

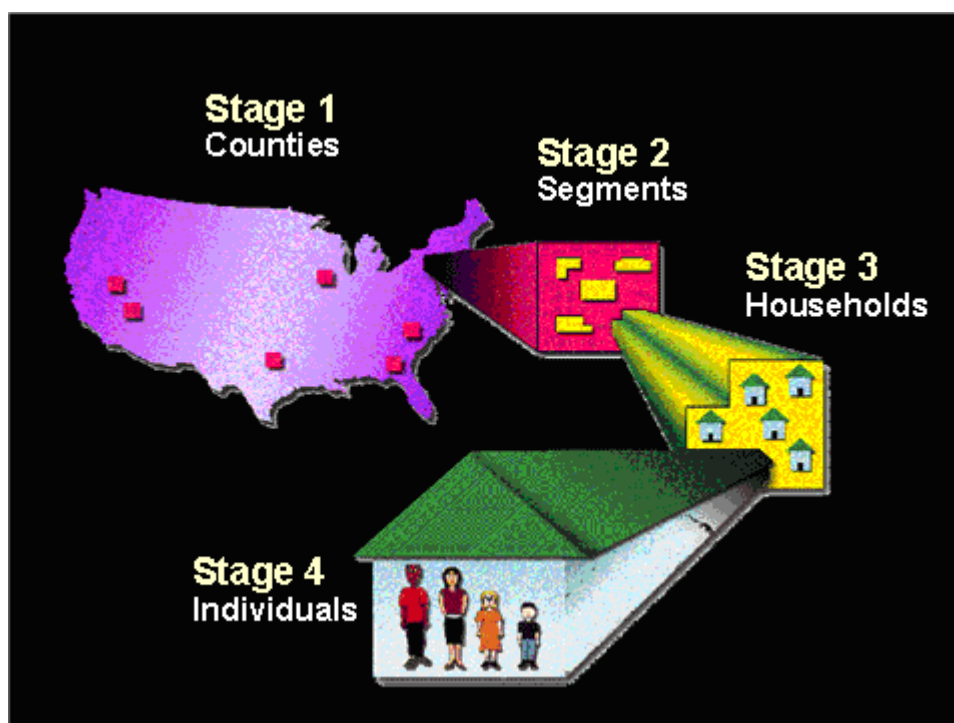
Key Concepts About NHANES Survey Design

NHANES data are NOT obtained using a simple random sample. Rather, a complex, multistage, probability sampling design is used to select participants representative of the civilian, non-institutionalized US population. The sample does not include persons residing in nursing homes, members of the armed forces, institutionalized persons, or U.S. nationals living abroad.

NHANES Sampling Procedure

The NHANES sampling procedure consists of 4 stages, shown and described below.

Four Stages of NHANES Sampling Procedure



- Stage 1: Primary sampling units (PSUs) are selected. These are mostly single counties or, in a few cases, groups of contiguous counties with probability proportional to a measure of size (PPS).
- Stage 2: The PSUs are divided up into segments (generally city blocks or their equivalent). As with each PSU, sample segments are selected with PPS.
- Stage 3: Households within each segment are listed, and a sample is randomly drawn. In geographic areas where the proportion of age, ethnic, or income groups selected for oversampling is high, the probability of selection for those groups is greater than in other areas.
- Stage 4: Individuals are chosen to participate in NHANES from a list of all persons residing in selected households. Individuals are drawn at random within designated age-sex-race/ethnicity screening subdomains. On average, 1.6 persons are selected per household.

What is a Sample Weight?

A sample weight is assigned to each sample person. It is a measure of the number of people in the population represented by that sample person in NHANES, reflecting the unequal probability of selection, nonresponse adjustment, and adjustment to independent population controls. When unequal selection probability is applied, as in the NHANES 1999-2002 sample, the sample weights are used to produce an unbiased national estimate. More information about sample weights and how they are created can be found in the Weighting module.

Oversampling

NHANES is designed to sample larger numbers of certain subgroups of particular public health interest. Oversampling is done to increase the reliability and precision of estimates of health status indicators for these population subgroups.

Examples of oversampled subgroups in the 1999-2004 surveys include:

- African Americans
- Mexican Americans
- Low income White Americans (beginning in 2000)
- Adolescents aged 12-19 years
- Persons age 60+ years

Different subgroups have been oversampled in other survey years. For example, during the late 1960s and early 1970s, there was concern that people of very low income and women of childbearing age were at greater risk of malnutrition than the general population. Therefore, during the first National Health and Nutrition Examination Survey (NHANES I), conducted in 1971-74, these subgroups were oversampled. In future surveys, different subgroups may be oversampled depending on public health trends.

WARNING

For your own analyses, it is critical to carefully review the documentation for each survey cycle to determine which subgroups were oversampled.

Strata and Masked Variance Units

The counties in PSUs from two panels of the 1995 National Health Interview Survey (NHIS) were used as the sampling frame for NHANES 1999-2001. The PSU samples for NHANES 2002-2006 and NHANES 2007-2010, were selected from a frame of all U.S. counties, using the 2000 census data and associated estimates and projections.

NHANES visited 12 PSUs in 1999 and 15 PSUs in each year from 2000 through 2006. For NHANES 2007-2010, NHANES will again visit 15 PSUs per year. For NHANES 1999-2010, each single year and any combination of consecutive years comprise a nationally representative sample of the U.S. population. However, in order to obtain stable estimates, two years of data are necessary for sufficient sample sizes, hence the data are released in two year cycles.

PSUs are selected from strata defined by geography and proportions of minority populations. Most strata contain two PSUs. Together, these strata and the PSUs represent the variance units (sampling units used to estimate sampling error).

To protect the confidentiality of data obtained from sample persons, masked variance units are constructed. Masked Variance Units (MVUs) are equivalent to Pseudo-PSUs used to estimate sampling errors in past NHANES. The 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 variables. These MVUs have been created for each two-year cycle of NHANES and have been created in a way that allows them to be used for any combination of data cycles without recoding by the user. These MVUs are used to define the strata and PSU variables on the public release files. The variable name for the stratum is *sdmvstra* and the variable name for the PSU is *sdmvpsu*.

Please refer to the [Analytic Guidelines](#) for more information on sampling and masked variance units.

 [Close Window](#)