National Health and Nutrition Examination Survey:

Analytic Guidelines, 2011-2014 and 2015-2016

December 14, 2018

Table of Contents

1. Introduction

2. Data considerations

2.1 Sample design changes for the 2011-2014 and 2015-2018 samples

2.1.1 Race and Hispanic origin variables

2.2 Disclosure assessment

2.2.1 Age

2.2.2 Place of birth

2.2.3 Pregnancy status

2.3 Survey subsamples

2.3.1 Fasting subsample and oral glucose tolerance test subsample

2.3.2 Environmental subsamples

2.4 Geography

2.5 Season

2.6 Combining NHANES survey cycles

2.7 Missing data

2.7.1 Unit or sample person non-response

2.7.2 Component or item non-response

2.7.3 Value codes

3. Analytic considerations

3.1 Sample weights

3.1.1 Determining the appropriate sample weights for analysis

3.1.2 Subsample weights

3.1.3 Combining survey cycles

3.1.4 1999-2002 NHANES

3.1.5 Computing population counts

3.2 Variance estimation

3.2.1 Methods

3.2.2 Other sources of variability

3.2.3 Subsetting data

3.3 Statistical precision of estimates

3.3.1 Sample size
3.3.2 Relative standard error
3.3.3 Reliability of the estimated standard error and degrees of freedom
3.3.4 Confidence intervals
4. Conclusion
5. References

Text Tables

Table A. Unweighted sample sizes (percentages) by race and Hispanic origin for examined participants, NHANES 2007-2010, 2011-2014 and 2015-2016
Table B. Age-related variables on public data files, NHANES 2007-2010 and 2011-2014
Table C. NHANES 2011-2016 subsamples
Table D. Overall unweighted survey response rates for all ages, NHANES 1999-2016
Table E. Distribution of MEC sample weights by race and Hispanic origin and survey years
Table F. Formulae for constructing weights for NHANES

1. Introduction

This report presents information regarding the analytic and reporting guidelines for the NHANES 2011-2012, 2013-2014, and 2015-2016 publicly released data. This report builds upon earlier published guidelines[1], but also includes revisions to the general guidelines for the statistical presentation of proportions[2]. The report will be updated to include the 2017-2018 cycle of NHANES upon the release of the public-use data in 2019. The statistical guidelines in this document and earlier published guidelines are not standards. Depending on the subject matter and statistical efficiency, specific analyses may depart from these guidelines. In conducting analyses, the analyst needs to use her or his subject matter knowledge (including knowledge of methodological issues), as well as information about the survey design.

Design-based statistical methods are primarily used for NHANES data when producing NCHS reports and other analytic products. Design-based methods explicitly take into account features of the survey design, such as differential selection probabilities and geographic clustering. Model-based approaches have been used with sample surveys; however, the specific application of these methods relate to the analysis objectives and are beyond the scope of this document. In all analyses, the less a method incorporates the sample design, the more important it is to evaluate the results carefully and to interpret the findings appropriately.

In addition to these and earlier Analytic Guidelines, specific data file documentation can be found via the link next to the respective data file on the NHANES website[3]. This documentation is always the most up-to-date source of information about the variables on each data file. Although not anticipated, some information about specific variables contained in this report may be updated. In addition, another resource for all analysts is the series of NHANES Tutorials[4] — a Web-based product designed to assist users in understanding and analyzing NHANES data.

2. Data considerations

The 2011-2014 and the 2015-2018 NHANES four-year sample designs allow for the production of aggregate-level national estimates. Data files are publicly released in 2-year cycles and the survey content within those years is fixed to the extent possible. While annual NHANES samples are nationally representative, estimates for single-year data are relatively unstable (have large variance estimates) since NHANES can only survey a small number of locations each year. Importantly, although data are released in 2-year cycles, data from any one NHANES cycle may not be sufficient for certain analyses, such as those that examine subsamples or outcomes with low prevalence. For this reason, combining cycles into samples of four years or more is recommended whenever possible. However, please note differences in survey content between 2-year cycles and between sample designs when combining 2 or more cycles of data.

Not all NHANES data are publicly released. Some restricted data described below are only available through the NCHS Research Data Center (RDC)[5]. For example, in addition to low precision, releasing only one year of data increases the possibility of disclosing a sample person's identity; therefore, annual samples are only available in the RDC.

2.1 Sample design changes for the 2011-2014 and 2015-2018 samples

NHANES is designed to produce stable prevalence estimates for population subgroups (domains) defined by age group, sex, low-income status, and race and Hispanic origin. The subgroups of particular public health interest have changed over time. Oversampling is done to increase the reliability and precision of estimates of health status indicators for these population subgroups. Sample weights allow estimates from these subgroups to be combined to obtain national estimates that reflect the relative proportions of these groups in the population as a whole.

For NHANES 2011-2014, the design was changed to oversample Asian persons, in addition to the ongoing oversample of Hispanic persons, non-Hispanic black persons, older adults, and low income white and “Other” persons. The “Other” race subgroup included those who reported a race other than black, white, or Asian or those who reported more than one race. Specifically, the oversampled subgroups, also called domains, in the 2011-2014 survey were as follows:

- Hispanic persons;
- Non-Hispanic black persons;
- Non-Hispanic non-black Asian persons;
- Non-Hispanic white persons and persons of “Other” races at or below 130 percent of the federal poverty level; and
- Non-Hispanic white persons and persons of “Other” races aged 80 years and over.

In the 2015-2016 sample design, these same groups continued to be oversampled. However, the income threshold for the low-income white and Other persons group was changed from at or below 130 percent of the federal poverty level to at or below 185 percent of the federal poverty level.

For the 2011-2014, and the 2015-2018 sample designs, the Hispanic category included all persons who reported to be of Hispanic ethnicity regardless of race. The non-Hispanic black category included all persons who were reported to be non-Hispanic black race (single race or in combination with any other race including Asian). The non-Hispanic non-black Asian category (single race or in combination with another race except black) included all persons having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. All other persons not falling into the categories above were assigned to the non-Hispanic white and Other category. Therefore, any Asian person who was also Hispanic or non-Hispanic black was considered to be in the respective latter categories. Please note that in the publicly released file the race and Hispanic origin of each sampled person is categorized to include all races reported as well as Hispanic origin.

Prior to 2007, Mexican-American persons were oversampled rather than all Hispanic persons. The current oversampling of all Hispanic persons leads to sample sizes sufficient to produce reliable estimates for Mexican-American persons in addition to Hispanic persons overall. However, sample sizes are insufficient for calculating estimates for other Hispanic subgroups. NCHS recommends that researchers not calculate estimates for all Hispanic persons for survey periods prior to 2007 or for Hispanic subgroups other than Mexican American, in any survey cycle through 2018.

2.1.1 Race and Hispanic origin variables

Due to the change in the sample design since 2011, an additional race/Hispanic origin variable, RIDRETH3, was included on the 2011-2014 and 2015-2016 public-use Demographics data files. As with the race/Hispanic origin variable RIDRETH1, which is available on previous and current survey data releases, the Mexican American and Other Hispanic categories may include persons who also reported two or more race groups while the non-Hispanic white, non-Hispanic black, and non-Hispanic Asian categories include only those reporting a single race. All non-Hispanic persons reporting two or more races are in the “other race - including multi-racial” category. The weighted percent distribution of the sample across the four major race/Hispanic origin categories is benchmarked to the corresponding U.S. distribution of these groups based on population estimates from the American Community Survey (ACS). However, the weighted percent distributions of the individual subgroups within a race/Hispanic origin group, such as Chinese within Asian persons, are not aligned with the corresponding percent distributions in the population. **Table A** shows the sample distributions for RIDRETH1 for 2007-2010, and RIDRETH3 for 2011-2014 and 2015-2016.

It is important to note that there is a distinction between the sampling race and Hispanic origin categories (see Section 2.1) and the publicly released race and Hispanic origin categories described in this section (Section 2.1.1). For example, during the screening stage of the survey (i.e., the stage in which a participant’s eligibility for participation is defined), if a non-Hispanic sampled person reports being black as a single race or in combination with any other race they are categorized as non-Hispanic black. However, for data release, only persons reporting being single race black are retained in the non-Hispanic black category. Those who report being black and another race are moved to the “other” race category.

Table A. Unweighted sample sizes (percentages) by race and Hispanic origin for examined participants by survey cycle, NHANES 2007-2016

Survey Years	**Hispanic**		**Non-Hispanic**				
	Mexican American	Other Hispanic	white, single race	black, single race	Asian[1], single race	Other race[2], including multiracial	Total
2007-2008	2,064 (21.1)	1,147 (11.8)	3,969 (40.7)	2,141 (21.9)	n/a[1]	441 (4.5)	9,762 (100.0)
2009-2010	2,305 (22.5)	1,103 (10.8)	4,317 (42.1)	1,903 (18.6)	n/a[1]	625 (6.1)	10,253 (100.0)
2011-2012	1,316 (14.1)	1,011 (10.8)	2,841 (30.4)	2,582 (27.7)	1,215 (13.0)	373 (4.0)	9,338 (100.0)
2013-2014	1,685 (17.2)	930 (9.5)	3,538 (36.1)	2,198 (22.4)	1,019 (10.4)	443 (4.5)	9,813 (100.0)
2015-2016	1,837 (19.2)	1,232 (12.9)	2,948 (30.9)	2,052 (21.5)	986 (10.3)	489 (5.1)	9,544 (100.0)

[1]Race and Hispanic origin data for 2007-2010 are from the "RIDRETH1" variable and for 2011-2016 are from the "RIDRETH3" variable provided on the publically released Demographic Files for the respective years.
[2]Non-Hispanic Asian persons were included in the "other race" category prior to 2011.
[3]Other race is non-Hispanic persons who reported a race other than white, black, or Asian or reported more than one race.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 2007-2016.

Because the total sample size in any year is fixed due to operational constraints, the increase in the Asian sample size resulted in a decrease in the percent examined for the Hispanic and non-Hispanic white and other race groups (**Table A**). Also, because the sample design is for four years, it is not unexpected to have differences between the two 2-year cycles that make up the design, such as more Hispanic and non-Hispanic white participants and fewer non-Hispanic black and Asian participants in 2013-2014 than in 2011-2012.

2.2 Disclosure assessment

NHANES data collection adheres to the requirements of Federal Law. The Public Health Service Act (42 USC 242k) authorizes data collection and Section 308(d) of that law (42 USC 242m), the Privacy Act of 1974 (5 USC 552A), and the Confidential Information Protection and Statistical Efficiency Act (PL 107-347) prohibit NCHS from releasing information that may identify any respondent or group of respondents. As a result, data edits are made to some variables to reduce the risk of disclosure.

With the addition of the Asian oversample, and the public release of a more detailed race and Hispanic origin variable, additional edits were necessary to other variables, such as age and country of birth, that were previously released with earlier survey cycles.

2.2.1 Age

Similar to previous 2-year data release cycles, the 2011-2012, 2013-2014, and 2015-2016 demographic files include a variable for age in years at screening (RIDAGEYR) for all participants. Age at screening was used to determine eligibility for examination components and should be used for most analyses. Age at examination and age in months for children may be useful for some analyses. However, because exact age, in combination with other information, can pose disclosure risks, these variables were changed for the 2011-2012, 2013-2014, and 2015-2016 files.

Table B. Age-related variables on the 2-year public data files, NHANES 2007-2016

Variable name	Description	2007-2008 data file	2009-2010 data file	2011-2012 data files	2013-2014 data files	2015-2016 data files
RIDAGEYR	Age in years at screening (for persons aged 0-80 years)	Yes	Yes	Yes	Yes	Yes
RIDAGEMN	Age in months at screening (for persons aged 0-79 years)	Yes	Yes	Yes - for children 24 months or younger	Yes - for children 24 months or younger	Yes - for children 24 months or younger
RIDAGEEX	Age in months at MEC examination (for persons aged 0-79 years)	Yes	Yes	No	No	No
RIDEXAGM	Age in months at MEC examination (for persons aged 0-19 years at screening)	No	No	Yes	Yes	Yes
RIDEXAGY	Age in years at MEC examination (for persons aged 2-19 years at screening)	No	No	Yes	No	No

SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 2007-2016.

2.2.2 Place of birth

The place of birth variable has changed over the years. Prior to 2007, DMDBORN contained three categories: "Born in 50 U.S. States or Washington, DC," "Born in Mexico," and "Others". In 2007-2010, the variable DMDBORN2 was available with the publicly released data and included categories of "Mexico" "Other Spanish Speaking Country" and "Other Non-Spanish Speaking Country". In 2011, the variable DMDBORN4 became available on the publicly released data. DMDBORN4 has only two categories: "Born in 50 U.S. States or Washington, DC" and "Others".

2.2.3 Pregnancy status

Pregnancy status (RIDEXPRG) at the time of examination is determined using urine pregnancy test results and self-reported pregnancy status for females 8-59 years of age, in part, to determine eligibility for other exam components. Persons who report being pregnant at the time of examination are assumed to be pregnant (RIDEXPRG = 1). Those who report they were not pregnant or did not know their pregnancy status are classified based on the results of the urine pregnancy test. If the respondent reported ‘‘no’’ or ‘‘don’t know’’ and the urine test result was positive, the respondent was coded as pregnant (RIDEXPRG = 1). If the respondent reported ‘‘no’’ and the urine test was negative, the respondent was coded not pregnant (RIDEXPRG = 2). If the respondent reported she did not know her pregnancy status and the urine test was negative, the respondent was coded ‘‘could not be determined’’ (RIDEXPRG = 3). Persons who were only interviewed were coded RIDEXPRG = 3 (pregnancy could not be determined). As a result of sample design changes during 2007–2010 that reduced the number of pregnant women sampled, pregnancy status was publicly released only for women aged 20–44, to reduce disclosure risk.

2.3 Survey subsamples

NHANES participants may be included in a variety of survey components that are statistically defined (or random) subsamples of the entire NHANES interviewed or examined sample. These components include a variety of lab and environmental assessments. Each component subsample usually has its own designated sample weight, which accounts for the additional probability of selection into the subsample component, as well as additional non-response. **Table C** provides information on specific survey subsamples from NHANES 2011-2014 and 2015-2016, and some are described below. Specific subsample file documentation can be found via the link next to the respective data file on the NHANES website[3]. Importantly, the documentation will provide detailed information on the component, the eligible sample, the laboratory method used, and it will include analytic notes containing information on the specific subsample weights that the analyst should use.

Although the 24-hour dietary recall is not considered a subsample, participants who completed this component also have special weights that incorporate day of the week of recall.

2.3.1 Fasting subsample and oral glucose tolerance test subsample

Fasting sample weights can take one of three values: non-zero and non-missing, zero, or missing depending on a number of eligibility criteria. Specifically, sampled participants twelve years and older who were examined in a morning session, who had fasted 8-23 hours before their MEC examination, and who have valid plasma fasting glucose readings have non-zero fasting sample weights. All other sampled participants examined in a morning session have zero values for the fasting sample weight. Sampled participants examined in an afternoon or evening session have missing values for the fasting sample weight.

Participants who have non-zero and non-missing fasting sample weights can have one of three possible values for the oral glucose tolerance (OGTT) sample weight; non-zero and non-missing, a weight equivalent to their fasting sampling weight, or zero depending on a number of eligibility criteria. Participants who have non-zero fasting sample weights and had fasted at least 9 hours, who did not report that they were pregnant, and who did not report diagnosed diabetes were eligible for the oral glucose tolerance test (OGTT). Participants who completed the OGTT and have valid readings have non-zero and non-missing OGTT sample weights. OGTT results for women stating that they were not pregnant, but were later determined to be pregnant from the lab results, were included and these women have non-zero and non-missing OGTT weights. Diagnosed diabetics, who are ineligible for the OGTT, have a non-zero and non-missing OGTT weight equal to their fasting weight. All other participants in a morning session have zero values for the OGTT sample weights. Sampled participants examined in an afternoon or evening session have missing values for the fasting sample weight, and therefore will have missing OGTT sample weight.

2.3.2 Environmental subsamples

Some NHANES environmental analytes are obtained on a full sample of participants; therefore, full sample examination weights can be used for analysis. However, most environmental analytes are measured in 1/3 subsamples. These subsamples are labeled A, B, and C for convenience. These labels do not correspond to particular analytes and the subsample could differ for a particular analyte between survey cycles. For example, urine phthalates were measured for participants in Subsample A in 2011-2012 and for those in Subsample B in 2013-2014 and 2015-2016 so the name of the sample weight variable needs to be changed to analyze the combined four or six years of data.

For the 2011-2012 cycle, blood lead was measured on all participants over one year of age, while in 2013-2014 and 2015-2016 blood lead was measured for all participants ages 1 to 11 years and a random one-half subsample of participants ages 12 years and over.

The CDC National Report on Human Exposure to Environmental Chemicals contains additional information on the background, data content, public health uses, and interpretation of the NHANES environmental chemicals[6].

Table C. NHANES Subsamples from 2011-2016

Category	Years	Description

24-hour urine	2014	A one-half subsample of all examined adults aged 20-69 years were eligible. These data are only available in the RDC.
Blood lead	2011-2016	All participants age 1 year and older were included in 2011-2012. A one-half subsample of examined participants aged 12 years old and over and all examined children aged 1 – 11 years are included for 2013-2014 and 2015-2016.
Environmental chemicals A[1]	2011-2016	A one-third subsample of examined participants aged 6 years and over were included.
Environmental chemicals B[1]	2011-2016	A one-third subsample of examined participants aged 6 years and over were included.
Environmental chemicals C[1]	2011-2016	A one-third subsample of examined participants aged 6 years and over were included.
Fasting	2011-2016	Participants aged 12 years old and over who were examined in a morning sessions and fasted 8-23 hours were eligible.
Oral Glucose Tolerance Test (OGTT)	2011-2016	Participants aged 12 years old and over who were examined in a morning sessions and fasted 9-23 hours were eligible except for women who stated they were pregnant. Women stating that they were not pregnant, but were later determined to be pregnant from the lab results, were eligible and are included. Diagnosed diabetics were not given the OGTT, but non-pregnant diagnosed diabetics are included in the OGTT file with a weight equal to their fasting weight.
Smoking	2011-2012	Participants in environmental chemicals subsample A plus all examined adults aged 20 years and over who were current smokers were included[2].
Smoking	2013-2016	Participants in environmental chemicals subsample A plus all examined adults aged 18 years and over who were current smokers were included[2].
VOC smoking	2011-2012	Participants in VOC subsample plus all examined adults aged 20 years and over who were current smokers were included[2].
VOC smoking	2013-2016	Participants in VOC subsample plus all examined adults aged 18 years and over who were current smokers were included[2].

Volatile organic compounds (VOC)	2011-2016	A one-half subsample of examined participants aged 12 years and over are included.

[1]Participants are randomly assigned to one of the three mutually exclusive 1/3 environmental subsamples. The analytes in each of the three subsamples vary by survey cycle. The names Subsample A, Subsample B, and Subsample C are used for convenience and are not based on the tested analytes. The proper subsample weights attached in the dataset should be used for analysis. As the same analytes might be in different subsamples in different survey releases, it is important to check the weight variable names, and rename if necessary, when combining cycles.
[2]Smokers are those who reported (in the interview) to have smoked more than 100 cigarettes in their life and reported being a current every day smoker.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 2007-2016.

2.4 Geography

Since 1999, NHANES has interviewed and examined a nationally representative sample of approximately 5,000 persons each year in counties across the country. During a single survey year, about 15 counties are selected out of approximately 3,100 counties in the United States. NHANES was not designed to produce regional or sub-regional estimates and no geographic data are released on the publicly available data files to protect the identification of NHANES participants.

However, research files for Los Angeles county and the state of California for two sets of combined survey cycles, 1999-2006 and 2007-2014[7,8], have been created and are available in the NCHS RDC. These files include sample weights and design variables developed for Los Angeles and California for the two eight year periods.

Other geographic data are available through the RDC. The U.S. Department of Housing and Urban Development (HUD) assign geographic codes (geocodes) to the NHANES data for analytical use in every two-year cycle. HUD geocodes include the following information: 1) census block group, census tract, county, state, and all other census codes normally provided by the HUD Geocoding Service Center for each residential address, and 2) latitude and longitude for each residential address[9,10].

Appropriate sample weights and design variables for sub-national estimates are not provided and would have to be developed by the analyst.

2.5 Season

The variable RIDEXMON, in the public release Demographics File provides the six-month timeperiod when the examination was performed and is categorized into two groups: November 1 through April 30 and May 1 through October 31. Due to operational considerations, the geographic scheduling of the MEC is restricted by consideration of weather. MEC operations avoid certain geographic areas during the winter. Thus, the statistical efficiency of the sample is diminished for any variable that may be related to seasonal variation that differs by region of the country. Most NHANES variables are not affected by season; however, this determination would need to be made by the analyst in the context of a specific research objective.

2.6 Combining NHANES survey cycles

Each 2-year cycle and any combination of 2-year cycles is a nationally representative sample. However, sometimes the sample size of a particular analytic cell is too small based on one 2-year cycle to produce statistically reliable estimates. The NHANES sample design makes it possible to combine two or more cycles in order to increase the sample size and analytic options. In general, any two-year data cycle in NHANES can be combined with adjacent two-year data cycles to create analytic data files based on four or more years of data to produce estimates with greater precision and smaller sampling error. However, when combining cycles of data, it is extremely important to:

1. be aware of sample design changes,

2. verify that data items collected in all combined years are comparable in wording, methods, and inclusion/exclusions (e.g., eligible age range),

3. select the proper weight to use for the combined dataset (see section 3.1), and

4. examine the inherent assumption of no trend in the estimate over the time period being combined.

2.7 Missing data

NHANES, like most population-based sample surveys, experiences both participant (unit) and component (item) non-response. In a statistical sense, non-response can be considered ignorable or non-ignorable. If the data are missing at random and the characteristics of the non-respondents are similar to the characteristics of the respondents, the non-response can be considered ignorable. However, non-respondents may have significantly different characteristics than respondents. In this case, the non-response mechanism may be non-ignorable with respect to the data analysis. Ignoring non-response in this case leads to biased estimates.

2.7.1 Unit or sample person non-response

Not all persons selected in the NHANES sample were interviewed and not all interviewed persons were examined. Unit or participant non-response, the failure to obtain any information on an individual selected to participate in an NHANES survey, can occur both at the interview and at the examination phase of the survey. Non-response bias resulting from this missing data can be an important source of survey error.

Like a number of other national probability based face-to-face surveys, NHANES has been experiencing a decline in response rates[11]. The overall response rates for 2011-2012, 2013-2014, and 2015-2016 were lower than previously experienced in recent years of NHANES (**Table D**). Briefly, the in-home interview response rate was 72.6% in 2011–2012, 71.0% in 2013–2014, and 61.3% in 2015–2016. The overall cumulative examination response rate was 69.5% in 2011–2012, 68.5% in 2013–2014, and 58.7% in 2015–2016. Sample sizes and response rates for all survey cycles, overall and by age and gender, are provided on the NHANES Website[12].

Adjustments made to the sample weights for survey non-response account only for interview or MEC exam non-response, but not for component/item non-response which can occur at the household interview or the exam (e.g., a participant declined to have their blood pressure measured in the examination component but completed all other examination components).

Table D. Overall unweighted survey response rates for all ages, NHANES 1999-2016

		Interviewed sample		Examined sample	
Survey years	Screened sample[1]	Sample size	Response rate (percent)	Sample size	Response rate (percent)
1999–2000	12,160	9,965	81.9	9,282	76.3
2001–2002	13,156	11,039	83.9	10,477	79.6
2003–2004	12,761	10,122	79.3	9,643	75.6
2005–2006	12,862	10,348	80.5	9,950	77.4
2007–2008	12,943	10,149	78.4	9,762	75.4
2009–2010	13,272	10,537	79.4	10,253	77.3
2011–2012	13,431	9,756	72.6	9,338	69.5
2013-2014	14,332	10,175	71.0	9,813	68.5
2015-2016	15,327	9,971	61.3	9,544	58.7

[1]Screener response rates across survey cycles up to 2013-2014 have varied from 98-100% and the loss of eligible respondents at this stage is considered negligible. For 2015-2016 cycle, the screener response rate was lower at 94.3%, therefore, the interviewed and examined response rates for 2015-2016 cycle were adjusted for sample loss at the screener.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 2007-2016.

Detailed non-response bias analyses performed for the 2013-2014 and 2015-2016 NHANES are planned for publication in the upcoming year.

2.7.2 Component or item non-response

In NHANES, there are a large number of examinations and tests that are conducted in the NHANES MEC and each component contains a number of items. Some examinees may not participate in all components of their designated examination or may not fully participate in a particular component, thus resulting in component or item non-response. If the component non-response varies substantially by demographic characteristics of the participants, the type of component, and survey cycle then these missing values may distort analysis results. Analysts should evaluate the extent of missing data in their dataset related to the outcome of interest as well as any predictor variables used in the analyses to determine whether the data are useable without additional re-weighting for item non-response. As a general rule, if 10% or less of data for the main outcome variable for a specific component is missing for eligible examinees, it is usually acceptable to continue analysis without further evaluation or adjustment. However, if more than 10% of the data for a variable are missing, the analyst may need to further examine respondents and non-respondents with respect to the main outcome variable, and decide whether imputation of missing values or use of adjusted weights is necessary. Even if the overall component non-response rate is <10%, non-response for a component within a subgroup may exceed 10% and may need to be further examined for statistical bias.

2.7.3 Value codes

When a respondent refuses to answer a question, a “refused” response is assigned a value of either “7,” “77,” or “777” depending on the number of digits in the variable value range. A “don't know” response is assigned a value of either “9,” “99,” or “999,” which is also dependent on the number of digits in the variable value range. Failing to identify these other types of missing data, and treating the assigned values for “refused” or “don't know” as numerical values, may distort analysis results; for categorical results, tabulating the number or percentage missing may be part of the analysis. Missing value and non-response codes are identified in the data dictionary for each variable when applicable.

Analysts are also encouraged to review codebooks to determine if a skip pattern affects the variables in their analysis. Failure in identifying skip patterns would erroneously lead the analyst to obtain data on a proportion of the population, instead of the entire study population.

3. Analytic considerations

The complex survey design used for NHANES, including oversampling, stratification, and clustering, must be considered when analyzing the data for appropriate variance estimation and to calculate statistics representative of the U.S. civilian non-institutionalized population.

3.1 Sample weights

The weighting of sample data permits analysts to produce estimates of statistics they would have obtained if all U.S. non-institutionalized civilians had been surveyed. The sample weights assigned to each record can be considered as measures of the number of persons represented by the particular survey respondent.

Weighting takes into account several features of the survey: the differential probabilities of selection for the individual domains described above, non-response to survey instruments, and differences between the final sample and the U.S. civilian non-institutionalized population. The sample weighting was carried out in three steps. The first step involved the computation of weights to compensate for unequal probabilities of selection given that some groups were over sampled.

The second step adjusted for participant non-response. Weights were adjusted for non-response to the in-home interview when creating the interview weights and further adjusted for non-response to the MEC exam when creating the exam weights. In the third step, the sample weights were post-stratified to match estimates of the U.S. civilian non-institutionalized population available from the U.S. Census Bureau. A detailed discussion of the sample weights can be found in the National Health and Nutrition Examination Survey: Sample Design, 2011-2014 report[13], and the National Health and Nutrition Examination Survey: Estimation Procedures, 2011-2014 report[14].

The oversampling of subgroups mentioned above, and some operational differences across survey locations, can cause the NHANES sample weights to be quite variable. Further, when sampling domains are combined for analysis, a wide range of sample weights may occur due to the different selection probabilities, which will lead to increased variance in the analytic results. For example, variable weights could be expected when combining 2011-2012 and 2013-2014 data for Asian persons with persons of any other race and Hispanic origin, since the distribution of sample weights for these groups differ; although the median sample weights are comparable, the 75th percentile and maximum sample weights for each of the other race/ethnicity groups are higher than those for the Asian group (**Table E**).

Analysts should examine the sample weights as an initial step in any analysis. Records with large sample weights can be influential in an analysis, especially when extreme weights are associated with extreme data points for the variable of interest. In addition to considering race and Hispanic origin**,** the following age categories are recommended for reducing the variability in the sample weights for estimates by age and race and Hispanic origin: 5 years and under, 6-11 years, 12-19 years, 20-39 years, 40-59 years, 60 years and over.

Table E. Distribution of MEC sample weights by race and Hispanic origin[1] and survey year

Survey years	Race/Hispanic origin	Minimum	25th percentile	Median	75th percentile	Maximum
2015-2016	Hispanic	3,419	10,970	15,556	23,919	57,063
	NH white	8,113	29,810	49,866	93,255	230,297
	NH black	3,833	11,963	17,255	23,777	49,591
	NH Asian	5,392	12,771	17,886	22,352	52,313
	All others[2]	4,799	12,234	17,390	30,978	242,387
2013-2014	Hispanic	3,748	12,536	16,596	24,755	77,534
	NH white	5,999	26,337	47,441	78,864	171,395
	NH black	3,986	10,008	14,581	23,491	49,931
	NH Asian	5,093	10,783	15,003	19,739	44,908
	All others[2]	4,933	10,183	15,630	27,288	127,207
2011-2012	Hispanic	4,344	12,994	17,533	30,195	72,577
	NH white	6,555	29,425	55,372	99,423	222,580
	NH black	3,522	9,333	12,886	18,840	53,078
	NH Asian	3,773	8,693	12,218	16,610	26,320
	All others[2]	4,027	9,258	14,392	25,331	176,993
2009-2010	Hispanic	3,364	9,236	12,342	17,846	45,218
	NH white	5,361	22,499	37,933	65,138	143,400
	NH black	5,977	12,868	17,086	24,715	52,703
	All others[2,3]	6,266	15,419	24,150	53,471	158,147
2007-2008	Hispanic	2,509	8,686	12,198	17,365	62,556
	NH white	5,680	26,622	40,877	71,231	167,686
	NH black	4,505	10,696	14,656	23,113	81,407
	All others[2,3]	4,698	18,224	32,778	63,077	192,771

[1]Race and Hispanic origin data for 2007-2010 are from the "RIDRETH1" variable and for 2011-2016 are from the "RIDRETH3" variable provided on the publically released Demographic Files for the respective years.
[2]The "other" race subgroup included those who reported a race other than black, or white, or reported more than one race in 2007-2010, and those who reported a race other than black, white, or Asian or those who reported more than one race in 2011-2016
[3]Includes non-Hispanic Asian.
NOTE: NH is non-Hispanic.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 2007-2016.

3.1.1 Determining the appropriate sample weight for analysis

Various sample weights are available on the data release files. Use of the correct sample weight for NHANES analyses depends on the variables being used. A good rule of thumb is to use "the least common denominator" where the variable of interest that was collected on the smallest number of respondents is the "least common denominator." The sample weight that applies to that variable is the appropriate one to use for that particular analysis.

Sampled participants who completed the interview and were eligible for the examination, but did not respond, were assigned non-zero interview weights and examination weights of zero. Records with a zero examination weight should be treated as missing when the exam data are analyzed. For example, if all variables come from the interview and exam, then the sample used in the analysis should reflect only those with non-zero exam weights and exam weights should be used in the analysis. Similarly, if any variable used comes from a specific subsample, then the sample used in the analysis should only represent those with a non-zero subsample weight and the subsample weights should be used in the analysis.

3.1.2 Subsample weights

As discussed in the ''Survey subsample'' section above, some NHANES participants are in survey components that include only random subsamples of the NHANES MEC-examined sample. Data collected from these participants include a variety of lab, nutrition or dietary, environmental, audiometry, and mental health components. Each subsample is selected in order to provide nationally representative estimates from that component. For example, some, but not all participants, were asked to participate in the Volatile organic compounds (VOC) subsample.

Each component subsample has its own designated sample weight, which accounts for the additional probability of selection into the subsample component and any additional non-response to the component.

When data collected via one of these subsamples are released, separate sample weights are constructed and included in the data file containing the subsample variables. These component subsample weights, which differ from the full examination sample weight must be used for statistical estimation of measures collected only in that subsample. For more details, see the "Subsample Weights" section and Table IV in Appendix II of the 2011-2014 NHANES estimation procedures report[14].

Although 24-hour dietary recall is not considered a subsample, special 24-hour dietary recall weights were assigned to participants who completed this component to incorporate day of the week of the recall.

Subsample weights from the same survey cycle are not designed to be combined within the data release cycle. In fact, many subsamples are mutually exclusive. Two or more subsamples can be combined if there is random overlap between the subsamples; appropriate sample weights need to be recalculated for the resulting combined subsample. For example, no sample weights are provided for the overlap in the fasting subsample with an environmental subsample; this overlap would be about a one-sixth sample. See the respective survey protocol or documentation for more specific information on each subsample.

There are instances that an analyte may be obtained for a subsample in one survey cycle and for the full sample in another (e.g. blood lead, as described above). When analyzing these data, sample weights can be adjusted to analyze the multi-year sample, as described below.

3.1.3 Combining survey cycles

Each two year data release file from 1999-2016, includes 2-year interview, exam, and subsample weights. Any 2-year survey cycle may be combined with adjacent 2-year releases to analyze data from multiple survey cycles. Use of the 2-year sample weights in analyses will lead to valid point estimates for means, variances, proportions and some other summary statistics, but will lead to invalid population totals.

A new sample weight can be calculated based on the sample weights of the combined survey cycles. When combining two or more 2-year cycles from 2001–2002 onward, new multi-year sample weights can be computed by simply dividing the 2-year sample weights by the number of 2-year cycles in the analysis. For example, to analyze data for 2011-2016, divide the three 2-year sample weights by three to obtain the 6-year combined sample weight. Table F provides the examples of formula for combining weights across survey cycles.

Table F. Formulae for constructing weights for NHANES

Number of Survey Years	Combined Survey cycles	Survey Cycle Code†	Formula for Combining Weights across Survey Cycles
Combining four survey years			
4 years	1999-2002		Provided on the Public-use Data Files
4 years	2001-2004	If sddsrvyr in (2,3) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2003-2006	If sddsrvyr in (3,4) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2005-2008	If sddsrvyr in (4,5) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2007-2010	If sddsrvyr in (5,6) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2009-2012	If sddsrvyr in (6,7) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2011-2014	If sddsrvyr in (7,8) then	MEC4YR = 1/2 * WTMEC2YR;
4 years	2013-2016	If sddsrvyr in (8,9) then	MEC4YR = 1/2 * WTMEC2YR;
Combining six survey years			
6 years	1999-2004	If sddsrvyr in (1,2) then	MEC6YR = 2/3 * WTMEC4YR; /*for 1999-2002*/
		If sddsrvyr=3 then	MEC6YR = 1/3 * WTMEC2YR; /*for 2003-2004*/
6 years	2001-2006	If sddsrvyr in (2,3,4) then	MEC6YR = 1/3 * WTMEC2YR;
6 years	2003-2008	If sddsrvyr in (3,4,5) then	MEC6YR = 1/3 * WTMEC2YR;
6 years	2005-2010	If sddsrvyr in (4,5,6) then	MEC6YR = 1/3 * WTMEC2YR;

6 years	2007-2012	If sddsrvyr in (5,6,7) then	MEC6YR = 1/3 * WTMEC2YR;
6 years	2009-2014	If sddsrvyr in (6,7,8) then	MEC6YR = 1/3 * WTMEC2YR;
6 years	2011-2016	If sddsrvyr in (7,8,9) then	MEC6YR = 1/3 * WTMEC2YR;

SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey, 1999-2016.

The sum of the combined multi-year sample weights should be reasonably close to an independent estimate of that midpoint population. The rules for combining surveys also apply to subsamples. Users should be aware of two assumptions made when combining sample weights for different years of data. First, that there are no differences in the estimates over the time periods being combined. Second, the estimate is the average over the time period.

3.1.4 NHANES 1999-2002

Including data for 1999-2000 requires an extra step. The NHANES 1999–2000 sample weights were based on information from the 1990 U.S. census. However, the NHANES 2001–2002 sample weights were based on the 2000 census. Because different population bases were used, the 2-year weights for 1999–2000 and 2001–2002 are not comparable. For this reason, 4-year sample weights were created to account for the two different reference populations in the 1999-2000 and the 2001-2002 NHANES. When combining data from the 1999-2000 NHANES cycle with other cycles, it is recommended that the 4-year sample weights be used for 1999-2002 and the 2-year sample weights be used for other cycles.

To use both the 4-year sample weight for 1999-2002 and 2-year sample weights for other cycles, the 4-year sample weight needs to be doubled prior to analysis so that the observations in 1999-2002 have weights similar in magnitude to the 2-year sample weights. This approach works for regression analyses and other summary statistics but, as above, a multi-year combined weight is needed for population totals. To create a multi-year combined sample weight for multiple survey cycles that include 1999-2002, for example, 1999-2004, first multiply the 4-year sample weight for 1999-2002 by 2, then divide the doubled 4-year 1999-2002 sample weight and the 2-year weights for the 2003-2004 cycles, by 3, the number of cycles; the resulting sample weight will be a 6-year weight.

3.1.5 Computing population counts

To understand the public health impact of a condition, it is often helpful to calculate population counts in addition to the prevalence of a health condition. By quantifying the number of people with a particular condition or risk factor, counts speak directly to the burden or magnitude.

Since NHANES is a nationally representative survey of the civilian noninstitutionalized U.S. population, population estimates are based on reliable estimates for this aspect of the U.S. population.

For NHANES 2011-2016, the sample weights were post-stratified to population totals obtained from the American Community Survey (ACS) and based on the 2010 Census. For NHANES 2001-2010, the sample weights were post-stratified to population totals obtained from the Current Population Survey (CPS) and based on the 2000 Census. For NHANES 1999-2000, the sample weights were post-stratified to population totals obtained from the Current Population Survey (CPS) and based on the 1990 Census. The different sources of these population totals could affect the interpretation of some results. Population totals used for each survey cycle are available at https://wwwn.cdc.gov/nchs/nhanes/responserates.aspx.

The 4-year sample weights (i.e., interview, examination, and all subsample weights) were created and included on the 1999-2000 and 2001-2002 data release files. It was later decided not to create

4-year weights for 2-year samples that crossed censuses. NHANES estimates of population totals will not match any published figures when combined 2-year samples are post-stratified to two different censuses.

The change from the CPS to the ACS was made, in part, as a result of the addition of the Asian oversample in 2011. With this addition, population totals that provided reliable estimates for Asian persons within age and sex categories were needed. While both the CPS and ACS are surveys, the sample size for the ACS is about 13 times larger than that of the CPS. This larger sample size resulted in more reliable estimates for the Asian population.

3.2 Variance estimation

The complex, multistage, probability cluster design of NHANES affects variance estimates (sampling error). Typically, individuals within a cluster (e.g., county, school, city, or census block) are more similar to one another than those in other clusters and this homogeneity of individuals within a given cluster is measured by the intra cluster correlation. When working with a complex sample, the ideal situation is to limit the correlation among sample persons within clusters by sampling more clusters with fewer people in each. However, because of operational limitations (e.g., the number of MECs, geographic distances between locations, etc.) NHANES currently samples only 30 locations (primary sampling units [PSUs]) within a 2-year survey cycle.

The design effect is a measure of the impact of the complex sample design on estimates of variance. It is defined as the ratio of the variance of a statistic which accounts for the complex sample design to the variance of the same statistic based on a hypothetical simple random sample of the same size. If the design effect is 1, the variance for the estimate under the complex sampling is the same as the variance under simple random sampling. For NHANES, the design effects are typically greater than 1. Design effects less than one may be due to variability in the estimate of the variance (see section 3.2.3). For NHANES 1999-2016, design effects differ among variables due to differences in variation by geography, by household intra-class correlation, and by demographic heterogeneity.

3.2.1 Variance estimation methods

For complex sample surveys, exact mathematical formulas for variance estimation are usually not available. Variance approximation procedures are required to provide reasonable, approximately unbiased, and design-consistent estimates of variance. Variance estimates computed using standard statistical software packages that assume simple random sampling are generally too low (i.e., significance levels are overstated) and biased because they do not account for the differential weighting and the correlation among sample persons within a cluster.

Two variance approximation procedures, which account for the complex sample design, are replication methods and Taylor Series Linearization. Currently NCHS uses the Taylor Series Linearization methods for variance estimation within survey software packages, such as SUDAAN, for most reports and data products from the 1999-2016 NHANES. Replication methods using either delete-one jackknife or balanced repeated replication (BRR) weights can also be used. Initially, for the NHANES 1999-2000 survey, the delete-one jackknife method was used to estimate variances and these weights are available on the public-use file. Jackknife weights are available for single year data in the RDC; BRR weights are available for the 2-year data releases in the RDC. In addition, BRR weights for the 24-hour urine subsample collected in 2014 and are available in the RDC[15]. If replication methods are to be used for any other survey years, replicate weights must be computed by the analyst.

For either linearization or replication methods, variance variables for strata and PSU must be available on the survey data file. To reduce risks of disclosure with a 2-year data release, the actual PSUs cannot be released. To use the Taylor Series Linearization approach for variance estimation in survey software packages, Masked Variance Units (MVUs) were created. These variables, the stratum (SDMVSTRA) and PSU (SDMVPSU), are included in the Demographic data file for each data release. These MVUs on the data file are not the "true" design PSUs. They are a collection of secondary sampling units aggregated into groups for the purpose of variance estimation. They produce variance estimates that closely approximate the variances that would have been estimated using the "true" design. MVUs have been created for all 2-year survey cycles from NHANES 1999–2000 through 2015–2016 and can be used for analyzing multi-cycle data sets. True stratum and PSU variables are available in the NCHS RDC[5].

Software such as SUDAAN, Stata, SPSS, SAS Survey procedures and R can all be used to estimate sampling errors by the Taylor Series Linearization method. Software packages or procedures that assume a simple random sample, should not be used for computing variances for NHANES.

3.2.2 Other sources of variability

As with any survey, quality control procedures are taken to ensure that sources of error are limited and that the data are of high quality. It is inherent to any measurement process that some sources of variation cannot be controlled and users should be aware of these. Some variables may be subject to within person variation. For example, outcomes from a 24-hour dietary intake interview will not be the same if taken on a different day. A person's blood pressure reading could be temporarily elevated due to personal stress and may not equal the average or usual blood pressure reading for that individual. The data collection protocols for each component contain important information on the procedures that can be used to interpret findings.

3.2.3 Subsetting data

3.2.3.1 Variance estimates

Sometimes an analyst may have a certain demographic subgroup of interest, such as a particular age range or sex, or a subsample of participants who received a particular laboratory test. For some variance estimation methods, including the Taylor Series Linearization, the entire set of data containing the appropriate weights for a particular survey cycle must be used to obtain the correct variance estimates. The estimation procedure must indicate which records are in the subgroup of interest. For example, to estimate mean body mass index and its standard error for men 20 years and over, the entire dataset of examined individuals who have an exam weight, including females and those younger than 20 years, must be read into the statistical software program. The SUBPOPN (or SUBPOPX) statement in SUDAAN, the STAT and DOMAIN statements in the SAS survey procedure, or comparable statements in other programs must be used to indicate the subgroup of interest (i.e., men aged 20 and over in the above example). Depending on the specifications for of the software, an indicator variable created by the analyst prior to the procedure may facilitate the identification of the subgroup in the procedure statements.

3.2.3.2 Degrees of freedom

The nominal degrees of freedom can be approximated using the stratum and PSU variables on the data file by subtracting the number of strata from the number of PSUs. If an analysis is performed on a subgroup of cases, the degrees of freedom should be based on the number of strata and PSUs containing the observations of interest. For example, if the standard error of the mean systolic blood pressure for non-Hispanic black persons is based on 25 PSUs and 13 strata then the degrees of freedom would be 25 minus 13, which is 12. The degrees of freedom are used in statistical tests and in the computation of confidence intervals.

The analyst should be aware how the software package being used determines degrees of freedom for subgroups as they can differ among packages. Many of these packages do not correct for the reduction in the degrees of freedom in analyses for subgroups where not all strata and PSUs are represented. Therefore, it is important to output the number of PSU's and stratum from the survey package procedures and calculate the correct degrees of freedom.

Some analysis packages will improperly calculate the degrees of freedom from a combined data set containing multiple NHANES survey cycles when only one NHANES survey cycle is being used in the analysis. Including only the survey cycles of interest in the analysis will produce the correct degrees of freedom.

3.3 Statistical precision of estimates

The issues of precision and statistical reliability should be addressed for each specific analysis. The statistical reliability of an estimate can be evaluated using several measures, including the sample size on which it is based, the effective sample size (the sample size adjusted for the design effect), the design effect, the width and relative width of its confidence interval, the relative standard error (RSE, defined as the ratio of the standard error of the estimate to the estimate itself and often multiplied by 100 and expressed as a percentage), and the degrees of freedom.

As mentioned above, although data are released in 2-year cycles, the accumulation of at least four years of data may be required to obtain an acceptable level of reliability. Thus, to create estimates for smaller 2-year samples, collapsing of some of the subgroups within the sample design may be necessary to produce adequate sample sizes (both in terms of the number of observations and the number of PSUs) for analysis purposes.

In 2017, NCHS published updated Data Presentation Standards for Proportions[2], a report describing statistical criteria for determining whether or not to publish a proportion in an NCHS report. Proportions, generally multiplied by 100 and expressed as percentages, are the most common estimates produced at NCHS and are commonly reported from NHANES. Criteria used for these Standards include sample size, confidence intervals and for some surveys, including NHANES, the degrees of freedom.

The Data Presentation Standards for Proportions are applied to proportions in NCHS reports, including reports that present estimates from NHANES. These Standards were developed for use with all NCHS data systems, not just NHANES. While research objectives of NHANES data users are diverse, the principles of the Data Presentation Standard for Proportions should be considered when making analytic decisions. The Data Presentation Standards for Proportions are not applicable for estimated means, percentiles, regression coefficients and other statistics. Importantly, additional criteria may be needed to meet assumptions for inference.

3.3.1 Sample size

Two general sample size considerations were used in the sample design for NHANES 1999-2016 and NHANES III. First, an estimated prevalence statistic should have a relative standard error of 30 percent or less; and second, the estimated (absolute) differences between population subgroups (domains) of at least 10 percent should be detectable with a Type I error rate (α) of $\leq$ 0.05 and a

Type II error rate (β) of $\leq$ 0.10. The population subgroups for which specified reliability was desired in NHANES are described in the Sample Design series reports [13]. As described earlier, to increase the precision of estimates for the subgroups of interest, oversampling was carried out (refer, for example, to the sections within regarding sample design changes dealing with race and Hispanic origin).

For presentation of proportions in NCHS reports, the NCHS Data Presentation Standards for Proportions require a minimum sample size and effective sample size (i.e. an actual sample size divided by the design effect) of 30 though estimated proportions must also meet other criteria. Prior NHANES analytic guidelines had recommend an effective sample size of 30 for proportions between 0.25 and 0.75 and for means of variables with symmetric distributions, with larger samples recommended for larger (>0.75) and smaller (<0.25) proportions.

For inference based on the normal approximation, the Central Limit Theorem guarantees that statistics based on a sufficiently large sample are approximately normally distributed. Rules of thumb for the Central Limit Theorem approximation vary, but typically, 5 or 10 events (or non-events) are suggested for the numerator when estimating proportions. As a result, for proportions based on rare or nearly universally occurring events (the extremes of the distribution), a much larger sample may be required to make inferences based on the Central Limit Theorem for some analyses. For proportions between 0.25 and 0.75 based on the NHANES surveys an effective sample size of at least 30 (in the denominator) should be sufficient to make inferences based on the normal approximation.

3.3.2 Relative standard error

The relative standard error (RSE) of an estimated statistic is defined as the ratio of the standard error of the estimated statistic to the estimated statistic and is usually expressed as a percentage.

% RSE = (Standard error of estimate / Estimate) * 100

An estimate with a very large relative standard error may be combined with other estimates to create an aggregate with a reasonably small RSE.

NCHS has often used thresholds based on the RSE in determining whether or not to show an estimate or to identify an estimate as unreliable in its reports, including NHANES reports. For proportions, the NCHS Data Standards for Proportions do not include criteria based on the RSE; other criteria are used to determine whether a proportion is sufficiently precise for publication. However, estimated means published in NCHS reports will continue to be evaluated based on the RSE and estimated means with RSE greater than or equal to 30% should be identified as unreliable.

3.3.3 Reliability of the estimated standard error and degrees of freedom

The variance of a statistic estimated from the NHANES data is also an estimate, and as such, is subject to its own variability. For complex surveys, such as NHANES, the precision of the estimated variance is approximately related to the square root of the degrees of freedom. As the number of degrees of freedom increases, the precision of the estimated variance increases. Conversely, variance estimates based on small numbers of degrees of freedom may not be reliable, in turn affecting the reliability of statistical tests and inferences.

The NCHS Data Presentation Standards for Proportions recommends that proportions based on fewer than 8 degrees of freedom be reviewed by a clearance official[2]. Depending on the purpose of the report and the particular analysis, this review could result in the presentation or suppression of the proportion. As the quality of the estimated standard errors for all estimates will depend on the degrees of freedom, this standard for proportions should be considered guidance for means and other statistics. Most population estimates from a public-use data file for a single NHANES cycle are based on 15 degrees of freedom (30 PSU – 15 strata). However, estimates for subgroups not represented in all locations and subnational estimates produced in the RDC may have fewer than 15 degrees of freedom.

3.3.4 Confidence intervals

Confidence intervals can be examined when assessing the reliability of an estimate. Interpreted based on the sample design, under repeated sampling from the same population, the true population parameter will be contained in, say, a 95 percent confidence interval in 95% of the repeated samples. For surveys, confidence intervals for proportions and means have often been computed using the Wald approach, with the degrees of freedom calculated, as above, as the number of PSU minus the number of strata. As mentioned above, survey software may not calculate the degrees of freedom accurately for NHANES subpopulations so extracting the necessary information and computing the interval may be needed.

When used for proportions, particularly proportions near 0 and 1, the Wald method may result in negative lower limits for small proportions and in upper limits that exceed 1 for large proportions; furthermore, in simulation studies, 95% confidence intervals for proportions using the Wald method generally do not contain the true parameter 95% of the repeated samples. Other methods for obtaining confidence intervals for proportions may be used. The properties of the proportion and the analytic goals should be considered when selecting an approach.

The NCHS Data Standards for Proportions include criteria based on the absolute and relative widths of the confidence interval. From a calculated CI, the absolute CI width is obtained by subtracting the value of the lower confidence limit from the value of the upper confidence limit. The relative CI width is calculated as the absolute CI width divided by the proportion and multiplied by 100%. For this purpose, confidence intervals standards are based on the Clopper-Pearson confidence interval[16], which was adapted for complex surveys by Korn and Graubard[17]. Although other confidence interval calculations may be appropriate for certain analyses, the thresholds for the NCHS Data Presentation Standards for Proportions were set using the Clopper-Pearson intervals and have not been evaluated for other intervals[2].

For proportions in NCHS publications, if the absolute confidence interval width is greater than 0 and less than or equal to 0.05, then the proportion can be presented if all other criteria (i.e. number of events, size of sample, relative standard error) are met. If the absolute confidence interval width is greater than or equal to 0.30, then the proportion should be suppressed. If the absolute confidence interval width is between 0.05 and 0.30 and the relative confidence interval width is more than 130% times the proportion, then the proportion should be suppressed.

4. Conclusion

In summary, these analytic guidelines represent the latest statistical procedures and analytic guidance for the continuous NHANES survey for the years 2011-2014 and 2015-2016.

As mentioned previously, another resource for all analysts is the series of NHANES Tutorials[4]—a Web-based product designed to assist users in understanding and analyzing NHANES data. The tutorial illustrates many of the topics described in this report, including preparing analytic datasets, understanding survey design features such as sample weights and variance estimation, and provides sample code using SUDAAN and other survey software.

5. References

1. Division of the National Health and Nutrition Examination Surveys. The National Health and Nutrition Examination Survey (NHANES) Analytic and Reporting Guidelines. https://wwwn.cdc.gov/nchs/nhanes/AnalyticGuidelines.aspx, 2018.
2. Parker J, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. 2017.
3. National Center for Health Statistics, Division of the National Health and Nutrition Examination Surveys. Website of the National Health and Nutrition Examination Surveys. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed August 6th, 2018.
4. Division of the National Health and Nutrition Examination Surveys. Continuous NHANES Web Tutorial,. https://www.cdc.gov/nchs/tutorials/Nhanes/index_continuous.htm, 2018.
5. The National Center for Health Statistics Research Data Center (RDC). https://www.cdc.gov/rdc/index.htm, 2018.
6. Centers for Disease Control and Prevention. Fourth Report on Human Exposure to Environmental Chemicals, Updated Tables, (March 2018). Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevent. http://www.cdc.gov/exposurereport/.
7. Division of the National Health and Nutrition Examination Surveys. 2007-2014 Los Angeles County, California - Demographic Variables & Sample Weights https://wwwn.cdc.gov/Nchs/Nhanes/limited_access/LDEMO_EH.htm. Accessed August, 7th, 2018.
8. Division of the National Health and Nutrition Examination Surveys. 2007-2014 California - Demographic Variables & Sample Weights https://wwwn.cdc.gov/Nchs/Nhanes/limited_access/CDEMO_EH.htm. Accessed August, 7th, 2018.
9. Division of the National Health and Nutrition Examination Surveys. Geocoded Data, NHANES 1999-2016, Census 2010 https://wwwn.cdc.gov/Nchs/Nhanes/limited_access/GEO_2010.htm. Accessed August, 7th, 2018.
10. Division of the National Health and Nutrition Examination Surveys. Geocoded Data, NHANES 1999-2016, Census 2000 https://wwwn.cdc.gov/Nchs/Nhanes/limited_access/GEO_2000.htm. Accessed August, 7th, 2018.
11. Williams D, Brick JM. Trends in US Face-to-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology.* 2017.
12. Division of the National Health and Nutrition Examination Surveys. Response Rates and Population Totals https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx. Accessed August, 7th, 2018.
13. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National health and nutrition examination survey: sample design, 2011-2014. *Vital and health statistics Series 2, Data evaluation and methods research.* 2014(162):1-33.
14. Chen TC, Parker JD, Clark J, Shin HC, Rammon JR, Burt VL. National Health and Nutrition Examination Survey: Estimation Procedures, 2011-2014. *Vital and health statistics Series 2, Data evaluation and methods research.* 2018(177):1-26.

15. Division of the National Health and Nutrition Examination Surveys. 24-Hour Urine Collection Data Processing - First Collection https://wwwn.cdc.gov/Nchs/Nhanes/limited_access/UR1_H_R.htm. Accessed August, 9th, 2018.
16. Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*. 1934;26(4):404-413.
17. Korn EL, Graubard BI. Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Survey Methodology*. 1998;24:193-201.