

2023 New York City Housing and Vacancy Survey

Sample Design, Weighting, and Error Estimation

May 1, 2024

Table of Contents

1. Overview.....	1
2. Sample Design	1
2.1. Eligible Universe	1
2.2. Sampling Frames.....	2
2.3. Sample Selection	2
2.3.1. Nonrespondent Subsampling	3
2.4. Interviews and Response Rates.....	3
3. Weighting.....	6
3.1. Base Weight.....	6
3.2. Subsampling Adjustment	6
3.3. Nonresponse Adjustment.....	6
3.4. Ratio Adjustment Factors for Housing Unit Weights	6
3.5. Ratio Adjustment Factors for Person Weights	7
4. Nonsampling Errors	8
4.1. Coverage Error	8
4.2. Nonresponse Error	9
4.3. Measurement Error from Missing Responses to Questions	10
4.4. Quality Validity Error	11
4.5. Processing Error	11
4.6. Additional Considerations.....	11
5. Sampling Errors and Replicate Weights.....	11
6. Comparability of Data	14
7. References	15
Appendix A: Example of Ratio Adjustments.....	1

1. Overview

This document describes the sample design, weighting, and error estimation for the 2023 New York City Housing and Vacancy Survey (NYCHVS). The U.S. Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Disclosure Review Board (DRB) approval number: **CBDRB-FY24-POP001-0068**). The NYCHVS is sponsored by the New York City Department of Housing Preservation and Development (HPD) and conducted by the Census Bureau.

The City of New York is required by law to conduct a survey periodically to determine if rent regulations should be continued. A primary tool in this decision is the "*vacant available for rent*" rate, which is defined as the ratio of the vacant available for rent units to the total number of renter-occupied and vacant available for rent units for the entire city. The NYCHVS measures rental and homeowner vacancy rates, as well as various household and person characteristics. The design requires the standard error of the rental vacancy rate for the entire city to be no more than one-fourth of one percent if the actual rate was three percent.

2. Sample Design

The NYCHVS is a longitudinal survey that is conducted about every three years. The main sample of the survey is selected every decade, and additional new units are selected in each subsequent NYCHVS cycle. The NYCHVS was conducted in 2021 and in 2023. The main sample was selected using the 2020 July Master Address File (MAF); additional sample units were selected for 2023.

2.1. Eligible Universe

The universe of interest for the NYCHVS consists of all residential housing units (HUs) located within the five boroughs of New York City (Bronx, Brooklyn, Manhattan, Queens, and Staten Island). The principal exclusions are living quarters at locations that are classified as group quarters. These include:

- Correctional facilities,
- Mental health institutions,
- Hospitals,
- Military installations,
- Convents, monasteries, and rectories,
- Shelters, group homes, communes, and halfway houses,
- Home for the aged, disabled, homeless, or needy, and
- Dormitories for students or workers.

2.2. Sampling Frames

Since the 2023 NYCHVS is a longitudinal sample, the bulk of the 2023 NYCHVS sample consists of the same HUs from 2021 selected from that frame. For 2023, a new growth frame was constructed using the January 2022 MAF extract. This frame consists solely of the set of valid HUs that were added between July 2020 and January 2022.

2.3. Sample Selection

From the 2021 design, HUs were selected based on subsidized program participation and Condo or COOP status. The sampling design used two variables, Strata 1 and Strata 2, each of which had several categories within them, resulting in 20 cross-sectional strata. For 2023, HPD requested an increased sample within the Other/Rent-Stabilized cross-sectional strata. Census had initially selected an oversample of these units, however, so these could be added in easily by selecting the whole oversample from the initial 2021 design. Additionally, new growth was added throughout the sample. All new growth falls within the "other" program types below.

The sample sizes for each category are presented in Tables 2.1 and 2.2 below, first shown by the marginal counts in each variable in Tables 2.1 and 2.2.

Table 2.1 Strata 1: First Housing Type

Program Type	2023 Sample Count*
Affordable Owner	450
Affordable Renter	500
Condo	1,100
COOP	1,500
Other	10,500
Public Housing	1,100
Rent Control	350

Source: U.S. Census Bureau, July 2020 and January 2022 Master Address Files; U.S. Census Bureau, American Community Survey, 2014-2018 5-year file; New York City Department of Housing Preservation and Development administrative files.

*Sample counts have been rounded.

Table 2.2 Strata 2: Second Housing Type

Program Type	2023 Sample Count*
Mitchell Lama COOP	500
Mitchell Lama Renter	450
Other	8,100
Rent Stabilized	6,500

Source: U.S. Census Bureau, July 2020 and January 2022 Master Address Files; U.S. Census Bureau, American Community Survey, 2014-2018 5-year file; New York City Department of Housing Preservation and Development administrative files.

*Sample counts have been rounded.

For the new growth, a systematic random sample of housing units was selected across only those HUs introduced onto the MAF between July 2020 and July 2022, sorting housing units by

- Borough
- Sub-borough
- Median Income by Tract based on 2016-2020 ACS
- Tract number
- Block number
- Binary variable indicating building size as “big” or “small”
- Basic street address
- Unit designation

Note that most new growth cases have not yet had block numbers assigned to them, so Census used median income by tract instead of block for new growth only. No cases were assigned county values if missing tract sample this time – this was found to have no effect since the sample is already sorted by county.

2.3.1. Nonrespondent Subsampling

In the last month of interviewing a systematic subsample of nonrespondents was applied to address unit nonresponse. The subsample selected half of the remaining nonrespondents in order to focus contact attempts and resources to complete the survey. The weighting includes a subsampling adjustment at the cross-sectional strata level. For each cross-sectional strata, the subsample base weights were adjusted upwards to account for the base weights of the units not included in the subsample. This results in an appropriate distribution of the base weights to units only within the same cross-sectional strata.

2.4. Interviews and Response Rates

The total number of sample HUs selected for the 2023 NYCHVS was 15,611. Table 2.3 provides the weighted and unweighted response rates by borough, as well as the distribution of completed interviews and noninterviews.

Table 2.3. Interview Activity for the 2023 New York City Housing and Vacancy Survey

Borough	Unweighted Response Rate	Weighted Response Rate	Selected	Completed Interviews	Type A Non-interviews	Type B & C Non-interviews	Subsample
Bronx	77%	75%	2,709	1,811	549	24	325
Brooklyn	73%	69%	4,619	2,946	1,110	86	477
Manhattan	69%	64%	4,623	2,642	1,203	72	706
Queens	67%	65%	3,109	1,842	921	27	319
Staten Island	74%	72%	551	356	124	16	55
Total	71%	68%	15,611	9,597	3,907	225	1,882

Source: U.S. Census Bureau, 2023 New York City Housing and Vacancy Survey Public User Files.

Note: The data are subject to error arising from a variety of sources.

In past cycles (prior to 2017), the NYCHVS conducted proxy or last resort interviews, where a proxy interview consisted of interviewing a real estate agent, building manager, or someone else knowledgeable about the HU and a last resort interview involved accepting an abbreviated questionnaire as complete for reluctant respondents. Starting in 2021, NYCHVS no longer conducts proxy or last resort interviews, which resulted in higher noninterviews.

In 2023, Type A noninterviews included occupied housing units where the occupants:

- Refused to be interviewed,
- Unable to locate,
- Were not at home after repeated visits,
- Were unavailable for some other reason, or
- Absent due to COVID and unable to locate.

Type A noninterviews also include vacant units. In these cases, an interview was not obtained if no informed respondent could be found after repeated visits.

Type B noninterviews indicate an interview cannot be conducted due to structural circumstances related to the sample unit. Type B outcome codes occur rarely. Type B noninterviews indicate sample units not currently eligible for interview but could be eligible for interview in future years (meaning the sample address should be visited in future survey years).

Type C noninterviews only occur when the sample address cannot be found or the sample unit does not meet the definition of a housing unit. Type C outcome codes occur rarely. Type C noninterviews indicate sample units not currently eligible for interview and will not be eligible for interview in future years (meaning the sample address should be removed from the survey).

The response rate is calculated as the total number of interviews divided by the total eligible sample, which can be written as:

$$\text{Response Rate} = \frac{\text{Total Sample} - \text{Type A noninterviews} - \text{Type B\&C noninterviews}}{\text{Total Sample} - \text{Type B\&C noninterviews}}$$

Note that the weighted response rate just applies the base weight of each HU to the counts. For calculating response rates, enough of the interview had to be completed for it to be considered a valid interview. For vacant interviews, the entire interview must be completed, including type of vacant unit and reason unit is not available for rent or sale. For non-vacant interviews, all of the following must be answered to be considered as a completed occupied interview:

- Occupancy status,
- Tenure,

AND two of the following five items answered from the household roster for each person:

- Sex,
- Age,
- Relationship to householder,¹
- Hispanic origin, and
- Race.

If these criteria were not met, the sampled unit was classified as a Type A noninterview, following the definitions above.

For evaluation of interviews, a second interview was conducted of all vacant units and five percent of all occupied units. The questions asked during the reinterview included information about the previous Field Representatives that collected data, the time, date, and length of that interview, tenure, and vacancy status.

¹ In very rare situations, having only relationship to householder was considered sufficient if the persons could be verified as real people through other methods.

3. Weighting

To estimate HU and person characteristics for the 2023 NYCHVS, sample weights are calculated for each sample HU and each sample person. The final weight for each HU is the product of the following weights and adjustments:

3.1. Base Weight

The base weight is the reciprocal of the probability of selecting the unit. This is simply the inverse of the rate at which HUs are selected within the cross-sectional breakdowns in Table 2.3. Note that this sample design resulted in an unequal probability of selection for HUs.

3.2. Subsampling Adjustment

To address the subsampling applied later in interviewing, a subsampling adjustment was applied to the base weights at the cross-sectional strata level. For units closed out prior to the subsampling, the adjustment is simply 1.0. For the remaining units, units were included in the subsample or not included. Within each cross-sectional strata, the subsample base weights were adjusted upwards to account for the base weights of the units not included in the subsample. This results in an appropriate distribution of the base weights to units only within the same cross-sectional strata.

3.3. Nonresponse Adjustment

The subsampling-adjusted base weight of each interviewed HU was further adjusted to account for the eligible units that did not respond (Type A noninterviews). This nonresponse adjustment – called a noninterview adjustment factor (NAF) – was applied to interviewed HUs to account for Type A noninterviews. The factor was calculated using the weighted sums of interviews and Type A noninterviews divided by the weighted sum of interviews within groups of similar responding and nonresponding HUs:

$$NAF = \frac{\textit{Interviews} + \textit{Type A noninterviews}}{\textit{Interviews}}$$

A new method of calculating the NAF was introduced for the 2021 survey cycle and used again in 2023. This involved estimating the probability of response from the responding and nonresponding HUs and grouping HUs with similar response propensities together for this adjustment. Note that some nonresponding HUs were excluded from this modeling because they could not be found within the 2020 Census.

3.4. Ratio Adjustment Factors for Housing Unit Weights

New methods for implementing ratio adjustment factors (RAFs) within NYCHVS were also introduced in 2021 and used again in 2023. The HU weights were adjusted using two main sources of known totals:

- The July 2023 MAF
- Totals by program type from HPD administrative files

At each step in the ratio estimation procedure, the factors were equal to the following ratio:

$$\frac{\textit{Known Totals}}{\textit{NYCHVS Sample Estimate}}$$

The denominators of the ratios are equal to the sum of the weights of HUs (or persons) with all previous factors applied. Appendix A includes more information on the ratio adjustment factors and examples on how the process works.

The updated process creates three partitions of the sample based on borough membership, subsidized program status, and Condo/COOP, similar but not identical to the sample design. The weighting program then iterates through these three partitions until the RAF factor at each iteration stabilizes and final estimated totals of the groups within these partitions equal their known totals. The partitions are listed below.

- Partition 1: Affordable Owners, Affordable Renters, Remainder – City-wide
- Partition 2: Mitchell Lama Renter, Mitchell Lama COOP, Remainder – by Borough
- Partition 3: Public Housing, Condo, COOP, Remainder – by Borough

Note that Partition 1 was done city-wide while Partitions 2 and 3 were done by borough. “Remainder” means any HU not fitting into the other categories in the partition. Estimates of total HUs made of these particular program types at the specified level of geography match their known totals.

3.5. Ratio Adjustment Factors for Person Weights

When calculating person weights, the final HU weight was used as the base weight for each person, then we added a ratio adjustment to account for sampling variability and known coverage deficiencies (as described in Section 4.1.) for persons other than reference persons, spouses, or unmarried partners within interviewed households. We computed this factor within each borough by age, race, Hispanic Origin, and sex (200 cells), but reduced this number of cells to 89 based on how high the initial factors were and how few cases there were in some of the original 200 cells.

- The numerator of the ratio equaled the independent estimate of 2023 total persons for the cell minus the NYCHVS sample estimate of reference persons and spouses or unmarried partners. The independent estimates were projected based on 2020 Census person totals to estimate the population of NYC in May, 2023.
- The denominator of the ratio equaled the NYCHVS sample estimate of persons other than reference persons, spouses, or unmarried partners for the cell. The

person ratio estimate factor was applied only to the persons other than reference persons, spouses, or unmarried partners.

The ratio estimation procedures, as well as the overall estimation procedure, reduced the sampling error for most statistics in comparison to what would have been obtained by simply weighting the sample by the base weight.

4. Nonsampling Errors

All numbers from the NYCHVS, except for sample size, are estimates. As in other surveys, two types of general errors occur: sampling errors and nonsampling errors. Sampling errors are discussed in Section 5. The definition of nonsampling errors is—

Nonsampling errors arise mainly due to misleading definitions and concepts, inadequate sampling frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units, and so on. These errors are unpredictable and not easily controlled. Unlike sampling error, this error may increase with increases in sample size. If not properly controlled, nonsampling error can be more damaging than sampling error for large-scale household surveys.²

The various types of nonsampling errors are discussed in the following sections.

4.1. Coverage Error

Coverage errors arise from the failure to give some units in the target population any chance of selection into the sample (undercoverage) or giving units more than one chance of selection (overcoverage). To calculate the coverage, we used the sample base weight, which is the weight prior to any sample adjustments. The sample adjustments, described in Section 3, help to mitigate the undercoverage identified in this section.

The coverage rate is the ratio of the survey population or HU estimate of a group or an area and the independent estimate (or the known totals). The undercoverage rate is calculated as:

$$\text{Undercoverage} = \left(1 - \frac{\text{NYCHVS Sample Estimate}}{\text{Known Totals}} \right) * 100$$

Table 4.1a indicates the estimated undercoverage for HUs and overcoverage for persons. Based on the July 2023 MAF, we missed about four percent of the housing units in the five

² https://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf

boroughs. Overall, we oversampled about four and a half percent of the people in sample households.

Table 4.1a Overall Coverage Error for Housing Units and Persons

	Estimated from 2023 NYCHVS Base Weights	Known Total	Undercoverage*
Housing Units	3,556,099*	3,705,251	4.03%
Persons	8,378,053	8,014,065	-4.54%

Source: U.S. Census Bureau, 2023 New York City Housing and Vacancy Survey Public User Files.

Note: The data are subject to error arising from a variety of sources.

* Sum of base weights for interviews and Type A noninterviews.

× Negative values indicate overcoverage.

The within-household undercoverage varied by age, race, sex, and borough. Table 4.1b gives the undercoverage of the various race-sex groups for the city as a whole.

Table 4.1b Coverage Error by Race/Ethnicity-Sex Group

Race/Ethnicity-Sex Group	Undercoverage*
White & Other Females	-1.48%
White & Other Males	-5.48%
African American Females	2.71%
African American Males	7.66%
Asian Females	-29.31%
Asian Males	-30.97%
Hispanic Females	-4.32%
Hispanic Males	7.28%

Source: U.S. Census Bureau, 2023 New York City Housing and Vacancy Survey.

Note: The data are subject to error arising from a variety of sources.

* Sum of base weights for interviews and Type A noninterviews.

× Negative values indicate overcoverage.

We adjusted for this coverage error through the HU and person ratio adjustment factors. These factors adjust the sample weights to population totals provided by the Census Bureau, so the resulting final weight accounts for the coverage error identified in Tables 4.1a and 4.1b. For more information on the sample adjustment process, see Sections 3.3 and 3.4. NYCHVS data users do not have to take any additional steps to account for coverage error.

4.2. Nonresponse Error

Some respondents refuse the interview or cannot be located. The Census Bureau mitigated the error due to nonresponse by applying the noninterview adjustment factors into the

weighting process, as discussed in Section 3.2. NYCHVS data users do not have to take any additional steps to account for nonresponse error.

4.3. Measurement Error from Missing Responses to Questions

Some respondents participate in an interview but refuse to answer questions or do not know a particular answer. For certain questions, the Census Bureau imputes missing responses. When imputing, the Census Bureau tries to find households or persons with similar characteristics to fill in missing data. For each imputation, records are divided into 'donors' and 'recipients'.

For the demographic items, the Census Bureau first tries to impute based on other household information or household members. Every household must have some demographic information for each person in the household or it would be made a Type A. It is rare that a household is missing all demographic information for one item.

For imputing the housing items (including housing quality questions), units with similar characteristics are grouped. For example, when imputing contract rent, a unit with a similar year moved, year-built range, number of bedrooms, and input control status (stabilized, public housing, unregulated, etc.) is found and unit's contract rent is used to impute the recipient's contract rent. If no such unit can be found, contract rent is imputed based on the median value for units in 2021 (adjusted for inflation) with the same input control status in the particular borough. In 2023, contract rent was imputed in 8.7 percent of the renter-occupied units. The 2021 and 2023 NYCHVS surveys had similar variables, so when no donor was found during hot deck imputation, median starting values for the variable were taken from 2021 data (adjusted for inflation, as needed).

For economic items, such as income and employment status, the best possible match between donors and recipients is achieved through a statistical match with key items. The items used for matching donors and recipients are public assistance/non public assistance, borough, tenure, gender, race, ethnicity, age, relationship, education, worked last week, hours worked, type of work, government/nongovernment, and rent/value. All of these criteria are used to get the best statistical match possible. There are 33 income variables in 2023; in rare cases where a suitable donor was not found, income amount is imputed based on the median value of that income category. All donors and recipients have the same borough, tenure, and either receive public assistance or do not.

Appendix B provides the list of variables being imputed. Variables that can be used to determine imputation rates are in the public use files (PUF) and are defined on the record layout and PUF User Guide and Codebook. Variables shown in Table B1 are for all units, Table B2 for vacant units, Table B3 for occupied units, and Table B4 for persons. For example, using these variables from the PUF, users can see that summer gas and electricity

costs were imputed for 12.2 percent of occupied units, age was imputed for 4.9 percent of all persons, and stories was imputed for 0.1 percent of vacant units.

The Census Bureau does not know how close the imputed values are to the actual values. For other items, “not reported” is used as an answer category. NYCHVS data users do not have to take any additional steps to account for measurement error from missing responses to questions.

4.4. Quality Validity Error

In order to design a survey question that accurately measures the constructs of interest, the Census Bureau carefully tests each new survey question to ensure it is measuring the construct of interest. While the questionnaire is provided in multiple languages, sometimes the respondent does not speak those languages. In these cases, the interview must be rescheduled so that a field representative (FR) that speaks the same language as the respondent can administer the interview. Although some respondents might misinterpret questions, the Census Bureau does not have any additional information to estimate validity error rates. NYCHVS data users do not have to take any additional steps to account for validity error.

4.5. Processing Error

After the data are collected, errors that can be introduced include data capture errors and data editing and imputation errors. The Census Bureau carefully tests all aspects of the data capture and the editing and imputation procedures. Although mistakes are possible, the Census Bureau believes they are minimal. If a processing error is discovered, the Census Bureau will let NYCHVS data users know and, in some cases, will publish revised estimates. NYCHVS data users do not have to take any additional steps to account for processing error.

4.6. Additional Considerations

The NYCHVS is a longitudinal survey conducted about every three years. Many NYCHVS users compare current year NYCHVS with prior year estimates. Users should be aware that HPD and the Census Bureau often make changes to the text of various questions between surveys and sometimes to the underlying weighting methodology or sample design. NYCHVS data users comparing estimates with prior year surveys should consult the ‘Overview’ document on the NYCHVS website (<https://www.census.gov/programs-surveys/nychvs/about/overview.html>).

5. Sampling Errors and Replicate Weights

Sampling error is a measure of how estimates from a sample vary from the actual value. By the term “actual value,” we mean the value we would have gotten had all HUs been interviewed, under the same conditions, rather than only a sample.

Users of NYCHVS PUF should use replicate weights to estimate errors for any estimate. This is different from prior survey cycles in 2010, where the Census Bureau provided replicate weights as well as generalized variance function (GVF) parameters as an alternative method to estimate variance.³ GVFs are not available for this release and will not be made available for future releases of NYCHVS.

Variance estimation for surveys refers to the variation of an estimate due to selecting a sample from the set of all possible samples for a given sample design. To estimate the variance, multiple samples are needed but only one is observed. Replication allows small changes to a single probability sample to create a set of replicate samples, which can then be used to measure the variation of the estimates. Replication is done through subsets selected from the original sample in a process that mimics the original sample design. Each replicate sample, r , is then fully weighted, using the same process as the original sample, to ensure it represents the population of interest. This process forms the set of final replicate weights $\{w_r \mid r = 1, \dots, R\}$. Considering a particular estimate of interest, each replicate weight, w_r , can be used to create a replicate estimate $\hat{\theta}_r$. The set of replicated estimates $\{\hat{\theta}_r \mid r = 1, \dots, R\}$ represents the variability, or dispersion, of the estimate of interest under multiple samples of the population.

The Hadamard matrix was used to derive replicate factors to apply to individual full sample weights in creating replicate weights. Please note that for 2023 NYCHVS, the weights in Replicate 1 equal full sample weights, the weight used to derive sample estimates.

The 2023 NYCHVS uses a replicate variance estimator derived from a variance equation called the successive differences estimator. This estimator was first introduced by Fay and Train (1995) and then expanded for replication by Ash (2014). Using the replicated estimates, data users can calculate an estimated variance of an estimate of interest using the replication variance estimator:

$$\hat{v}(\hat{\theta}) = \frac{4}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2$$

where $\hat{\theta}$ is the weighted point estimate of the statistic of interest, such as a total, median, mean, proportion, regression coefficient, or log-odds ratio, using the weight for the full sample and $\hat{\theta}_r$ is the replicate estimate for replicate r of the same statistic using the

³ 2017 was the first NYCHVS survey cycle to include replicate weights upon release. 2011 and 2014 had replicate weights released after their initial release.

replicate weights. $\hat{\theta}_0$ is the full sample estimate. The value of 80 in $\hat{v}(\hat{\theta})$ is the number of replicates used (NYCHVS uses 80 replicates).

There are two sets of replicate weights. One set of replicate weights is used for computing standard errors of housing unit characteristics, and the second set is used for computing standard errors of person characteristics.

To calculate a standard error, the measure of dispersion when parameter estimates are calculated through repeated sampling from the population, obtain the square root of the variance estimate.

The following example illustrates how a statistic would be estimated, replicated, and combined to form a variance estimate. The goal of this example is to estimate the total number of renter-occupied HUs in Queens for 2023 and its corresponding estimate of variance.

For 2023, there are 1,065 completed interviews that are renter-occupied HUs in Queens (sample cases in Queens with responses to tenure status question as renters). Table 6.1 displays the first four and last one renter-occupied sample units in Queens. Note that the ordering in Table 6.1 is based on the variable *CONTROL*.

Table 6.1: Example of Estimating Variances with Replication of Renters in Queens

Sample HU	Full Sample Weight	Replicate Weights				
		Replicate 1	Replicate 2	Replicate 3	...	Replicate 80
1	640.2806	640.2806	187.8265	631.6027	...	643.2387
2	255.0803	255.0803	253.7466	251.3302	...	261.7359
3	62.6224	62.6224	62.8740	19.3173	18.0409
4	250.6189	250.6189	426.4146	250.5737	...	75.7650
...
1,065	380.6192	380.6192	379.3571	380.5505	...	392.8585

Source: U.S. Census Bureau, 2023 New York City Housing and Vacancy Survey Public User File.

Note: The data are subject to error arising from a variety of sources.

In NYCHVS, the full sample weight and the full sample estimate are referred to as replicate weight 0 (w_0) and replicate estimate 0 ($\hat{\theta}_0$), respectively.

Step 1: Calculate the full sample weighted survey estimate.

The statistic of interest is the total number of renter-occupied housing units in Queens for 2023. Add the full sample weights of the sample cases that meet your criteria of interest. Therefore, the total number of renter-occupied housing units in Queens is calculated as follows:

Full Sample Renter-Occupied HUs in Queens Estimate:

$$\hat{\theta}_0 = \sum_{i=1}^{1,065} w_{0,i} = 640.2806 + 255.0803 + \dots + 380.6192 = 481,021$$

Step 2: Calculate the weighted survey estimate for each of the replicate samples.

The replicate survey estimates are as follows:

$$\begin{aligned} \text{Rep 1 Rent-Occ Estimate } \hat{\theta}_1 &= \sum_{i=1}^{1,065} w_{0,i} = 640.281 + 255.080 + \dots + 380.619 = 481,021 \\ \text{Rep 2 Rent-Occ Estimate } \hat{\theta}_2 &= \sum_{i=1}^{1,065} w_{0,i} = 187.827 + 253.747 + \dots + 379.357 = 478,029 \\ \text{Rep 3 Rent-Occ Estimate } \hat{\theta}_3 &= \sum_{i=1}^{1,065} w_{0,i} = 631.603 + 251.330 + \dots + 380.551 = 490,030 \\ &\vdots \\ &\vdots \\ &\vdots \\ \text{Rep 80 Rent-Occ Estimate } \hat{\theta}_{80} &= \sum_{i=1}^{1,065} w_{0,i} = 643.239 + 261.736 + \dots + 392.858 = 473,243 \end{aligned}$$

Step 3: Use the replicate estimates $\hat{\theta}_r$ in the formula below to calculate the variance estimate for the total renter-occupied HUs in Queens.

$$\begin{aligned} \hat{v}(\hat{\theta}) &= \frac{4}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2 \\ &= 0.05 \times [(481,021 - 481,021)^2 + (478,029 - 481,021)^2 \\ &\quad + (490,030 - 481,021)^2 + \dots + (473,243 - 481,021)^2] \\ &= 0.05 \times [0 + 8,949,366 + 81,163,637 + \dots + 60,485,722] \\ &= 174,818,181 \end{aligned}$$

The survey estimate for total renter-occupied population in Queens is 481,021 housing units, with an estimated variance of 174,818,181 or a standard error of 13,221.9 housing units.

6. Comparability of Data

Data obtained from the NYCHVS are not necessarily comparable to other surveys or data sources due to differences in survey processes, question text, how the information were

obtained, mode of data collection, or timing of data collection, among other reasons. These differences are examples of nonsampling variability and are not reflected in the standard errors. Caution should be used when NYCHVS comparing results to other data sources.

The NYCHVS was redesigned for the longitudinal panel that began with 2021. The redesign included the selection of housing units for the new longitudinal panel, improved training on field operations, and revised questionnaire content. Notably, beginning in 2021, the survey was conducted via computer-assisted personal interviewing (CAPI), whereas previously it was a paper survey questionnaire completed by personal interview. The transition to CAPI allowed for response category edit checks and consistency checks among questionnaire items within the laptop instrument.

7. References

- Fay, R. E. and Train, G. F. (1995). Aspects and Survey and Model-based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. *Proceeding of the Sections on Government Statistics*, American Statistical Association, 154-159.
- Ash, S. E. (2014) Using Successive Difference Replication for Estimating Variances. *Survey Methodology*, Statistics Canada, Catalogue no.12-001-X Business Survey Method Division, Vol. 40, No.1, pp.47-59.

Appendix A: Example of Ratio Adjustments

This appendix provides one hypothetical example that demonstrates how the sample weights are ratio-adjusted so that they are consistent with a set of control totals.

For this example, assume weights were calculated for a sample, including all weighting adjustments up to a nonresponse adjustment. With these weights, totals by two categories – simply identified as A or B for Category 1 and C or D for Category 2 – were created. Table A1 summarizes the estimated totals resulting from the hypothetical sample and weights, and Table A2 shows the hypothetical control totals.

Table A1: Example Estimated Totals

Cat1\Cat2	C	D	Total
A	110	91	201
B	97	107	204
Total	207	198	405

Table A2: Example Control Totals

Cat1\Cat2	C	D	Total
A	115	105	220
B	95	105	200
Total	210	210	420

The control totals of Table A2 are used to improve the weights by making the estimates from the weights consistent with the control totals. Table A3 shows the Ratio Adjustment Factor (RAF) that will make the estimated totals consistent with the control totals.

Table A3: Example Ratio Adjustment Factors

Cat1\Cat2	C	D
A	115/110 = 1.0455	105/91 = 1.1583
B	95/97 = 0.9794	105/107 = 0.9813

If the factors from Table A3 are multiplied to the weights of the sample units, then the estimates from the revised weights will be consistent with the totals of Table A2.

For example, the ratio-adjusted weights for the combination of Category 1 = A and Category 2 = C is the product of the original weight and the RAF for the A/C combination:

$$\text{Ratio-adjusted weight} = \text{original weight} \times 1.0455$$

The ratio-adjusted weights for the other combinations of Categories 1 and 2 are calculated in the same way, using the corresponding RAF for each combination.