

## NYC HANES 2013-14 ANALYTIC GUIDELINES

The New York City Health and Nutrition Examination Survey (NYC HANES) 2013-14 dataset include data from five data collection streams:

- **CAPI (Computer Assisted Personal Interview)**
- **ACASI (Audio Computer-Assisted Self-Interview)**
- **PHYSICAL EXAM**
- **LABORATORY TEST RESULTS FOR URINE**
- **LABORATORY TEST RESULTS FOR BLOOD**

### Weighting Methods and Variables

To ensure that the 1,527 NYC HANES participants are representative of all non-institutionalized adult New York City residents age 20 years and older (as per the 2010 U.S. Census, N=6,285,749), statistical weights were created to:

- 1) Account for the *probability of selection* into the survey—the chance that a participant will be selected into the survey based on the probability of selection for the neighborhood or segment, the household, and the person(s) within each household.
- 2) Account for *non-response* to the survey—adjustment that attempts to reduce the bias that occurs when characteristics of survey non-respondents differ from those of respondents.
- 3) *Post-stratify* to reflect aggregate population demographics—this ensures that the weighted survey totals among demographic subpopulations match those of 2010 U.S. Census subpopulation totals.

Using these statistical weights, analysts can produce citywide estimates that accurately reflect the true population-wide burden. Five sets of weights have been constructed for the NYC HANES 2013-14 dataset.

- **CAPI\_WT**—The CAPI weight should be used to analyze participants' responses to interview questions, including demographic questions, questions about health and nutrition, and ACASI (self-interview) questions. All 1527 survey participants have a CAPI\_WT.
- **ACASI\_WT**—The ACASI weight should be used to analyze participants' responses to questions in the ACASI section including sexual behavior and drug use questions. 1260 participants have an ACASI\_WT greater than 0.
- **EXAM\_WT**—The EXAM weight should be used to analyze the body measurements collected during the physical examination: height, weight, waist circumference, blood pressure, and pulse. 1500 participants have an EXAM\_WT greater than 0.
- **BLOOD\_WT**—The BLOOD weight should be used to analyze the results of lab tests of blood: glucose, glycohemoglobin, total cholesterol, HDL, LDL (estimated), triglycerides, hepatitis B, hepatitis C, blood total mercury, lead, cadmium, serum cotinine, and serum creatinine. 1210 participants have a BLOOD\_WT greater than 0.
- **URINE\_WT**—The URINE weight should be used to analyze the results of lab tests of urine: mercury, albumin, creatinine, and albumin to creatinine ratio. 1450 participants have a URINE\_WT greater than 0.
- **SALIVA\_WT**—The SALIVA weight should be used to analyze the results of oral HPV testing; however, oral HPV lab results are not yet available for analysis. 1458 participants have a SALIVA\_WT greater than 0.

Each of the six weights represents the non-institutionalized population of New York City residents who are at least 20 years old, and each of the six weights sums to 6,285,749. The determination of the most appropriate weight to use for a specific analysis depends upon the variables selected by the data analyst. When an analysis involves variables from different components of the survey, the analyst should decide whether the outcome is inclusive or exclusive. For example, analysis of an outcome that uses two variables—one that was part of the health interview and another that was the result of a lab test on blood—should determine the following:

- If the outcome is inclusive (for example, a self-reported diagnosis variable from the CAPI **OR** an objective diagnosis based on lab test on blood), then the analyst should use the weight that corresponds to the largest component, which is CAPI\_WT (the most inclusive approach).
- If the outcome is exclusive (a self-reported diagnosis from CAPI **AND** a diagnosis based on the blood test), then use the weight that corresponds to the smallest component, which is BLOOD\_WT (the most exclusive approach).

Please see the sample SAS programs for examples of when and how to use weights, also available under the *Data* tab.

### Subsample Adjusted Weights

Although the dataset is generally weighted to account for participant non-response, analysts can consider further adjusting the survey weights to account for missing data in specific variables being used. For some variables, data are missing because of component or item non-response. For example:

- Some participants did not complete the ACASI component because of language or physical capacity issues, and thus did not provide data on illicit drug use or sexual behavior.
- Some participants did not provide an adequate blood sample to yield results for all of the laboratory tests.
- Some participants chose not to answer certain specific CAPI questions.

In these cases, participants who are missing data on a primary outcome of interest (for example, heroin use, hepatitis C) would not be included in the main analysis or in cross tabulations with other variables. The weighted sample size would be less than 6,285,749, which is the population of New Yorkers represented in the survey.

One approach to account for missing data due to item non-response is to adjust the weight to ensure that the subsample being analyzed reflects the target population of 6,285,749 New York City residents. The assumption underlying the adjusted weight is that selection bias has occurred, and readjustment based on age, gender, and race/ethnicity will minimize any effects of that selection bias.

To adjust the weighting for analysis of items with large non-response, analysts can create a subset or subsample excluding those with missing data, and can then compute appropriate

adjustment weights, assuming that response vs. non-response to the item within each age-gender-race/ethnicity cell is random. If response is thought to be related to another variable, for example, mental health status, then that variable should be used in addition to or instead of age, gender, and race/ethnicity.

Weights for subsamples may be obtained as follows:

1. Depending on the outcome of interest, determine which weight is appropriate: CAPI\_WT, ACASI\_WT, EXAM\_WT, BLOOD\_WT, URINE\_WT, or SALIVA\_WT.
2. Subset the file, so that the subset includes only those with responses to the outcome of interest.
3. Within each category of (1) age, (2) gender, and (3) race/ethnicity, sum the appropriate weights of those with responses. Call this **Sum B**.
4. Having chosen either the capi, acasi, exam, blood, urine, or saliva weights, use the appropriate control file, which include the sums for the appropriate age, gender, and race/ethnicity categories. Call this **Sum A**.
  - control\_capi
  - control\_acasi
  - control\_exam
  - control\_blood
  - control\_urine
  - control\_saliva
5. For each participant, adjust the weight by the adjustment factor variable **Sum A/Sum B** to create the new adjusted weight:
  - New weight = Old weight \* (adjustment factor variable **Sum A/Sum B**).

The new weight should sum to 6,285,749, the population of non-institutionalized New York City residents age 20 years and older, who are represented in the survey. Please see sample SAS programs for examples of how to create adjustment weights, available under the *Data* tab.

### Age Adjustment

Analysts can refer to the following articles for information on when and how to age adjust estimates:

- Age Adjustment Using the 2000 Projected U.S. Population, <http://www.cdc.gov/nchs/data/statnt/statnt20.pdf>.
- Age adjustment of diabetes prevalence: use of 2010 U.S. Census data, <http://www.ncbi.nlm.nih.gov/pubmed/24393518>

## Reporting Guidelines

Analysts are advised to consider the reliability of estimates that they are reporting. The relative standard error (RSE) of an estimate can be used to determine if an estimate is unstable. The RSE is equal to the standard error divided by the percent, and is often expressed as a percentage. Typically, an estimate with an RSE > 30% is considered unreliable, and a note is included in the report that the estimate should be interpreted with caution.

For further details on measures of reliability, including minimum sample size requirements, and on other issues concerning the analyzing NYC HANES data, please refer to the National Health and Nutrition Examination Survey (NHANES) website: Survey Methods and Analytic Guidelines ([http://www.cdc.gov/nchs/nhanes/survey\\_methods.htm](http://www.cdc.gov/nchs/nhanes/survey_methods.htm)), which includes the following reports:

- National Health and Nutrition Examination Survey: Analytic Guidelines, 2011-2012 [http://www.cdc.gov/nchs/data/nhanes/analytic\\_guidelines\\_11\\_12.pdf](http://www.cdc.gov/nchs/data/nhanes/analytic_guidelines_11_12.pdf)
- Previous versions that serve as supplemental guidelines include:
  1. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999-2010 [http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_161.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_161.pdf)
  2. The National Health and Nutrition Examination Survey (NHANES) Analytic and Reporting Guidelines [http://www.cdc.gov/nchs/data/nhanes/nhanes\\_03\\_04/nhanes\\_analytic\\_guidelines\\_dec\\_2005.pdf](http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf)