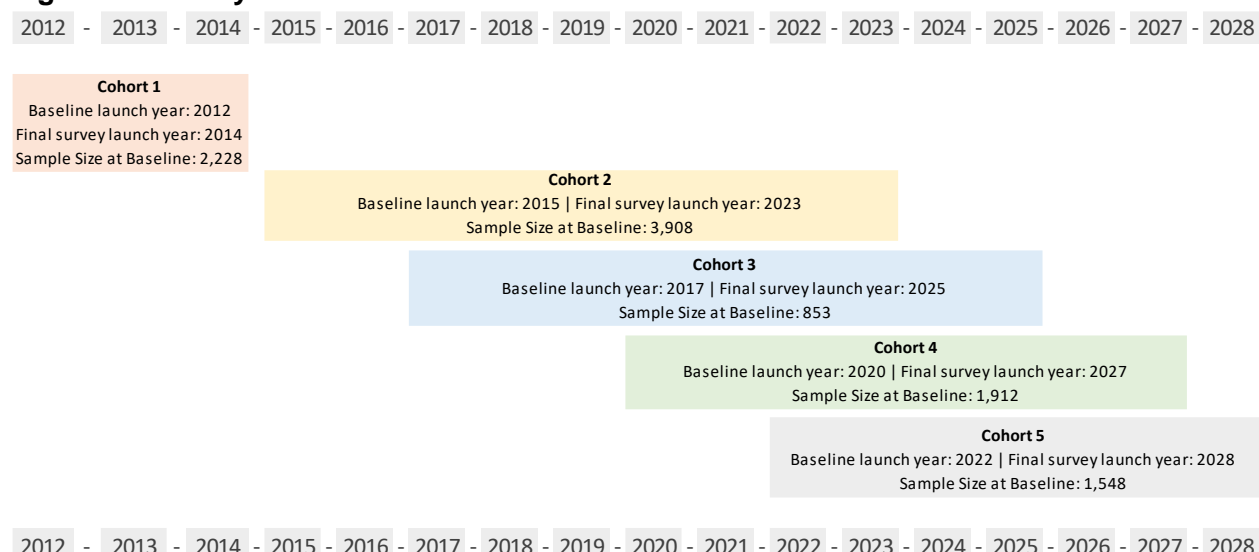
Cohort 1 | 2012 to 2014

Under its original design, the Poverty Tracker was a two-year longitudinal study of poverty and disadvantage in New York City. The first cohort (n=2,228) consisted of two samples. A survey research firm recruited the first subsample of 2,002 respondents using a Random Digit Dial (RDD) methodology.[1] The RDD recruitment strategy also included a method for oversampling New Yorkers in high-poverty zip codes (defined as zip codes where more than 20% of the population lived in poverty when measured using the Official Poverty Measure). Note that this was part of the overall RDD recruitment strategy for Cohort 1, so we cannot identify the specific number of participants recruited due to this oversampling method. The second subsample included a group of randomly selected individuals who used social service agencies funded by Robin Hood, such as soup kitchens and job training programs. With a sample size of 226, the social service user sample was smaller than the RDD sample; its purpose was to provide the study with an oversample of social service users who may be disproportionately low-income. Surveys with the first cohort were conducted in English and Spanish. After the survey research firm recruited the RDD sample, they delivered information about the respondents to the Poverty Tracker team at Columbia. The Columbia-based team oversaw all subsequent surveys.[2] The study followed this approach of having a survey research firm manage RDD recruitment and a Columbia-based team manage subsequent surveying for all cohorts recruited since Cohort 1. The Columbia-based Poverty Tracker team conducted recruitment of the social service user subsample.

---

[1] See: Dillman, D. A. (1978). *Mail and telephone surveys: The total design method* (Vol. 19). New York: Wiley.

[2] Surveys are conducted by the Columbia team online or by phone based on respondents' preference. Hard to reach respondents are mailed paper survey documents.

Cohort 2 | 2015 to 2023

The initial survey cycle with the first cohort concluded in 2014. Soon after, in 2015, surveys with a second cohort (n=3,908) commenced. Again the bulk of the sample was recruited by a survey research firm, but this time from a pool of approximately 10,000 individuals who had completed the NYC Department of Health and Mental Hygiene's 2015 Community Health Survey (CHS). The study recruited a total of 3,403 participants from the CHS sample. As with the first cohort, the study recruited a smaller subsample of 505 social service users to supplement the Cohort 2 sample.[3] Surveys with Cohort 2 were conducted in English and Spanish. Unlike in the initial cohort, the study surveyed Cohort 2 respondents for up to six years instead of two years. In their first four years of study participation, Cohort 2 completed four surveys per year. In their final two years, they completed one survey per year. See the **Note on the timing of cohort recruitment and length of study participation (page 4)** for additional details.

Cohort 3 | 2017 to 2024

2017 saw the move to a rotating cohort design. Rather than waiting for a cycle to conclude before recruiting the next cohort, the study team decided to recruit a new group of respondents every two years to achieve a steady-state sample of roughly 2,000 New Yorkers. Thus, multiple cohorts began to run concurrently. Figure 1 illustrates the overlap between the various cohorts. The study recruited a third cohort of New Yorkers (n=853) in 2017. Like previous cohorts, a survey research firm recruited Cohort 3 through an RDD design.[4] Surveys with the third cohort were conducted in English and Spanish.

Cohort 4 | 2020 to 2026

Recruitment of the fourth cohort began in 2020 (n=1,912). Like previous cohorts, the study recruited the bulk of this sample using an RDD methodology (n=1,491). Notably, this cohort included an additional oversample of New Yorkers of Chinese origin with a sample size of 421 (termed the Chinese-origin oversample). The intention was to capture better the experience of Asian New Yorkers, a group often underrepresented in social science research and one previously underrepresented in Poverty Tracker cohorts due to insufficient sample sizes and language limitations in surveying. The study recruited this oversample through both RDD and community sampling methods, specifically recruiting through WeChat community groups.[5] The study also began conducting phone surveys in Mandarin and releasing online surveys in both simplified and traditional Chinese to aid in recruiting this oversample.

---

[3] This group was recruited from a different set of social service agencies than those where outreach for the first cohort was conducted.

[4] The third and subsequent cohorts do not include a subsample of social service users, as the study team learned from the first two cohorts that after controlling for demographic characteristics, baseline poverty, and service utilization, the subsample of social services users was no more likely to experience persistent economic disadvantage compared to the RDD subsample.

[5] For additional information on these recruitment methods, please reach out to the Poverty Tracker team at povertytracker@columbia.edu.

Cohort 5 | 2022 to 2028

Recruitment of the fifth cohort (n=1,548) took place in 2022. Again, the study recruited the bulk of this sample through an RDD methodology (n=1,287), and this refresh cohort included an oversample of New Yorkers of Chinese origin recruited through both an RDD and a Respondent-Driven Sampling (RDS) method (n = 261).[6] Like Cohort 4, the study conducted phone surveys in English, Spanish, and Mandarin and released online surveys in English, Spanish, and simplified and traditional Chinese. The recruitment of this cohort also coincided with the transition to a redesigned survey schedule; the next section of this guide provides additional details on the redesigned survey schedule.

Survey Schedule

A key feature of the Poverty Tracker is its longitudinal design. By reaching out to families year after year and multiple times each year, the study team can determine whether respondents are moving out of poverty and hardship or falling into it. In addition, the study collects data on several factors related to economic stability and well-being, including physical and mental health, assets and debts, employment, and service use. The Poverty Tracker collects this rich data through a series of "annual" and "interim" surveys.

Survey Schedule for Cohorts 1, 2, 3, and Year 1 of Cohort 4

The initial survey completed by respondents after joining the study is called the **baseline annual** survey. The annual survey collects data on the study's three core measures: income poverty, material hardship, and health (see more information on these **core measures** on page 11. The questions in this survey are repeated annually thereafter on the Poverty Tracker *annual* surveys, providing insights into the dynamics of poverty and disadvantage in the city over time.

In addition to the annual surveys, respondents complete interim surveys on topics like wealth, employment, and service use. Table 2 highlights the general survey schedule as it applied to Cohorts 1, 2, and 3 as well as to Cohort 4's first year of study participation. These cohorts completed approximately four surveys per year (one annual and three interim surveys per year) for up to four years, passing through a similar survey schedule (Table 2). There were, however, some differences in the survey schedule across cohorts. For example, Cohort 1 respondents only completed two years of Poverty Tracker surveys, while the study surveyed subsequent cohorts for a longer period (six years). In their final two years of participation, respondents in Cohorts 2 and 3 only completed the annual survey. Cohort 4 only worked through this survey sequence for one year until the study underwent a redesign, discussed in the next section of this guide. Other differences in the surveys across cohorts are the inclusion of specific modules and questions in surveys for some cohorts and not for others. Table 2 highlights the topical modules completed by Cohorts 1, 2, and 3. However, data users should consult the codebooks and question library for additional details about modules and questions included in the surveys.

---

[6] For additional information on these recruitment methods, please reach out to the Poverty Tracker team at povertytracker@columbia.edu.

Survey Schedule for Cohorts 4 and 5 after 2022 Redesign

In 2022, the study underwent a redesign that applied to Cohorts 4 and 5 and will be the general study structure moving forward. The redesign included three major changes.

1. First, the survey cycle was changed so respondents complete three surveys per year instead of four.

2. Second, all active cohorts began to cycle through two years of surveys (named Year A and Year B) which then repeat in sequence once completing the first two-year cycle. This survey schedule differs from the previous structure, where the second two years of surveys did not mirror the first two years. Table 3 details the redesigned survey schedule.

3. And finally, all active cohorts (Cohorts 4 and 5 as of this writing) complete the same surveys at a similar time. The study made this switch so that data users can pool the interim survey data collected within the same calendar year.

It is important to note that Tables 2 and 3 **do not capture all** differences between surveys fielded to the various cohorts. For more information on when specific questions and modules were fielded, please refer to the Poverty Tracker question library, public use codebooks, and survey instruments. In addition to these resources, **Appendix B** documents other major topics covered in Poverty Tracker surveys (and in which surveys those topics were asked about).

Table 2. Poverty Tracker Survey Schedule for Cohorts 1, 2, and 3 and Cohort 4 Year 1*, **

	Annual Surveys	**Interim Surveys and Topics**		
Year 1 in Survey Cycle *Launch years* Cohort 1: 2012 Cohort 2: 2015 Cohort 3: 2017 Cohort 4: 2020**	**Baseline Annual:** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks	**3-Month Survey:** Adult Health	**6-Month Survey:** Service Use, City Services	**9-Month Survey:** Assets, Debts
Year 2 in Survey Cycle *Launch years* Cohort 1: 2013 Cohort 2: 2016 Cohort 3: 2018	**12-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks	**15-Month Survey:** Housing, Consumption	**18-Month Survey:** Work and Employment, Volunteering	**21-Month Survey:** Adult and Child Health, immigration, Sick Leave
Year 3 in Survey Cycle *Launch years* Cohort 1: 2014 Cohort 2: 2017 Cohort 3: 2019	**24-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks *(Last survey for Cohort 1)*	**27-Month Survey:** Housing, Assets & Debts, Sick Leave	**30-Month Survey:** Housing, Service Use, City Services, Employment and Wages	**33-Month Survey:** Housing, Present and Future, Fairness and Opportunity, Experiences of Discrimination, Religion, Minimum Wage
Year 4 in Survey Cycle *Launch years* Cohort 2: 2018 Cohort 3: 2021	**36-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks	**39-Month Survey:** Housing, Adult and Child Health & Well-Being, Caretaking, Minimum Wage	**42-Month Survey:** Housing, Work and Employment, Commuter benefits, Volunteering	**45-Month Survey:** Housing, Sleep, Education History, Student Loans, Adult Children, Family History
Year 5 in Survey Cycle *Launch years* Cohort 2: 2019 **Cohort 3: 2023**	**48-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks			
Year 6 in Survey Cycle *Launch years* Cohort 2: 2021	**60-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks			
Year 7 in Survey Cycle *Launch years* Cohort 2: 2023	**72-Month Annual Survey** Housing | Adult Health & Well-Being | Hardship | Employment | Poverty & Income | Personal Household Shocks			

Note: Table 2 outlines the survey topics that cohorts listed in column 1 completed at different periods of their study participation. Questions on income, poverty, hardship, and health are repeated annually.

*See the Poverty Tracker question library, codebooks, and survey instruments for additional details on survey content.

**Note that after Cohort 4 completed their first year of surveys, the Poverty Tracker underwent a redesign whereby active cohorts still in the field were put on the same survey schedule (see Table 3),

Table 3. 2022 Poverty Tracker Redesign for Cohorts 4 and 5*

	Annual Surveys	**Interim Surveys and Topics**	
Year A in Survey Cycle *Launch* Cohort 4: Late 2021 Cohort 5: Early 2022	**Year A, Survey 1 (Annual):** Income **|** Poverty Status | Material Hardship **|** Health	**Year A, Survey 2:** Employment, job search, job training, educational history, student loans	**Year A, Survey 3:** Housing conditions, adult health, caretaking, safety
Year B in Survey Cycle *Launch* Cohort 4: Early 2023 Cohort 5: Early 2023	**Year B, Survey 1 (Annual):** Income **|** Poverty Status | Material Hardship **|** Health	**Year B, Survey 2** *(tentative):* Assets and debts, alternative financial services, affording food, child health, child development, parenting, healthcare utilization	**Year B, Survey 3** *(tentative):* Service use, city services, fairness and opportunity, discrimination, immigration, family history

Note: Table 3 shows the survey sequence for Cohorts 4 and 5 after the 2022 redesign (see Survey Schedule for Cohorts 4 and 5 after 2022 Redesign). Note that for the first year of Cohort 4's survey participation, they followed the survey sequence Table 2. After its initial year, Cohort 4 entered the survey cycle beginning with Year A in this table. *Beginning in 2022, all active cohorts began to cycle through two years of surveys (Year A and Year B) which will repeat once completing the two-year cycle. All active cohorts (Cohorts 4 and 5 of this writing) complete the same surveys at roughly the same time in the calendar year.

Core Measures

The Poverty Tracker annually tracks three core measures of disadvantage:

1. **Income Poverty:** Annual resources (post-tax cash income plus in-kind benefits minus necessary expenditures for medical care and work expenses) that fall below a poverty line specific to New York City and based on the Census Bureau and Bureau of Labor Statistics' [Supplemental Poverty Measure](#).[7]
2. **Material Hardship:** Chronic or acute inability to make ends meet, e.g., running out of food or having utilities cut off for failure to pay bills). Below, see the five categories of hardship captured.
3. **Health Problems:** Self-reported poor health or a health condition that limits the kind or amount of work one can do.

Five Categories of Hardship

The Poverty Tracker measures of material hardship complement the poverty estimates, revealing another dimension of disadvantage that is often overlooked. These material hardship measures include:

1. **Food Hardship:** Often or sometimes running out of food or worrying food would run out without enough money to buy more.
2. **Bills Hardship:** Having utilities cut off because of a lack of money or falling behind on utility payments.
3. **Financial Hardship:** Often or sometimes running out of money between paychecks or pay cycles.
4. **Housing Hardship:** Having to stay in a shelter or other places not meant for regular housing, moving in with others because of costs, or falling behind on rent or mortgage payments.
5. **Medical Hardship**: Not being able to see a medical professional because of cost.

On annual surveys, the surveys ask about experiences of hardship in the past 12 months; interim surveys ask about a subset of hardships experiences in the past three months.[8]

In addition to the core measures, the Poverty Tracker includes a series of questions about changes and shocks that a respondent might have experienced, including:

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 guidance counselor

These questions are included on **all surveys** for the first three cohorts and are asked about in reference to the previous three months, tracking various changes in the lives of New Yorkers.[9]

---

[7] The study also collects all requisite information to measure poverty under the Official Poverty Measure.
[8] When respondents are hard to reach and have not completed an interim survey after an extended period of time, they are sent a "late" survey that does not include these questions about hardship.
[9] For the first three cohorts, "shock" questions are also omitted from late interim surveys (see footnote 8). For the fourth and fifth cohorts, the "shock" questions are not included on the interim surveys but are included on annual surveys with a reference period of the past 12 months.

Imputation and Weighting

Before the Poverty Tracker data can be analyzed as a representative sample, missing data for core annual measures must be imputed and the survey responses must be weighted. In any survey, some data will be missing due to subject non-response, unintentionally omitted due to subject or surveyor error (e.g., leaving a required question blank), or discarded due to error or implausibility in the response (e.g., saying that one's birth year was 1856). Imputation refers to retroactively filling in ("imputing") the missing data using a model that relates non-respondents to respondents. For the Poverty Tracker datasets, the study only imputes a subset of essential variables: those concerning income (of any form), hardship, health, assets, debts, and key demographics. See **Poverty Tracker Technical Notes** for additional information on the study's imputation methods.

Survey weights are used to account for the sampling, coverage, and non-response, and to calculate appropriate estimates of population parameters. Weighting is essential to ensure that statistical inferences drawn from the sample data apply to the general population. Poverty Tracker data is weighted to the American Community Survey (ACS) data[10] to create a sample representative of all New Yorkers ages 18 and over[11]. See the **Poverty Tracker Technical Notes** for additional information on the study's weighting methods for the first, second, and third Poverty Tracker cohorts.[12]

Accessing and Using the Data

Accessing the Poverty Tracker Data
As of this writing, data from Poverty Tracker Cohort 1 and the **first four years** of data from Cohort 2 are available for users to download.

Data users can access the data through the Center on Poverty and Social Policy's website: https://www.povertycenter.columbia.edu/poverty-tracker-data.

---

[10] The ACS is an annual survey conducted nationwide by the Census Bureau. The ACS samples between 26,000 and 27,000 households in New York City alone, comprised of approximately 60,000 individuals.
[11] Cohort 1 is weighted using the New York City samples from the 2011 to 2013 ACS; Cohort 2, using the 2014 to 2016 ACS, and Cohort 3, using the 2016 to 2018 ACS.
[12] For Cohorts 4 and 5, the weighting methodology are different since the study oversampled respondents with Chinese origin. Therefore, a separate technical note will be provided to discuss weights for the fourth cohort when finalized.

You'll first need to register as a user to access the data. Simply click on the **"REGISTER HERE"** button to do so.

New Users

Click button to register for
public-use data access.

REGISTER HERE

You'll be prompted to enter your name, email address, and some additional information in the form below.

REGISTRATION - POVERTY TRACKER DATA

New Users

For any questions regarding the
registration or data, please email
povertytracker@columbia.edu

Name *

First Name   Last Name

Email Address *

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 *
☐ I agree not to make any attempt to identify individuals in the datasets.

SUBMIT

Once you've completed the form, you'll be given a password to access the data. It's recommended that you store this password somewhere! Then in the future, you can click the **"LOGIN HERE"** button on the main Poverty Tracker data page and re-enter the password. If you don't store/remember it, you must re-register when you return to download the data.

Returning Users

Click login button to access
the public-use data.

LOGIN HERE

Once you enter the password, you'll be sent to the data portal containing the study's datasets and supplemental documents, including:

1) The data
2) The codebooks
3) The survey questionnaires
4) Additional supporting documentation, including the **Poverty Tracker Technical Notes**, the **Poverty Tracker Question Library**, and this **Data User Guide** (you can also find these documents on the Poverty Tracker data launch page)

Pointers for Using the Poverty Tracker Data

Here are some helpful pointers for using the Poverty Tracker data, including an overview of how the variables are named and a discussion of the different weights in the dataset.

Naming convention

All Poverty Tracker variables have a **prefix** indicating the survey on which the data stored in the variable was collected. Nearly all variables begin with a "***q***" for questionnaire, followed by a number corresponding to a particular survey. Table 3 maps the variable prefixes to the corresponding survey wave.

Table 4. Survey waves and variable prefixes for Poverty Tracker data

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.

The prefix also identifies if the variable was imputed. Imputed variables have the prefix "***imp_***" preceding the prefixes in Table 3.

The variables also have a **suffix** that indicates one of three things: (1) if the variable directly matches a survey question and has not been imputed;[13] (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.

The best way to understand the variable naming convention is with some examples. Let's take a look:

(1) Fatigue

The 3-month and 21-month quarterly surveys ask respondents, *How would you rate your fatigue on average?*[14] On both surveys, this is question B9 (the 9th question in Module B[15]).

In the public dataset, the variables representing these questions are:
q1b9 for the 3-month survey
q7b9 for the 21-month survey

You'll see that the prefix is *q1* for the responses from the 3-month survey and *q7* for those from the 21-month survey. Because this variable matches a survey question, the suffix ("*b9*") indicates the question number.[16]

(2) Imputed disability income

As discussed above, the study imputes variables concerning income (all forms), hardship, health, assets, debts, or demographics. Specific names are assigned to all of the imputed variables. These names are not associated with their survey question number. Instead, the variable names for imputed variables are descriptive. You can see the complete list of imputed variables in **[Poverty Tracker Technical Notes](#)**, but one example is disability income. All annual surveys ask about respondents' disability income, and this variable is top-coded in the public dataset.

In the public dataset, the variables representing these questions 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

---

[13] Variable names in the surveys may not line up exactly with those in the public datasets. Instead you should reference the codebooks to confirm the meaning of each variable in the dataset. Please reference the variable crosswalk document for information on how survey questions are mapped to final variable names.

[14] The response categories are: *None, Mild, Moderate, Severe, Very severe*

[15] To aid data users in identifying comparable variables across survey waves, most variables associated with the same survey questions have been assigned matching names in the public data. This means that the variable name does not always align with the survey question number. Please consult codebooks and the question library to confirm placement of questions in the relevant survey.

[16] See footnote 15 for additional details.

The **imp_q**, **imp_q4**, and **imp_q8** prefixes tell you that it's an imputed variable and what survey the data comes from. The suffix, **incdis_tc** in this case, is a descriptive name for the variable and lets you know it's top coded.

(3) SPM poverty status

The Poverty Tracker datasets also include several variables *constructed* using responses to the individual survey questions. For example, the variable indicating if a respondent is living in poverty draws on all of the annual survey income questions. All of the constructed variables also have descriptive names.

For example, in the dataset, the variables representing this 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

Types of weights

The Poverty Tracker dataset includes two types of weights: person- and family-level weights.

The person-level weights make the data representative of all adults (age 18 and above) in New York City in the year the sample was recruited (i.e., 2012 for the first cohort, 2015 for the second cohort, and so on). There are weights for every survey wave, which, like the other variables, can be differentiated by their prefix.

For example:

qweight_p is the person-level weight for the baseline survey

q1weight_p is the person-level weight for the 3-month survey

q2weight_p is the person-level weight for the 6-month survey… and so on for each survey

The family-level weights make the data representative of all families in New York City in the sample recruitment year. The only difference in the variable names between the person-level and family-level weights is that the family-level weights have the suffix **_pu** instead of **_p**. The suffix **_pu** stands for "poverty unit," which is essentially the family, though with a somewhat broader definition that matches the definition used in the SPM. For example, under the SPM, cohabiters and their relatives are treated equivalently to married couples rather than as separate family units (which they are under the Official Poverty Measure).

For example:

qweight_pu is the family-level weight for the baseline survey

q1weight_pu is the family-level weight for the 3-month survey

q2weight_pu is the family-level weight for the 6-month survey… and so on for each survey

Requesting Secure Data

In addition to downloading public-use data files, researchers and data users interested in working with sensitive or unreleased data can do so by requesting access and agreeing to a secure data contract. Sensitive variables of interest often include location data and non-top-coded variables, among other potentially sensitive data points.[17] The data access process involves contacting povertytracker@columbia.edu with a description of your project and the variables of interest. The initial step involves some back and forth with the Poverty Tracker team, who will help you navigate the available variables and set up a timeline for sharing specific data points.

Once finalizing the set of variables that you are interested in, the Poverty Tracker team will share the contract with you. You will need to specify information about the project, who will have access to the data, and the data security measures you will have in place. The study's Secure Data Officer will review the data security measures to ensure compliance with Columbia's standards. The rest of the contract will be reviewed by a Poverty Tracker representative and a representative from Columbia's Office of Sponsored Projects. In addition to submitting this contract, you will need to provide an extended abstract and obtain IRB approval for your project.

---

[17] Variables not included in public use data files because of risk of disclosure include: location data, non-top-coded incomes, expenses, assets, debts, age, open-ended text responses, country of origin, immigration status, family immigrations status/history, number of incarcerated family members, history of being stopped by police, places respondents sought help with personal/family issues, home value, year of home purchase, cost of home purchase, car make/model/value, drug use, occupation and industry, hourly wage rate/salary, focal child name, number of doctor visits in a year, height, weight, name of focal child's school, and number of children receiving free school lunch.

Appendices

Appendix A. Cohort Retention

Cohort retention rates are discussed below. Note that for all cohorts, a survey research firm conducted the initial recruitment of the RDD samples. For the first cohort, the survey research firm conducted the full baseline survey; for subsequent cohorts, the survey research firm conducted a short intake survey while the Columbia-based team conducted the baseline survey.

Cohort 1 (2012)

The survey research firm conducted the first cohort's baseline survey with 2,002 respondents recruited through the RDD. The Columbia-based team conducted face-to-face baseline surveys with 226 respondents in the social service user subsample at 14 Robin Hood funded agencies. After the baseline surveys, the Columbia-based team completed all follow-up surveys with both the RDD and social service user subsample. Of the 2,228 original respondents, 2,009 agreed to be re-contacted (1,789 from the RDD sample and 222 from the service-user sample). Of those 2,009 respondents, 72% completed their first follow-up survey (n = 1,442). Subsequently, for those who completed their first 3-month follow-up, retention rates were high for their second annual survey (83%) and third annual survey (79%). Of the 2,009 who agreed to join the cohort, 1,627 (81%) completed at least one follow-up survey. Table A1 lists the response rate by survey wave.

Cohort 2 (2015)

Respondents first completed New York City's Community Health Survey (CHS) administered by the Department of Health and Mental Hygiene. Of the 10,172 respondents who completed the CHS, 8,134 were offered the opportunity to participate in a short screener survey, and 4,752 did so (58%).

Of the 4,752 respondents that completed the intake survey, 3,403 completed the full baseline survey conducted by Columbia staff (72%). The Columbia-based team recruited 505 additional respondents from 28 Robin Hood agencies. These respondents did not complete an intake survey but instead completed their baseline survey at the time of recruitment. Altogether, Columbia staff conducted the second cohort's baseline survey with 3,908 total respondents. Retention rates were again high after their second annual survey (81%) and third annual survey (76%). Of the 3,908 respondents who joined the second cohort, 3,601 (92%) completed at least one follow-up survey.

Cohort 3 (2017)

The survey research firm completed an intake survey with 1,251 respondents, of whom 853 completed the full baseline survey conducted by Columbia staff (68%). As of this writing, surveys with this cohort are ongoing; Table A1 lists response rates for this cohort by survey wave for all surveys no longer in the field.

Cohort 4 (2020)

The study completed a total of 2,730 intake surveys with Cohort 4, but the survey research firm did not complete all intake surveys. The survey research firm completed 2,047 intake surveys in English and Spanish and 334 in Mandarin. The Columbia-based team completed Another 349 intake surveys with respondents of Chinese origin recruited through WeChat.

Of the 2,730 respondents who completed the intake survey, 1,912 (67%) completed the baseline survey:

- 1,491 of the 2,047 (73%) who completed their intake with the survey research firm in either English or Spanish,
- 196 of the 334 (57%) who completed their intake with the survey research firm in Mandarin, and
- 225 of the 349 (64%) recruited through WeChat.

Cohort 5 (2022)

The study completed a total of 2,175 intake surveys with Cohort 5, but again, the survey research firm did not complete all intake surveys. The survey research firm completed 1,883 intake surveys in English and Spanish. An additional 20 respondents called by the survey firm were forwarded to Columbia staff to complete an intake survey in Mandarin. The Columbia-based team conducted another 272 intake surveys with respondents of Chinese origin recruited through a Respondent-Driven Sampling (RDS) method.

Of the 2,175 respondents who completed the intake survey, 1,548 (71%) completed the baseline survey:

- 1,287 of the 1,883 (68%) who completed their intake with the survey research firm,
- 11 of the 20 (55%) Mandarin-speaking respondents forward from the survey research firm to the Columbia-based team, and
- 250 of the 272 (92%) recruited through the RDS method.

Table A1. Sample size by cohort and survey wave

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**
Intake screener	N/A	4,752*	1,251	2,730	2,175
Baseline	2,228	3,908	853	1,912	1,548
3-Month Quarterly	1,422	3,507	766	1,714	*Not applicable*
6-Month Quarterly	1,397	3,346	735	1,637	
9-Month Quarterly	1,367	3,254	718	1,582	
12-Month Annual	1,328	3,182	696	1,487	
15-Month Quarterly	1,300	3,108	677	*Not applicable*	
18-Month Quarterly	1,264	2,960	659		
21-Month Quarterly	1,260	2,931	649		
24-Month Annual	1,253	2,952	637		
27-Month Quarterly	*Not applicable*	2,836	636		
30-Month Quarterly		2,841	611		
33-Month Quarterly		2,820	570		
36-Month Annual		2,722	581		
39-Month Quarterly		2,711	571		
42-Month Quarterly		2,662	542		
45-Month Quarterly		2,610	517		
48-Month Annual		2,535	*Forthcoming*		
60-Month Annual		2,222			

Note: Table A1 lists response rates for this cohort by survey wave for all surveys that are **no longer in the field.**

*The service-user subsample in the second cohort did not take the intake survey. Instead, this group took the baseline survey at the time of recruitment.

Appendix B. Supplementary Topics Covered in Poverty Tracker Surveys

Table B1, below, highlights some additional topics covered in Poverty Tracker surveys and lists the surveys including these topics.

Table B1. Other topics covered in Poverty Tracker Surveys

Topic	Survey and Cohort
Affording Food	**Second cohort** (*surveys* – 9-month, 27-month) **Third cohort** (*surveys* – 9-month) **Fourth cohort** (surveys – 9-month)
Aging	**First cohort** (*surveys* – 18-month)
COVID-19	**Second cohort** (*surveys* – 60-month) **Third cohort** (*surveys* – 27-month, 30-month, 33-month, 36-month, 45-month, COVID standalone survey*) **Fourth cohort** (*surveys* – Baseline, 3-month, COVID standalone survey*)
EITC	**Second cohort** (*surveys* – 6-month, 9-month)
Government shutdown	**Second cohort** (*surveys* – 39-month, 42-month) **Third cohort** (*surveys* – 15-month, 18m-month
Health Care Consumption	**Third cohort** (*surveys* – 30-month, 45-month) **Fourth cohort** (*surveys* – 3-month, 14-month standalone survey**)
Minimum Wage	**Second cohort** (*surveys* - 6-month, 9-month, 15-month, 21-month, 27-month, 33-month, 39-month, 45-month) **Third cohort** (*surveys* – 3-month, 9-month, 15-month, 21-month, 27-month, 33-month, 39-month, 45-month) **Fourth cohort** (*surveys* – 3-month, 6-month, 9-month)
Moving/Evictions/Housing Quality	**Second cohort** (*surveys* 21-month, 33-month, 48-month) **Third cohort** (*surveys* 9-month, 21-month)
Paid Sick Leave	**First cohort** (*surveys* – 12-month, 18-month, 21-month) **Second cohort** (*surveys* – 9-month, 18-month, 21-month, 27-month, 39-month, 42-month) **Third cohort** (*surveys* – 9-month, 18-month, 21-month, 27-month, 42-month)

Policing	**Second cohort** (*surveys* – 48-month) **Third cohort** (surveys – 24-month, 33-month)
Student Debt	**Second cohort** (*surveys* – 45-month) **Third cohort** (*surveys* – 21-month, 45-month)

Note: Table B1 supplements Tables 2 and 3 and outlines additional topics covered in the Poverty Tracker surveys.

* The study administered a rapid-fire COVID standalone survey in the fall of 2020 to capture New Yorkers' experiences during the pandemic. This survey is not counted among the standard slate of interim surveys (3m, 6m, 9m, etc.).

** The study administered a short standalone survey on health and healthcare to Cohort 4 in spring 2022 as the cohort transitioned between the original survey schedule and the redesigned survey schedule (see Tables 2 and 3 for more details).

Appendix C. Record of Updates to the Data User Guide

July 27, 2023: Change the title for table A1 from "Response rates by cohort and survey wave" to "Sample size by cohort and survey wave."