

National Health and Nutrition Examination Survey

2005-2006 Data Documentation, Codebook, and Frequencies

Demographic Variables & Sample Weights (DEMO_D)

Data File: DEMO_D.xpt

First Published: November 2007

Last Revised: September 2009

Component Description

The Demographics public release file includes information that was collected using the Sample Person and Family Demographics questionnaires. An educational attainment variable (DMDEDUC3) was added to this updated Demographics file to provide consistency with the NHANES 1999-2004 Demographics files. An additional variable, an indicator variable for the month of exam, has been included.

In addition to demographics variables, the Demographics File includes the household interview and examination status codes, interview and examination sample weights, languages of interview for the household and examination interviews, proxy respondent codes, and information about the household reference person.

Disclosure risks and issues pertaining to confidentiality protection prevent NCHS from releasing all of the NHANES demographics variables publicly. Additional information may be publicly accessible through the NCHS Research Data Center (RDC). The RDC website has information about special request data files.

Eligible Sample

All survey participants who have a household interview record have a Demographics file record. The Demographics questions provide family-level and individual-level information. The target age groups for the Demographics questions vary. Please review the NHANES 2005-2006 Family and Sample Person Demographics questionnaires to determine the target age groups for the demographics questions.

Interview Setting and Mode of Administration

Persons 16 years of age and older and emancipated minors were interviewed directly. A proxy provided information for survey participants who were under 16 years of age and for individuals who could not answer the questions themselves.

The household interview was conducted in-person with an interviewer. The survey participants selected the language of interview (English or Spanish) or requested that a translator be used. Computer-assisted personal interview methods were used for the interviews. Several questions required the use of printed hand cards that listed the response choices or provided information that survey participants needed to answer the questions. The hand cards were printed in English and Spanish and the interviewer directed the respondent to the appropriate hand card during the interview. When necessary, the interviewers assisted survey participants by reading the response choices listed on the hand card.

The NHANES 2005–2006 interview and examination procedures manuals and questionnaires are posted on the NHANES website.

Quality Assurance & Quality Control

The NHANES computer-assisted personal interview (CAPI) software program, that was used to collect the interview data, had built-in data edit and consistency checks. The data edit checks alerted the interviewer when unusual or potentially erroneous data values were recorded. The consistency checks were used to alert the interviewer when information was recorded that was inconsistent with previous data entries or respondent characteristics such as the respondent's age. Questionnaire "skip" patterns were pre-programmed in the questionnaires to reduce respondent burden. Online information screens provided the interviewers with standardized descriptions of the terminology and concepts that were used in the questionnaires.

After data collection, interview records were reviewed by the NHANES field office staff for accuracy and completeness. A subset of the household interviews was verified by re-contacting the survey participants. The interviewers were required to record several interviews and the audio-taped interviews were reviewed by NCHS and contractor staff. The NHANES quality assurance and quality control procedures are described in the field procedures manuals that are posted on the NHANES website.

Data Processing and Editing

SDDSRVYR represents the two-year data release cycle number. A value of "4" denotes NHANES 2005–2006.

RIDSTATR is a recoded variable representing interview/examination status.

RIDEXMON is a variable indicating the six month time period when the examination was performed. A value

of "1" indicates November 1st through April 30th; a value of "2" indicates May 1st through October 31st.

RIDAGEYR: Age in years at the time of the screening interview is reported for survey participants between the ages of 1 and 84 years of age. All adults who were 85 years and older have RIDAGEYR values of '85'. The reporting of age in single years for adults 85 years and older was determined to be a disclosure risk. In NHANES 2005-2006, the weighted mean age for participants 85 years and older is 89 years.

The respondents' actual or imputed date of birth was used to calculate RIDAGEYR. The procedure that NCHS used to impute age at screener when the date of birth is missing or refused, but the respondent's age in years is provided is as follows: 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 the year of birth is missing or not given, it is imputed as the year of the screening interview minus the age in years provided by the respondent during the screening interview. Corrections are made to this imputed information for respondents who are less than 1 year of age at the time of screening.

RIDAGEMN is the respondent's age in months at the time of the household screening interview. RIDAGEMN is provided for respondents who are less than 85 years of age (RIDAGEYR<85). If the exact date of birth is not provided by the respondent, the age in months is calculated based on the imputed age at the time of the screener interview.

RIDAGEEX is age in months at the Mobile Examination Center (MEC) examination. RIDAGEEX is provided for subjects under 85 years of age at the time of household screening (RIDAGEYR<85). A total of 25 respondents who were under 85 years of age at the time of the household screening interview have missing data on RIDAGEEX because their exact date of birth was not available. RIDAGEEX was not calculated for individuals who have imputed values for age.

RIDRETH1: RIDRETH1 was derived from responses to the survey questions on race and Hispanic origin. Respondents who self-identified as "Mexican American" were coded as such (i.e., RIDRETH1=1) regardless of their other race-ethnicity identities. Otherwise, self-identified "Hispanic" ethnicity would result in code "2, Other Hispanic" in the RIDRETH1 variable. All other non-Hispanic participants would then be categorized based on their self-reported races: non-Hispanic white (RIDRETH1=3), non-Hispanic black (RIDRETH1=4), and other non-Hispanic race including non-Hispanic multiracial (RIDRETH1=5).

RIDRETH1 is the only race/ethnicity variable in the 2005-2006 Demographics data file. The Demographics files that were released for NHANES 1999-2004 had a variable called 'RIDRETH2' that provided an analytic link to the NHANES III race/ethnicity variable. RIDRETH2 is not included in the NHANES 2005-2006 Demographics file. The NHANES Analytic Guidelines provide additional information.

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

DMDCITZN: Citizenship status is reported using two codes: 1) citizen by birth or naturalization or 2) Not a citizen of the U.S. Persons who were born in the U.S. or US territories who acquire citizenship at birth were coded as US citizens.

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

RIDEXPRG: Pregnancy status at the time of the health examination was ascertained for females 8–59 years of age. The information used to code RIDEXPRG values included urine pregnancy test results and self-reported pregnancy status. Urine pregnancy tests were performed prior to the dual energy x-ray absorptiometry (DXA) exam. Menstruating girls 8–11 years of age and all females 12 years and over received a urine pregnancy test. Persons who reported they were pregnant at the time of exam were assumed to be pregnant. If the urine test was negative, but the subject reported they were pregnant, the status was coded as 'pregnant at exam' (RIDEXPRG = 1). If the urine pregnancy results were negative and the respondent stated that they were not pregnant, the respondent was coded 'not pregnant at examination' (RIDEXPRG = 2). Persons who were interviewed, but not examined have an RIDEXPRG value = 3 (could not be determined).

DMDYRSUS: This variable is the number of years the respondent has lived in the United States. Respondents who were born outside the U.S. were asked the month and year when they came to the U.S. to stay (DMQ.160). A small number of records were imputed because the respondent did not report the month of their arrival. A month value of 7 (July) was used to impute DMDYRSUS for these respondents. The responses to the question were recoded into 9 categories ranging from less than one year to 50 years or more.

DMDFMSIZ: This variable is the number of people in the respondent's family. The definition of family is based on the Current Population Survey (CPS) criteria. 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. Due to disclosure concerns, families that are comprised of 7 or more people are included the category that is labeled '7 or more'.

DMDHHSIZ: This variable is the number of people in the respondent's household. The values for this variable range from 1 to 7 with 7 being the code used for households comprised of 7 or more members.

DMDEDUC3: This variable provides information on the highest grade or level of education completed by respondents 6-19 years of age. The responses were recoded by NCHS as follows: single years of education (grades 1-12), high school graduate/GED, and post-high school. Code '66' (less than 9th grade) was used to categorize older youth who had very low education levels.

DMDEDUC2: This variable is the highest grade or level of education completed by adults 20 years and older. The response categories are: less than 9th grade education, 9-11th grade education (includes 12th grade and no diploma), High school graduate/GED, some college or associates (AA) degree, and college graduate or

higher. DMDEDUC2 provides more detailed information on education levels of adults compared to the categories released with the NHANES 1999-2004 Demographics files. The NHANES 1999-2004 Demographics data files will be revised to include DMDEDUC2. Please refer to the Analytic Notes for additional information on education level variables.

MDMSCHOL: This variable is school attendance status. It is asked for respondents 6-19 years of age.

DMQMILIT: This variable provides information on veteran's/military status. Veterans include persons who served in the armed forces of the United States, specifically, the U.S. Army, Navy, Air Force, Marine Corps, and Coast Guard.

INDFMINC: This variable is the total family income variable and is released as an income range value. NCHS used the U.S. Bureau of the Census Current Population Survey definition of "family" to group household members into one or more families (1). 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.

The income section of the household interview includes several questions about sources of income including wages, retirement income, disability payments, interest income, and assistance programs, but the amounts of income from each of the income sources were not obtained. The respondent was asked to report total 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 ≥ \$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 value was used as the total family income value.

Family income was used to calculate the ratio of income to poverty (INDFMPIR) and estimated total household income (INDHHINC).

INDHHINC: This variable is the estimated total household income. 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 the estimated total household income. The preferred method was to use income data reported by each CPS family that was interviewed (INQ200). A second method was used with multi-family households in which the total household income information (INQ200) was asked of the household reference person.

When income information was obtained from all 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 was not calculated for multi-family households when income range data were reported by any of the families in the household and in these instances, INDHHINC was coded as missing.

INDFMPIR: This variable is an index for the ratio of family income to poverty. The Department of Health and Human Services' (HHS) poverty guidelines were used as the poverty measure to calculate this index. These guidelines are issued each year, in the Federal Register, for determining financial eligibility for certain federal programs such as Head Start, Supplemental Nutrition Assistance Program (SNAP) (formerly Food Stamp Program), Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), and the National School Lunch Program.

The variable INDFMPIR was calculated by dividing family income by the poverty guidelines, specific to family size, as well as the appropriate year and state. The values were not computed if the income screener information (INQ 220: < \$20,000 or ≥ \$20,000) was the only family income information reported. If family income was reported as a range value, the midpoint of the range was used to compute the variable. Values at or above 5.00 were coded as 5.00 or more because of disclosure concerns. The values were not computed if the family income data was missing.

Household Reference Person: The household reference person is defined as the first household member 18 years of age or older listed on the Screener household member roster who owns or rents the residence where members of the household reside. The household reference person is comparable to "family reference person" in previous NHANES. Analysts frequently use information about the reference persons to characterize the socioeconomic status of the households where survey participants reside. The Demographics file includes information on the household reference person's gender (DMDHRGND), age (DMDHRAGE), country of birth (DMDHRBRN), education level (DMDHREDU), and marital status (DMDHRMAR). Additionally, information on the education level of the household reference person's spouse is included (DMDHSEDU).

SIALANG is the language used during the Sample Person Questionnaire interview

SIAPROXY denotes if a proxy respondent was used during the Sample Person Questionnaire interview.

SIAINTRP denotes if an interpreter was used during the Sample Person Questionnaire interview. The language spoken by the respondent is not reported.

FIALANG is the language used during the Family Questionnaire interview.

FIAPROXY denotes if a proxy respondent was used to complete the Family Questionnaire interview

FIAINTRP denotes if an interpreter was used to complete the Family Questionnaire interview. The language spoken by the respondent is not reported.

MIALANG is the language used during the Mobile Examination Center (MEC) interview.

MIAPROXY denotes if a proxy respondent was used during the Mobile Examination Center (MEC) interview.

MIINTRP denotes if an interpreter was used for the Mobile Examination Center (MEC) interview. The language spoken by the respondent is not reported.

AIALANG is the language used for the audio-computer-assisted self interview (A-CASI) portion of the MEC interview.

Analytic Notes

Income variables: None of the income values reported were imputed.

Education Level: Analysts who wish to include education level information in analyses that use multiple 2-year NHANES cycles may recode DMDEDUC2 (for ages 6-19) or DMDEDUC3 (for ages 20+) to create the education level variable that was released with the NHANES 1999-2004 data (variable name: DMDEDUC) as follows:

```
if (sddsrvyr = 4) then do;
  if dmdeduc2 in (1,2)      then dmdeduc = 1;
  else if dmdeduc2 = 3      then dmdeduc = 2;
  else if dmdeduc2 in (4,5) then dmdeduc = 3;
  else if dmdeduc3 in (0,1,2,3,4,5,6,7,8,9,10,11,12,55,66)
    then dmdeduc = 1;
  else if dmdeduc3 in (13,14) then dmdeduc = 2;
  else if dmdeduc3 = 15      then dmdeduc = 3;
  else                      dmdeduc = .;
end;
```

Masked Variance Units: Masked Variance Strata and Masked Variance Units or "MVUs" are included in the Demographics data file. The MVUs are a collection of secondary sampling units that are aggregated into groups for the purpose of variance estimation. The variance estimates that are produced using the Masked strata and MVUs closely approximate the variances that would have been estimated using the "true" sample design variance units that are based on the actual survey sample strata and primary sampling units (PSUs). MVUs are used to protect the confidentiality of information provided by survey participants and to reduce disclosure risks. The MVUs are described in the NHANES Analytic Guidelines. Analysts should review the Guidelines carefully prior to analyzing the survey data.

Sample Weights: The 2-year sample weights (WTINT2YR, WTMEC2YR) should be used for all NHANES 2005-2006 analyses. There are no 4-year weights in this file. The 4-year weights were only provided with the NHANES 2001-2002 release file. Detailed instructions for linking earlier datasets (1999-2000, 2001-2002, and 2003-2004) are provided in the NHANES 2005-2006 Analytic Guidelines.

Please refer to the Analytic Guidelines and the on-line NHANES Tutorial for further details on the use of sample weights and other analytic issues. Both of these are available on the NHANES website.

References

- U.S. Census Bureau. Current Population Survey (CPS) Definitions and Explanations. Revised 1/20/04. <http://www.census.gov/population/www/cps/cpsdef.html>
- U.S. Department of Health & Human Services. Poverty Guidelines, Research, and Measurement. Washington, DC: U.S. Department of Health & Human Services, January 2011. <http://aspe.hhs.gov/POVERTY/>

SEQN - Respondent sequence number

Variable Name:	SEQN
SAS Label:	Respondent sequence number
English Text:	Respondent sequence number.
Target:	Both males and females 0 YEARS - 150 YEARS

RIAGENDR - Gender

Variable Name: RIAGENDR
SAS Label: Gender
English Text: Gender of the sample person
Target: Both males and females 0 YEARS - 150 YEARS

Code or Value	Value Description	Count	Cumulative	Skip to Item
1	Male	5080	5080	
2	Female	5268	10348	
.	Missing	0	10348	

RIDAGEYR - Age at Screening Adjudicated - Recode

Variable Name:

RIDAGEYR

SAS Label:

Age at Screening Adjudicated - Recode

English Text:

Best age in years of the sample person at time of HH screening. Individuals 85 and over are topcoded at 85 years of age.

Target:

Both males and females 0 YEARS - 150 YEARS

Code or Value	Value Description	Count	Cumulative	Skip to Item
0 to 84	Range of Values	10178	10178	
85	>= 85 years of age	170	10348	
.	Missing	0	10348	

RIDRETH1 - Race/Ethnicity - Recode

Variable Name: RIDRETH1

SAS Label: Race/Ethnicity - Recode

English Text: Recode of reported race and ethnicity information.

Target: Both males and females 0 YEARS - 150 YEARS

Code or Value	Value Description	Count	Cumulative	Skip to Item
1	Mexican American	2847	2847	
2	Other Hispanic	349	3196	
3	Non-Hispanic White	3928	7124	
4	Non-Hispanic Black	2710	9834	
5	Other Race - Including Multi-Racial	514	10348	
.	Missing	0	10348	