

New York City Health and Nutrition Examination Survey (NYC HANES) 2004 and 2013-14:

Part II – Weighting and Age Adjustment

Why do we need to weight our data?

- You can analyze the data as it is without weighing, but:
 - Your calculation will be applicable only to those who participated in the study, not to the entire adult non-institutionalized NYC population.
 - Your calculation might be biased. Some demographic groups were overrepresented in the sample, because of both unequal probability of selection and non-response.

Weighting

- NYC HANES data are weighted in order to compensate for unequal probability of selection

	2004	2013-2014
Baseline Population	American Community Survey (ACS), 2004 , Non-institutionalized NYC adult residents aged 20+ years	American Community Survey (ACS), 2013 , Non-institutionalized NYC adult residents aged 20+ years
Population Represented	5,827,719	6,285,749
Survey Weights	1. Clinic + Home: WTSF1CH 2. Clinic Only: WTSF1C 3. Fasting: WTSF1F	1. CAPI: CAPI_WT 2. Physical Exam: EXAM_WT 3. Blood Lab results: BLOOD_WT 4. Urine Lab results: URINE_WT 5. Saliva Lab results: SALIVA_WT
Sample Size for each set of weights	1. Clinic + Home: 1999 2. Clinic Only: 1861 3. Fasting: 1350 Each sample is a subset of the larger one	1. CAPI: 1527 2. Physical Exam: 1500 3. Blood Lab results: 1210 4. Urine Lab results: 1450 5. Saliva Lab results: 1458

2004 Weighting: Clinic + Home

- Variable name: **WTSF1CH**
- **2004 total sample size: 1,999**
- Populated for all respondents
- Should be used for components measured at the clinic OR in the home:
 - CAPI
 - Anthropometrics

2004 Weighting: Clinic Only

- Variable name: **WTSF1C**
- **Total sample size: 1,861** (subset of Clinic + Home)
- Populated for all respondents who went to the clinic
- Should be used for components measured at the clinic, including:
 - ACASI
 - CIDI
 - Laboratory tests (non-fasting tests only, for urine, serum, or whole blood)

2004 Weighting: Fasting

- Variable name: **WTSF1F**
- **Total sample size: 1,350** (subset of Clinic Only)
- Populated for all respondents who fasted for at least 8 hours before their blood test (some respondents were randomly assigned to fast, and other respondents fasted voluntarily)
- Should be used for components measured at the clinic which required fasting:
 - Fasting plasma glucose
 - *Can* also be used for lipids

More about 2004 Weighting

- Use the Clinic + Home weight if:
 - Your outcome of interest is in CAPI or anthropometrics
 - Examples: medical history, demographics, social support, Body Mass Index (BMI), waist circumference
- Use the Clinic Only weight if:
 - Your outcome of interest is only in ACASI or CIDI, or is only a non-fasting laboratory test
 - Examples: drug use, sexual behavior, incarceration, domestic violence, mental health
- Use the Fasting weight if:
 - Your outcome of interest is only fasting plasma glucose or fasting lipids (if your outcome is non-fasting glucose or non-fasting lipids, use the Clinic Only weight)

More about 2004 Weighting

- When an outcome is defined using variables that are collected in different survey components:
 - If the definition is inclusive (e.g., self-reported diagnosis from CAPI OR diagnosis based on blood tests), then use the weight that corresponds to the largest component (the most inclusive approach).
 - If the definition is exclusive (e.g., self-reported diagnosis from CAPI AND diagnosis based on blood tests), then use the weight that corresponds to the smallest component (the most exclusive approach).
- Reminders:
 - Clinic + Home (n=1,999) is largest
 - Clinic Only (n= 1,861)
 - Fasting (n=1,350) is smallest

Choosing a weight (2004): Example 1

- Outcome of interest: Depression
 - Component: CIDI
 - Weight: Clinic Only
- Stratified by: Age, height, and weight
 - Component: Anthropometric
 - Weight: Clinic + Home
- Since the outcome of interest is measured only at the clinic, you should use the Clinic Only weight for this analysis

Choosing a weight (2004): Example 2

- Outcome of interest: Cholesterol
 - Component: Laboratory
 - Weight: Clinic Only
- Stratified by: Fasting Plasma Glucose
 - Component: Laboratory
 - Weight: Fasting
- If you analyze cholesterol levels only among those who were tested for fasting plasma glucose, then you should use the **Fasting** weight. By using the Fasting weight, you will exclude all participants who did not provide fasting blood samples (the most exclusive approach).
- **BUT**, if fasting plasma glucose is just one part of your analysis, then you should use the **Clinic Only** weight. By using the Clinic Only weight, your analysis will include all participants with a valid cholesterol measurement, regardless of whether they also have a valid fasting plasma glucose (the most inclusive approach).

2013-2014 Weighting

- NYC HANES 2013-14 uses weights to compensate for unequal probabilities of selection
 - Weighted to the American Community Survey (ACS) 2013
 - Weights sum to total population of **6,285,749**
- 4 survey weights – different from 2004 because of differences in field logistics and fasting rules
 - CAPI weight: CAPI_WT
 - Physical weight : EXAM_WT
 - Blood Lab results weight: BLOOD_WT
 - Urine Lab results weight: URINE_WT
 - Saliva Lab results weight: SALIVA_WT

2013-14 Weighting: CAPI

- Variable name: **CAPI_WT**
- **2013 total sample size: 1,527**
- Populated for all respondents
- Should be used for the sections of the CAPI and ACASI components:
 - Demographics
 - Current Health Status, Physical Functioning
 - Occupation
 - Medical Conditions
 - Blood Pressure – **Self Reported**
 - Diabetes – **Self Reported**
 - Immunization
 - Smoking and Tobacco Use – **Self Reported**
 - Oral Health
 - Health Insurance, Hospital Utilization, and Access to Care
 - Physical Activity, Physical Fitness
 - Nutrition and Dietary Behavior
 - Weight History – **Self Reported**
 - Alcohol Use
 - Income
 - Reproductive Health
 - Sexual Behavior
 - Mental Health
 - Social Support

2013-14 Weighting: Exam

- Variable name: **EXAM_WT**
- **Total sample size: 1,500** (subset of CAPI)
- Populated for all respondents who completed at least one portion of the Physical Exam
- Should be used for measurements from the Physical Exam:
 - Height
 - Weight
 - Waist circumference
 - Blood pressure
 - Pulse

2013-14 Weighting: Blood

- Variable name: **BLOOD_WT**
- **Total sample size: 1,210** (subset of CAPI)
- Populated for all respondents who provided blood samples. If your outcome of interest requires the respondent to have fasted for at least 8 hours before providing blood, you need to adjust the weighting for non-response. Among those who provided blood, 96% fasted.
- Should be used for measurements from laboratory analysis of blood:
 - Diabetes: glucose, glycohemoglobin (A1c)
 - Heart disease: cholesterol, HDL, triglycerides
 - Infectious diseases: hepatitis B, C, and E
 - Environmental exposures: mercury, lead, cadmium, cotinine (nicotine byproduct)
 - Kidney function: creatinine

2013-14 Weighting: Urine

- Variable name: **URINE_WT**
- **Total sample size: 1,450** (subset of CAPI)
- Populated for all respondents with lab results from urine analysis
- Should be used for measurements from laboratory analysis of urine:
 - Environmental exposure: urine mercury
 - Kidney function: urine albumin to creatinine ratio

2013-14 Weighting: Saliva

- Variable name: **SALIVA_WT**
- **Total sample size: 1,458** (subset of CAPI)
- Populated for all respondents with lab results from saliva analysis
- Should be used for measurements from laboratory analysis of saliva:
 - Infectious disease: oral HPV
- **NOTE: No laboratory tests on saliva have been performed as of August 2016**

Recap: 2013-14 Weighting (1)

- Use the **CAPI** weight if:
 - Your outcome of interest is in CAPI or ACASI
 - Examples: medical history, demographics, social support, mental health
- Use **Exam** weight if:
 - Your outcome of interest was measured in the physical exam
 - Examples: Body Mass Index (BMI), Blood Pressure
- Use the **Blood** weight if:
 - Your outcome of interest was measured in blood or serum
 - Examples: glycohemoglobin (A1c), cholesterol, hepatitis C
 - If your outcome requires fasting prior to the blood draw (such as glucose), you will need to adjust the Blood weight
- Use the **Urine** weight if:
 - Your outcome of interest was measured in urine
 - Urine mercury or UACR (urine albumin/creatinine ratio)
- Use the **Saliva** weight if:
 - Your outcome of interest was measured in saliva
 - Oral HPV (not yet measured as of August 2016)

Recap: 2013-14 Weighting (2)

- When an outcome is defined by variables that are collected in different survey components:
 - If the definition is inclusive (e.g., self-reported diagnosis **OR** diagnosis based on lab tests), then use the most inclusive approach.
 - If the definition is exclusive (e.g., self-reported diagnosis **AND** diagnosis based on lab tests), then use the most exclusive approach.

Choosing a weight (2013-14): Example 1

- Outcome of interest: Breast cancer
 - Component: CAPI
 - Weight: CAPI_WT
- Stratified by: Age and weight
 - Components: CAPI (age) and Physical Exam (weight)
 - Weight: CAPI_WT and EXAM_WT
- Your analysis will include breast cancer outcomes for participants who were weighed and will also include outcomes for participants who reported their age. You should use CAPI_WT (the most inclusive approach).
- HOWEVER, if you decide to restrict your analysis only to those who have a valid weight measurement, then you should use the EXAM weight. By using EXAM_WT, you will exclude from the analysis all participants who did not take part in the physical (the most exclusive approach).

Choosing a weight (2013-14): Example 2

- Outcome of interest: Urine Albumin/Creatinine Ratio
 - Component: Urine and Saliva collection
 - Weight: URINE_WT
- Stratified by: Age, prior diagnosis of kidney disease, prior diagnosis of diabetes
 - Component: CAPI
 - Weight: CAPI_WT
- You will analyze only the participants who provided urine samples, and those who did not provide urine will be excluded from the analysis. You should choose the most exclusive weight, in this case **URINE_WT**.

Choosing a weight (2013-14): Example 3

- Outcome of interest: Hypertension
 - Components: Physical Exam and CAPI
 - Weights: EXAM_WT and CAPI_WT
- Stratified by: Exposure to nicotine
 - Component: CAPI (self-reported smoking)
 - Weight: CAPI_WT
- If you define hypertension either as the prior diagnosis of high blood pressure **OR** as measured high blood pressure, then use the most inclusive weight, in this case CAPI_WT.
- NOTE: If you restrict your analysis only to those with a valid blood pressure measurement, then you should use EXAM_WT. Be aware that by using the physical exam weight, you exclude those participants who did not complete the physical exam but who did report a prior diagnosis of hypertension.

Creating New Variables

- Check the “Recode programs” folder in the 2004 or the 2013-14 folder. For 2004, be sure to review the “Common Recodes” SAS program. Chances are the 2004 variable has already been created!
- Check to see if the variable you want to create for NYC HANES is available on EpiQuery. If so, the variable is already in the dataset.
- Recodes will sometimes use variables from more than one component. Use the correct weight for your new outcome.

Creating new variables: Skip patterns

- Some recodes need to be constructed using multiple variables. For example, current smoker in 2004:
- Have you smoked at least 100 cigarettes in your entire life? (Question SMQ020)
 - Yes (n = 826; skip to SMQ040)
 - No (n = 1157)
 - Missing (n = 16)
- Do you now smoke cigarettes? (Question SMQ040)
 - Every day (n = 313)
 - Some days (n = 164)
 - Not at all (n = 349)
 - Missing (n = 1173): Sum of SMQ020 No's and SMQ020 Missing's

Adjusting existing weights for item non-response (1)

- In many cases, data are missing, because survey participants did not respond to a particular item.
- **Survey participants who are missing data on the outcome of interest should be excluded from the analysis.**
- Survey participants who have complete data should be re-weighted, so that they can represent participants who are similar but have missing data.

Adjusting existing weights for item non-response (2)

- Weights are adjusted based on age, gender, and race/ethnicity.
 - Weight adjustment for non-response is an imputation method.
 - The assumption is that in each age category, gender category, and race/ethnicity category, the people who responded to the item are similar to those people who did not respond to the item.
 - If this assumption is not met, then adjusting for item non-response using weight adjustment is invalid, and you should use a different method for non-response analysis and the adjustment of estimates and standard errors. Choose a different imputation method based on the level of non-response and on why non-response occurred.

Adjusting existing weights for item non-response (3)

- Weights are adjusted so that:
 - 2004 weighted sample size = 5,827,719
 - 2013-14 weighted sample size = 6,285,749
- These weighted sample sizes are the total population of non-institutionalized New Yorkers ages 20 years and older, who are represented in the 2004 and 2013-14 surveys

EpiQuery outcomes and weights for 2004

VARIABLE	OUTCOME	WEIGHT
htnper2*	Hypertension prevalence	htnwt
hclper2*	High cholesterol prevalence	hclwt
dbtper2*	Diabetes prevalence	dbtswt
bmicat4^	BMI category	bmiadjwt
herpes2	HSV-2 Prevalence	hsvwt
hepcv02	HCV Prevalence	hcvwt
majdep2	MDD Prevalence	mhwt
elevld2	Elevated blood lead levels	metwt
elevhg2	Elevated blood mercury levels	metwt

*Prevalence estimates include both diagnosed and undiagnosed disease

^Pregnant women were excluded from this recode

Age adjustment

- Age adjustment is a method used to **compare** the same health outcome between two populations, or between the same population over time. It is:
 - Used to estimate what the prevalence would be if the age distribution of two populations were the same
 - A relative measure, not an actual measure of risk
- Age-adjusted estimates are only comparable to other age-adjusted estimates that use the same standard population.
- The [US 2000 Standard Population](#) is used to adjust NYC HANES data, both 2004 and 2013-14.

Weighting vs. age adjustment

- Weighting (**required** to obtain prevalence estimates): allows the survey sample to be representative of everyone in NYC
- Age adjustment (**not required**): facilitates the comparison of
 1. NYC estimates to estimates elsewhere
 2. Estimates between groups within NYC
 3. Estimates in NYC over time

Check your data for reliability

- Before reporting any numbers, you must:
 1. Calculate the relative standard error (RSE) and confidence interval for each estimate, and then;
 2. Label estimates according to our reliability guidelines if necessary.
- The relative standard error can be obtained by dividing the standard error percent by the total percent: $RSE = (se_percent/percent)$
- Guidance on how to do this is available at:
 - DOH Agency Share Drive:
 1. <R:\SUDAAN\Data Use Policy and Guidance\BES Data release policy and flowchart Jan 2013.pdf>
 2. <R:\SUDAAN\Community Health Survey\Analysis tips and data use policies\example code for calculating RSE and reliability.sas>
 - Open Access to the Public:
 1. http://www.nyc.gov/html/doh/downloads/pdf/episrv/bes_data_reliability.pdf

NHANES Guidelines

- The National Health and Nutrition Examination Survey (NHANES) Analytic and Reporting Guidelines:

http://www.cdc.gov/nchs/nhanes/analytic_guidelines.htm

Appendix

Some notes about the NYC HANES 2004 public use datasets

- Data are divided into 6 datasets:
 - SPFile – Demographic information and all weights
 - CAPI - Computer-Assisted Personal Interview
 - ACASI - Audio Computer-Assisted Personal Self- Interview
 - CIDI - Composite International Diagnostic Interview
 - EXAM
 - LABS
- Use unique identifier 'SP_ID' to merge datasets

2004 Public use datasets merge example

```
/*Download datasets and save somewhere*/

libname hanes 'C:\Data';

/*Each dataset must be sorted on merging variable
   SP_ID*/
proc sort data=hanes.spfile out=spfile; by sp_id;
proc sort data=hanes.capi out=capi; by sp_id;
proc sort data=hanes.exam out=exam; by sp_id;
data example;
  merge spfile (in=a) capi exam;
  by sp_id;
  if a; /*Chooses obs that are in the
        SPfile */

run;
```