

# NHANES 2003–2004 Public Data General Release File Documentation

Years of Coverage: 2003–2004

First Published: June 2005

Last Revised: N/A

---

**Introduction** The National Center for Health Statistics (NCHS), Division of Health Examination Statistics (DHES), part of the Centers for Disease Control and Prevention (CDC), has conducted a series of health and nutrition surveys since the early 1960's. The National Health and Nutrition Examination Surveys (NHANES) were conducted on a periodic basis from 1971 to 1994. Details of the design and content of each survey, and the public use data files are available (<http://www.cdc.gov/nchs/nhanes.htm>).

The most recent NHANES began in 1999. Every year, approximately 7,000 individuals of all ages are interviewed in their homes; of those interviewed, approximately 5,000 complete the health examination component of the survey. A majority of the health examinations are conducted in mobile examination centers (MECs); the MECs provide an ideal setting for the collection of high quality data in a standardized environment. In addition to the MEC examinations, a small number of survey participants (SPs) receive an abbreviated health examination in their homes because they are unable to come to the MEC.

**Target Population** The NHANES target population is the civilian, noninstitutionalized U.S. population. NHANES 2003–2004 includes over-samples of low-income persons, adolescents 12–19 years, persons 60+ years of age, African Americans, and Mexican Americans.

**Survey Objectives** The major objectives of NHANES are:

- To estimate the number and percent of persons in the U.S. population and designated subgroups with selected diseases and risk factors;
- To monitor trends in the prevalence, awareness, treatment, and control of selected diseases;
- To monitor trends in risk behaviors and environmental exposures;
- To analyze risk factors for selected diseases;
- To study the relationship between diet, nutrition, and health;
- To explore emerging public health issues and new technologies;
- To establish a national probability sample of genetic material for future genetic research; and
- To establish and maintain a national probability sample of baseline information on health and nutritional status.

## **Data Collection Procedures**

The NHANES 2003–2004 data collection was carried out under a contractual agreement. First, the eligible sample for the survey and tasks related to survey operations and data management are performed. The NHANES survey design is a stratified, multistage probability sample of the civilian noninstitutionalized U.S. population. The stages of sample selection are: 1) selection of Primary Sampling Units (PSUs), which are counties or small groups of contiguous counties; 2) segments within PSUs (a block or group of blocks containing a cluster of households); 3) households within segments; and 4) one or more participants within households. A total of 15 PSUs are visited during a 12-month time period. A brief description of the data collection procedures follows.

## **Household Interview Data Collection Procedures**

Initially, households are identified for inclusion in the NHANES sample, and an advance letter is mailed to each address informing the occupant(s) that an NHANES interviewer will visit their home. The household interview component is comprised of Screener, Sample Person, and Family interviews, each of which has a separate questionnaire (please refer to the data file codebooks). Trained household interviewers administer all of the questionnaires. In most cases, the interview setting is the survey participant's home. The interview data are recorded using a Blaise format computer-assisted personal interview (CAPI) system.

When the interviewer arrives at the home, he or she shows an official identification badge and briefly explains the purpose of the survey. If the occupant has not seen the advance letter, a copy is given to the occupant to review. The interviewer requests that the occupant answer a brief questionnaire (Household Screener Questionnaire Module 1) to determine whether any household occupants are eligible to participate in NHANES. If eligible individuals are identified, the interviewer proceeds with efforts to recruit these individuals. Initially, the interviewer explains the household questionnaires to all eligible participants 16 years of age and older, informs the potential respondents of their rights, and provides assurances about the confidentiality of the survey data (reiterating what is stated in the advance letter).

A majority of the household interviews are conducted during the first contact. If this is inconvenient for the survey participant, an appointment is made to administer the household interview questionnaires later. Household interviews for survey participants under 16 years of age are conducted with a proxy (usually the SP's

parent or guardian). If there is no one living in the household who is over 16, participants under 16 years of age are permitted to self-report. Respondents are asked to sign an Interview Consent Form to sign, agreeing to participate in the household interview portion of the survey. For participants 16–17 years of age, a parent or guardian consents and the child gives his or her assent.

After the household interview is completed, the interviewer reviews a second informed consent brochure with the participant. This brochure contains detailed information about the NHANES health examination component. All interviewed persons are asked to complete the health examination component. Those who agree to participate are asked to sign additional consent forms for the health examination component. The interviewer telephones the NHANES field office from the participant's home to schedule an appointment for the examination. The interviewer informs the participants that they will receive remuneration as well as reimbursement for transportation and childcare expenses, if necessary.

**Question Sources**

Many of the NHANES 2003–2004 questions were also asked in NHANES II 1976–80, Hispanic HANES 1982–84, NHANES III 1988–94, and NHANES 1999–2002. New questions were added to the survey based on recommendations from survey collaborators, NCHS staff, and other interagency work groups.

**Questionnaire Target Populations**

Please note that there are different target population groups for the topics within and between NHANES questionnaire sections. For example, in the Nutrition and Diet Behavior section, questions pertaining to infant nutrition and breast-feeding were asked of proxy respondents for children 6 years of age and younger; alcohol consumption frequency questions were asked of persons 20+ years of age; and senior meal program participation questions were asked of respondents 60+ years of age. Data users should review the survey questionnaire codebooks thoroughly to determine the target populations for each NHANES questionnaire section and sub-section.

**Health Examination Component**

When a participant arrives at the MEC, the MEC Coordinator greets the participant and verifies all pertinent identifier information. Each participant receives a disposable paper gown and a pair of slippers to

wear during their examination. Persons 6 years of age and older are asked to provide a urine specimen. MEC staff direct participants to the rooms where the examination components are conducted. In addition to the MEC Coordinator, each MEC survey team consists of one physician, one dentist, two dietary interviewers, three certified medical technologists, five health technicians, one phlebotomist, two interviewers and one computer data manager. Upon completion of the examination, each examinee is remunerated. Some of the medical findings from the examination are given to the examinees before they leave the MEC. The other reportable survey findings are mailed to participants after the laboratory assays and special tests are completed.

## **MEC Operations**

Three MECs are equipped for use in NHANES. Each MEC consists of four large, inter-connected trailer units. An advance team sets up the MECs prior to the start of the survey examinations; water, sewer, electrical, and communications lines are connected during set-up. The MEC equipment and data collection systems must be checked and calibrated prior to the start of survey data collection. The MECs are open a total of 5 days per week; the non-operational days change on a rotating basis so that appointments can be scheduled on any day of the week. Two examination sessions are conducted daily. For the convenience of the survey participants, appointments can be scheduled during morning, afternoon, or evening hours. The examinations require up to 3 hours to complete. At any given time during the survey, examinations are conducted at two survey locations simultaneously. Staff vacations are scheduled for periods of about 1 month at New Year's and about 2 weeks during the summer; leaving 10½ months to conduct examinations.

## **Guidelines for NHANES Data Users**

NHANES 2003–2004 survey design and demographic variables are found in the demo\_b.xpt file in this release. All of the NHANES public use data files can be linked by using the common survey participant identification number (variable name: SEQN). Merging information from multiple NHANES 2003–2004 data files using SEQN ensures that the appropriate information for each survey participant is linked correctly. All data files should be sorted by SEQN.

The NHANES 2003–2004 data files do not have the same number of records in each file. For example, there are different numbers of subjects in the Interview and Examination samples of the survey.

Additionally, the number of records in each data file varies depending on gender and age profiles for the specific component(s).

Confidential and administrative data are not being released. Some variables have been recoded or top- or bottom-coded to protect the confidentiality of survey participants.

## **Sample Person Demographic File**

The sample person demographic file is composed of a limited set of recoded core variables that are required to analyze NHANES 2003–2004 data. Most data analyses require either the interviewed sample weight (variable name: WTINT2YR) or examined sample weight (variable name: WTMEC2YR). The 2-year sample weights (WTINT2YR, WTMEC2YR) should be used for NHANES 2003–2004 analyses. Many variables that are listed in the Demographic questionnaire sections of the Household Interview were omitted from this data release due to concerns about participant confidentiality.

Demographic data file variables are grouped into three broad categories:

- Status Variables: Provide core information on the survey participant. Examples of the core variables include interview status, examination status, and sequence number. (Sequence number [SEQN] is a unique ID number assigned to each sample person and is required to match the information on this demographic file to the rest of the NHANES 2003–2004 data.)
- Recoded Demographic Variables: The variables include age (age in months for persons through age 19 years, 11 months; age in years for 1–84 year olds; and a top-coded age group of 85+ years), gender, a race/ethnicity variable, an current or highest grade of education completed, (less than high school, high school, and more than high school education), country of birth (United States, Mexico, or other foreign born), Poverty Income Ratio (PIR), income, and a pregnancy status variable (adjudicated from various pregnancy-related variables). Some of the groupings were made due to limited sample sizes for the 2-year data set.
- Interview and Examination Sample Weight Variables: Sample weights are available for analyzing NHANES 2003–2004 data. Most data analyses require either the interviewed sample weight (variable name: WTINT2YR) or examined sample weight (variable name: WTMEC2YR). The 2-year sample weights (WTINT2YR, WTMEC2YR) should be used for NHANES 2003–2004 analyses.

Use of the correct sample weight for NHANES analyses is extremely important and depends on the variables being used. A good rule of thumb is to use "the least common denominator" approach. With this approach, the analyst checks the variables of interest. The variable that was collected on the smallest number of persons is the "least common denominator," and the sample weight that applies to that variable is the appropriate one to use for that particular analysis. Please refer to the NHANES 2003–2004 Analytic Guidelines provided with the data release files to determine the appropriate analytic methodology.

**Getting  
Started with  
NHANES  
2003–2004  
Data Analysis**

The National Health and Nutrition Examination Survey, 2003–2004 (NHANES 2003–2004) contains data for 10,122 individuals of all ages. Data were collected between January 2003 and December 2004. The data and corresponding documentation for the survey interview and examination components are found in several files.

NHANES data in this release is in SAS transport file format. SAS transport file format can be read by many software packages. To access this data in any version of SAS, you should use the XPORT engine. NCHS recommends that data users copy the transport files to a permanent SAS library. For example, assuming "C:" is your hard drive, you can use the following SAS code to copy the Body Measurements Examination Data:

```
LIBNAME XP XPORT "C:\NHANES\bmx_b.xpt";  
PROC COPY IN=XP OUT=SASUSER;  
RUN;
```

NHANES documentation and codebooks are in Adobe Acrobat PDF. To obtain a free copy of the current version of Acrobat Reader, go to the Adobe website:  
<http://www.adobe.com/products/acrobat/readstep2.html>.

## NCHS Locator Fields

**Title:** Public Data General Release File Documentation

**Contact Number:** 1-866-441-NCHS

**Years of Content:** 2003-2004

**First Published:** June 2005

**Revised:** N/A

**Access Constraints:** None

**Use Constraints:** None

**Geographic Coverage:** National

**Subject:** National Health and Nutrition Examination Surveys (NHANES)

**Record Source:** NHANES 2003-2004

**Survey Methodology:** NHANES 2003-2004 is a stratified multistage probability sample of the civilian non-institutionalized population of the U.S.

**Medium:** NHANES Web site; Sybase relational database, SAS datasets