

**NHANES 2001-2002 Data Release
July 2004
Sample Person Questionnaire**

Demographics (DEMO)

Survey Years Included in this File: 2001-2002

Questionnaire Section Characteristics:

Questionnaire Section Description:

The NHANES 2001-2002 Sample Person Demographics File provides the interview and MEC examination status variable, two-year and four-year sample weights, and selected demographic variables such as gender, age, race/ethnicity, education, marital status, country of birth, pregnancy status (for women), total family and household income, and poverty income ratio. Some variables found in the Sample Person Questionnaire Demographics section are not included in the NHANES 2001-2002 data release file for a variety of reasons including concerns about confidentiality protection.

Interview Setting:

Home Interview

Mode of administration:

In-person

The Sample Person section of the household interview was done using a computer assisted personal interview (CAPI) methodology. The CAPI was conducted face-to-face by a trained interviewer in the household in either English or Spanish. Some of the questions in this section required the use of printed hand cards. In most cases, the respondent selected a response choice (or choices) after looking at the hand card (s). When necessary, the household interviewers read the hand card response categories to the respondent.

Topics Included in the Section:

Gender, age, race/ethnicity, income and other demographic information

Eligible sample and any section-specific exclusion criteria:

All sample persons (0-150 yrs) who took the household interview have a Demographics file record. Some questions are asked for all sample persons while others are asked for a particular subgroup.

Persons 16 years of age and older and emancipated minors were interviewed directly. A proxy provided information for sample persons less than 16 years of age and for individuals

who could not self-report.

Quality Control During Data Collection:

The computer assisted personal interview (CAPI) had built-in consistency checks. All interviews were edited by the NHANES field office staff for accuracy and completeness. Some interviews were further validated by recontacting the survey participants.

Data Processing and Editing:

SSDSRVYR represents the data release number (2 for NHANES 2001-2002).

RIDSTATR is a recoded variable representing interview/examination status. Sample persons who took only the household interview and not the MEC examination are not included in the NHANES 2001-2002 data release.

RIDAGEYR: Age in years is top coded at 85 years. This is the age of the sample person at the time of the screening interview. If exact date of birth is provided then this is used to calculate the exact age on the date of screening, otherwise an imputed date of birth is created with the best information provided by the sample person and the age given by the sample person is used. The following procedure is used to impute age at screener when date of birth information is missing or refused but age in years is provided by the sample person: If month of birth is missing or not given it is imputed as 7; if day of birth is missing or not given it is imputed as 1; if year of birth is missing or not given it is imputed as the year of the screening interview less the age in years provided by the sample person at screening. Corrections are made to this imputed information for sample persons who are less than 1 year of age at the time of screening.

RIDAGEMN is age in months at household screening, provided only for those <20 years of age (**RIDAGEYR**<20). If exact date of birth is not provided by the sample person, then the age in months is calculated based on the imputed age at screener to allow the sample person to proceed with the questionnaire and examination.

RIDAGEEX is age in months at MEC examination, provided only for those <20 years of age at household screening (**RIDAGEYR**<20). Six sample persons <20 years of age at household screening have missing data on **RIDAGEEX** because their exact date of birth was not available. **RIDAGEEX** was not calculated for these individuals with an imputed age.

RIDRETH1: This race/ethnicity variable is derived by combining responses to questions on race and Hispanic origin. The recode category "5" (Other race – including multiracial) includes following groups: those with single racial/ethnic identity other than Mexican American, Other Hispanic, Non-Hispanic White, and Non-Hispanic Black; those who report more than one racial identity (multiracial); and those with missing values on race/ethnicity.

RIDRETH2 is the race/ethnicity recode that can be linked to the NHANES III race/ethnicity variable. Those who indicated more than one race (multiracial) and then selected a main race as black (non-Hispanic) or white (non-Hispanic) were recoded into those respective categories. The recode category "4" (Other race – including multiracial) includes all remaining single race responses, those who indicated more than one race but did not select a main race, those who indicated a verbatim response to non-specific multiracial heritage (e.g., multiracial, Mulatto), and those with missing values on race.

DMDBORN: Country of birth was recoded into three categories: 1) born in 50 U.S. states or Washington, D.C.; 2) born in Mexico; and 3) born in any other location or foreign country.

INDFMINC: This variable is the total family income variable. NCHS used the U.S. Bureau of the Census Current Population Survey (CPS) definition of “family” to group household members into one or more families (US Census 2003). The CPS defines a family as: “a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together;” all such people (including related subfamily members) are considered to be members of one family. Over eighty percent of the NHANES households were single-family households; the remaining households were comprised of 2 or more CPS families.

After the information about sources of income was obtained in the Family Interview Income section questionnaire, the respondent was asked to report total combined family income for themselves and the other members of their family in dollars (question INQ200). If the respondent refused to answer INQ200 or did not know the total combined family income, an income screener question was asked (question INQ220) to query if the total family income was $< \$20,000$ or $\geq \$20,000$. If the respondent answered INQ220, a follow-up question asked the respondent to select an income range (question INQ230) from a list of income ranges listed on a printed hand card. The midpoint of the income range was then used as the total family income value. Family income values were used to calculate the Poverty Income ratio (INDFMPIR) and estimated total household income (INDHHINC) as described below. Total family income is reported as a range in the NHANES data file.

INDHHINC: This variable is the estimated total household income. The estimated household income was derived from family income data. If a household was comprised of a single CPS family, the family income value was used as previously described. When more than one CPS family resided in the household, two methods were used to compute estimated total household income. One method was to use income data reported by each CPS family that was interviewed (INQ200). The second method for multi-family households used total household income information (INQ200) provided by a household reference person. The income information provided by respondents for each CPS family (method 1) was used whenever possible because this information was considered to be more reliable than information provided by a household reference person who may or may not have had firsthand knowledge of the total household income.

When income information was obtained from each of the CPS families in a household, the reported CPS family income values were summed to compute total household income (INDHHINC). When information was missing for any of the CPS families in the household, the estimated household income value provided by a family reference person was used to compute estimated total household income. Total household income could not be calculated for multi-family households when income range data were reported by any of the families in the household and thus the household income data are coded as missing.

INDFMPIR is the poverty income ratio (PIR) variable (top coded at 5). The U.S. Census Bureau uses a set of money income thresholds that vary by family size and composition to determine who is poor in accordance with Office of Management and Budget (OMB) Statistical Policy Directive Directive 14 (OMB 1978). The poverty income ratio (PIR) is the ratio of income to the family's appropriate poverty threshold (US Census Bureau, 2003a). If a family's total income is less than that family's threshold income value, then that family, and

every individual in it, is considered poor. Thresholds are updated annually for inflation with the Consumer Price Index. The official poverty definition counts money income before taxes and excludes capital gains and noncash benefits (such as public housing, medicaid, and food stamps). PIR values below 1.00 are below the official poverty threshold while PIR values of 1.00 or greater indicate income above the poverty level. (Census Bureau 2003b).

PIR values for NHANES participants were computed using family income data. PIR values were not computed if income screener information (INQ 220) was the only income information reported by the respondent.

DMDMARTL is the derived marital status variable. The marital status question was asked for sample persons 14 years of age and older. If this item was missing for sample persons 14 years of age older, marital status was imputed from other questionnaire items that made reference to their marital status.

Analytic Notes:

RIDRETH2 race/ethnicity recode should be used to compare NHANES 2001-2002 estimates of health measures with those of NHANES III.

Income variables: Income information was not obtained from all of the families in the survey sample. Some respondents refused to provide this information and others had little or no knowledge of family income. No attempt was made to assign or impute income in these instances. Incomplete information was reported to the extent possible.

Questionnaire section files, MEC examination files, and laboratory files can be linked to sample person demographics variables using the unique survey participant identifier SEQN. RIDSTATR provides the MEC examination status of a sample person.

The two-year sample weights (WTINT2YR, WTMEC2YR) should be used for NHANES 2001-2002 analyses and the four-year sample weights (WTINT4YR, WTMEC4YR) for combined analyses of NHANES 1999-2000 & NHANES 2001-2002 data. Please refer to the analytic guidelines provided with the data release files to determine the appropriate analytic methodology for this dataset.

Data Access:

The DEMO data are publicly available at www.cdc.gov/nchs/nhanes.htm

References

U.S. Census Bureau. Current Population Survey (CPS) Definitions and Explanations. 1003. Posted on the CPS website URL: <http://www.census.gov/population/www/cps/cpsdef.html>. Accessed 4/21/04.

Office of Management and Budget. Statistical Policy Directive No. 14. Definition of poverty for Statistical Purposes. May 1978. Posted on U.S. Bureau of the Census

website: URL: <http://www.census.gov/hhes/poverty/povmeas/ombdir14.html> Checked 4/21/04.

U.S. Census Bureau. Ratio of income to poverty level.
<http://www.census.gov/hhes/income/defs/ratio.html> 2003a. Checked 4/21/04.

U.S. Census Bureau. How the Census Bureau Measures Poverty
<http://www.census.gov/hhes/poverty/povdef.html> 2003b. Checked 4/21/04.
