NSDUH: Technical Guidance for Analysts

The purpose of this document is to provide guidance on using NSDUH restricted-use data in the RDC setting.

Recommendations for Pre-analysis Preparation

- Review the data file code book and code book appendices to become familiar with how specific variables were derived and their limitations. There can be differences occasionally for year to year, and so the questionnaire and codebooks across the different years should be reviewed. The restricted-use variable comparability tracking sheet can be used for this purpose as well. The questionnaires are available on the CBHSQ website.
- Review the detailed tables to determine what estimates are generally output. Review the first release report(s) to determine how estimates are generally described within text. Replication of the text and standard errors will ensure correct analysis and interpretation of the NSDUH data. For a comprehensive list of all 2017 Detailed Tables, see https://www.CBHSQ.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf
- Review the statistical inference report and other methodological documentation from the Methodological Resource Book (see the link, below). https://www.CBHSQ.gov/data/sites/default/files/cbhsq-reports/NSDUHmrbStatInference2016.pdf
- Review NSDUH reports on the partial redesign of the NSDUH in 2015 in order to learn which variables had breaks in trend (see references below).

Analysis Guidelines

Below are guidelines on the appropriate analyses of NSDUH data.

Precision/Disclosure Suppression

- Full precision and disclosure based suppression (i.e. overwriting estimates) should be used unless otherwise requested.
- NSDUH Precision-based suppression rules should always be used (see the Statistical Inference Report).
- Disclosure Rules
  a. Unweighted sample sizes must be rounded to the nearest hundred.
  b. Any other unweighted data distributions should not be released.

Design/Nesting Variables

- VESTR and VEREP are the appropriate design/nesting variables that account for the complex design of the NSDUH data. VESTR VEREP should be used for 2002-2017 NSDUH data.

Using the Correct Degrees of Freedom (DOF)

- 2014-2017
  o   National Estimates: DOF=750
  o   California: DOF=144

- o   Florida, New York, Texas: DOF=120
- o   Illinois, Michigan, Ohio, Pennsylvania: DOF=96
- o   Georgia, New Jersey, North Carolina, Virginia: DOF=60
- o   Remainder 39 states: DOF=48

- 2002-2013
  - o   National Estimates: DOF=900
  - o   Small State Estimates: DOF=48
  - o   Large State Estimates: DOF=192
  - o   NOTE: Large States are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas

- DOF must be calculated in either of these two cases:
  (1) Confidence Intervals are requested.
  (2) Significance Testing is requested.

  Cell-Specific DOF (CSDOF) must be calculated for the 2 cases above if the tables present averages. Starting in 2014, estimates for averages are treated differently due to the possibility of smaller sample sizes and the relaxed suppression threshold on sample size for averages (minimum of 10 rather than 100 used for prevalence rates). CSDOF accounts for the reduced variance estimation strata.

  To calculate the CSDOF or degrees of freedom for geographic domains other than states, use the VESTR VEREP variables to identify variance strata and PSUs. In general DOF = #PSUs - #strata.

  Note there were sample design changes in 2005 and 2014, so when combining years across these time frames the DOF change. The change is additive, for example, combining 2013 and 2014 data to calculate confidence intervals for a California estimate will use DOF=336 (144+192).

  Note that a maximum of 900 degrees of freedom should be used. Also, in performing significance testing between levels of a certain variable (e.g. year groups) where that variable has multiple DOF values (e.g. certain year groups require doubling the DOFs and others don't) it is appropriate to use the lower DOF value for all levels of that variable. The lower DOF provides a more conservative test.

Standard Errors
NSDUH uses 2 types of methods for producing standard errors: Controlled and Uncontrolled.
- SE estimate calculated as product of SE Mean and Weighted Population Size (assumes domain size fixed). (Controlled)
- Estimate taken directly from SUDAAN. (Uncontrolled)

  Note that SETOTCALC only affects the SE of the numbers in thousands, so if a only percentage estimates are being created then SETOTCALC can be set to 1 or 2 - it does not

matter in this case. It is important to remember that the commonly used race variable, NEWRACE2, is already a two-way interaction of race and hispanicity.

This chart will help you determine if your domain is controlled or not:

Demographic and Geographic Domains Treated as Controlled from 2002-2017 Data		
Main Effects		
Age Groups (*12-17, 18-25, 26-34, 35-49, 50-64, 12 or Older, 18 or Older, 26 or Older, 35 or Older, 50 or Older, 65 or Older, all other combinations (ex. 12-34, 35-64, etc.)*) **Gender** (*Male, Female*) **Hispanicity** (*Hispanic, Non-Hispanic*) **Race** (*White, Black*) **Region** (*Northeast, Midwest, South, West*) **Division** (*New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific*) **State** (*50 US States*)		
Two-Interactions		
Gender	**X Hispanicity**	
Race	**X Hispanicity**	
Age Group	**X Gender** **X Hispanicity** **X Race** **X Region** **X Division**	
State	**X Age Group** (**12-17, 18-25, 26**+)*, all other combinations (12+, 12-25, 18+)* **X Gender** **X Hispanicity** **X Quarter**	

Subpopulations
- Use appropriate methods for subpopulation analyses (instead of using a subset of the data file, use a SUBPOPN statement in SUDAAN, for example).

Weights
- When creating estimates for single years of data, use ANALWT. When combining years of data to produce an annual average estimate over that time period, an adjusted weight variable will need to be created by averaging ANALWT over the number of years. e.g. An adjusted weight for combined 2012 to 2014 data (3 years)→ ADJWT3=ANALWT/3.

Missing data
- Use recoded/imputed variables when available. Guidance on how to accommodate missing data are available here: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUH%20Methods%20for%20Handling%20Missing%20Item%20Values%202018.pdf

Coverage
- When focusing on small special populations assess coverage, compare with authoritative external source when possible. Consider developing special weights as appropriate since the weights were not specifically targeted to control for all special populations. A forthcoming report on coverage of the NSDUH data may be requested from SAMHSA.

Trend Breaks
- Note that data prior to 2002 are not comparable to data from 2002 and beyond. There were larger breaks in trends in 2004, 2008 and 2015, with small breaks between years. Please refer to the variable crosswalk file (see link, below) to see variable comparability. Crosswalks and codebooks should be used in the determination of comparability.

- The 2015 NSDUH involved a redesign of the questionnaire that led to a break in trends for many variables. The NSDUH includes information on which variables can be trended and which cannot.

- References on the NSDUH trend breaks in estimates in 2015: For more detailed information, see the Summary of the Effects of the 2015 NSDUH Questionnaire Redesign. https://www.CBHSQ.gov/data/sites/default/files/NSDUH-TrendBreak-2015.pdf https://www.CBHSQ.gov/data/sites/default/files/NSDUHmrbQxRIAR2015.pdf

- For a comprehensive list of all 2015 variables along with trending status and new variable name (when applicable), see the variable crosswalk. Note that all trend breaking variables were given new variables name for 2015 and later data to acknowledge the differences across years. The crosswalks for the public use files are located in the link below: https://datafiles.CBHSQ.gov/study-series/national-survey-drug-use-and-health-nsduh-nid13517

For further questions on the appropriate analyses of NSDUH data, please send an e-mail request to the following address: rdca@cdc.gov

4