NHANES Analytic Guidelines

June 2004 Version

Beginning in 1999, the National Health and Nutrition Examination Survey

(NHANES) became a continuous, annual survey rather than the periodic survey that it

had been in the past. For a variety of reasons, including disclosure and reliability issues,

the survey data are released on public use data files every two years. Thus, the data

release cycle for the ongoing (and continuous) NHANES is described as NHANES

1999-2000, NHANES 2001-2002, NHANES 2003-2004, etc.

Beginning in 2003, the survey content for each two year period is held as constant

as possible so as to be consistent with the data release cycle. This was not always the

case in the first four years of the continuous survey and some special data release and

data access procedures were developed and used for selected survey content collected in

"other than two-year" intervals.   In addition to the analysis of data from any two year

cycle, it is possible to combine two or more "cycles" (e.g. 1999-2000 and 2001-2002) to

create NHANES 1999-2002, thus increasing sample size and analytic options. In order to

produce estimates with greater statistical reliability, combining two or more 2-year cycles

of the continuous NHANES is encouraged and strongly recommended. The user should

verify that data items collected in all combined years were comparable in wording and

methods.

These analytic guidelines are the most current recommendations from the

National Center for Health Statistics and the NHANES Program for use with the

NHANES data from the continuous survey. For the most part, the recommendations

provided herein are also appropriate for the past (periodic) NHANES surveys. Although

similar in many ways to the previously provided NHANES III analytic guidelines

(http://www.cdc.gov/nchs/data/nhanes/nhanes3/nh3gui.pdf) and the NHANES

1999-2000 Addendum to the NHANES III Analytic Guidelines

(http://www.cdc.gov/nchs/data/nhanes/guidelines1.pdf) , there are a number of

modifications and these latest guidelines should be used for all future analyses of the

continuous NHANES survey and for comparisons with previous NHANES surveys.

These guidelines will be updated on a periodic basis as more information is learned from

the analyses of the NHANES survey data and as new and important statistical procedures

are developed for use in complex surveys like NHANES. Users should regularly visit the

NHANES website to see if a new version of these analytic guidelines has been released.

At this time, with the release of data from NHANES 1999-2000 and NHANES

2001-2002, the recommended procedure for analysis is to consider the time period 1999-

2002 as one survey. Data files for 1999-2000 (interview sample size 9,965 and MEC

examined sample size of 9,282) and 2001-2002 (interview sample size 11,039 and MEC

examined sample size of 10,477) should be concatenated to form a single analytic file

(interview sample size of 21,004 and MEC examined sample size of 19,759).

When analyzing the combined four year data set, the correct sampling weights

must be used to produce unbiased estimates. Unfortunately, there is no simple procedure

for combining the 1999-2000 and 2001-2002 sample weights (like dividing each

individual sample weight by two and using that number as the four-year sample weight).

The sample weights for NHANES 1999-2000 were based on population estimates

developed by the Bureau of the Census before the Year 2000 Decennial Census counts became available. The two-year sample weights for NHANES 2001-2002 are based on population estimates that incorporate the year 2000 Census counts. The two population estimates are not comparable.

Therefore, to facilitate analysis of these first two cycles of the continuous NHANES, the appropriate four-year sample weights have been calculated and added to the demographic data files for both 1999-2000 and 2001-2002. Thus, users of the earlier release of the NHANES 1999-2000 demographic file must use the updated demographic file to appropriately analyze the combined four-year data. These four-year sample weights have the same variable name in each two-year demographic file. For example, for the sample persons for whom there are MEC data items, the variable name for the four year weight is WTMEC4YR.

Income and Poverty Income Ratio (PIR) have also been added to both the NHANES 1999-2000 and NHANES 2001-2002 demographic data files. This is an additional reason to utilize the recently updated (May 2004) demographic files for both two-year time periods. In the future, additional variables (e.g. marital status) will be added to these demographic files, so users should periodically check the NHANES website for notice of such updates.

In addition, the procedure for variance estimation (sampling errors) has been changed. When NHANES 1999-2000 data were released as public use files,

confidentiality and disclosure avoidance principles prohibited the release of a PSU

variable (as was done in past NHANES). A new approach was needed. For NHANES

1999-2000, 52 replicate weights were produced and a jackknife technique recommended

for variance estimation. This procedure is less feasible for a four-year data set or for

multiple combinations of various two-year NHANES data sets. An alternative method

for variance estimation that protects confidentiality and allows the use of "PSU's" was

developed and is now the recommended approach for analysis on the ongoing and

continuous NHANES. This method creates Masked Variance Units (MVU's) which can

be used as if they were Pseudo-PSU's to estimate sampling errors (similar to past

NHANES). The Pseudo-PSU's on the data file are not the "true" design PSU's. They

are a collection of secondary sampling units aggregated into groups (called Masked

Variance Units) for the purpose of variance estimation. They produce variance estimates

that closely approximate the variances that would have been estimated using the "true"

design variance estimates. These MVU's have been created for both NHANES 1999-

2000 and NHANES 2001-2002 and added to the demographic data files for both two-

year periods. They can also be used for the combined four-year data set.

The stratum variable is SDMVSTRA and the PSU variable is SDMVPSU.

Software such as SUDAAN, STATA and SAS can be used to estimate sampling errors by

the Taylor series (linearization) method. Typically, the data set should first be sorted by

SDMVSTRA and SDMVPSU. There are no replicate weights provided for NHANES

2001-2002. Replication techniques can still be used to estimate sampling errors if the

software, such as WESVAR, computes its own set of replicate weights based on the nested PSU within stratum design.

Some data components of NHANES are only available for 1999-2000 or 2001-2002. For these components, the two-year sample weights and the two-year MVU's can be used for analysis. For 2001-2002, the two-year weights and MVU's are provided on the demographic data file. For 1999-2000, the previously released demographic file has been updated to add the MVU's and four-year sample weights At this time, the preferred approach for calculating sampling errors is to use the MVU's and to ignore the JK-1 technique utilized as an interim approach with the release of the NHANES 1999-2000 data.

On occasion, there may be a particular issue that requires comparison of results from NHANES 1999-2000 with NHANES 2001-2002. For summary statistics such as means and proportions, the appropriate two-year sample weights and MVU's can be used for reasonably valid inferences (although caution should be used when producing estimates for any detailed population subgroup). Calculation of totals (e.g. estimates of the number of persons with a particular characteristic such as BMI greater than 30 or with impaired glucose tolerance) is not appropriate for NHANES 1999-2000 unless the numbers are ratio-adjusted to population counts based on year 2000 Census figures. This is not a issue for NHANES 2001-2002 which was linked to the 2000 Census counts.

To summarize, there are a number of changes and updates to previous NHANES Analytic Guidelines and these revised recommendations should be used for all analyses of NHANES 1999-2002 data. They are also generally applicable to previous NHANES surveys. NHANES 1999-2000 and NHANES 2001-2002 data files should be combined and analyzed as a four-year data set where possible. The demographic file for NHANES 1999-2000 was updated with four-year weights and MVU's (in addition to the previously existing two-year weights). The NHANES 2001-2002 demographic file contains 2-year and 4-year weights along with MVU's. The 4-year weights and MVU's must be used when the two data sets are merged to create an NHANES 1999-2002 data file. With the four-year combined file, one should ignore the two-year sample weights and the 1999-2000 JK-1 replicate weights. Finally, additional information and recommendations will be added to these guidelines on a periodic basis. Users are encouraged to check the NHANES website on a regular basis to be aware of the latest version of these NHANES analytic guidelines.