NCHAT Weights

The National Couples' Health and Time Study (NCHAT) sampling weights minimize bias in survey-based estimates. To appropriately account for the design effect from weighting, and to obtain accurate standard errors, an analysis package that can take into account the sampling design should be used (e.g., Stata, MPlus).

Purpose of Weights

Virtually all survey data have sampling designs that result in the unequal probability sampling of population units. Datasets are weighted so that they more adequately represent the study's target population. Weights can correct for the over- or under-sampling of specific cases in the target population. For NCHAT, of the target population of individuals aged 20 to 60 who were cohabiting or married, respondents in same-gender unions were oversampled. Additional oversamples included Black, Latina/o/x/Hispanic, and Asian Americans, and respondents who were not heterosexual who were in a different-gender union. Further, weights can correct for non-response bias. Secondary data users vary in whether or not they apply weights to their estimates. **We strongly urge NCHAT users to weight their data for all analyses.**

About the NCHAT Weights

Weighting the NCHAT data was uniquely difficult due to the lack of high-quality, population-representative data in the U.S. that measures sexual identity, marital and cohabitation status, and gender identity of spouses or partners. The 2019 American Community Survey (ACS; United States Census Bureau 2019) was used to construct the same-gender and different-gender weighting targets and the National Health Interview Survey (NHIS; Center for Disease Control and Prevention 2019) was used to construct the non-heterosexual weighting targets – specifically to correct for the oversample of individuals who were not heterosexual but were in different-gender relationships. Weights were constructed for the target partner if more than one household romantic partner was reported.

The weights adjust the demographic distributions within same-gender and different-gender couples, and within heterosexual and non-heterosexual persons. The 2019 ACS population that fit the NCHAT target population (20 to 60 years old; cohabiting or married to same or different-gender partner) was used to subset the NCHAT data based on age, gender, sex, race, ethnicity, education, and marital status within same-gender and different-gender couples. Individuals in NCHAT who did not report a man or woman gender identity were set to missing and were imputed through multiple imputation to avoid very small groups and because the ACS does not fully capture gender identity. Because the ACS does not measure sexual identity, the 2019 NHIS was used to construct the weights by sexual identity for cohabiting and married couples aged 20 to 60. LGB in the 2019 NHIS was an identity-based measure about whether the respondent identified as lesbian, gay, bisexual, or something else. These data were used to construct weights to correct for the overrepresentation of bisexual individuals in different-gender couples. Again, the data were used to subset the NCHAT data based on age, gender, sex, race, ethnicity, education, and marital status within heterosexual and non-heterosexual persons.

After the base weights were constructed, they were adjusted using a multiple iterative raking procedure to post-stratify by age, sex, education, race, ethnicity, and marital status. For cases with missing data on raking variables, multiple imputation was used to impute missing values. Ten datasets were raked, and weights were averaged across the ten datasets to produce a single weight value for each case. The poststratification weighting was conducted in two stages. In the first stage, the heterosexual and non-heterosexual groups were weighted to the NHIS targets. In the second stage, the weights from the first stage were adjusted by same-gender and different-gender couple type using the ACS targets. The NHIS was used for the first stage and the ACS for the second stage because the NHIS had a small sample size of non-heterosexual individuals (< 400). The weights were trimmed. Couple weights were generated for respondents in which the main respondent and the partner both completed the main survey. A similar procedure was used. The time diary data was also weighted, again using a similar procedure.

More detailed information about the weighting process and how the demographic composition of the NCHAT sample compares to other population estimates can be found in the NCHAT Methodology Report on the NCHAT website.

Weighting Variables in NCHAT

There are currently four weights available in NCHAT, two for the main respondent and partner surveys and two for the main respondent and partner time diaries. Details about the weights and when to use each weight are described below.

Main Surveys

Weight_mainrespondent is the constructed weight variable to be used for the analysis of the main respondent file. This weight should be used when analyzing data from primary respondents and will generate nationally representative estimates of individuals aged 20 to 60 who are living or cohabiting with a partner. This weight can also be used to generate representative estimates when analyzing sub-groups, such as individuals who are in a same-gender or different-gender couple, or individuals who are gay, lesbian, or bisexual.

Weight_couple is the constructed weight variable to be used for the analysis of couple-level data from primary respondents and their partners. This weight should be used when analyzing the dyadic data (primary and partner responses). Cases from the primary respondents and partner respondents should not be treated as individual cases and combined to create a larger sample of individual adults. Weights are not valid for this purpose. Further, this type of analysis would require accounting for clustering within couples and potentially highly correlated responses (intraclass correlation). It is recommended that researchers do not take this approach.

Time Diary

Weight_Main is the constructed weight variable to be used for the analysis of the main respondent time diary file.

Weight_Couple is the constructed weight variable to be used for analysis of couple level data from primary respondents and their partners. This weight should be used when analyzing the dyadic data (primary and partner responses).

Using Weights

There are many statistical packages available that allow for the use of survey weights such as Stata, SAS, and MPlus. The current overview focuses on Stata, but we provide links to helpful resources for other statistical programs at the end of this section. In addition to applying weights, when performing analyses limited to a specific subpopulation (e.g., only parents, only same-gender couples), it is important to consider using subpopulation commands (in stata referred to as "subpop."

Stata

In Stata, you will use the svy command to fit statistical models for complex survey data. Detailed information about the svy command can be found by typing "help svy" into the command window or navigating to https://www.stata.com/manuals/svysvy.pdf

You will first need to set the survey weight using the command "svyset". If using data from the main survey for the main respondents only, set this command by typing:

 svyset Weight_mainrespondent [pweight=Weight_mainrespondent]

into the command window.

Once your survey weight is set, you then can use svy followed by any estimation command (e.g., "svy: regress…"; "svy: mean…"). For example:

svy: tab "variable name"
svy: mean "variable name"

We also recommend you use the "subpop" command in Stata to define separate populations of interest rather than the simply selecting the sample with a "keep" "drop" or "if" statement.

   Other helpful resources using Stata:
   ● Information about the subpopulation command in Stata can be found here or by typing "help subpopulation" in the Stata command window: https://www.stata.com/manuals13/svysubpopulationestimation.pdf

   ● Analyzing Complex Survey Data: Some key issues to be aware of: https://www3.nd.edu/~rwilliam/stats3/SvyCautionsX.pdf

SAS
   ● SAS Documentation on applying weights: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/proc/p0ctnx21fdgs7qn1jgolu1ihl7kf.htm

- SAS Blogs: How to understand weight variables in statistical analyses: https://blogs.sas.com/content/iml/2017/10/02/weight-variables-in-statistics-sas.html

MPlus
- Mplus Documentation: https://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_web.pdf