**2001 National Health Interview Survey (NHIS)**
**Public Use Data Release**

# NHIS Survey Description

**Division of Health Interview Statistics**
**National Center for Health Statistics**
**Hyattsville, MD**

**Centers for Disease Control and Prevention**
**U.S. Department of Health and Human Services**

**January, 2003**

**2001 National Health Interview Survey (NHIS)**

**Public Use Data Release**

**Table of Contents**

## The NCHS Web Page and NHIS Electronic Mail List

Data users can obtain the latest information about the National Health Interview Survey by periodically checking our web site:

**<http://www.cdc.gov/nchs/nhis.htm>**.

The web site features downloadable public use data and documentation for the 2001 NHIS, as well as important information about any modifications or updates to the data and/or documentation. Published reports from previous years' surveys are also available, as are updates about future surveys and datasets.

Researchers may also wish to join the NHIS electronic mail list. To do so, scroll down to "Related Links" on the NHIS web page, and then click on "NHIS Listserve". Fill in the appropriate information, and click on the "National Health Interview Survey (NHIS) researchers" box, followed by the "Subscribe" button at the bottom of the page. The listserve is made up of over 3,000 NHIS data users located around the world who receive e-news about NHIS surveys (e.g., new releases of data or modifications to existing data), publications, workshops, and conferences.

The Division of Health Interview Statistics also maintains a "help line" phone number and email address. Users may contact us with their data requests and questions at 301-458-4901, or send email to us at nhislist@cdc.gov.

## New NHIS Topics/Questions for 2001

The 2001 NHIS supplements are in the Core files. The terms "supplement" and/or "supplementary questions" refer to any co-sponsored questions that are in the NHIS for a year at a time. Beginning in 1997, co-sponsored questions were referred to as "topical module" and "periodic module", but these terms proved to be neither mutually exclusive nor exhaustive of the possible types of supplements. Therefore, effective 2001, we are using the terms "supplement" or "supplementary questions" to describe co-sponsored questions.

A supplement or supplementary questions may be merged among core questions, or may be placed following the end of a section of the core. For 2001, supplements are released in the same file as other Core data. The existence of three extra digits (.xxx) at the end of the variable name helps to identify supplementary questions. A list of all co-sponsored supplements is on page 18. A list of changes to existing questions in the NHIS Core is located on page 15.

# 2001 National Health Interview Survey (NHIS)
# Public Use Data Release

## Introduction

The National Health Interview Survey (NHIS) is a multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and is the principal source of information on the health of the civilian, noninstitutionalized, household population of the United States. The NHIS has been conducted continuously since its beginning in 1957. Public use microdata files are released on an annual basis.

The NHIS Core questionnaire items were revised every 10-15 years, with the last major revisions occurring in 1982 and in 1997. The NHIS that was fielded from 1982-1996 consisted of two parts: (1) a set of basic health and demographic items (known as the Core questionnaire) that remained stable from one survey year to the next, and (2) one or more sets of questions on current health topics that varied with each survey, referred to as Supplements. Despite periodic revisions to the Core questionnaire, Supplements played an increasingly important role in the survey as a means of enhancing topic coverage in the Core. Eventually, certain Supplements, such as "Family Resources" and "Childhood Immunization", were incorporated in the NHIS Core on an annual basis.

The redesigned NHIS introduced in 1997 consists of a Basic Module or Core as well as variable Supplements. The Basic Module, which remains largely unchanged from year to year, consists of three components: the Family Core, the Sample Adult Core, and the Sample Child Core. The Family Core component collects information on everyone in the family, and its sample also serves as a sampling frame for additional integrated surveys, as needed. Information collected on the Family Core for all family members includes: household composition and socio-demographic characteristics, tracking information, information for matches to administrative data bases, and basic indicators of health status, activity limitations, injuries, health insurance coverage, and access to and utilization of health care services.

From each family in the NHIS, one sample adult and one sample child (if any children under age 18 are present) are randomly selected, and information on each is collected with the Sample Adult Core and the Sample Child Core questionnaires. Because some health issues are different for children and adults, these two questionnaires differ in some items, but both collect basic information on health status, health care services, and behavior. These sections of the survey yield the Sample Adult, Sample Child, and Child Immunization files.

The Family Core yields several data files, including the Household-Level file, the Family-Level file, the Person-Level file, and two data files pertaining to injuries. Because these files contain the same or comparable variables from one survey year to the next, they are suitable for trends analysis;

moreover, multiple years of these data may be easily pooled to increase the sample size for analytic purposes.

## Data Collection Procedures

The U.S. Census Bureau, under a contractual agreement, is the data collection agent for the National Health Interview Survey. NHIS data are collected through a personal household interview by Census interviewers. Nationally, the NHIS uses about 400 interviewers, trained and directed by health survey supervisors in the 12 U.S. Census Bureau Regional Offices. The supervisors are career Civil Service employees whose primary responsibility is the NHIS, and they are selected through an examination and testing process. Interviewers (also referred to as Field Representatives, or "FRs") receive thorough training on an annual basis in basic interviewing procedures and in the concepts and procedures unique to the NHIS.

For the Family Core component of the Basic Module, all adult members of the household 17 years of age and over who are at home at the time of the interview are invited to participate and to respond for themselves. For children and those adults not at home during the interview, information is provided by a knowledgeable adult family member (18 years of age or over) residing in the household. For the Sample Adult questionnaire, one adult per family is randomly selected; this individual responds for him/herself to the questions in this section. Information for the Sample Child questionnaire is obtained from a knowledgeable adult residing in the household.

The NHIS is conducted using computer-assisted personal interviewing (CAPI). The CAPI version of the NHIS questionnaire is administered using laptop computers, which allow interviewers to read questions from and enter responses directly into the computer during the interviews.

## Sample Design

Traditionally, the sample for the NHIS is redesigned and redrawn every ten years to better measure the changing U.S. population and to meet new survey objectives. The fundamental sample design structure of the 1995-2004 NHIS is similar to that of the 1985-1994 NHIS; however, there were two major changes to the 1995-2004 sample design. First, a state-level stratification increased the number of primary sampling units (PSUs) from 198 to 358. This enhanced the capability of using the NHIS for state estimation and *future* dual-frame surveys at the state level. (Users should note that the NHIS is currently not designed to provide state-level estimates; however, in some cases this can be done, particularly for those states with larger populations. Contact the NCHS Research Data Center for more information, or visit their web page: **<http://www.cdc.gov/nchs/r&d/rdc.htm>**.) Second, both the black and Hispanic populations are now oversampled to allow for more precise estimation of health in these growing minority populations. In the previous design, only black Americans were oversampled.

Two other important features first implemented in the 1985-1994 design continue. NCHS survey integration and followback surveys are facilitated by an area frame with independent address lists; while the area frame is based on the preceding decennial Census, the address lists are obtained in a separate listing activity, explicitly for the NHIS. Also, the NHIS sample is divided into four representative panels to further facilitate integration with other NCHS surveys. See NCHS Series 2, Number 130, for a description of the 1995-2004 survey design, the methods used in estimation, and general characteristics of the data obtained from the survey. This publication is available on-line at **<http://www.cdc.gov/nchs/data/series/sr_02/sr02_130.pdf>**. (Users may also be interested in another Series 2 (number 126) report, *National Health Interview Survey: Research for the 1995-2004 Redesign*, which is available at **<http://www.cdc.gov/nchs/data/series/sr_02/sr02_126.pdf>**.)

## Weighting Information

The sample is chosen in such a way that each person in the covered population has a known non-zero probability of selection. These probabilities of selection, along with adjustments for nonresponse and post-stratification, are reflected in the sample weights that are provided in the accompanying data files.

Since the NHIS uses a multistage sample designed to represent the civilian noninstitutionalized population of the United States, it is necessary to utilize the person's basic weight for proper analysis of person record data. In addition to the design and ratio adjustments included in the calculation of the Person file's basic weights, the person weights are further modified by adjusting them to Census control totals for sex, age, and race/ethnicity population (post-stratification).

Each file has one or more sets of weights based on the unit of analysis. Two sets of weights are provided on the Person-Level file:

> Weight - Final Annual (WTFA) is based on design, ratio, non-response and post-stratification adjustments. This should be used in most analyses of the Family/Person data. National estimates of all person-level variables can be made using these weights.
>
> Weight - Interim Annual (WTIA) does not include the post-stratification adjustment (age-sex-race/ethnicity adjustment to Census population control totals). It is required by some software packages for variance estimation for surveys with complex sample designs.

The Sample Adult data file contains two sets of weights:

> Sample Adult Weight - Final Annual (WTFA_SA) includes design, ratio, non-response and post-stratification adjustments for sample adults. National estimates of all adult sample variables can be made using these weights.

Sample Adult Weight - Interim Annual (WTIA_SA) does not include the post-stratification adjustment (age-sex-race/ethnicity adjustment to Census population control totals). It is required by some software packages for variance estimation for surveys with complex sample designs.

Two sets of weights are also included on the Sample Child data file:

Sample Child Weight - Final Annual (WTFA_SC) includes design, ratio, non-response and post-stratification adjustments for sample children. National estimates of all sample child variables can be made using these weights.

Sample Child Weight - Interim Annual (WTIA_SC) does not include the post-stratification adjustment (age-sex-race/ethnicity adjustment to Census population control totals). It is required by some software packages for variance estimation for surveys with complex sample designs.

Two sets of weights are provided on the Immunization (Child) data file from the Sample Child Core:

Weight - Final Annual (WTFA_IM) includes design, ratio, non-response and post-stratification adjustments for sample children who are all under 18 years of age and for additional children aged 12-35 months.

Weight - Interim Annual (WTIA_IM) does not include post-stratification adjustment (age-sex-race/ethnicity adjustment to Census population control totals). It is required by some software packages for variance estimation for surveys with complex designs.

In addition, two sets of weights are provided on the Household file:

Weight - Final Annual Household (WTFA_HH) includes the probability of selection and non-response adjustments. This weight does not include a post-stratification adjustment to Census control totals for the number of civilian, non-institutionalized households in the U.S. because suitable control totals do not exist. Non-responding households have a zero weight in this field. WTFA_HH is the appropriate weight to use when analyzing only responding households.

Weight - Interim Annual Household (WTIA_HH) reflects the probability of household selection. It does not include non-response or post-stratification adjustments. WTIA_HH is the appropriate weight to use when analyzing all households in the file.

Lastly, the Family-Level weight is discussed in greater detail in the section of this document pertaining to the Family file (see page 23).

**NOTE**: Analysts should be aware that 265 persons are on the Person-Level file who were active duty members of the Armed Forces at time of interview, despite the fact that NHIS covers only the civilian noninstitutionalized household population. The value of WTFA for these persons is zero, so they will not be counted when making national (i.e., weighted) estimates. Data for these Armed Forces members are included in all relevant files in order to aid any analyses pertaining to family structure or relationships. No active duty Armed Forces members were selected as sample adults.

Recall Period and Weights

Some questions for particular events have recall periods referring to, for example, the "last 2 weeks" or "last 3 months". In general, annual estimates can be made using these types of variables. For example, for a variable that counts events within a two-week recall period, (variable)(26)(WTFA) = annual estimate; for a variable with a three-month recall period, (variable)(4)(WTFA) = annual estimate. This assumes that the average rate of occurrence is the same over the last year as over the last two weeks (or three months). Analysts are cautioned to check the accompanying file documentation and the questionnaire in order to insure that annual estimates for these kinds of event variables are possible and have intrinsic meaning.

Variance Estimation

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. Because of this complex design and adjusted sampling weights, the direct application of standard statistical analysis methods for estimation and hypothesis testing may yield misleading results. If data are not weighted, severely biased estimators may result. For this reason, as indicated previously, it is necessary to use the weights that are included in the accompanying data file for analyses.

Weighted data used in standard software packages may provide unbiased point estimates for commonly computed first-order statistics like means or regression coefficients, but the computed standard errors of the estimates may be too small. Also, standard packages may produce hypothesis test results (such as p values) that are incorrect. Hence, it is recommended that users of NHIS data utilize computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs. NCHS uses Taylor series linearization methods for NHIS variance estimation. Appendix III provides SUDAAN code and a description of its use to compute standard errors of means, percentages and totals with the 2001 NHIS database.

Analyses of large NHIS domains usually produce reliable estimates but analyses of small domains may yield unreliable estimates, as indicated by their large variances. The analyst should pay particular attention to the coefficient of variation (relative standard error) for estimates of

means, proportions and totals. In addition, small sample sizes, or small numbers of primary sampling units containing targeted data, may be an indication of estimates lacking precision.

## General Information About the 2001 Data

The interviewed sample for 2001 consisted of 38,932 households, which yielded 100,761 persons in 39,633 families. The interviewed sample for the Sample Adult component, which required self-response to all questions, was 33,326 persons 18 years or age and older. The interviewed sample for the Sample Child component, by proxy response from a knowledgeable adult in the family, was 13,579 children 0-17 years old. Lastly, the interviewed sample for the Immunization section, again, by proxy response from a knowledgeable adult in the family, was 14,709 children aged 17 and younger. Data were not collected on any infant who was born during the assignment week of the interview.

The total household response rate was approximately 88.9%: 7.3 percentage points of the noninterview rate (11.1%) were the result of respondent refusal and unacceptable partial interviews. The remaining 3.8 percentage points were primarily the result of failure to locate an eligible respondent at home after repeated calls.

The conditional response rate for the Sample Adult component was 84.2% of persons identified as sample adults. The final response rate for the Adult Sample Person component is calculated as (Overall Family Response Rate)(Sample Adult Response Rate), or (87.6%)(84.2%) = 73.8%.

The conditional response rate for the Sample Child component was 92.0%, which was calculated by dividing the number of completed Sample Child interviews (13,579) by the total number of eligible sample children (14,766). The unconditional or final response rate for the Sample Child component was calculated by multiplying the conditional rate by the overall family response rate of 87.6%, yielding a rate of 80.6%.

The total noninterview rate for the Immunization (Child) section of the Sample Child component was 1.9% of persons asked to respond to the Immunization (Child) section. The final response rate for the Immunization (Child) file was 80.3%.

Additional information about NHIS response rates can be found in Appendix I.

## Information About the 2001 Data File Documentation

Along with the redesign of the NHIS questionnaire in 1997, other aspects of the data production process were also modified. As with the previous design, each data file has its own data documentation file (also known as a “codebook” or “data dictionary”), which is available on

the NHIS web site, **<http://www.cdc.gov/nchs/nhis.htm>**. Beginning with the 1997 data (and continuing with subsequent years), the format and content of these files, henceforth referred to as Dataset Documentation, have changed. For most variables, the Dataset Documentation provides the actual question that generated the data, questionnaire location information, universe, values, value labels, and frequency counts. Additional specific information is provided under "Sources", "Recodes", "Keywords", and "Notes". These terms are defined as follows:

Sources - If the variable in question is a recode, then all variables that were used to make this recode are listed.

Recodes - A *recode* is a variable derived from the reordering or collapsing of another variable, such as the family income recode (INCGRP) found in the Person-Level file. Alternatively, a recode may be constructed from two or more variables, as is the body mass index (BMI) variable included in the Sample Adult file. If a particular variable was used in making other recode variables, then those recode variables are listed. Users will note that a number of standardized variables appear in the dataset. A *standardized variable* is a particular type of recode based on time unit information obtained during the course of the interview. When respondents are asked any questions pertaining to time – for example, how long the respondent has worked at his/her job – the answer is typically obtained in two parts. The respondent provides the number of time units ("1-365"), followed by the type of time unit (days, weeks, months, or years). During the course of data editing, this information is standardized into an appropriate time unit (in the case of job history, years on the job). Some of the standardized time unit recodes may also be top-coded for confidentiality reasons.

Keywords - *Keywords* are descriptive words or phrases relevant to the topic of the variable; these can be used for word searches.

Notes - *Notes* provide unique information that analysts need to know about a particular variable, such as assumptions, limitations, caveats, differences between instrument versions, or other important information. Analysts are encouraged to read the notes for variables of interest. Currently, there are two generic notes that can appear in addition to specific information:

> 1) If the original questionnaire item was asked at the family level but resulted, after the editing process, in a person-level variable, this note is added: Family/person variable conversion
>
> 2) If other questions in the instrument ask about the same topic, or if similar questions appear in other sections of the instrument, this note is added: Refer to {variable name and section number} for a {family/person/child} level question on a related topic.

The universe refers to those respondents deemed eligible to answer a given question. For example, the universes for all Sample Adult variables are specified as "ASTATFLG = 1 and AGE = 18+", or "ASTATFLG eq 1 and AGE ge 18 ", followed by any other universe descriptors

specific to the variable. ASTATFLG = 1 refers to a variable on the Person file and indicates that the respondent was selected as a sample adult and answered at least the first three sections of the Sample Adult component (constituting a completed interview or an acceptable partial interview). Responses for persons who stopped answering key questions after the first three sections will appear as 8's (not ascertained) for the corresponding variables throughout the remaining Sample Adult variables.

The universes for all Sample Child variables are specified as CSTATFLG = 1 and AGE < 18 (or AGE lt 18), followed by any other universe descriptors specific to the variable. CSTATFLG = 1 refers to a variable on the Person file that indicates a selected Sample Child with a completed interview or an acceptable partial interview (completion at least through the CHS section, or about half the questions). Again, responses from acceptable partial interviews have a code of 8, meaning "not ascertained", throughout the remaining, unanswered Sample Child sections.

The universes for the Immunization (Child) file from the Sample Child Core are specified as IMMUNFLG = 1 and AGE < 18 (or relevant sub-grouping thereof, to indicate younger children), followed by any other universe descriptors specific to the variable, including two categories: with a shot record (SHOTRC = 1) or without a shot record (SHOTRC = 2, 7-9). IMMUNFLG = 1 indicates a classification for a good immunization record for children under 18 years of age. Additional variables include ICAGEMR, which refers to a recoded variable for age in months, and IMRESPNO, which refers to the person who was originally recorded as the respondent for the sample child. The NHIS does not record, specifically, whether this adult respondent answered the Child Immunization questions for the sample child, or for the additional children aged 12-35 months included in the file.

Within the NHIS, the same codes are used across all files to designate "Refused" and "Don't know" responses: refusals are coded as "7" (with leading 9's to the length of the field, as in 7, 97, 997, etc.), while "don't know" responses are "9" (again, with leading 9's to the length of the field, such as 9, 99, 999, etc.). A code of "8" is used to indicate "Not ascertained" responses, which typically occur when an in-the-universe respondent had a blank field or the field contained an impossible code. Lastly, in some limited situations (primarily recodes), the "Refused", "Don't know", and "Not ascertained" categories are collapsed into a single category called "Unknown", which is typically designated with a "9" (with leading 9's to fill out the field, if necessary).

## Information About the 2001 CAPI Questionnaire

The NHIS CAPI questionnaire, also referred to as the CAPI Reference Questionnaire or CRQ, is an integral part of the data documentation and should be consulted when analyzing data. Users desiring greater detail should also consult the 2001 NHIS Field Representative's Manual (both the CRQ and FR Manual are available on the NHIS web site, **<http://www.cdc.gov/nchs/nhis.htm>**). Every effort was made to insure that the variable names

in the data are consistent with the question items in the instrument. In a few cases, this was not possible. When there is a question about variable names, matching the question number in the instrument to the variable number in the Dataset Documentation can resolve any discrepancies.

Because the questionnaire for the NHIS is administered by computer, the questionnaire exists as a long and complex computer program. While stringent quality control measures were applied, a few errors are known to have occurred in the program. The most common errors caused questions to be asked that were inappropriate in view of the respondent's previous answers. These errors are commonly referred to as "skip pattern errors". Various other instrument problems were identified over the course of the year, and efforts were taken to correct these errors. Some of these problems were resolved through correction of skip patterns, question wording changes, addition of questions, and other internal instrument corrections.

When errors were detected and diagnosed, and time permitted, the instrument was changed to correct for the errors. In 2001, instrument changes were kept to a minimum, so that there was basically one version of the NHIS in the field across all four quarters of the survey year. Analysts are encouraged to read the notes in the Dataset Documentation for important information pertaining to specific variables that may have changed across quarters.

## Questionnaire Sections

The 2001 NHIS contained the annual Basic Module, which is broken into various sections that group questions into broad and specific categories. Each section is designated by a section title and corresponding three-digit acronym (or section code); questionnaire items are numbered sequentially (but not consecutively) within their respective sections, with the section acronym making up part of the item number. For example, the first item in the Household Composition section is identified as HHC.010; note that HHC.010 also has an associated variable name, RPNAME. The list on the following page details the various questionnaire sections, their acronyms and description titles.

## Table 1. 2001 NHIS Core Questionnaire Sections and Topics

### A. Household

| Section No. | Section Code | Description |
|---|---|---|
| **I** | HHC | Household Composition |

### B. Family Core

| Section No. | Section Code | Description |
|---|---|---|
| **I** | FID | Family Identification and Verification |
| **II** | FHS | Health Status and Limitation of Activity |
| **III** | FIJ | Injury |
| **IV** | FAU | Health Care Access and Utilization |
| **V** | FHI | Health Insurance |
| **VI** | FSD | Socio-demographic |
| **VII** | FIN | Income and Assets |

### C. Sample Adult Core

| Section No. | Section Code | Description |
|---|---|---|
| **I** | AID | Identification and Verification |
| **II** | ACN | Conditions |
| **III** | AHS | Health Status and Limitation of Activity |
| **IV** | AHB | Health Behaviors |
| **V** | AAU | Health Care Access and Utilization |
| **VI** | ASD | Demographics |
| **VII** | ADS | AIDS |

**D. Sample Child Core**

| Section No. | Section Code | Description |
|---|---|---|
| **I** | CID | Identification and Verification |
| **II** | CHS | Conditions, Limitation of Activity and Health Status |
| **III** | CAU | Health Care Access and Utilization |
| **IV** | CIM | Immunization |

**E. Recontact**

| Section No. | Section Code | Description |
|---|---|---|
| **I** | RCI | Recontact Information and Follow-up |

In addition to the three Core sections comprising the Basic Module, the 2001 NHIS contains several other data files: the Household- and Family-level files, the Injury and Poisoning Episode file, and the Injury and Poisoning Verbatim file. The Household file is derived largely from the Household composition section of the Basic Module and describes characteristics of each household. The variables contained in the Family-level file are reconstructions of the person-level data from the Basic Module sections at the family level. The Injury and Poisoning files are derived from the information obtained from the injury/poisoning questions in the Family Core section.

## Changes/Additions/Deletions to 2001 Core

A number of changes were introduced to the Core sections of the 2001 NHIS, resulting in new, changed, or deleted variables (relative to 2000). Following is a brief summary. Users are *strongly* encouraged to check the notes in the Dataset Documentation, as well as the relevant sections in this document, for more information.

### Family Core

| Section | Variable | Brief Description of Change |
|---|---|---|
| FHS | PSPEDEM | New variable (child receives services because of emotional or behavioral problem) |
| FHS | LAHCC12 - LHCCLY12 | New answer category and related time variables (learning disability) |
| FHS | LAHCC13 - LHCCLY12 | New answer category and related time variables (ADD/ADHD) |
| FHI | MCNAME | New variable (type of medicare HMO) |

| | | |
|---|---|---|
| FSD | HEADSTV1 | New variable (replaces HEADSTEV) |
| FSD | DOINGLW1 | New variable (replaces DOINGLW) |
| FSD | WHYNOWK1 | New variable (replaces WHYNOWRK) |
| FSD | WRKLYR1 | New variable (replaces WRKLYR) |
| FIN | ELIGPWIC | New variable (indicates if person was a member of a family with a WIC age-eligible person) |
| FIN | WIC_FLAG | New variable (indicates if person who received WIC benefits was age-eligible for the WIC program) |

**Sample Adult Core**

| Section | Variable | Brief Description of Change |
|---|---|---|
| ACN | AASSTILL | New variable (still have asthma) |
| ACN | ARTH | New variable (ever told had arthritis) |
| AHS | WRKLYR3 | New variable (replaces WRKLYR2) |
| AHB | ALC7STAT | New variable in 2001 (recode of alcohol consumption) |
| AAU | ADNLONG2 | Replaces ADNLONGR (change in question wording) |
| ASD | BUSINC | Deleted in 2001 |
| ASD | LOCPRTNO | Deleted in 2001 |
| ASD | ALL_SA | New recode (employment status, all sample adults) |
| ASD | EVERWRK | New variable (has sample adult ever worked?) |
| ASD | INDSTR1A | New recode (detailed industry recode for sample adults who have ever worked) |
| ASD | INDSTR2A | New recode (simple industry recode for sample adults who have ever worked) |
| ASD | OCCUP1A | New recode (detailed occupation recode for sample adults who have ever worked) |
| ASD | OCCUP2A | New recode (detailed occupation recode for sample adults who have ever worked) |
| ASD | WRKCATA | New variable (category of worker for sample adults who have ever worked) |
| ASD | BUSINC1 | New variable (is business incorporated, for currently employed sample adults) |
| ASD | BUSINC1A | New variable (is business incorporated, for sample adults who have ever worked) |
| ASD | LOCALL1A | New variable (number of employees at workplace, for sample adults who have ever worked) |
| ASD | YRSWRKPA | New recode (number of years on the job, for sample adults who have ever worked) |
| ASD | HOURPDA | New variable (paid by the hour for sample adults who have ever worked) |
| ASD | PDSICKA | New variable (paid sick leave for sample adults who have ever worked) |

| | | |
|---|---|---|
| ADS | WHYTST_R | New variable (replaces WHYTST_C) |
| ADS | REATST_C | New variable (replaces REATST_R) |
| ADS | REASWHOR | New variable (replaces REASWHO) |

**Sample Child Core**

| Section | Variable | Brief Description of Change |
|---|---|---|
| CHS | CASSTILL | New question (child still have asthma) |
| CHS | CMHAGF22-25 | Deleted in 2001 (replaced by the SDQ) |
| CHS | CMHAGF32-35 | Deleted in 2001 (replaced by the SDQ) |
| CHS | CMHAGM22-25 | Deleted in 2001 (replaced by the SDQ) |
| CHS | CMHAGM32-35 | Deleted in 2001 (replaced by the SDQ) |

**Immunization**

| Section | Variable | Brief Description of Change |
|---|---|---|
| CIM | INFMOR | New variable (ever rec'd an additional influenza shot) |
| CIM | INFMNO | New variable (did -- receive influenza shot past 12 months) |
| CIM | HEPAMOR | New variable (ever rec'd an additional hepatitis A shot) |
| CIM | HEPAMNO | New variable (number of additional hepatitis A shots received) |
| CIM | OTHRAY | New variable (received any other immunizations) |
| CIM | OTHREV01 | New variable (influenza vaccine) |
| CIM | OTHREV02 | New variable (hepatitis A vaccine) |
| CIM | OTHREV03 | New variable (other vaccine) |
| CIM | INFENO | New variable (has -- received influenza shot past 12 months) |
| CIM | HEPAENO | New variable (number of hepatitis A vaccines rec'd) |
| CIM | OTH1ENO | New variable (number of other shots received) |
| CIM | ROT | Deleted in 2001 |
| CIM | ROTDT | Deleted in 2001 |
| CIM | ROTMOR | Deleted in 2001 |
| CIM | ROTMNO | Deleted in 2001 |
| CIM | ROTEV | Deleted in 2001 |
| CIM | ROTENO | Deleted in 2001 |

## Supplemental Question Co-Sponsors, 2001 NHIS

| Topic | Agency | Title | Survey Section/Numbers |
|---|---|---|---|
| Healthy People 2010 | Health Resources and Services Administration (HRSA) | Emergency Access Care Questions | AAU.061.010<br>AAU.061.020 |
| Healthy People 2010 | Agency for Healthcare Research and Quality (AHRQ) | Access to Care Questions | FAU.040.010-FAU.040.020; AAU.050.010-AAU.050.040; AAU.050.090; AAU.050.120-AAU.050.150 |
| Healthy People 2010 | National Institute of Neurological Disorders and Stroke (NINDS), National Institute of Health (NIH) | Stroke Warning Signs Questions | ACN.031.030 |
| Child Mental Health Supplement | National Institute of Mental Health (NIMH), National Institute of Health (NIH) | Questions on Child Mental Disorders, Symptoms | FHS.065; CAU.20.010; CAU.345.005-CAU.345.100 |
| **Centers for Disease Control and Prevention (CDC) Agreements** | | | |
| Healthy People 2010 | National Center for Environmental Health (NCEH) | Disability and Secondary Conditions: Social Support | ACN.530.020-ACN.530.050 |
| Healthy People 2010 | National Center for Chronic Disease Prevention and Health Promotion ( NCCDPHP) | Heart Disease, CPR and Stretching Exercise Questions | **Heart Disease:**<br>ACN.031.010-ACN.031.020<br>**CPR:**<br>ACN.031.040-ACN.031.050<br>**Stretching:**<br>AHB.130.010-AHB.130.020 |
| Immunization | National Immunization Program (NIP) | NHIS Immunization Questions | CIM.010-CIM.790 |
| **Agencies Providing General Support** | | | |
| n/a | Substance Abuse and Mental Health Services Administration (SAMHSA); Center for Mental Health Service (CMHS) | Collection and Analysis of Mental Health Data using NHIS | |

## 2001 National Health Interview Survey
## Household-Level File

Each record on the Household file represents a unique household included in the NHIS sample or sampling frame. Each household can be identified by using the household identification variable (HHX). Note that the Household file is considered as the base file from which all other files are built. That is, the main sampling unit in the NHIS is the household, and each record on the Household file represents an eligible sampling unit.

Some of the variables found only in this file include: month and year of interview completion, nature/reason for "Type A" non-responses, household telephone service, and number of responding and non-responding families and persons. (For information about Type A non-response, see Appendix I.) Variables in other NHIS data files that may be appropriately analyzed at the household level can be merged with this file for analysis.

The universe for the Household file is all eligible households, including both responding households and non-responding (Type A) households. The Household file contains information on 43,797 households: 38,932 households were interviewed, while 3,181 refused to participate. The nature of refusals for Type A households is detailed in the variable NON_INTV.

The total non-interview rate for the Household file was 10.2% of households. The response rate for the Household file is calculated as the number of responding households divided by the total number of eligible households (responding + non-responding households), or 89.8%.

## 2001 National Health Interview Survey
## Family-Level File

The Family-Level file contains variables that describe characteristics of the 39,633 families living in households that participated in the 2001 NHIS. Thus, each record in the file represents a unique family. The universe for all variables in this file is limited to all responding families in those households participating in the 2001 survey; this is specified as FM = ALL in the Family file Dataset Documentation. Users wishing to determine the number of responding and non-responding families in each household are referred to ACPT_FAM and REJ_FAM in either the Family or Household files.

As Table 2 indicates, most NHIS families consist of a group of two or more related persons who are living together in the same occupied housing unit (i.e., household) in the sample. (In some instances, unrelated persons sharing the same household may also be considered as one family, such as unmarried couples who are living together.) All relationships in the family are recorded relative to a household reference person, who is the person who owns or rents the housing unit. If more than one person owns or rents the housing unit, the oldest among them is designated the household reference person; if none of the family members owns or rents the unit, then the oldest person in the family is designated the reference person. Note that when there is only one family per household, all household and family relationships (as indicated by the Person file variables RRP and FRRP, respectively) will be identical.

**Table 2. Number of Families per Household, 2001 National Health Interview Survey (unweighted counts)**

| Families per household | Frequency | Percent |
|---|---|---|
| 1 | 38,859 | 98.1 |
| 2 | 668 | 1.7 |
| 3 | 77 | 0.2 |
| 4 | 22 | 0.1 |
| 5-8 | 7 | 0.0 |

In the small number of instances where there is more than one unrelated family living in a single household, the various NHIS questionnaires (e.g., Family Core, Sample Adult Core, etc.) will then be administered separately to *each* family within the sampled household. Moreover, one household reference person is chosen for the housing unit (again, this will be the person who owns or rents the unit), *and* one family reference person is chosen for each distinct family within the household. Each family in the household will thus have the same household reference person but a *different* family reference person, and all relationships in both the household and the family will be described relative to these two persons. Examples of multi-family households include several unrelated roommates sharing a house or apartment; a family with an unrelated lodger and his/her child; a family with a live-in housekeeper and his/her spouse; etc.

Family size may vary considerably. Table 3 shows a breakdown of the 39,633 families by number of family members. Again, note that multiple families may share one household; this information can be determined from ACPT_FAM and REJ_FAM, or HHX and FMX, in the Family file.

**Table 3. Size of Family, 2001 National Health Interview Survey (unweighted counts)**

| Number of Members | Frequency | Percent |
|---|---|---|
| 1 | 11,645 | 29.4 |
| 2 | 11,879 | 30.0 |
| 3 | 6,248 | 15.8 |
| 4 | 5,563 | 14.0 |
| 5 | 2,655 | 6.7 |
| 6 | 996 | 2.5 |
| 7 | 351 | 0.9 |
| 8 | 165 | 0.4 |
| 9 | 59 | 0.2 |
| 10 | 36 | 0.1 |
| 11 | 20 | 0.1 |
| 12 | 6 | 0.0 |
| 13 | 4 | 0.0 |
| 14 | 2 | 0.0 |
| 15 | 2 | 0.0 |
| 16 | 1 | 0.0 |
| 25 | 1 | 0.0 |

The first part of the Family file contains the technical variables that identify or describe the record type (all observations in this file have a record type value of "60"), the survey year, the household and family numbers, the interview quarter and week, characteristics of the family's housing unit, geographic information associated with the housing unit, variables used for variance estimation, and a family-level weight variable. The second part of the file consists of a series of recodes derived from the six family core sections of the NHIS that collapse the 100,000+ individual level observations into information about their respective families. Generally, two types of recodes are possible. The first is a simple "yes-no" measure that indicates whether *any* family member falls into a particular category or exhibits a particular characteristic. These variables are equivalent to, but not directly derived from, the family-level questions in the instrument.

Every yes-no measure also has a corresponding counter that indicates the number of family members in that category or with that characteristic. Note that counters always consist of values from zero to 30; in addition, blanks are also possible if a family is not contained in the universe for a specific question. For example, FSALYN and FSALCT, two recodes from the Income and Assets section of the Family Core, are limited to families with at least one member aged 18 or older; families consisting solely of emancipated minor(s) are coded as blanks to indicate that they are out of the universe. The Family file also contains some counters that lack corresponding yes-no indicators.

For example, FHSTATEX, FHSTATVG, FHSTATG, FHSTATFR, and FHSTATPR (all derived from PHSTAT, FHS.310) provide counts of the number of family members in excellent, very good, good, fair, and poor health, respectively. Counters were also constructed to indicate the number of working adults in the family, the number of adults in the family looking for work, the number of adults working full time, the number of children (under age 18) in the family, and the number of family members aged 65 and older.

Because most of the variables in the Family file are recodes of the person-level variables in the family core, the sum of the number of persons across all families in each family-level counter should be equal to the number of "yes" responses in its person-level source. Returning to our previous example, consider FSALCT: 15,646 families have one member receiving income from wages/salary, 11,939 families have two members (or 2(11,939) = 23,878 persons) with wage/salary income, 1,818 families have three members (or 3(1,818) = 5,454 persons), 436 families have four members (or 1,744 persons), 79 families have five members (or 395 persons), and 28 families have six members (168 persons) with wage/salary income in 2000. Thus the sum of persons across the 28,465 families answering "yes" to FSALYN, the associated yes-no indicator, is 47,285 (15,646 + 23,878 + 5,454 + 1,744 + 395 + 168), which is equal to the 47,285 "yes" responses to the person-level source variable, PSAL. Users are advised to check the Dataset Documentation for each Family file recode in order to determine its person-level source variable.

The 2001 NHIS Family file contains three variables describing family type and structure in both general and detailed terms. FMTYPE consists of just four categories, and represents an initial classification of families according to the numbers of adults and children that are present. Two additional variables, FMSTRCT2 and FMSTR2, categorize families according to familial relationships. FMSTRCT2 is consistent with the variable of the same name that was included on the 1998 NHIS Family file, while FMSTR2 is consistent with the variable of the same name that was included on the 1999 and 2000 NHIS Family files. Users will note that these variables differ in the way they treat families with cohabiting couples. In FMSTRCT2, married and unmarried parents living with their biological child (or children) are included in one category, while parent/cohabiting partner/child families (where the partner is unrelated to the child) are in a separate category. In contrast, FMSTR2 includes all cohabiting couple families in the *same* category (FMSTR2 = 42), regardless of the adults' relationships to the child(ren) in the family. In the case of both FMSTRCT2 and FMSTR2, families that could not be classified are coded "99". Finally, please note that emancipated minors are treated as adults with respect to FMTYPE, FMSTRCT2, and FMSTR2, despite the fact that they may be under 18 years of age.

Other Changes in the 2001 Family file

The 2001 Family file contains ten additional recodes *for 2001 only* derived from variables in the Family Access and Utilization section of the Person file. The source questions asked whether any member of the family 18 years of age or older needed a particular type of medical care, but did not receive it. Accordingly, the yes-no recodes included in the Family file also indicate whether any

family members needed but did not receive this type of medical care, followed by an associated counter that indicates the number of family members who needed but did not receive this care. Also, families lacking any members 18 years of age and older (i.e., families consisting of *only* emancipated minors) are excluded from the universe for these recodes.

## The Family File Weight

The 2001 NHIS Family file can be thought of as a household-level file of sorts for all families. The ideal situation for creating weights for such a family file would be to use independent estimates of the number of households or families from a reliable source, such as the U.S. Census Bureau, to perform post-stratification adjustments in a manner similar to what is done for the NHIS Person file weight. Unfortunately, no suitable independent estimates exist.

Due to the lack of appropriate independent estimates, a variation of the "principal person" method is used to create the 2001 NHIS Family file weight (WTFA_FAM). Our method is similar to that used in the Current Population Survey to create their household- and family-level weights. Briefly, a person-level ratio adjustment is used as a proxy for the NHIS family-level ratio adjustment. Use of the person weight with the *smallest* ratio adjustment within each family (that is, the smallest post-stratification factor between the interim and final person weights within each family) is believed to provide a more accurate estimate of the total number of U.S. families than either the use of other person weights in the family or the use of no ratio adjustments whatsoever.

Accordingly, the weight provided with the 2001 NHIS Family file, WTFA_FAM, corresponds to the 2001 NHIS person weight for one of the persons in the family. As a result, the Family weight contains factors for selection probabilities at the household level, household nonresponse adjustment, and several ratio adjustment factors that are applied to all person weights.

**2001 National Health Interview Survey**
**Person-Level File**

The Person-level variables are derived from the six substantive sections making up the Family Core of the 2001 NHIS. The information in the Family Core questionnaire is collected for all household members. Any adult household members who are present at the time of the interview may take part; information regarding adults not participating in the interview, as well as all household members under age 18, is provided by a knowledgeable adult member of the household. (If there is more than one family in the household, then these procedures are followed for each family in the household. See page 20 for more information.) The six sections comprising the Family Core are discussed in greater detail below.

The 2001 NHIS Person file contains three variables describing family type and structure in both general and detailed terms. FM_TYPE consists of just four categories, and represents an initial classification of families according to the numbers of adults and children that are in the family. Two additional variables, FM_STRCP and FM_STRP, categorize families according to familial relationships. FM_STRCP is consistent with the variable of the same name that was included on the 1998 NHIS Person file, while FM_STRP is consistent with the variable of the same name that was included on the 1999 and 2000 NHIS Person files. Users will note that these variables differ in the way they treat families with cohabiting couples. In FM_STRCP, married and unmarried parents living with their biological child (or children) are included in one category, while parent/cohabiting partner/child families (where the partner is unrelated to the child) are in a separate category. In contrast, FM_STRP includes all cohabiting couple families in the *same* category (FM_STRP = 42), regardless of the adults' relationships to the child(ren) in the family. In the case of both FM_STRCP and FM_STRP, families that could not be classified are coded "99". Finally, please note that emancipated minors are treated as adults with respect to FMTYPE, FM_STRCP, and FM_STRP, despite the fact that they may be under 18 years of age.

## I. Health Status and Limitation of Activity Section (FHS)

The Health Status and Limitation of Activity (FHS) section of the Family Core for the 2001 NHIS contains information addressing respondent-assessed disabilities, disability-associated conditions, and overall health status for all family members. Users should note that additional information on health status and disability is also included in other sections of the Sample Adult file, as well as the Sample Child file.

### Limitation of Activity at the Person Level

Information on activity limitations, including questions about work limitations, the need for personal assistance with personal care needs such as eating, bathing, dressing, getting around inside the home, and the need for personal assistance with handling routine needs such as everyday household chores, doing necessary business, shopping or running errands, is collected for each family

member (with some exclusions for children and youth). If any limitations are identified, the respondent is asked to specify the health condition(s) causing the limitation(s) and indicate how long he or she has had each condition. The questions in the 2001 NHIS Family Core regarding activity limitations due to physical, mental or emotional problems are comparable to the 1997-2000 NHIS (with minor exceptions discussed below) and are substantively comparable to previous NHIS surveys (despite some noteworthy changes in question wording).

Since cognitive impairment is increasingly recognized as a source of activity limitations among older adults, a new question was added (beginning in 1997) to determine if anyone in the family is limited because of difficulty remembering or periods of confusion. Other indicators in this section identify family members who have difficulty walking without any special equipment, plus the section contains an array of questions about limitations with specific personal care needs. In addition, the section also contains measures indicating children who receive special education or early intervention services. Limitations in play activities are also collected for young children.

The 2001 data on activity limitations were processed in the same manner as similar data from the 1999 and 2000 NHIS, although there were changes in how the conditions causing limitation were processed. The 1999-2001 data processing differed from the 1997 and 1998 protocols in minor ways. (See the NHIS Survey Description documents from these years for more information.)

Conditions

For each family member with a previously mentioned limitation, the respondent is asked about the condition or health problem associated with that limitation, as well as the length of time he/she has had the condition. Respondents are then handed one of two flash cards listing various condition categories. These categories are broad in scope, and vary according to age. Family members under age 18 were subject to the following fixed condition categories listed on the first flash card: “vision/problem seeing”, “hearing problem”, “speech problem”, “asthma/breathing problem”, “birth defect”, “injury”, “mental retardation”, “other developmental problem (e.g., cerebral palsy)”, “other mental, emotional, or behavioral problem”, “bone, joint, or muscle problem”, “epilepsy or seizures”, “learning disability”, attention deficit/hyperactivity disorder”, and two instances of “other impairment problem”. In 2001 the epilepsy category was expanded to include seizures, and two new categories, “learning disability” and “attention deficit/hyperactivity disorder”, were added. The fixed response categories in the instrument for adults were equally broad, and comprise the conditions listed on the second flash card: “vision/problem seeing”, “hearing problem”, “arthritis/rheumatism”, “back or neck problem”, “fractures, bone/joint injury”, “other injury”, “heart problem”, “stroke problem”, “hypertension/high blood pressure”, “diabetes”, “lung/breathing problem”, “cancer”, “birth defect”, “mental retardation”, “other developmental problem (e.g., cerebral palsy)”, “senility”, “depression/anxiety/emotional problem”, “weight problem”, and two instances of “other impairment problem”. Respondents could supply a verbatim response for the “other impairment problem”. In 1999 and continuing in 2000, verbatim responses could be entered in a longer, 50-character field. Starting in 2001, the FR was able to view (but not read aloud) an additional list of 17 conditions on

the laptop screen to help categorize the "other" responses. The "other" categories remained as a final option. Respondents could list any number of applicable conditions.

The verbatim responses recorded by FRs in one or both of the 50-character fields indicating "other impairment problem", as well as those in the 17 additional general categories seen by the FRs, were subsequently analyzed during data processing. While most respondents named conditions that did not fall into the 18 fixed response categories as originally specified in the instrument, some respondents named conditions that should have been included in one of the fixed categories. In the latter case, these "other" responses were assigned codes during data processing corresponding to the original response categories (the first 13 for children, and the first 18 for adults). For adults, an additional 16 *ad hoc* categories were created during coding to group responses that fell outside the original 18 condition categories included in the instrument: these *ad hoc* categories were assigned numbers 19-34. Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the new categories remained in the "other impairment problem" categories, and were renumbered "90" and, if necessary, "91" for both children and adults. In addition, responses in the 17 general categories seen only by the FR were also back-coded and "grouped" into 8 of the *ad hoc* categories. The resulting 36 categories for adults and 13 categories for children were generally informed by the International Classification of Diseases, Ninth Revision, Clinical Modification (see Table 4, below; note that the ICD-9-CM codes shown in this table are *not* included on the data file). These specific condition categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any activity (a mention/not-mention format). Because the 16 adult *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions causing limitations will be underestimated. Therefore, these variables should be analyzed with care. Moreover, none of the FHS condition variables (the 11 child variables, LAHCC1 through LAHCC11, and the 34 adult variables, LAHCA1 through LAHCA34) should be used to estimate prevalence for the conditions they represent, because only those persons with a previously reported limitation were eligible for the condition questions that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult and Child Cores.

## Recodes

The recode LA1AR is a summary measure that indicates household members reporting *any* limitation regarding one or more of the activities discussed during the course of the FHS section of the interview. In other words, individuals who answered "yes" to PLAPLYLM, PSPEDEIS, PLAADL, PLAIADL, PLAWKNOW, PLAWKLIM, PLAWALK, PLAREMEM, *or* PLIMANY are coded "1" for LA1AR. LACHRONR is based on LA1AR but adds the additional criterion of whether at least one of the reported causal conditions is a chronic condition. This recode corresponds most closely with the pre-1997 NHIS recode for Activity Limitation, although it has fewer response categories and does not allow for levels of limitation.

In response to analysts' requests that the LA1AR recode distinguish persons who are not limited from those with unknown disability status, this variable includes three response levels: "1" for limited, "2" for not limited and "3" for unknown if limited. (For comparability with previous years level 3 may be collapsed into level 2.) Users can also utilize the information contained in LA1AR to control for "unknown if limited" cases with respect to LACHRONR (that is, when LACHRONR = 0).

Also, a series of age-group-specific recodes (e.g., under 18 versus 18 and over) regarding conditions limiting activity and duration of limiting conditions have been created. Because the questions about limitation of activity in the redesigned NHIS are asked differently for different age groups, and because the questions are more general (in some cases) or more specific (in other cases) than in pre-1997 years, the degree to which a respondent is limited cannot be determined.

Chronic Conditions

Each condition reported as a cause of an individual's activity limitation has been classified as "chronic", "not chronic", or "unknown if chronic" based on the nature of the condition and/or the duration of the condition. Conditions that are generally not cured, once acquired (such as heart disease, diabetes, and birth defects in the original response categories, and amputee and "old age" in the *ad hoc* categories) are considered chronic, while conditions related to pregnancy are always considered not chronic. Additionally, other conditions must have been present for three months or longer to be considered chronic. An exception is made for children less than one year of age who have had a condition "since birth", and these conditions are considered chronic. Because the presence of a limitation determined whether persons were eligible for the condition questions and the chronicity recodes, we caution data users that these variables should *not* be used to produce estimates of the prevalence rates of chronic conditions.

**Table 4. FHS Categories with Approximate ICD-9-CM Ranges**

***A. Codes for Adults (ages 18+)***

| **NHIS Category** | **ICD-9-CM Codes** |
|---|---|
| **1 - Vision or seeing problem** | 360-379 |
| **2 - Hearing problem** | 387-389 |
| **3 - Arthritis / rheumatism** | 711-712, 714-716, 720.0, 721, 729.0 |
| **4 - Back or neck problem** | 722-724, 732.0, 737 |
| **5 - Fractures, bone or joint injury**<br>*Injury with specific mention of bone or joints* | 800-848, 850-999 |

| | |
|---|---|
| **6 - Other injury**<br>*Injury without specific mention of bone or joints* | 850-999 |
| **7 - Heart problem** | 410-417, 420-429, 745, 746, 785.0-785.3 |
| **8 - Stroke problem** | 430-438 |
| **9 - Hypertension or high blood pressure** | 401-405 |
| **10 - Diabetes** | 250 |
| **11 - Lung or breathing problem** | 460-461, 465-466, 470-471, 473, 477, 480-487, 490-496, 500-508, 510-519 |
| **12 - Cancer** | 140-208 |
| **13 - Birth defect**<br>*Excludes Down's syndrome and microcephalus* | 740-742.0, 742.2-744, 747-757.9, 758.1-759 |
| **14 - Mental retardation**<br>*Includes Down's syndrome and microcephalus* | 317-319, 742.1, 758.0 |
| **15 - Other developmental problem**<br>*Includes learning disabilities* | 315, 343, 783.4 |
| **16 - Senility (and other cognitive problems)** | 290 |
| **17 - Depression, anxiety or emotional problem**<br>*Includes neurotic disorders, personality disorders, and other nonpsychotic mental disorders, excluding alcohol and drug related problems and developmental problems* | 300-302, 306-314, 799.2 |
| **18 - Weight problem**<br>*Indicates a problem with being overweight or obese* | |
| **19 - Missing limbs (any part) / amputee**<br>*Indicates loss of a limb or digit* | |
| **20 - Other musculoskeletal system conditions**<br>*Diseases of the musculoskeletal system and connective tissue not coded to 3, 4, 5* | 710-739 |
| **21 - Other circulatory system conditions**<br>*Any diseases of the circulatory system not coded to 7, 8, 9* | 390-459 |
| **22 - Other endocrine system, etc. conditions**<br>*Any Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders not coded to 10 or 18* | 240-279 |
| **23 - Other Nervous system conditions**<br>*Diseases of the nervous system and sense organs not coded to 1, 2, 15, 16* | 320-389 |
| **24 - Digestive system conditions** | 520-579 |
| **25 - Genitourinary system conditions** | 580-629 |

| | |
|---|---|
| **26 - Skin & subcutaneous system conditions** | 680-709 |
| **27 - Blood & blood-forming organ conditions** | 280-289 |
| **28 - Tumors & cysts, benign & unspecified**<br>*Any mention of "tumor" without cancer, malignancy, etc.* | 210-239 |
| **29 - Alcohol & drug related problems**<br>*Any mention of "alcohol", "drugs" (or specific drug types), or substance abuse* | 291-292, 303-305 |
| **30 - Other mental conditions**<br>*Any mental disorders not coded to 14 or 15 or 17* | 293-299 |
| **31 - After effects of surgery or other medical treatment**<br>*Any mention of "surgery" or "operation" or other treatment as the causal condition; includes ongoing or recent treatment (1 year or less) or specific and sole mention of surgery/medical procedure as specific cause of limitation.* | |
| **32 - Old age**<br>*Any mention of age as the only specified cause* | |
| **33 - Fatigue/Tiredness**<br>*Any mention of tiredness, stiffness, or weakness without referring to any specific part of the body* | |
| **34 - Pregnancy related conditions**<br>*Any mention of "pregnancy" or "childbirth"* | |
| **90 - Others Not Elsewhere Classified**<br>*1st other-specify verbatim, not elsewhere classified* | |
| **91 - Others Not Elsewhere Classified**<br>*2nd other-specify verbatim, not elsewhere classified* | |

## *B. Codes for Children (ages < 18)*

| | |
|---|---|
| **1 - Vision or seeing problem** | 360-379 |
| **2 - Hearing problem** | 387-389 |
| **3 - Speech problem** | 307.0, 307.9, 315.3, 784.3, 784.5 |
| **4 - Asthma or breathing problem** | 460- 461, 465-466, 470-471, 473, 477, 480-487, 490-496, 500-508, 510-519 |
| **5 - Birth defect**<br>*Excludes Down's syndrome and microcephalus* | 740-742.0, 742.2-757.9, 758.1-759 |
| **6 - Injury** | 800-999 |
| **7 - Mental retardation**<br>*Includes Down's syndrome and microcephalus* | 317-319, 742.1, 758.0 |

| | |
|---|---|
| **8 - Other developmental problem**<br>*Includes learning disabilities* | 315, 343, 783.4 |
| **9 - Other mental, emotional, or behavioral problem**<br>*Includes ADD, ADHD, and hyperactivity* | 290-314, 799.2, V15.4 |
| **10 - Bone, joint or muscle problem** | 710-739 |
| **11 - Epilepsy and seizures** | 345, 779.0, 780.3 |
| **90 - Others Not Elsewhere Classified**<br>*1st other-specify verbatim that does not fit in any other category* | |
| **91 - Others Not Elsewhere Classified**<br>*2nd other-specify verbatim that does not fit in any other category* | |

## II. Injury/Poisoning Section (FIJ)

Injury and Poisoning data are no longer included on the NHIS Person file. All injury and poisoning data can be found on two stand-alone data files: the Injury/Poisoning Episode file and the Verbatim Injury/Poisoning Episode file. Please refer to page 38 for a description of these files.

## III. Health Care Access and Utilization Section (FAU)

The Health Care Access and Utilization (FAU) data from the Family Core of the 2001 NHIS contain information addressing access to health care and utilization of health care services. The FAU section has remained largely unchanged since 1997 and consists of three parts: Part A, Access to Care; Part B, Hospital Utilization; and Part C, Health Care Contacts. In 2001, five supplementary questions that address the 2010 Healthy People Objectives were added to the section.

Since 1997 questions that ask about delay of health care because of worry about the cost, overnight hospital stays, home care, calls to health professionals, and office visits have been included in the survey; there is also an expanded list of health care professionals, and respondents were instructed to consider "care from ALL types of medical doctors, such as dermatologists, psychiatrists, ophthalmologists, and general practitioners", as well as nurses, physical therapists, and chiropractors. Lastly, a question asking about 10 or more visits to doctors or other health care professionals in the last 12 months has been included.

2001 Healthy People Questions

In 2001, five supplementary questions were included that address the 2010 Healthy People Objectives and were limited to persons 18 years of age and over. These questions asked respondents what type of long term care they needed but did <u>not</u> get. These questions included:

home health care, adult day care, assisted living, hospice care or care for the terminally ill, and nursing home care.

Technical Notes

A few extreme values were found for hospitalizations (HOSPNO) and hospital nights (HPNITE). In addition, large numbers exist for home care visits (PHCHMN2W), doctor visits (PHCDVN2W), and calls to health professionals (PHCPHN2W). Analysts should be aware that although the variables HOSPNO and HPNITE have been edited for consistency, none of the above-mentioned variables have been revised.

Due to a question wording change that occurred in 2000, a minor modification has been made to the universe description for the variable PHCPH2W. Analysts are advised to read the notes in the Dataset Documentation for further information pertaining to any changes that may have occurred and to compare the 2001 Dataset Documentation to documentation from the 2000 (and earlier) NHIS for any other changes that may have occurred over time to the variables in this section.

**IV. Health Insurance Section (FHI)**

The Health Insurance section of the 2000 NHIS Family Core has a full range of data items addressing health insurance. The questions pertaining to health insurance programs covered by this section are similar to those asked on the 1993-96 NHIS Health Insurance Supplements and the 1997-2000 NHIS Family Core.

The health insurance section (FHI) covers several different topic areas:

Type of health care coverage (Medicare, Medicaid, Children's Health Insurance Program (CHIP), military/VA, CHAMPUS/TRICARE/CHAMP-VA, State-sponsored health plan, Indian Health Service, other government programs, private insurance and single service plans);

Managed care arrangement for those covered by Medicare and Medicaid, and need for referrals;

Medicare managed care model types (new for 2001);

Private insurance characteristics reported by the family respondent, including HMO, PPO, and POS status, source of coverage, existence of employer subsidies for premiums, amount paid by individual/family, managed care detail information, need for a referral;

Private insurance plan types, including HMO model types coded from private plan names;

Types of single service plans;

Periods of time without health insurance and reasons for no health insurance;

Out-of-pocket costs in the past year for medical expenses (excluding health insurance premiums).

Beginning in 2000, FHI data contain several modifications, as well as some new variables. The beginning of the FHI portion of the instrument was restructured to allow an active response to the question pertaining to health insurance coverage status for each individual family member. As a result, the HIKIND item now includes a category (HIKINDN) that allows the respondent to indicate whether each family member has health insurance coverage. Health insurance coverage status is subsequently verified for all family members (instead of just persons deemed not to have coverage) with HICHANGE (FHI.075). (HICHANGE is not available on the public use data file, but HIKINDN is included.) Moreover, a new question was added to the Medicare detail (MCCHOICE), which provides information on whether the Medicare recipient is enrolled in a Medicare + Choice Plan. Beginning in 2001 Medicare managed care plan names were coded to model types. These are included in the MCNAME field. Lastly, the variables HICHECK and PHICOV have been removed from the 2000 instrument.

Technical Notes

During the course of data editing, it was determined that some respondents indicated plans (in response to the questions HIPNAM, NEXTPNM, NEXTPNM2, and NEXTPNM3) that were not private health insurance plans, or were single service plans that were excluded from the private health insurance coverage category. These respondents were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, CHIP, IHS, MILITARY, OTHERPUB, OTHERGOV and SINGLE. Similarly, in looking at the verbatim responses to the question STNAME that asks respondents for the name of their CHIP, state sponsored or other government coverage, it was found that some respondents indicated plans and names of programs that were clearly private health insurance, Medicare, Medicaid, military coverage, Indian Health Service, single service plans or no coverage at all. Persons with these forms of coverage were reassigned to the appropriate enrollment recodes for MEDICARE, MEDICAID, PRIVATE, IHS, MILITARY and SINGLE. Respondents who answered “other state sponsored” or “other government coverage” who were subsequently determined through the STNAME field to be covered by the Children’s Health Insurance Program were assigned to the CHIP recode. Beginning in 2001, in looking at the verbatim responses to the question MCHMO_NA that asks respondents for the name of their Medicare managed care plan, it was found that some respondents indicated a plans or programs which were clearly private health insurance, Medicaid, military coverage, Indian Health Service, single service plans or no coverage at all. Persons with these forms of coverage were reassigned to the appropriate enrollment recodes for MEDICAID, PRIVATE, IHS, MILITARY and SINGLE. Additionally, beginning in 2001, in looking at the verbatim responses to the questions MACHMD_1 and MACHMD_2 that ask respondents for the name of their Medicaid managed care plan, it was found that some

respondents indicated plans or programs which were clearly private health insurance, Medicare, CHIP, military coverage, Indian Health Service, single service plans or no coverage at all. Persons with these forms of coverage were reassigned to the appropriate enrollment recodes for MEDICARE, CHIP, PRIVATE, IHS, MILITARY and SINGLE. Also, some respondents offering an "other" response to the survey item (HISTOP@SPC) that inquired about the reason(s) their coverage stopped subsequently indicated in their verbatim responses that they did in fact have health insurance. These persons were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, CHIP, PRIVATE, IHS, MILITARY, OTHERPUB and OTHERGOV. Analysts are therefore strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, CHIP, IHS, MILITARY, OTHERPUB, OTHERGOV, and SINGLE for types of health care coverage, because these take into account the above-mentioned back edits. In contrast, the data contained in HIKINDA-HIKINDN were not back-edited and reflect the respondents' original replies. In addition, a recode (NOTCOV) is included in the data file that reflects the definition of non coverage as used in *Health, United States* (in which persons with *only* Indian Health Service coverage are considered uninsured).

As a result of the complicated editing process that takes place in the FHI section, the variables HILAST and HINOTYR, which reflect periods of non coverage, cannot be used to estimate the rate of uninsurance. Users should derive such estimates from NOTCOV (if they do not count IHS as coverage) or, alternatively, the health insurance recodes (MEDICARE, MEDICAID, PRIVATE, CHIP, IHS, MILITARY, OTHERPUB, and OTHERGOV). Using the most conservative estimate of the uninsured (which would exclude persons with IHS coverage only), a total of 1537 persons did not receive the HILAST question during the course of the interview because they indicated that they had health care coverage. It was subsequently established during the course of editing that they lacked coverage (given the information that they provided about their insurance plan(s)). NHIS staff elected not to edit these people out of the universe for HINOTYR. In addition, a total of 1100 respondents were not asked either the HILAST or the HINOTYR questions.

## V. Socio-demographic Section (FSD)

The Socio-demographic (FSD) section of the Family Core in the 2001 NHIS collects information on place of birth, citizenship status, and educational attainment for all family members, regardless of age. In addition, family members 18 years of age or older are asked if they were working last week, and if not, their main reason for not working. For those working, additional questions inquired about the number of hours they worked during the previous week, how many months they worked in 2000, an estimate of their earnings from wages in 2000, and whether their employer provided health insurance.

The questions in the instrument pertaining to work status during the previous week (DOINGLW), reason for not working during the previous week (WHYNOWRK), ever attended Head Start (HEADSTEV), and whether the respondent worked for pay during the last year (WRKLYR) were modified in the 2001 NHIS. The 2001 DOINGLW question has a new response category ("working, but not for pay, at a job or business") that is specifically intended to

represent persons who are performing unpaid work in a family-owned farm or business. Note that FRs were instructed to *exclude* volunteer activities from this category. The 2001 WHYNOWRK has three new response categories ("On a planned vacation from work", "On family or maternity leave", and "Have job/contract; off-season") that are intended primarily for those respondents who said they were with a job or business but not at work during the previous week. The 2001 HEADSTEV and WRKLYR have expanded universes: the universe for HEADSTEV now includes *all* children under 18 years of age not currently attending Head Start, while the universe for WRKLYR includes all persons aged 18 and over. As a result of these changes in response categories or universes, these variables have different names in 2001 (i.e., DOINGLW1, WHYNOWK1, HEADSTV1, and WRKLYR1) to indicate to users that they are substantively different indicators.

The 2001 NHIS contains a "region of birth" recode called REGIONBR, which categorizes all respondents into one of 12 categories, depending on their country of origin. The CIA on-line World Factbook was used to place countries into the regional categories shown below (see **<http://www.cia.gov/cia/publications/factbook/index.html>**). Note that respondents born in Canada were included in the "Elsewhere" category in order to satisfy NCHS confidentiality requirements.

| **Category** | **Countries included** |
|---|---|
| United States | All persons born in one of the 50 states or the District of Columbia |
| Mexico, Central America, Caribbean Islands | All countries in Central America and the Caribbean Island area, including Puerto Rico |
| South America | All countries on the South American continent |
| Europe | Albania, Austria, Azores Islands, Belgium, Bosnia, Bulgaria, Corsica, Crete, Croatia, Czechoslovakia, Denmark, Finland, France, Germany, Great Britain, Greece, Herzegovina, Holland, Hungary, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Macedonia, Majorca, Malta, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Prussia, Romania, Scotland, Serbia, Sicily, Slovakia, Spain, Sweden, Switzerland, Yugoslavia |
| Russia (and former USSR areas) | Lithuania, Latvia, Russia, Ukraine, and all places formerly a part of the USSR |
| Africa | All countries on the African continent, plus the Canary Islands, Comoros, Madagascar, Madeira Islands |
| Middle East | Aden, Arab Palestine, Arabia, Armenia, Bahrain, Cyprus, Gaza Strip, Iran, Iraq, Israel, Jordan, Kuwait, Syria, Lebanon, "Middle East", Oman, Palestine, Persia, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, West Bank, Yemen |

| | |
|---|---|
| Indian Subcontinent | Afghanistan, Bangladesh, Bhutan, British Indian Ocean Territory, Ceylon, East Pakistan, India,, Maldives, Nepal, Pakistan, Sri Lanka, Tibet, West Pakistan |
| Asia | Asia, Asia Minor, China, Japan, Mongolia, North Korea, South Korea |
| SE Asia | Borneo, Brunei, Burma, Cambodia, Christmas Island, Hong Kong, Indonesia, Laos, Malaysia, Myanmar, North Vietnam, Philippines, Singapore, South Vietnam, Taiwan, Thailand |
| Elsewhere | Bermuda, Canada, Greenland, Oceania, as well as "At sea", "High seas", "International waters", "North America" |
| Unknown | Places that could not be classified in the above categories |

Users seeking more detailed information on respondents' place of birth may gain limited, supervised access to this information (in the case of persons born in the United States, state of birth; in the case of persons born outside of the U.S., country of birth) through the NCHS Research Data Center. For more information, please refer to the Research Data Center web page (**<http://www.cdc.gov/nchs/r&d/rdc.htm>**).

Analysts may also refer to the Adult Core socio-demographic section (ASD) for additional occupational and employment data regarding those individuals selected as sample adults.

Technical Notes

Data editing procedures have reconciled inconsistencies between DOINGLW1 and the relevant occupation and work-related variables in the Sample Adult Demographics section (ASD), but no such corrections are possible for non-sample adults. Moreover, data captured in DOINGLW1 have not been reconciled with other variables pertaining to the same subject matter within other sections of the data file. The "major activity" variable available in pre-1997 versions of the NHIS is now roughly approximated by a recode, MAJR_ACT, which sorts adults (persons aged 18 and older) into five categories: working at a job or business, keeping house, going to school, "something else" (e.g., retirement), and unknown. Users are advised that MAJR_ACT is derived in an entirely different fashion than the major activity variable available prior to 1997; comparable frequencies should not be expected.

Respondents who were not born in the United States were asked the year in which they came to the United States to stay. Respondents who could not recall or refused to answer were subsequently asked to estimate the number of years they had been in the United States. This information was combined to create a recode that indicates how long these respondents have been living in the United States (YRSINUS). Regarding USBORN_P, respondents born in Puerto Rico, Guam, and other outlying territories of the United States are included in response category "2"; that is, they were not born in one of the fifty United States or the District of Columbia. The 2001 data also contain a citizenship recode that distinguishes between U.S. citizens and non-citizens.

## VI. Income and Assets Section (FIN)

The Income and Assets (FIN) section of the Family Core contains information regarding a variety of income sources, as well as estimates of total combined family income and home tenure status. Respondents are asked whether anyone in the family received income from a variety of sources; if so, the respondent is then asked to name the member(s) receiving income from that source. The section also includes questions about the family's total income from all sources in 2000, and their home tenure status. The basic universe for most questions is "All families"; however, note that universes for several questions (most importantly, PSAL, PSEINC, and PWIC) are further limited with respect to age (of family members). All variables in the Income and Assets section were converted from the family level to the person level during the editing process (i.e., the information in the respondent's record was transferred to the records for each family member).

### Sources of Income

The first two questions in the section ask about income from wages and salary, and from self-employment (business or farm) for family members 18 years of age and older. Subsequent questions are not limited to adult family members. Respondents were asked about income from Social Security or Railroad Retirement (including that which was received as a disability benefit); other pensions; Supplemental Security Income (SSI); Welfare/Temporary Assistance to Needy Families (TANF); other kinds of government assistance (e.g., job training or placement, transportation assistance, or child care); interest from savings or other bank accounts; dividends from stocks, mutual funds, and/or net rental income from property, royalties, estates or trusts; child support payments; and other income sources (the question asked of respondents specifically mentioned alimony, contributions from family or friends, VA payments, Worker's Compensation, and Unemployment Compensation as possible sources of "other" income). Respondents are told at the start of the Income and Assets section that all questions are seeking information about possible income sources in the previous calendar year (i.e., 2000).

### Amounts and Home Ownership

In previous years, NHIS obtained information about the amount of income received from each financial source, but that was dropped in the redesigned NHIS in favor of a single overall estimate of combined family income. And, unlike previous NHIS surveys, the redesigned instrument contained three questions to identify the family's combined income from all sources during the previous calendar year, including a question (FIN.250) that allowed the respondent to supply a specific dollar amount (up to $999,995). Any family income responses greater than $999,995 were entered as $999,996. Respondents who did not know or refused to give a dollar amount to this question were then asked if their total combined family income for the previous year was $20,000 or more, OR less than $20,000 (FIN.260). If the respondent answered this question, he/she was then given a flash card and asked to indicate which income group listed on the card best represented the family's combined income during the previous year (FIN.270). Information from these three

variables was combined into an income recode (INCGRP) that uses 13 categories to describe the family's income, as well as a second recode (AB_BL20K) that indicates whether the family's income was at or above $20,000 or below $20,000.

Additionally, a more detailed indicator of poverty status was created by utilizing published information from the U.S. Census Bureau regarding 2000 poverty thresholds (see *Poverty in the United States, 2000*; U.S. Census Bureau). A ratio of the 2000 income value reported by respondents to the poverty threshold for the same year was constructed, given information on the family's overall size as well as the number of children aged 17 and under present in the family. The resulting ratio was subsequently ordered into a poverty gradient consisting of 14 categories (RAT_CAT). Users should note that the universe for this variable is considered to be all families, because the initial income question was asked of all families. However, the income-to-poverty ratios and resulting RAT_CAT values could not be calculated in two cases: when families simply did not supply adequate income information (e.g., those who would only indicate that their income was above or below $20,000, as well as those who declined to give any income information), or those families where the number of children aged 17 or under *equaled* the overall number of family members (these observations are coded "99" and "96", respectively, on RAT_CAT). Respondents were also asked whether the family's house or apartment was owned or being bought, rented, or occupied by some other arrangement. If the family was renting the current residence, the question asked was whether the family was paying lower rent due to governmental rental assistance (FIN.282) .

Program Participation

Respondents were asked in the final part of the FIN section if any family members were authorized to receive food stamps in 2000, and if so, which members. In addition, respondents were asked whether any family member(s) had *ever* applied for Supplemental Security (SSI) or Social Security disability benefits (even if the claim(s) had been denied). It should be noted that most of the respondents who refused/didn't know if they were receiving SSI were not asked if they have ever applied for SSI (FIN.310). This was due to a CAPI initialization error, and these individuals are included in the "not ascertained" response category. Lastly, if one or more family members had received food stamps or Temporary Assistance to Needy Families (TANF), the respondent was asked, in two separate questions, for how many months during the last calendar year food stamps and/or TANF were provided.

There are also two new variables in 2001 relating to the Women, Infants, and Children (WIC) program. The first of these variables, ELIGPWIC, indicates if the person was in a family with at least one WIC age-eligible person (children 0-5 years of age and women 12-55 years of age). If there is at least one WIC age-eligible person in the family, the family respondent is asked if anyone in the family received WIC benefits in 2000 (PWIC). Finally, the second new variable for 2000, WIC_FLAG, indicates if persons who received WIC benefits were age-eligible for the WIC program. While there were 42 WIC recipients deemed age-ineligible for the WIC program, their PWIC responses were not changed (since there were a small number of families where the only

person receiving WIC benefits was age-ineligible for the WIC program). This flag allows analysts to determine how to analyze these individuals.

## 2001 National Health Interview Survey
## Injury and Poisoning Episode Files

The Family Core portion of the 2001 survey included questions about medically attended injuries and poisonings that occurred to any member of the family within a three-month reference period. All injury and poisoning information was provided by the family respondent. Two data files containing injury and poisoning information were created from these data: the Injury/Poisoning Episode file, and the Verbatim Injury/Poisoning Episode file.

Between 2000 and 2001, only one change was made to the Family Core Injury/Poisoning Section of the NHIS. The question that ask respondents "How many times in the past three months did you seek medical advice because you were injured or poisoned?" was changed back to "How many different times in the past three months were you injured or poisoned seriously enough to seek medical advice or treatment?". Users familiar with the NHIS injury/poisoning data are no doubt aware of the continual decline in the overall number of injuries and poisonings reported since the injury and poisoning section was added to the NHIS in 1997. Evidence from the data indicates that the number of injuries and poisonings is continuing to decline, despite a return to the preferred question wording, and that injuries and poisonings may be under-reported. Other factors may be involved, such as declining response rates, other changes in the section, and a reluctance on the part of respondents to answer detailed questions about injuries and poisonings.

Consequently, NHIS staff made the decision not to include the injury and poisoning data on the 2001 Person file. However, injury and poisoning data will continue to be available to users on the Injury/Poisoning Episode file and the Verbatim Injury/Poisoning Episode file. **Users are strongly cautioned that national prevalence estimates derived from the 2001 NHIS injury and poisoning episode data may be underestimated.** Despite the significant decline in the overall number of injury and poisoning episodes in 2001, the percentage distributions of episodes for many injury-related variables are comparable across 1997-2001, such as the external cause of the injury, the nature of the injury, the injured person's activity at the time of injury, where the injury happened (e.g., inside the home, outside the home, at work, etc.) and any limitation of activity resulting from the injury.

### Injury/Poisoning Episode File

The Injury/Poisoning Episode file is an episode-based file: each injury and poisoning episode occurred during the three months prior to the interview and resulted in one or more conditions. An injury episode refers to the traumatic event in which the person was injured one or more times from an external cause (e.g., a fall, a motor vehicle traffic accident). An injury condition is the acute condition or the physical harm caused by the traumatic event. Likewise, a poisoning episode refers to the traumatic event resulting from ingestion of or contact with harmful substances, as well as overdoses or wrong use of any drug or medication, while a poisoning condition is the acute condition or the physical harm caused by the traumatic event. A person may record up to a total of ten injury and/or poisoning episodes and will be represented in this file as many times as he/she had unique

injury and/or poisoning episodes. Each episode must have at least one injury condition or poisoning classified according to the nature-of-injury codes 800-999 in the Ninth Revision of the International Classification of Diseases (ICD-9-CM). Other health conditions that were reported as occurring with the injury or poisoning, even if they are not classified according to the nature-of-injury code numbers 800-999 (e.g. mononeuritis of unspecified site (355.9), other symptoms referable to back (724.8)), are also included in the Injury/Poisoning Episode file. The resulting file contains information about the cause of the injury or poisoning episode, what the person was doing at the time of the injury or poisoning episode, the date and place of occurrence, the elapsed time between the date of the injury or poisoning episode and the date of the interview, whether the person was hospitalized, whether the person missed any days from work or school due to the injury or poisoning, whether the injury or poisoning episode caused any limitation of activity, ICD-9-CM diagnostic codes, and ICD-9-CM external cause codes. Information about how the injury or poisoning happened, the body part injured or poisoned, the type of injury or poisoning, along with responses to questions about specific types of injury or poisoning episodes, place of occurrence, and activity were used to assign ICD-9-CM diagnostic and external cause codes for all injury and poisoning episodes.

During the 2001 data editing process, some injury and poisoning episodes were removed. These included episodes with no information regarding cause, date and place of occurrence, etc., episodes that did not occur within the reference period, and duplicate episodes. In addition, injury episodes were removed if they consisted solely of health conditions that could not be classified according to nature-of-injury codes 800-999 of the ICD-9-CM.

As in previous years, respondents reported episodes that they considered poisonings (e.g., food poisoning and allergic reactions) that are not considered poisonings based on the Ninth Revision of the International Classification of Diseases (ICD-9-CM). These types of episodes are covered by question FIJ.195 (POITP) under categories “06 ” (food poisoning) and “07 ” (allergic reaction) and are not removed from the file. However, since these episodes are not poisonings according to nature-of-injury codes 800-999 of the ICD-9-CM, it is suggested that these episodes be removed prior to calculating national estimates of poisoning.

This file only contains information about injury, poisoning, food poisoning, and allergic reaction episodes. Other person-level information can be obtained by linking the Injury/Poisoning Episode file to other 2001 NHIS data files using the household serial number (HHX), family serial number (FMX), and person number (PX).

Recall Period and Weights

Questions in the Injury/Poisoning section of the 2001 NHIS have a recall period of the “last 3 months”. To calculate estimates of the number of injury or poisoning episodes occurring annually, each three-month estimate should be multiplied by 4. It is not possible to estimate the number of different *people* injured or poisoned annually using the NHIS. Although the number of different persons who were injured or poisoned during the three-month recall period is known, individuals may or may not have additional episodes over the rest of the year. On the other hand, it is generally appropriate to estimate the number of injuries or poisonings over the twelve-month period (by

multiplying the 3-month estimate by 4) because that figure is the same whether or not individuals had multiple injuries or poisonings. Analysts are cautioned to check the Dataset Documentation and the specific item in the questionnaire in order to insure that annual estimates for these kinds of injury or poison episodes have intrinsic meaning.

## Variance Estimation

This file does not contain the design variables used in variance estimation. To obtain the design information, the Injury/Poisoning Episode file must be linked to the Person file.

## Technical Notes

Two variables on the Injury/Poisoning Episode file, CAUSNEW and ECAUSNEW, describe the cause of the episode. CAUSNEW is the actual item found in the questionnaire. For each unique episode, the interviewer selected the category of CAUSNEW that he/she felt best described the episode based on responses that were given to questions FIJ.050 (IJTYPE) and FIJ.070 (IJHOW). ECAUSNEW is a recoded variable that describes the cause of the episode using categories based on ICD-9-CM external cause codes. The category into which an episode was placed was based entirely on the first ICD-9-CM external cause code listed for that episode. Appendix I in the Injury/Poisoning Episode Dataset Documentation contains a list of the ICD-9-CM external cause codes found in each category.

Analysts are cautioned regarding their use of the variable RPCKDM, which indicates the elapsed time between the date of the injury or poisoning episode and the date of the interview. The date of the interview used in the calculation of this variable is actually the last date when the interview was opened for examination or input of data. This means that if the interviewer was unable to complete the interview in one visit and had to return at a later date, the injury and poisoning questions may have been completed earlier than indicated by the date of the interview recorded by the CAPI instrument. If this occurred, the actual time between the date of the injury or poisoning episode and the date of the interview would be less than indicated by variable RPCKDM. It is for this reason that value “92 ” (92-105 days) was created and that four months before the date of the interview was included in value “96 ” (3 or 4 months before interview).

Additionally, in the case of some injury and poisoning episodes, the respondent was only able to provide the month and year of occurrence, but not the day. In these cases, it was only possible to determine whether the injury or poisoning episode occurred during the same month as the interview or within a certain number of months. For injury and poison episodes that occurred during the same month as the interview, this means that the time between the date of the injury or poisoning episode and the date of the interview could be anywhere from zero to 30 days. For injury and poisoning episodes that occurred during the month before the interview, this means that the amount of time between the date of the injury or poisoning episode and the date of the interview could be anywhere from zero to 60 days. The time between the date of the injury or poisoning episode and the date of the interview in months is only provided when the day of the injury or poisoning episode was not given.

## Verbatim Injury/Poisoning Episode File

The Verbatim Injury/Poisoning Episode file contains edited narrative text descriptions of the injury or poisoning provided by the respondent and includes the body part injured or poisoned, the kind of injury or poisoning, and a description of how the injury or poisoning happened. (The pre-edited responses are "verbatim " only insofar as the interviewer could type the information and condense it to fit the field size.) Editing was done only to protect the injured or poisoned person's confidentiality. Text descriptions used to replace original text that could result in a breach of confidentiality are surrounded by arrows (<>). Grammatical and/or spelling errors were not corrected. The codes of "R", which represents "Refused"; "D" or "DK", which represent "Don't know"; and "N", which represents "No more information" have also been left in the file. The following types of changes were made to the file in order to protect the injured or poisoned person's confidentiality:

Person names (first, middle, and/or surnames or initials) were replaced with <He> or <She>;

Names of commercial operations were replaced with a general category (i.e., the name of a restaurant that serves fast food would be replaced with <fast food restaurant>);

All place names including cities, counties, states, and street addresses were removed;

The detailed description of an occupation was replaced with a more general category using the Standard Industrial Classification as a guide;

Brand names were replaced with a generic term for the product (i.e., the brand name of a car would be replaced with <motor vehicle>);

Text that indicated unusual personal behavior or events was modified to make it less remarkable;

Any group or organization that was known to have a register of its members was replaced with a generic term.

## Technical Notes

Due to the way in which IJBODY1, IJBODY2, IJBODY3, IJBODY4 (all referring to the body part injured or poisoned), IJKIND1, IJKIND2, IJKIND3, IJKIND4 (all referring to the kind of injury or poisoning), and IJHOW1, IJHOW2, IJHOW3, IJHOW4 (all referring to how the injury or poisoning happened) were recorded, the information contained in these variables may not correspond exactly to the ICD-9-CM codes (ICD9_1, ICD9_2, ICD9_3, ICD9_4) and E codes (ECODE_1, ECODE_2, ECODE_3) associated with a given injury or poisoning episode. For example, the body part or body parts listed in IJBODY1 (question FIJ.050) may not be the one or ones listed in ICD9_1. When comparing verbatim injury/poisoning episode information and ICD-9-CM codes and E codes, it is better to look at all the information together rather than to try to match

first body part field with first ICD-9-CM code field, second body part field with second ICD-9-CM code field, etc.

## 2001 National Health Interview Survey
## Sample Adult File

The Sample Adult section of the 2001 NHIS covers many of the subject areas included in the Family Core. However, the questions in the Sample Adult section are more specific, and are intended to gather more detailed information. In addition, sample adults generally respond for themselves, although in a small number of cases, proxy responses are allowed if the selected adult had a physical or mental condition prohibiting him/her from responding. The variable PROXYSA indicates those cases where information was obtained from a proxy respondent. The six sections comprising the Sample Adult section are discussed below.

### I. Adult Conditions Section (ACN)

The ACN section of the 2001 NHIS obtains information from the sample adult as to whether he/she has cardiovascular disease, emphysema and asthma, ulcers, cancer, diabetes, other respiratory conditions, renal conditions, liver conditions, joint symptoms, pain, hearing loss, vision loss, or has lost any permanent (or natural) teeth. In most instances, sample adults were asked whether a doctor or other health professional had told them that they had the condition in question (joint symptoms, pain, hearing, vision, and tooth loss are the exceptions). The section also contains information about the sample adult's current mental health (whether he/she experienced feelings of sadness, nervousness, restlessness, hopelessness, worthlessness, or that everything was an effort in the past 30 days), and the extent to which these feelings interfered with their life or daily activities. Table 5 shows the specific health-related conditions covered in the NHIS, as well as the various reference periods covered by the questions. With the exception of head or chest colds, and stomach or intestinal illnesses, no question in the ACN section refers to a two-week reference period.

Regarding the ACN data on colds and intestinal illnesses, analysts should keep in mind that the questions are measuring fairly broad symptoms and illnesses. Furthermore, these may be a result of either acute or chronic conditions (e.g., irritable bowel syndrome or respiratory allergies). These data are best used to measure trends over time. Also, users are advised that the condition data in the Person and Sample Adult files have not been compared for consistency of reported conditions.

The conditions data in the 2001 NHIS are considerably different than the conditions data found in the pre-1997 NHIS. Users should consult the ACN section of the 1997 NHIS Survey Description document to obtain more information.

**Table 5. Sample Adult File: Conditions and Reference Periods**

| Reference Period in 2001 NHIS | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CRQ #** | **Condition** | **Ever** | **12 months** | **3 months** | **30 days** | **2 weeks** | **Now** |
| **ACN.010** | High blood pressure | X | | | | | |
| **ACN.031** | Coronary heart disease | X | | | | | |

| Reference Period in 2001 NHIS | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CRQ #** | **Condition** | **Ever** | **12 months** | **3 months** | **30 days** | **2 weeks** | **Now** |
| **ACN.031** | Angina | X | | | | | |
| **ACN.031** | Heart attack | X | | | | | |
| **ACN.031** | Other heart condition | X | | | | | |
| **ACN.031** | Stroke | X | | | | | |
| **ACN.031** | Emphysema | X | | | | | |
| **ACN.080**<br>**ACN.085**<br>**ACN.090** | Asthma ever<br>Still have asthma<br>Episode attack | X | <br><br>X | | | | <br><br>X |
| **ACN.110-ACN.120** | Ulcer | X | X | | | | |
| **ACN.130** | Cancer (Any + list max. 3 of 30 specific types) | X | | | | | |
| **ACN.160** | Diabetes | X | | | | | |
| **ACN.201** | Hay fever | | X | | | | |
| **ACN.201** | Sinusitis | | X | | | | |
| **ACN.201** | Chronic bronchitis | | X | | | | |
| **ACN.201** | Weak or failing kidneys | | X | | | | |
| **ACN.201** | Liver condition | | X | | | | |
| **ACN.250** | Joint pain | | X | | | | |
| **ACN.300** | Neck pain | | | X | | | |
| **ACN.310-ACN.320** | Low back pain, with or without leg pain | | | X | | | |
| **ACN.331** | Facial pain | | | X | | | |
| **ACN.331** | Severe headaches | | | X | | | |
| **ACN.350** | Head or chest cold | | | | | X | |
| **ACN.360** | Stomach or intestinal illness | | | | | X | |
| **ACN.370** | Pregnancy | | | | | | X |
| **ACN.410-ACN.420** | Hearing impairment | X | | | | | X |
| **ACN.430-ACN.440** | Vision impairment | | | | | | X |

| Reference Period in 2001 NHIS | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CRQ #** | **Condition** | **Ever** | **12 months** | **3 months** | **30 days** | **2 weeks** | **Now** |
| **ACN.451** | Lost all upper/lower teeth | | | | | | X |
| | Emotional Health: | | | | | | |
| **ACN.471** | Sad | | | | X | | |
| **ACN.471** | Nervous | | | | X | | |
| **ACN.471** | Restless | | | | X | | |
| **ACN.471** | Hopeless | | | | X | | |
| **ACN.471** | Everything an effort | | | | X | | |
| **ACN.471** | Worthless | | | | X | | |

The cancer questions were asked in a format that allowed a respondent who reported having had cancer to specify up to three kinds of cancer or to indicate that he/she had had more than three kinds. This is referred to as a mention/not-mention format. The responses were recorded with the codes indicated in the questionnaire and were then transformed into mention /not-mention variables during editing. These variables assign to every sample adult who reported having had cancer either a "mentioned", if he/she specified that particular type of cancer, a "not mentioned", if he/she did not specify that type of cancer, or a "refused", "don't know", or "not ascertained", if there was no information for any of the cancers. Thus, a sample adult may have a code in each of the cancer variables, but can have only up to three "mentions", with a fourth mention possible for the variable "More than 3 kinds".

In addition to the variables comprising the Core portion of the ACN section, 25 supplementary variables are included in the 2001 ACN section. Eleven variables represent "knowledge" questions that asked whether the sample adult knew the symptoms of heart attack or stroke, and what is the best thing that should be done for a heart attack victim. Additionally, two variables provide information regarding any training that the sample adult may have received regarding cardio-pulmonary resuscitation (CPR), and when that training took place. Lastly, the ACN section contains a number of new variables measuring the sample adult's satisfaction with life, his or her support network, and the types of social activities that the sample adult has engaged in during the past 2 weeks.

## II. Adult Health Status and Limitation of Activity Section (AHS)

The Adult Health Status and Limitation of Activity component of the Sample Adult file contains information from respondents on illness behavior, health status, special equipment, limitations in functional activities, and the conditions underlying such limitations. While the AHS section may seem similar to the FHS section in the Person file, the questions in these sections have a somewhat different focus. For example, both sections asked about the ability to walk without special

equipment. However, the walking limitation question in the FHS section (FHS.210) only captured whether a person has difficulty walking without using special equipment. In contrast, the Sample Adult question on walking (AHS.091) asked about the degree of difficulty the respondent has walking a specified distance (a quarter mile, or about three city blocks) by him/herself and without using any special equipment.

The 2001 NHIS survey includes some minor modifications in the way information on functional limitations and the conditions that cause them is collected and processed. These revisions are highlighted below.

Health Indicators: Illness Behavior and Health Status

The first questions in this section determined the number of days off from work and days spent in bed due to illness or injury during the 12 months prior to the interview. In addition, respondents were asked to compare their health now (is it better, worse, or the same) to their health 12 months ago.

Limitation of Functional Activities

The functional limitation questions in the AHS section asked the respondent to indicate the *degree of difficulty* he/she would have in performing specific physical tasks (e.g., walking a quarter of a mile, walking up ten steps, standing for two hours, carrying a ten pound object, etc.), and engaging in social activities and recreation (e.g., going shopping, attending club meetings, visiting friends, sewing, reading, etc.) without the assistance of another person or using special equipment. This is in sharp contrast to the questions in the FHS section, which allow only "yes" or "no" responses to questions inquiring whether household members needed help from another person with personal care needs (e.g., bathing, dressing, eating, etc.) or in handling routine tasks (doing everyday chores or shopping).

Beginning in 2000, for *all* activities (not just shopping, participating in social activities, and relaxing at home as in previous years), respondents had the opportunity to respond in the interview that they "do not do this activity". This response was added to certain functional activities (related to walking, climbing, standing, sitting, stooping, reaching, grasping, carrying, and pushing) in the 2000 NHIS. In prior years, respondents were not permitted to use this response during the course of the interview, but might have been reassigned to "do not do this activity" in the course of data editing based on information obtained by the FR. As in FHS, if the sample adult reported difficulty with any of these 12 activities, he/she was then asked what condition(s) cause the difficulty, as well as how long he/she has had the condition. The format of this condition data is similar to what is found in the FHS section.

Conditions

Each sample adult indicating any functional limitation (regardless of the degree of the limitation) is asked about the condition(s) or health problem(s) associated with that limitation, as well

as the amount of time he/she has had the condition. Sample adults were given the following fixed response categories: "vision/problem seeing", "hearing problem", "arthritis/rheumatism", "back or neck problem", "fractures, bone/joint injury", "other injury", "heart problem", "stroke problem", "hypertension/high blood pressure", "diabetes", "lung/breathing problem", "cancer", "birth defect", "mental retardation", "other developmental problem (e.g., cerebral palsy)", "senility", "depression/anxiety/emotional problem", "weight problem", and two instances of "other impairment problem". Respondents could supply a verbatim response (entered in an expanded up-to-50-character field) for the "other impairment problem". Starting in 2001, the FR was able to view an additional list of 17 conditions to help categorize the "other" responses. The "other" categories remained as a final option.

The AHS condition data were edited very much like the condition data in FHS. The verbatim responses recorded by FRs in one or both of the 50-character fields indicating "other impairment problem", as well as those in the 17 additional general categories seen by the FRs, were subsequently analyzed during data processing. While most respondents named "other" conditions that did not fall into the 18 fixed response categories as originally specified in the instrument, some respondents named conditions that should have been included in one of the fixed categories. In the latter case, these "other" responses were assigned codes during data processing corresponding to the one of the original 18 response categories. An additional 16 *ad hoc* categories were created; these were assigned numbers 19-34. Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the new categories remained in the "other impairment" categories, and were renumbered "90" and, if necessary, "91". In addition, responses in the 17 general categories seen only by the FR were also back-coded and "grouped" into 8 of the *ad hoc* categories. The resulting 36 categories were generally based on the International Classification of Diseases, Ninth Revision, Clinical Modification (see Table 4 in the FHS section). These specific condition categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any functional activity (a mention/not-mention format). Because the 16 *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions may be underestimated. Therefore, these variables should be analyzed with care. Moreover, none of the AHS condition variables (AFLHCA1 through AFLHCA34) should be used to estimate the prevalence rates for the conditions they represent, because only those sample adults with a previously reported functional limitation were eligible for the condition question that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult Conditions (ACN) section (above).

Recodes

The recode FLA1AR is a summary measure that indicates sample adults who reported *any* difficulty with *any* one or more of the functional activities discussed during the course of the AHS section of the interview. In other words, individuals who indicated *any* degree of difficulty to FLWALK, FLCLIMB, FLSTAND, FLSIT, FLSTOOP, FLREACH, FLGRASP, FLCARRY, FLPUSH, FLSHOP, FLSOCL, *or* FLRELAX are coded "1" for FLA1AR. This variable includes three response levels: "1" for limited, "2" for not limited and "3" for unknown if limited.

ALCHRONR is based on FLA1AR but adds the additional criterion of whether at least one of the reported causal conditions is a chronic condition. The section also includes time recodes (ALANTR1-34, ALANTR90 and ALANTR91) and chronic recodes (ALCNDR1-ALCNDR34, ALCNDR90 and ALCNDR91) for each of the 36 categories. These recodes for conditions that cause functional limitation(s) are similar to those used in the FHS section and described above.

## III. Adult Health Behaviors Section (AHB)

The AHB section of the NHIS Sample Adult questionnaire contains questions related to cigarette smoking, leisure-time physical activity, alcohol use, height, and weight. This series of health behavior questions has been in the NHIS Sample Adult core questionnaire since 1997. The section also includes two supplementary items related to leisure-time stretching that are not part of the ongoing Sample Adult Core. For details concerning the history of the Adult Health Behavior section, refer to the Survey Description documents for 1997-2000.

### Smoking

Current smokers are defined as persons who have ever smoked 100 cigarettes and currently smoke every day or some days. Smoking status recodes (SMKSTAT1 SMKSTAT2 and SMKSTAT3) are provided for easier data analysis

### Leisure time physical activities

The section on leisure-time physical activity is introduced with the following statement: "The next questions are about physical activities (exercise, sports, physically active hobbies...) that you may do in your LEISURE time". In this section, respondents are asked to summarize their *usual* leisure-time physical activity – both in terms of frequency and duration. This requires some mental calculations by the respondent. Responses can be offered in terms of any time unit the respondent volunteers (times per day, per week, per month, or per year). A recode, converting all responses into frequency in times per week, is provided for each type of activity. The set of leisure-time physical activity questions that are included every year in the sample adult core module are: frequency and duration of vigorous activities, frequency and duration of light-moderate activities, and frequency of strengthening activities. In 2001, for one year only, two supplementary questions concerning frequency and duration of stretching activities were included for the purpose of tracking Healthy People 2010 Objectives. The other leisure-time physical activity questions are also used for tracking Healthy People 2010 Objectives, although they are included every year.

### Alcohol use

Lifetime drinking status was assessed for all sample adults. Questions related to current drinking were asked of all respondents who had had at least 12 drinks in their lifetimes. Respondents were permitted to answer in terms of days per week, per month, or per year. Standardized variables

convert the various time unit responses to a single standardized unit (i.e. days per week, days per month, days per year).

The question asking how often the respondent had five or more drinks in one day during the past year was asked of all adults who had anything to drink in the past year. The responses were not edited for consistency with usual quantity or frequency because there was no basis for deciding which one might be the most accurate. Notice that the quantity questions in the alcohol section are phrased in terms of drinks *per day*, not drinks at a sitting.

The variable ALCSTAT1 was edited slightly differently in 2001 compared to earlier years. In data years 1997-2000, respondents with unknown frequency and unknown amount of drinking in the past year were classified as current drinkers based on the logic that they had not explicitly said they did not drink (which was a response option). This logic followed the skip pattern in the questionnaire. Beginning in 2001, the editing logic was changed to *exclude* from the current drinker category those respondents (numbering 165) who provided neither frequency nor amount of drinking in the past year. These cases, which represent one-half of one percent of adult respondents, were assigned to unknown drinking status (ALCSTAT1 = 9) in 2001. Persons who reported either frequency or quantity or both continued to be classified as current drinkers.

A new variable, ALC7STAT, was added to the AHB section of the data file in 2001. This variable reflects the classification of lifetime and current drinking status shown annually in *Health, United States*. Persons who reported frequency or quantity, but not both, were classified as "current drinker, level unknown" in this variable.

Body weight and height

Sample adults were asked to estimate their current height and weight. In cases where very large or very small values were reported for either height or weight, the data for both variables were changed to "96" or "996" ("Not available") on public use data files. This was done in order to protect the confidentiality of NHIS respondents who might be identifiable by their unusual physical characteristics. No physical measurements were taken. Estimates based on physical measurements, such as those available from NCHS's National Health and Nutrition Examination Survey, are likely to differ from those available from the NHIS.

BMI is a recode of the height and weight data, calculated using the formula:
BMI = kilograms /meters$^2$. Although respondents had the option of answering in metric units, fewer than 1% did so. For purposes of calculating BMI, responses provided in U.S. Customary units were converted to metric units using the following factors: 1 kilogram = 2.205 pounds; 1 meter = 39.37 inches. BMI was calculated for all persons who provided height and weight, including those for whom specific height and weight values were changed to "96" (not available) on the public use file to protect confidentiality. The values for the BMI include two implied decimals. Consistent with criteria established by the World Health Organization, the following classification of body weight status is

suggested for both men and women: underweight (BMI < 18.5); healthy weight 18.5< BMI <25; overweight (BMI > = 25); obese (BMI > = 30).

Desirable Body Weight (DESIREWT) is another measure of relative body weight that was constructed using self-reported height and weight. DESIREWT is based on the Metropolitan Life Insurance Company (MLIC) standards of desirable body weight (1983). The MLIC standards are based on the mortality experience of life insurance policy holders. The Desirable Body Weight variable is restricted to persons whose height and weight are within the range provided by the MLIC chart of desirable weight for height. Because of this, the number of unknowns is substantially larger in the DESIREWT than in the BMI.

## IV. Adult Health Care Access and Utilization Section (AAU)

The core Adult Health Care Access and Utilization (AAU) section of the 2001 NHIS has remained largely unchanged since 1997 and consists of four parts: Access to Care, Dental Care, Health Care Provider Contacts, and Immunizations. In 2001, eleven supplementary questions that address the 2010 Healthy People Objectives were added to the section.

The "Access to Care" section includes questions that ask respondents whether they had a usual place for sick care, a usual place for routine/preventive care, whether they experienced any changes in their place of care, any delays in getting medical care, and any instances when they were unable to afford medical care. The multi-part question about the reason for delaying care focused on such access issues as transportation difficulties, getting an appointment, and waiting time prior to actually seeing the doctor.

The "Dental Care" section includes only one question: length of time since last dental visit.

The "Health Care Provider Contacts" section contains questions that ask about doctor contacts during the past 12 months. Doctor visit probe questions allow for visits not only from medical doctors but from a variety of other health care professionals, including chiropractors. Questions about home care are included as well as a question asking about the number of visits to a hospital emergency room in the past 12 months, and a question that asks how long it has been since the respondent saw or talked to a doctor.

The "Immunizations" section includes several questions related to adult immunizations: flu shots, pneumonia vaccinations, and the hepatitis B vaccine. Four additional questions inquire whether the respondent had ever had chickenpox or had chickenpox in the past 12 months, had ever had hepatitis, or had ever lived with someone with hepatitis.

### 2001 Healthy People Questions

In 2001, eleven supplementary questions were added that address the 2010 Healthy People Objectives. These questions included type of particular health care provider seen most often, and whether advice was received about diet and nutrition; physical activity and exercise; alcohol

consumption; cigarette smoking; or menopause. Also included was a question asking about delaying, having trouble, or inability to get care from a hospital emergency room and the reason(s).

Technical notes

Three of the 2001 Healthy People questions (AQHP2, AQHPKND2, and AQHPVI2) were similar to three supplementary questions included in the 1999 AAU section (AQHP, AQHPKIND, AQHPVI). As a result of some minor modifications, however, the data may not be comparable. The AAU alcohol and smoking questions were based on responses to questions from the Adult Health Behaviors (AHB) section. For easier interpretation, the universe for the alcohol question has been modified to exclude records that included "unknown" in any one of the AHB variables included in the universe. The universe has also been modified to be consistent with the 2010 Healthy People Objectives. For easier interpretation, persons who had quit smoking in the past year but did not know when were excluded from the universe for the AAU smoking questions.

The question wording for ADNLONG2 was modified in the 2001 instrument. In 2000, respondents were asked how long since they had last seen or talked to a dentist (ADNLONGR). In 2001, respondents were asked how long since they had last seen a dentist. Because this wording change could conceivably reduce frequency counts, the name of the output variable was changed from ADNLONGR to ADNLONG2. Additionally, users will note slight changes in question wording for the variable SHTPNUYR; because these changes were comparatively minor, the output variable name remained the same across the 2000 and 2001 data files. Analysts are advised to compare the notes in the NHIS Dataset Documentation across survey years for information regarding other changes that may have occurred over time to variables in the AAU section.

## V. Adult Demographics Section (ASD)

The Adult Socio-Demographics (ASD) section was revised in 2001, and now contains information regarding the occupation, industry, workplace, and employment conditions of currently employed sample adults as well as those who have ever worked (e.g., retired persons).

Sample adults aged 18 years and older who were "working at a job or business", "with a job or business but not at work", or "working, but not for pay, at a job or business" during the week prior to their interview (DOINGLW1 = 1, 2, 4) were asked a series of questions about their job and work status during the week prior to the interview. Those sample adults who said that they were "looking for work" or "not working and not looking for work" during the week prior to the interview (DOINGLW1 = 3, 5) were asked if they had "ever held a job or worked at a business" (EVERWRK). Sample adults who responded affirmatively were then asked the occupation, industry and work status questions in the ASD section. Note that sample adults who had ever worked and were either retired *or* 65 years of age or older were asked about the job they had held the longest, whereas sample adults who had ever worked, were younger than 65 years of age *and* not retired were asked about their most recently held job.

The 2001 ASD section contains two sets of variables. One set is entirely consistent with previous years' ASD variables: they have the same universes (i.e., sample adults who were working during the week prior to the interview), the same response categories, and consequently the same variable names as in the 1997-2000 NHIS. The second set of variables has the expanded universe – currently employed sample adults as well as those who have ever worked – and all end in the letter "A".

Industry and Occupation Coding

During the course of the interview, verbatim responses were obtained from each respondent regarding his/her industry and occupation. These were subsequently recoded into two two-digit industry recodes that are consistent with the 1995 revisions to the Standard Industrial Classification (SIC) system. A detailed recode (INDSTRY1) indicates 42 possible industries, while a more simple recode (INDSTRY2) distinguishes 14 industries. A similar pair of recodes was created from the occupation information, consistent with the 1995 revisions to the Standard Occupational Classification (SOC) system. OCCUP1 distinguishes 41 separate occupations, while OCCUP2 indicates 13 occupations. These coding categories are provided in the Industry and Occupation Appendices (following the Dataset Documentation for the Sample Adult file).

Other Employment Questions

Subsequent questions in the ASD section ask sample adults to describe their current/most recent/longest-held employment situation (whether they were an employee of a private company or business, the federal government, a state or local government, self-employed in their own business or professional practice, or working without pay in a family business or farm), the number of full and part time employees at their workplace, how long they had worked at their current/most recent/longest-held job, whether they were paid by the hour, and whether they received paid sick leave. Currently employed sample adults were also asked whether they were working more than one job.

The variables LOCPRTNO and BUSINC, which were included in the 1997-2000 ASD section, have been eliminated from the 2001 NHIS (these questions are no longer asked). In 2001, respondents who indicated that they were self-employed at their current/most recent/longest-held job were asked whether they had an incorporated business. These new outcome variables are called BUSINC1 and BUSINC1A.

## VI. Adult AIDS Section (ADS)

This section contains a series of questions related to testing for the AIDS virus. Respondents were asked whether they had ever donated blood and whether they had had a blood test for HIV, the virus that causes AIDS; their main reasons for getting or not getting tested; when they had their last test; and where (the location/facility) the testing was done. The section also contains questions on respondents' plans for being tested in the future and their reasons for those plans, as well as their perceived personal risk for getting AIDS. With some modifications and additions, these questions

are similar to those asked in the AIDS Knowledge and Attitudes Supplements that were included in the NHIS from 1987 to 1995.

Three output variables in the 2000 NHIS ADS section, WHYTST_C, REATST_R, and REASWHO, were renamed to WHYTST_R, REATST_C, and REASWHOR, respectively, in the 2001 ADS section because of changes in their response categories. Analysts are advised to compare the notes in the NHIS Dataset Documentation across survey years for information regarding other changes that may have occurred over time to variables in the ADS section.

# 2001 National Health Interview Survey
# Sample Child File

The Sample Child section of the 2001 NHIS covers additional subject areas not included in the Family Core. Moreover, the questions in the Sample Child section are more specific, and are intended to gather more detailed information, than those in the Family Core. Sample children do not speak for themselves; instead a knowledgeable adult (typically a parent or guardian) answers questions in the sample child's behalf. The sections comprising the Sample Child section are discussed below.

## I. Child Conditions, Limitation of Activity and Health Status Section (CHS)

The Child Health Status (CHS) section of the 2001 NHIS consists of two parts: "Conditions, Limitations of Activity, and Health Status" and "Child Behavior". The section on Conditions, Limitations of Activity, and Health Status includes questions on the following health conditions: mental retardation, developmental delays, Attention Deficit Hyperactivity Disorder (ADHD) or Attention Deficit Disorder (ADD), Down's syndrome, cerebral palsy, muscular dystrophy, cystic fibrosis, sickle cell anemia, autism, diabetes, arthritis, congenital and other heart disease, asthma, allergies, colitis, anemia, ear infections, seizures, headaches, stuttering, and stammering. A new question about whether the sample child still has asthma was added. This section also contains a question used to determine the number of school-loss days reported during the 12 months prior to the interview. In addition, respondents were asked about hearing and vision loss; if a health problem requires the sample child to use special equipment such as a brace, wheelchair, or hearing aid; whether the sample child's health is better, worse, or the same compared with 12 months ago; and whether the sample child currently has a problem that has required prescription medication for at least three months. Lastly, there are questions about the sample child's height and weight.

In 2001 the Strengths and Difficulties Questionnaire (SDQ), a supplement on the mental health of children ages 4-17 was added to the CAU section, and the Child Behavior Checklist for children ages 4-17 by Achenbach and Edelbrock was dropped from the CHS section. However, the questions derived from the Child Behavior Checklist for Ages 2-3 remain in the CHS section. These questions on child behavior were designed to serve as a global mental health indicator. The items in the checklist were chosen for their ability to discriminate between children who have not received mental health services in the preceding 12 months and those who have, by using demographically-matched normative and clinical samples for boys and girls. Each set of items can be viewed as comprising a scale with each item scored as either "0", "1", or "2". More information on the scale derived from the Child Behavior Checklist is included in Appendix IV of this document. More information about the SDQ can be found in the CAU section (following page) and in Appendix V.

Technical Notes

Several questions pertaining to child behavior are used to create mental health indicator recodes; only the recodes are included in the Public Use file. The background and usage of the mental health indicators can be found in Appendix IV.

Regarding the CHS data on colds and intestinal illnesses, analysts should keep in mind that the questions are measuring fairly broad symptoms and illnesses. Furthermore, these may be a result of either acute or chronic conditions (e.g., irritable bowel syndrome or respiratory allergies). These data are best used to measure trends over time.

## II. Child Health Care Access and Utilization Section (CAU)

The Child Health Care Access and Utilization (CAU) section of the 2001 NHIS consists of three parts: "Access to Care", "Dental Care", and "Health Care Provider Contacts". The questions pertaining to "Access to Care" include: having a usual place for sick care; having a usual place for routine/preventive care; change in place of care; reasons for a delay in getting medical care; and the inability to afford medical care. The section on "Dental Care" includes only one question: length of time since last dental visit.

Questions regarding "Health Care Provider Contacts" include visits to or from medical doctors and other health care professionals (such as chiropractors) in the past 12 months. As with the FAU section discussed previously, the category of "health care professional" has been expanded to include chiropractors, various types of therapists, psychiatrists, psychologists, and social workers; moreover, contacts or visits are not restricted to medical doctors or professionals working with/for a medical doctor. Note that questions about home care are asked independently of other types of health care visits. In addition, the reference period for all health care contacts is the past 12 months. Lastly, a separate question is asked about the number of visits to a hospital emergency room in the past 12 months.

In addition, the Sample Child Core contains a new variable, CHCSYREM, which obtains information about any contact the sample child has had with a general doctor because of an emotional or behavioral problem.

In the 2001 NHIS, as a part of a collaborative agreement with the National Institute of Mental Health (NIMH), the Strengths and Difficulties Questionnaire (SDQ) was added as a supplement on children's mental health. The SDQ is a brief behavioral screening questionnaire for children ages 4 to 17 with extended questions that provide information on the duration of a child's problem and the impact that problem has on the child and his/her family. It is copyrighted by Dr. Robert Goodman, London, England, and used in the NHIS with Dr. Goodman's permission. More detailed information on the SDQ is provided in Appendix V.

## III. Child Immunization File (CIM)

The Child Immunization file of the 2001 Sample Child Core involves questions on the vaccination status of children under 18 years of age and within two age groupings (under 7 years, and 7-17 years) for one randomly selected sample child per family in a household, along

with any non-sample children aged 12-35 months within families of the household. The inclusion of additional children in the younger age ranges increases the precision of estimates of vaccination coverage for young children. The age split at 7 years reflects a differential focus on vaccinations by age. Among younger children, the focus is on the standard shots for which NHIS has previously obtained information (in earlier surveys). Among older children, vaccines such as hepatitis, measles, and diphtheria-tetanus booster are emphasized.

Using the child's shot record, if available, the NHIS interviewer transcribes information on type of shot, number of shots, and shot dates for specific shot types according to the child's current age, or alternatively, a knowledgeable adult in the family uses the shot record to report the same information to the interviewer. In the absence of a shot record, information on shot type and number (but not date) is obtained from the adult respondent in the family. In addition, information is also obtained about shots not listed on the shot record, other immunizations, and booster shots. This information is appended to the Child Immunization file in the form of shot type and date matrices.

Several new questions were added in 2001; they include questions on receiving additional influenza, and hepatitis shots for children under 18 years of age (whether with or without shot records), and an "other" vaccine question was asked of the latter group.

Technical notes

Since 2000, the variable ICSTAT has been derived from a check item within the questionnaire instead of a question. The variables IMRESPNO, ICRELTIV, and ICAGEMR were first included on the 1999 CIM file and remain on the 2001 file. IMRESPNO indicates the person who answered the questions on behalf of the child, while ICRELTIV defines the (proxy) respondent's relationship to the child whose immunization status is assessed. Lastly, ICAGEMR indicates the child's age in months.

## Guidelines for Citation of Data Source

With the goal of mutual benefit, the National Center for Health Statistics (NCHS) requests that recipients of data files cooperate in certain actions related to their use. Any published material derived from the 2001 data should acknowledge NCHS as the original source. The suggested citation to appear at the bottom of all tables and graphs is as follows:

Data Source: National Center for Health Statistics (2002)

In a bibliography, the suggested citation should read:

National Center for Health Statistics (2002). Data File Documentation, National Health Interview Survey, 2001 (machine readable data file and documentation). National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.

The published material should also include a disclaimer that credits any analyses, interpretations, or conclusions reached to the author (recipient of the data file) and not to NCHS, which is responsible only for the initial data. Users who wish to publish a technical description of the data should make a reasonable effort to insure that the description is consistent with that published by NCHS.

## References

Achenbach and Edelbrock. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. University of Vermont, Department of Psychiatry.

Dept. of Commerce, U.S. Census Bureau. (2001). *Poverty in the United States, 2000*. Washington, D.C.: U.S. Dept. of Commerce, Census Bureau.

Kessler, R.C., Andrews, G., Mroczek, D., Ustun, T.B., Wittchen, H.U. (1997). The World Health Organization's Composite International Diagnostic Interview Short-Form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, 7, 171-185.

Metropolitan Life Insurance Company. (1983). 1983 Metropolitan height and weight tables. *Stat Bull Metrop Insu Co 64*:1.

National Center for Health Statistics. (1999). National Health Interview Survey: Research for the 1995-2004 redesign. Vital Health Stat 2(126).

Shah, B.V., Barnwell, B.G. and Bieler, G.S. (1997). *SUDAAN User's Manual; Release 7.5*, Research Triangle Institute, Research Triangle Park, NC.

Stephens, T. and Craig, C.L. (1989). Fitness and Activity Measurement in the 1981 Canada Fitness Survey. In: National Center for Health Statistics. *Assessing Physical Fitness and Physical Activity in Population-Based Surveys*. T.F. Drury (ed.) DHHS Pub. No. (PHS) 89-1253. Public Health Service. Washington. U.S. Government Printing Office.

U.S. Department of Agriculture, Agricultural Research Service, Dietary Guidelines Advisory Committee, Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans, 1995, to the Secretary of Health and Human Services and the Secretary of Agriculture, p. 23-24.

U.S. Department of Health and Human Services. Healthy People 2010. 2nd ed. With Understanding and Improving Health and Objectives for Improving Health. 2 vols. Washington: U.S. Government Printing Office, November, 2000.

# Appendix I

## Calculation of Response Rates for the 2001 NHIS

The redesigned NHIS incorporated a change from the previous paper and pencil questionnaire to a new computer assisted personal interviewing (CAPI) system. The response rates calculated here pertain to the Basic Module questions in the 2001 NHIS.

The Basic Module collects basic information on the household and all family members. In addition, for each family, more detailed information is collected on one sample adult, one sample child, if any, and any child within the age guidelines for the immunization section.

### Household Response Rate

$$\frac{\text{(Interviewed Households)}}{\text{(Interviewed Households + Type A Non - Response Households)}}$$

The Household (HH) response rate is calculated by dividing the number of responding households by the number of households that are in-scope or eligible for the survey. Note that Type A non-response households are eligible households that were not interviewed for a variety of reasons: language problems; no one was at home after repeated contact attempts; family temporarily absent; refusal; household records rejected for insufficient data; household records rejected for other CAPI related problems; or other reasons for no interview.

### Conditional Family Core Response Rate

$$\frac{\text{(Interviewed Families)}}{\text{(Interviewed Families + Rejected Families (from Interviewed HH))}}$$

This file was created from Family Core data collected from the respondent about all persons in the family. Because all eligible persons in the family are included, the response rates for the Person file are identical to the response rates for the Family file. The response rates for the Family Core can be calculated in two ways. The conditional Family response rate is the rate only for those families identified as eligible and does not take into account household non-response. The conditional Family response rate is calculated by dividing the number of responding families by the number of families that are eligible for the survey, that is, from interviewed households. Note that a household can have multiple families, and rejected families are families that were deleted from interviewed households because of insufficient data.

**Final Family Core Response Rate**

$$\frac{\text{(Interviewed Families)}}{\text{(Interviewed families + Rejected Families (from Interviewed HH))}}\left[\text{Household Response Rate}\right]$$

The final Family response rate is the rate for those families identified as eligible that takes into account household non-response. The final Family response rate is calculated by dividing the number of responding families by the number of families that are eligible for the survey, that is, from interviewed households, and then multiplying this quotient by the Household response rate.

**Conditional Sample Adult Response Rate**

$$\frac{\text{(Interviewed Sample Adults)}}{\text{(Eligible Sample Adults)}}$$

The response rates for the Sample Adult section can be calculated in two ways. The conditional Sample Adult response rate is the rate only for those sample adults identified as eligible and does not take into account household or family non-response. The conditional Sample Adult response rate is calculated by dividing the number of responding sample adults by the number of eligible sample adults from interviewed families.

**Final Sample Adult Response Rate**

$$\frac{\text{(Interviewed Sample Adults)}}{\text{Eligible Sample Adults from Interviewed families}}\left[\begin{array}{l}\text{Final Family}\\ \text{Response Rate}\end{array}\right]$$

The final Sample Adult response rate is the rate for those sample adults identified as eligible that takes into account household and family non-response. The final Sample Adult response rate is calculated by dividing the number of responding sample adults by the number of sample adults who are eligible for the survey, that is, from interviewed families, and then multiplying this quotient by the final Family response rate.

**Conditional Sample Child Response Rate**

$$\frac{\text{(Interviewed Sample Children)}}{\text{(Eligible Sample Children)}}$$

The response rates for the Sample Child section can be calculated in two ways. The conditional Sample Child response rate is the rate only for sample children and does not take into account household or family non-response. The conditional Sample Child response rate is calculated by dividing the number of responding sample children by the number of eligible sample children from interviewed families.

**Final Sample Child Response Rate**

$$\frac{\text{(Interviewed Sample Children)}}{\text{(Eligible Sample Children from Interviewed families)}}\left[\begin{array}{l}\text{Final Family}\\ \text{Response Rate}\end{array}\right]$$

The final Sample Child response rate is the rate for sample children that takes into account household and family non-response. The final Sample Child response rate is calculated by dividing the number of responding sample children by the number of sample children who are eligible for the survey, that is, from interviewed families, and then multiplying this quotient by the final Family response rate.

**Conditional Immunization Response Rate**

$$\frac{\text{(Interviewed Certainty Children + Interviewed Non-certainty Sample Children)}}{\text{(Eligible Certainty Children + Eligible Non-certainty Sample Children)}}$$

The response rates for the Immunization section can be calculated in two ways. The conditional immunization response rate is the rate only for the Immunization section and does not take into account household or family non-response. The conditional immunization response rate is calculated by dividing the respondents to the immunization section by the number of eligible children from interviewed families. Note that certainty children are children who were selected to participate in the Immunization section with certainty based on their age, regardless of whether they were selected as the sample child. Non-certainty sample children were selected, because all sample children (under 18 years) who responded to the Sample Child section were eligible for the Immunization section.

**Final Immunization Response Rate**

$$\frac{\begin{array}{l}\text{[(Interviewed Certainty Children)(Final Family Response Rate) +}\\ \text{(Interviewed Non-certainty Sample Children)(Final Sample Child Response Rate)]}\end{array}}{\text{(Eligible Certainty Children + Eligible Non-certainty Sample Children)}}$$

The final immunization response rate is the rate for the Immunization section that takes into account household and family non-response. The final immunization response rate is calculated by adding the product of the number of responding certainty children and the final Family response rate to the product of the number of responding non-certainty sample children and the final Sample Child response rate, and then dividing this sum by the sum of the number of certainty children and non-certainty sample children who are from interviewed families and eligible for the survey. Note that certainty children are children who were selected to participate in the Immunization section with certainty based on their age, regardless of whether they were selected as the sample child. Non-certainty sample children were selected, because all sample

children (under 18 years) who responded to the Sample Child section were eligible for the Immunization section.

**Appendix I, Table 1. Response Rates for the 2001 NHIS**

| | |
|---|---|
| Household | 88.9% |
| Family/Person (Core) - Conditional | 98.5% |
| Family/Person (Core) - Final | 87.6% |
| Sample Adult - Conditional | 84.2% |
| Sample Adult - Final | 73.8% |
| Sample Child - Conditional | 92.0% |
| Sample Child - Final | 80.6% |
| Immunization - Conditional | 98.1% |
| Immunization - Final | 80.3% |

### Calculation of Response Rates for Combined NHIS Data Years

The response rates for combined NHIS data years are calculated in the same basic way as for a single year, but the sum of the numerators for all combined data years is used for the combined numerator and the sum of the denominators for all combined data years is used for the combined denominator. The following examples are shown for two years of data. The same methods apply for multiple years of data with the same sample design used in 1997-2004.

### Household Response Rate for Combined Data Years

$$\frac{\text{(Interviewed Households for Years 1 and 2)}}{\text{(Interviewed Households for Years 1 and 2 + Type A Non-Response Households for Years 1 and 2)}}$$

The Household (HH) response rate for combined data years is calculated by dividing the number of responding households for Years 1 and 2 by the number of households that are in-scope or eligible for the survey for Years 1 and 2. Note that Type A non-response households are eligible households that were not interviewed for a variety of reasons: language problems; no one was at home after repeated contact attempts; family temporarily absent; refusal; household records rejected for insufficient data; household records rejected for other CAPI related problems; or other reasons for no interview.

### Conditional Family Core Response Rate for Combined Data Years

$$\frac{\text{(Interviewed Families for Years 1 and 2)}}{\text{(Interviewed Families for Years 1 and 2 + Rejected Families for Years 1 and 2)}}$$

This file was created from Family Core data collected from the respondent about all persons in the family. Because all eligible persons in the family are included, the response rates for the Person file are identical to the response rates for the Family file. The response rates for the Family Core can be calculated in two ways. The conditional Family response rate is the rate only for those families identified as eligible and does not take into account household non-response. The conditional Family response rate for combined data years is calculated by dividing the number of responding families in Years 1 and 2 by the number of families that are eligible for the survey in Years 1 and 2, that is, from interviewed households in Year 1 and Year 2. Note that a household can have multiple families, and rejected families are families that were deleted from interviewed households because of insufficient data.

**Final Family Core Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Families for Years 1 and 2)}}{\left[\begin{array}{c}\text{Interviewed Families for Years 1 and 2 +}\\ \text{Rejected Families from Interviewed HH for Years 1 and 2}\end{array}\right]}\left[\begin{array}{c}\text{Household Response}\\ \text{Rate for Years 1 and 2}\end{array}\right]$$

The final Family response rate is the rate for those families identified as eligible that takes into account household non-response. The final Family response rate for combined data years is calculated by dividing the number of responding families for Years 1 and 2 by the number of families that are eligible for the survey for Years 1 and 2 , that is, from interviewed households for Year 1 and Year 2 , and then multiplying this quotient by the Household response rate for the combined data years.

**Conditional Sample Adult Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Sample Adults for Years 1 and 2)}}{\text{(Eligible Sample Adults for Years 1 and 2)}}$$

The response rates for the Sample Adult section can be calculated in two ways. The conditional Sample Adult response rate is the rate only for those sample adults identified as eligible and does not take into account household or family non-response. The conditional Sample Adult response rate for combined data years is calculated by dividing the number of responding sample adults for Years 1 and 2 by the number of eligible sample adults from interviewed families for Years 1 and 2.

**Final Sample Adult Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Sample Adults for Years 1 and 2)}}{\text{(Eligible Sample Adults from Interviewed families for Years 1 and 2)}}\left[\begin{array}{c}\text{Final Family Response}\\ \text{Rate for Years 1 and 2}\end{array}\right]$$

The final Sample Adult response rate is the rate for those sample adults identified as eligible that takes into account household and family non-response. The final Sample Adult response rate for

combined data years is calculated by dividing the number of responding sample adults for Years 1 and 2 by the number of sample adults who are eligible for the survey, that is, from interviewed families for Year 1 and Year 2, and then multiplying this quotient by the final Family response rate for the combined data years.

**Conditional Sample Child Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Sample Children for Years 1 and 2)}}{\text{(Eligible Sample Children for Years 1 and 2)}}$$

The response rates for the Sample Child section can be calculated in two ways. The conditional Sample Child response rate is the rate only for sample children and does not take into account household or family non-response. The conditional Sample Child response rate for combined data years is calculated by dividing the number of responding sample children for Years 1 and 2 by the number of eligible sample children from interviewed families for Years 1 and 2.

**Final Sample Child Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Sample Children for Years 1 and 2)}}{\text{(Eligible Sample Children from Interviewed families for Years 1 and 2)}}\left[\begin{array}{l}\text{Final Family Response}\\ \text{Rate for Years 1 and 2}\end{array}\right]$$

The final Sample Child response rate is the rate for sample children that takes into account household and family non-response. The final Sample Child response rate for combined data years is calculated by dividing the number of responding sample children for Year 1 and Year 2 by the number of sample children who are eligible for the survey, that is, from interviewed families for Year 1 and Year 2, and then multiplying this quotient by the final Family response rate for the combined data years.

**Conditional Immunization Response Rate for Combined Data Years**

$$\frac{\text{(Interviewed Certainty Children for Years 1 and 2 + Interviewed Non-Certainty Sample Children for Years 1 and 2)}}{\text{(Eligible Certainty Children for Years 1 and 2 + Eligible Non-Certainty Sample Children for Years 1 and 2)}}$$

The response rates for the Immunization section can be calculated in two ways. The conditional immunization response rate is the rate only for the Immunization section and does not take into account household or family non-response. The conditional immunization response rate for combined data years is calculated by dividing the respondents to the immunization section for Year 1 and Year 2 by the number of eligible children from interviewed families for Year 1 and Year 2. Note that certainty children are children who were selected to participate in the Immunization section with certainty based on their age, regardless of whether they were selected as the sample child. Non-certainty sample children were selected, because

all sample children (under 18 years) who responded to the Sample Child section were eligible for the Immunization section.

**Final Immunization Response Rate for Combined Data Years**

$$\frac{[(\text{Interviewed Certainty Children for Years 1 and 2})(\text{Final Family Response Rate for Years 1 and 2}) + (\text{Interviewed Non-certainty Sample Children for Years 1 and 2})(\text{Final Sample Child Response Rate for Years 1 and 2})]}{(\text{Eligible Certainty Children for Years 1 and 2} + \text{Eligible Non-certainty Sample Children for Years 1 and 2})}$$

The final immunization response rate is the rate for the Immunization section that takes into account household and family non-response. The final immunization response rate for combined data years is calculated by adding the product of the number of responding certainty children for Years 1 and 2 and the final Family response rate for Years 1 and 2 to the product of the number of responding non-certainty sample children for Years 1 and 2 and the final Sample Child response rate for Years 1 and 2, and then dividing this sum by the sum of the number of certainty children and non-certainty sample children who are from interviewed families and eligible for the survey for Years 1 and 2. Note that certainty children are children who were selected to participate in the Immunization section with certainty based on their age, regardless of whether they were selected as the sample child. Non-certainty sample children were selected because all sample children (under 18 years) who responded to the Sample Child section were eligible for the Immunization section.

**Appendix I, Table 2. Number Eligible/Interviewed 2001 NHIS**

| File | Eligible | Interviewed |
|---|---|---|
| **Household** | 43,797 | 38,932 |
| **Family/Person** | 40,227 | 39,633 |
| **Sample Adult** | 39,564 | 33,326 |
| **Sample Child** | 14,766 | 13,579 |
| **Immunization** | 15,000 | 14,709 |

**Appendix I, Table 3. Number Eligible/Interviewed 2000 NHIS**

| File | Eligible | Interviewed |
|---|---|---|
| **Household** | 43,437 | 38,632 |
| **Family/Person** | 39,998 | 39,264 |
| **Sample Adult** | 39,201 | 32,374 |
| **Sample Child** | 14,711 | 13,376 |
| **Immunization** | 14,890 | 14,618 |

**Appendix I, Table 4. Number Eligible/Interviewed 1999 NHIS**

| File | Eligible | Interviewed |
|---|---|---|
| **Household** | 42,882 | 37,573 |
| **Family/Person** | 38,845 | 38,171 |
| **Sample Adult** | 38,117 | 30,801 |
| **Sample Child** | 14,217 | 12,910 |
| **Immunization** | 14,178 | 13,881 |

**Appendix I, Table 5. Number Eligible/Interviewed 1998 NHIS**

| File | Eligible | Interviewed |
|---|---|---|
| **Household** | 42,440 | 38,209 |
| **Family/Person** | 39,559 | 38,773 |
| **Sample Adult** | 38,729 | 32,440 |
| **Prevention Sample Adult** | 32,440 | 31,882 |
| **Sample Child** | 14,619 | 13,645 |
| **Prevention Sample Child** | 13,645 | 13,610 |
| **Immunization** | 15,041 | 14,775 |

**Appendix I, Table 6. Number Eligible/Interviewed 1997 NHIS**

| File | Eligible | Interviewed |
|---|---|---|
| **Household** | 43,370 | 39,832 |
| **Family/Person** | 41,291 | 40,623 |
| **Sample Adult** | 40,552 | 36,116 |
| **Sample Child** | 15,244 | 14,290 |
| **Immunization** | 15,558 | 15,402 |

# Appendix II

## Race and Hispanic Origin in the 2001 NHIS

### Background

For over 20 years, the National Health Interview Survey (NHIS) has collected information on the race and ethnicity of its respondents, following guidelines set forth by the Office of Management and Budget in a policy known as OMB Directive 15 (Office of Management and Budget, 1977). The NHIS has relied on respondents to provide self-identified race and ethnicity information (proxy information is reported for children and non-present household members), although interviewer-observed race was also recorded through 1996, the last year of the paper questionnaire. NHIS data are routinely tabulated by race and ethnicity in NCHS publications such as Current Estimates, Health U.S., and Advance Data reports.

In response to the changing demographics of the U.S. population, the OMB revised Directive 15 in 1997 after an extensive period of research and public commentary. The new race and ethnicity standards allow respondents to the Census and federal surveys to indicate more than one race group in answering questions on race. A complete description of the new OMB guidelines on the collection of racial and ethnic data, including descriptions of the new race categories, the ordering of race and ethnicity questions, and guidelines for the tabulation and publication of data under the new standards, can be found on the OMB web site: **<http://www.whitehouse.gov/OMB/inforeg>**. Although this policy is not expected to be fully implemented across the federal statistical system until 2003, surveys like the NHIS that are reviewed by OMB for renewal on a yearly basis are expected to implement changes to their survey instruments when they apply for their first OMB clearance after the policy's effective date. In accordance with this requirement, the NHIS became fully compliant with the new race and ethnicity standards with the fielding of the 1999 questionnaire, although the NHIS had been following some aspects of the new guidelines for many years.

### Race and Hispanic Origin Questions in the National Health Interview Survey

The 2001 NHIS included two questions about Hispanic Origin:

Do/Does {you/name} consider {yourself / himself / herself} Hispanic / Latino?" [HHC.170],

and

"Please give me the number of the group that represents your Hispanic Origin or ancestry" [HHC.180].

There were no changes in the wording of the 2001 Hispanic origin question, but some responses were imputed and the variable name was changed from HISPANCR to HISPAN_I to indicate this fact (see section on the imputation of race and ethnicity later in this document).

The 2001 NHIS included two race questions to obtain information on a respondent's race:

"What race {does/do} {name/you} consider {himself/herself/yourself} to be?" [HHC.200],

and

"Which one of these groups, that is (FR: READ GROUPS) would you say BEST represents {your/name's} race?" [HHC.220].

The first question is asked of all respondents, while the second question is asked only of those respondents who give more than one race in response to the first question. Although the wording and placement of these two questions are basically the same as they had been in the NHIS for the past several years, there were changes made in the response categories effective 1999. In compliance with the new race and ethnicity standards, the category "Asian and Pacific Islander" is now split into two categories, "Asian" and "Native Hawaiian and Other Pacific Islander", in data collection. Because confidentiality regulations on minimum sample size do not permit the NHIS to release data for Native Hawaiians and Other Pacific Islanders or some Asian subgroups separately, data are provided for the three largest Asian subpopulation groups, while the "Other Pacific Islander" and "Other Asian" categories combine the remaining groups that cannot be shown separately.

Data users are strongly urged to read carefully the 2001 public use documentation, where details on the specific response categories for the race questions can be found. The following table summarizes the changes made to the Hispanic origin and race variables in the 2001 data file. Additional details on these variables can be found in the survey documentation, and users are urged to read the variable descriptions carefully to determine how and when the variables should be used in analysis. Data users are also encouraged to examine frequencies of the unweighted data for these variables before computing weighted estimates.

**Appendix II, Table 1. Description of the 2001 NHIS Race and Ethnicity Variables**

| 2001 Variable Name | Description |
|---|---|
| **ORIGIN_I** | Hispanic origin/ancestry with imputed values for some records |
| **ORIGIMPT** | Hispanic origin imputation flag |
| **HISPAN_I** | Type of Hispanic origin/ancestry with imputed values for some records |
| **HISPIMPT** | Type of Hispanic origin imputation flag |
| **RCDT1P_I** | Detailed race variable; multiple race persons in separate category with imputed values for some records. |
| **RC_SMP_I** | Summary race variable (i.e., no detailed groups); contains 4 of 5 OMB race groups and "Other race"; multiple race persons in separate category; imputed values for some records |
| **RACERP_I** | Contains 4 of 5 OMB race groups and "Other race"; multiple race persons coded differently than in RCDT1P_I and RC_SMP_I (see documentation), and values were imputed for some records. |
| **MRACRP_I** | Detailed race variable; only multiple race persons not selecting a primary race group in separate category. Values were imputed for some records. |
| **MRACBP_I** | See section below on bridging; values were imputed for some records. |
| **RACREC_I** | All persons not coded in a race category are imputed to a race category on this variable. |
| **RACEIMPT** | Imputation flag for use in determining which cases were imputed for the race variables |
| **HISCOD_I** | Same categories as RACREC_I, crossed with ORIGIN_I (Hispanic/non-Hispanic); values were imputed for some records. |
| **ERIMPFLG** | Summary race/ethnicity imputation flag - indicates that either race or ethnicity or both race and ethnicity were imputed |

## Procedures For Imputation of Ethnicity And Race in the 2001 NHIS

In an effort to improve the quality of data on ethnicity and race in the NHIS, hot-deck imputation of selected race and ethnicity variables was done for the first time in the 2000 NHIS. These procedures were replicated for the 2001 NHIS data. Prior to the 2000 NHIS, a crude imputation method that assigned a race to persons with missing values on the variable MAINRACE for race recodes #1 and #2 was used. Under these procedures, if an observed race was recorded by the field representative, it was used to code a race value. If there were no observed race value, all persons who had a missing value for MAINRACE and were identified as Hispanic on the Hispanic origin question were coded as "white". In all other cases, non-Hispanic persons were coded as "other race".

The variables ORIGIN (whether or not the respondent is of Hispanic origin), HISPTY01-HISPTY10 (type of Hispanic origin), RACE1-RACE5 (each of 5 possible race mentions), and

MAINRACE (primary race selection for persons reporting more than one race) with missing values were imputed (note that the pre-imputation variable names are used in this description because the names were not changed until the imputation was completed). The imputation was carried out in two stages: within households at the first stage, and between households at the second stage. Hot-deck imputation procedures developed for the Decennial Census Dress Rehearsal (conducted in 1998) were adapted for use on the 2000 and 2001 NHIS data. These specifications formed the basis of the first stage of the imputation (within households), although they were adapted to utilize NHIS family relationship variables. However, the specifications obtained from Census did not contain information on the imputation of race and ethnicity between households. Staff in DHIS and NCHS's Office of Research and Methodology developed the specifications for the between-household imputation, using the secondary sampling unit (SSU) as the geographic unit for selecting donors.

**1. Stage 1 Imputation** - this was done for households in which some persons had missing values, and some persons had valid entries for ethnicity and race variables.

Step 1. Generate datasets based on NHIS household files for within-household imputation.

Step 2. Preview the frequency distributions of the variables to be imputed.

Step 3. Reclassify donors based on variables RRP (relationship of person to household reference persons) and DEGREE1-DEGREE7 (relationship variables - e.g., whether person is biological, step, foster, or in-law child of reference person).

Step 4. Load donors' data to hot decks within each household, and conduct imputation for each donee in the same household. Donees are classified in twenty-six categories based on the relationship of the donees to the Reference Person in the household (see following section ). The allocation sequence of donors for each type of donee is different, depending on the type of the donee, and the relationship between the donor and the donee.

Step 5. Review the distributions of the imputed variables after imputation for comparison and analysis. Combine all records, and reclassify households for Stage 2 imputation.

**2. Stage 2 Imputation** - this was done for households that had all persons with missing values.

A. The imputation was divided into three parts:

1). Imputation among Hispanic households (ORIGIN=1).

2). Imputation among Non-Hispanic households (ORIGIN=2).

3). Imputation for households with unknown Hispanic origin (ORIGIN=7, 8, 9).

B. Each part of the imputation complied with certain rules that are outlined in further detail in the Stage 2 imputation specification (not provided here). The combinations of imputed variables in each part are different.

C. After all imputations were completed, datasets from Stage 1 and Stage 2 were combined, records that were imputed flagged for the in-house and public use data files, and comparisons of the distributions of the variables before and after imputation were examined.

## Bridging to the Old OMB Standards

The OMB tabulation guidelines for the new race and ethnicity standards recognize that the complete transition from the old standards to the new standards will take some time, and that many federal statistical systems have a primary mission to track data trends over time. During this transitional period, known as the "bridge", it has been recommended that data systems tabulate data for publication under the new standards, while also providing a means for data users to bridge the new data back to the old standards. This will allow data users to examine differences, if any, in tabulating the data under the old and new standards, assist in the maintenance of data trends, and allow users to become accustomed to data tabulated under the new standard before the transition is complete. In the NHIS, the second race question (commonly known as the "follow-up question") is used to create the bridge between data collected under the old standards and data collected under the new ones. The 2001 NHIS public use data release contains one bridge race variable to allow comparisons of 2001 data with data from previous years, and to enable merging the 2001 data with 1997, 1998 , 1999, and/or 2000 data.

There is one major change to the race and ethnicity data in the 2000 NHIS (and now in 2001) that occurs as a result of the creation of a bridge variable. NCHS confidentiality standards do not permit NCHS to release data that might lead to the inadvertent identification of individual respondents to the survey (e.g., a combination of demographic, geographic and other characteristics of persons in relatively small population groups could lead to identification of an individual respondent). Beginning with the 2000 survey (and now in 2001), data on "Asian" persons and "Native Hawaiian and Other Pacific Islander (NHOPI)" persons were collected separately according to the new OMB guidelines. Ideally, these two groups could be combined to recreate the old category "Asian and Pacific Islander (API)" as a bridge back to data collected under the old race standards. However, the NCHS Disclosure Review Board (DRB), consulting with DHIS analysts, determined that releasing data using an all-inclusive "Other Pacific Islander" category (which would include the Native Hawaiian, Samoan, Guamanian, and Other Pacific Islander groups) would pose a disclosure risk, especially when used in combination with other demographic and geographic information available on the file. For this reason, the decision was made to suppress the "Other Pacific Islander" category on all public use bridge variables. **This is important for data users to know because this change makes it impossible to bridge back to the old "Asian and Pacific Islander " category that existed in the 1998 and earlier NHIS surveys**. Data users who need this information for their analyses will have to contact the NCHS Research Data Center to obtain controlled access to non-released data.

## Creation and Editing of 2001 Race Variables

The variables RACEEC_I and MRACRP_I correspond to the old OMB guidelines for collecting racial and ethnic data (see the survey documentation for further descriptions of these variables). They were created in the same fashion as their previous NHIS counterparts (National Center for Health Statistics 1996), with two exceptions. First, since observed race is no longer collected in the NHIS (beginning in 1997), it was not used to help classify persons with "Unknown" race on the RACREC_I recode. Second, the recodes "White/Non-White" and "Black/Non-Black" were not created because they are no longer used in the weighting and tabulation of NHIS data. As in the past, smaller subgroups have been collapsed for confidentiality reasons.

Since the NHIS is now required to collect racial and ethnic data under the new OMB guidelines, new variables have been created to allow users to tabulate NHIS data by race variables that correspond to the new OMB guidelines. These variables conform to the new OMB race standards; therefore they are created independently of the follow-up race question (see the section of this appendix on Race and Hispanic Origin Questions in the National Health Interview Survey). The variable RACERP_I was created using an algorithm that first coded the five race mentions from the survey into the single and multiple race group combinations (shown in bold/italicized and regular font, respectively) included in Table 2, below. All of the multiple race categories in the table were then collapsed into a single "Multiple race" category, and along with 4 of the 5 OMB single race categories and the category "Other race", the variable RACERP_I was created. The full algorithm is provided below so that our data users can better understand how this variable is derived.

## SAS Code for Single and Multiple Race Groups

This SAS algorithm takes into account the new OMB categories: White, Black, American Indian/Alaskan Native (AIAN), Asian, and Native Hawaiian and Other Pacific Islander (NHOPI). In addition, it includes an "Other race" category for persons who could not be classified elsewhere. In the NHIS, data are collected in 16 race categories: White, Black, Indian (American), Alaska Native, Native Hawaiian, Guamanian, Samoan, Other Pacific Islander (a verbatim mention which is back-coded to this category), Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian (a verbatim mention which is back-coded to this category). Persons that cannot be classified elsewhere are put in the "Other race" category. These can all be collapsed back to the OMB categories in the following fashion: *White*, *Black*, *AIAN* (includes Indian (American) and Alaska Native), *Asian* (includes Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese and Other Asian), *NHOPI* (includes Native Hawaiian, Guamanian, Samoan and Other Pacific Islander), plus the "Other race" category

**Step 1**: In the NHIS there are 5 possible mentions of race, which, when edited and cleaned, will become 5 race variables called RACE1, RACE2, RACE3, RACE4 and RACE5.

**Step 2**: Create and initialize the following variables to 0:

```
RACEW=0;
```

RACEB=0;
RACEAIAN=0;
RACEASIA=0;
RACENHPI=0;
RACEOTHR=0;

**Step 3**: Set non-mutually exclusive conditions for recoding the 5 race variables, and set each of the above categories to the number designated:

IF ((RACE1=1) or (RACE2=1) or (RACE3=1) or (RACE4=1) or RACE5=1)) then RACEW=1;
** This sets RACEW to 1 if there is any mention of the race "White" in any of the 5 race variables;*

IF ((RACE1=2) or (RACE2=2) or (RACE3=2) or (RACE4=2) or RACE5=2)) then RACEB=2;
**This sets RACEB to 2 if there is any mention of the race "Black" in any of the 5 race variables;*

IF ((RACE1=3) or (RACE2=3) or (RACE3=3) or (RACE4=3) or RACE5=3)) then RACEAIAN=4;
*******This sets RACEAIAN to 4 if there is any mention of the race "AIAN" in any of the 5 race variables;*

IF ((RACE1=4) or (RACE2=4) or (RACE3=4) or (RACE4=4) or RACE5=4)) then RACEASIA=8;
*******This sets RACEASIA to 8 if there is any mention of the race "Asian" in any of the 5 race variables;*

IF ((RACE1=5) or (RACE2=5) or (RACE3=5) or (RACE4=5) or RACE5=5)) then RACENHPI=16;
*******This sets RACENHPI to 16 if there is any mention of the race "NHOPI (Native Hawaiian and Other Pacific Islander)" in any of the 5 race variables;*

IF ((RACE1=6) or (RACE2=6) or (RACE3=6) or (RACE4=6) or RACE5=6)) then RACEOTHR=32;
*******This sets RACEOTHR to 32 if there is any mention of the race "OTHR" in any of the 5 race variables;*

**Step 4:** RACEFULL=SUM(OF RACEW RACEB RACEAIAN RACEASIA RACENHPI RACEOTHR);

The variables RACEW, RACEB, RACEAIAN, RACEASIA, RACENHPI, and RACEOTHR thus take on the numbers 1,2,4,8,16, and 32, which add up to a series of unique numbers corresponding to specific combinations of races. The value of RACEFULL tells which races (RACEW through RACEOTHR) combined to give that number. For example, if RACEFULL=3, the only the sum of the values for RACEW=1 and RACEB=2 could have produced the number 3. Therefore anyone with the value RACEFULL=3 falls into the "White/Black" race category. If RACEFULL=1, then those persons fall into the "White" category. This scheme accurately allocates persons with multiple API and AIAN mentions. The full listing of categories and the numbers to which they correspond are included in the following table:

**Appendix II, Table 2. Coding Scheme for SAS Algorithm**

| **Coding Scheme for OMB Race Category Data (including single and multiple race mentions)** | | |
|---|---|---|
| **# of Category (reported in SAS frequency distribution of RACEFULL)** | **Sum of Coding (breakdown of RACEFULL= SUM(OF RACEW+RACEB+ RACEAIAN+RACEASIA+RACENHPI+RACEOTHR))** | **Resulting Category (used in the PROC FORMAT statement to label the categories in SAS)** |
| 1 | 1+0+0+0+0+0 | ***White*** |
| 2 | 0+2+0+0+0+0 | ***Black*** |
| 3 | 1+2+0+0+0+0 | White/Black |
| 4 | 0+0+4+0+0+0 | ***AIAN*** |
| 5 | 1+0+4+0+0+0 | White/AIAN |
| 6 | 0+2+4+0+0+0 | Black/AIAN |
| 7 | 1+2+4+0+0+0 | White/Black/AIAN |
| 8 | 0+0+0+8+0+0 | ***Asian*** |
| 9 | 1+0+0+8+0+0 | White/Asian |
| 10 | 0+2+0+8+0+0 | Black/Asian |
| 11 | 1+2+0+8+0+0 | White/Black/Asian |
| 12 | 0+0+4+8+0+0 | AIAN/Asian |
| 13 | 1+0+4+8+0+0 | White/AIAN/Asian |
| 14 | 0+2+4+8+0+0 | Black/AIAN/Asian |
| 15 | 1+2+4+8+0+0 | White/Black/AIAN/Asian |
| 16 | 0+0+0+0+16+0 | ***NHOPI*** |
| 17 | 1+0+0+0+16+0 | White/NHOPI |
| 18 | 0+2+0+0+16+0 | Black/NHOPI |

| Coding Scheme for OMB Race Category Data (including single and multiple race mentions) | | |
|---|---|---|
| **# of Category (reported in SAS frequency distribution of RACEFULL)** | **Sum of Coding (breakdown of RACEFULL= SUM(OF RACEW+RACEB+ RACEAIAN+RACEASIA+RACENHPI+RACEOTHR))** | **Resulting Category (used in the PROC FORMAT statement to label categories in SAS)** |
| 19 | 1+2+0+0+16+0 | White/Black/NHOPI |
| 20 | 0+0+4+0+16+0 | AIAN/NHOPI |
| 21 | 1+0+4+0+16+0 | White/AIAN/NHOPI |
| 22 | 0+2+4+0+16+0 | Black/AIAN/NHOPI |
| 23 | 1+2+4+0+16+0 | White/Black/AIAN/NHOPI |
| 24 | 0+0+0+8+16+0 | Asian/NHOPI |
| 25 | 1+0+0+8+16+0 | White/Asian/NHOPI |
| 26 | 0+2+0+8+16+0 | Black/Asian/NHOPI |
| 27 | 1+2+0+8+16+0 | White/Black/Asian/NHOPI |
| 28 | 0+0+4+8+16+0 | AIAN/Asian/NHOPI |
| 29 | 1+0+4+8+16+0 | White/AIAN/Asian/NHOPI |
| 30 | 0+2+4+8+16+0 | Black/AIAN/Asian/NHOPI |
| 31 | 1+2+4+8+16+0 | White/Black/AIAN/Asian/NHOPI |
| 32 | 0+0+0+0+0+32 | ***Other*** |
| 33 | 1+0+0+0+0+32 | White/Other |
| 34 | 0+2+0+0+0+32 | Black/Other |
| 35 | 1+2+0+0+0+32 | White/Black/Other |
| 36 | 0+0+4+0+0+32 | AIAN/Other |
| 37 | 1+0+4+0+0+32 | White/AIAN/Other |
| 38 | 0+2+4+0+0+32 | Black/AIAN/Other |
| 39 | 1+2+4+0+0+32 | White/Black/AIAN/Other |
| 40 | 0+0+0+8+0+32 | Asian/Other |
| 41 | 1+0+0+8+0+32 | White/Asian/Other |
| 42 | 0+2+0+8+0+32 | Black/Asian/Other |
| 43 | 1+2+0+8+0+32 | White/Black/Asian/Other |

| Coding Scheme for OMB Race Category Data (including single and multiple race mentions) | | |
|---|---|---|
| **# of Category (reported in SAS frequency distribution of RACEFULL** | **Sum of Coding (breakdown of RACEFULL= SUM(OF RACEW+RACEB+ RACEAIAN+RACEASIA+RACENHPI+RACEOTHR))** | **Resulting Category ( used in the PROC FORMAT statement to label categories in SAS)** |
| 44 | 0+0+4+8+0+32 | AIAN/Asian/Other |
| 45 | 1+0+4+8+0+32 | White/AIAN/Asian/Other |
| 46 | 0+2+4+8+0+32 | Black/AIAN/Asian/Other |
| 47 | 1+2+4+8+0+32 | White/Black/AIAN/Asian/Other |
| 48 | 0+0+0+0+16+32 | NHOPI/Other |
| 49 | 1+0+0+0+16+32 | White/NHOPI/Other |
| 50 | 0+2+0+0+16+32 | Black/NHOPI/Other |
| 51 | 1+2+0+0+16+32 | White/Black/NHOPI/Other |
| 52 | 0+0+4+0+16+32 | AIAN/NHOPI/Other |
| 53 | 1+0+4+0+16+32 | White/AIAN/NHOPI/Other |
| 54 | 0+2+4+0+16+32 | Black/AIAN/NHOPI/Other |
| 55 | 1+2+4+0+16+32 | White/Black/AIAN/NHOPI/Other |
| 56 | 0+0+0+8+16+32 | Asian/NHOPI/Other |
| 57 | 1+0+0+8+16+32 | White/Asian/NHOPI/Other |
| 58 | 0+2+0+8+16+32 | Black/Asian/NHOPI/Other |
| 59 | 1+2+0+8+16+32 | White/Black/Asian/NHOPI/Other |
| 60 | 0+0+4+8+16+32 | AIAN/Asian/NHOPI/Other |
| 61 | 1+0+4+8+16+32 | White/AIAN/Asian/NHOPI/Other |
| 62 | 0+2+4+8+16+32 | Black/AIAN/Asian/NHOPI/Other |
| 63 | 1+2+4+8+16+32 | White/Black/AIAN/Asian/NHOPI/Other |

***Data users should be aware that the variable RACEFULL and others derived from it, are not available on public use data files for confidentiality reasons***. The recode RACERP_I is a recode based on RACEFULL. Analysts who wish to use more detailed race data in their analyses should submit a proposal to the NCHS Research Data Center.

## Further Information

Although the race variables included in the 2000 file have been edited and tested, analytic and methodological work with these variables continues. NCHS is also evaluating other recodes for possible public release at a later date. If these analyses should result in changes to the 2000 NHIS race data, information about this will be placed on the NCHS web site (see page 3).

## References

Office of Management and Budget. Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity, December 15, 2000. **<http://.whitehouse.gov/OMB/inforeg>** (Statistical Programs and Standards section)

Office of Management and Budget. Interagency Committee for the Review of Standards for Data on Race and Ethnicity. Draft Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity, February 15, 1999.

National Center for Health Statistics (1996). Computer Processing Procedures Calendar Year 1995, Volume 1 of 2, National Health Interview Survey, 1995. National Center for Health Statistics, Hyattsville, MD (Producer). National Technical Information Service, U.S. Department of Commerce, Springfield, VA. 22161 (Distributor).

Office of Management and Budget. Office of Management and Budget Circular No. A48. Standards and Guidelines for Federal Statistics. Section 7H, Exhibit F, May 12, 1977.

Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Federal Register 62(210):58782-58790. 1997.

# Appendix III

## Variance Estimation and Other Analytic Issues, NHIS 2001

## Introduction

The data collected in the NHIS are obtained through a complex, multistage sample design that involves stratification, clustering, and oversampling of specific population subgroups. The final weights provided for analytic purposes are adjusted in several ways to yield estimates for the civilian, noninstitutionalized population of the United States. As with any variance estimation methodology, those presented here involve several simplifying assumptions about the design and weighting scheme applied to the data. This appendix provides guidelines for data users based on simplified concepts of the NHIS sample design structure so that users may compute reasonably accurate standard errors.

There are several available software packages for analyzing complex samples. The web site, *Summary of Survey Analysis Software,* currently located at

**<http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html>**

provides references for and a comparison of different software alternatives for the analysis of complex data. Analysts at NCHS use the software package SUDAAN® (Shah et al. 1997) to produce accurate standard errors. In this appendix, examples of SUDAAN computer code are provided for illustrative purposes. However, the appropriate application of these procedures is the ultimate responsibility of data users and the example command code is *not* "guaranteed". Both the computer command code and methods are subject to change without notification to the user. NCHS strongly recommends that NHIS data are analyzed under the direction of or in consultation with a statistician who is cognizant of sampling methodologies and techniques for the analysis of survey data.

☹ **CAUTION.** Users are reminded that the use of standard statistical procedures, which are based on the assumption that data are generated via simple random sampling (SRS), will produce incorrect estimates of variances and standard errors when used to analyze data from the NHIS. The clustering protocols that are used in the multistage selection of the NHIS sample require other analytic procedures described below. Analysts who apply SRS techniques to NHIS data will produce standard errors that are, on average, too small, and are likely to produce results that are subject to Type I error.

## Conceptual NHIS design for 1995-2004

Thorough discussions of the NHIS design, the methods used for weighting data, and the methods used for variance estimation are beyond the scope of this appendix, but are provided elsewhere (NCHS 1999; NCHS 2000). This appendix outlines the basic technical ideas published in these technical reports (NCHS 1999; NCHS 2000).

To achieve sampling efficiency and to keep survey operations manageable and cost-effective, the NHIS survey planners used multistage sampling techniques to select the sample of persons and households

for the NHIS. These multistage methods partition the target universe into several nested levels of strata and clusters. The NHIS target universe is defined as all dwelling units in the U.S. that contain members of the civilian noninstitutionalized population. As the NHIS is conducted in a face-to-face interview format, a simple random sample of dwelling units would be too dispersed throughout the nation; as a result, the costs of obtaining a simple random sample of 50,000 households would be prohibitive. Also, specific population subgroups, such as black and/or Hispanic households, would not be sampled sufficiently under a simple random sample design. To achieve survey objectives subject to resource constraints, the NHIS uses methods of clustering, stratification, and oversampling of specific population subgroups.

First, the target universe was partitioned into approximately 1900 Primary Sampling Units (PSUs), single counties or groups of adjacent counties (or equivalent jurisdictions) and/or metropolitan areas. These PSUs vary in population size and number of jurisdictions. The PSUs with the largest populations (e.g., the New York metropolitan area) support cost-effective sampling and are sampled with certainty; consequently, they are designated as self-representing (SR) PSUs. Resource constraints required that the remaining smaller PSUs be sampled for data collection. These smaller PSUs are called non-self-representing (NSR) or non-certainty PSUs. The universe of NSR PSUs is stratified using multiple criteria consistent with NHIS objectives. The NSR PSUs were stratified first at the state level according to metropolitan status (metro or non-metro). If a particular NSR stratum in a given state contained a large population, then it was further stratified by aggregate-level poverty rates. Thus, the number of NSR strata varies from state to state, and the number of PSUs varies from stratum to stratum. Once these strata were defined, a sample of PSUs was selected; within each NSR stratum, two PSUs were selected without replacement with probability proportional to population size, and the SR PSUs were selected with certainty. For some stratum with smaller population size, only one NSR PSU was drawn from a stratum.

The U.S. Census Bureau partitioned each selected NSR or SR PSU into substrata of Census blocks or combined blocks based on the concentrations of black and Hispanic populations. These race and ethnicity density substrata were defined according to the population concentrations from the 1990 Decennial Census. New housing within a PSU was included as its own substratum in order to produce the most current sample of households. Each PSU could be partitioned into up to 21 substrata of dwelling units. Large metropolitan SR PSUs tend to have many substrata, while the NSR PSUs tend to have only a few.

Sampling within the PSU substrata is complex and involves clustering dwelling units within each substratum. These clusters form a universe of Secondary Sampling Units (SSUs). A systematic sample of SSUs is selected to represent each substratum. Each race and ethnicity density substratum has its own sampling rate for SSU selection.

Within each selected SSU all households containing black or Hispanic persons are selected for interview, while only a sample of other households are selected. These non-black, non-Hispanic households are sampled at different rates within the 21 substrata. For selected households, the NHIS collects some information on all household members, and additional information is obtained for randomly selected persons in each household. For example, one adult per family is randomly selected for interview with the sample adult questionnaire.

This hierarchy of sampling allows the creation of household- and person-level base weights. Each base weight is the product of the inverse probabilities of selection at each sampling stage. Roughly speaking, the base weight is the number of population units a sampled unit represents. Under ideal sampling conditions, a base-weighted sample total will be an unbiased estimator for the true total in the target population. In practice, however, the base weights are adjusted for non-response, and ratio-adjusted to create final sampling weights. The final weights are adjusted according to a quarterly poststratification by 88 age/sex/race and/or ethnicity classes based on Census control totals.

Internally, NCHS uses the design and weighting information to formulate appropriate variance estimators for NHIS statistics. While recognizing the need to provide accurate information, NCHS also must adhere to the Public Health Service Act (Section 308(d)) that forbids the disclosure of any information that may compromise the confidentiality promised to its survey respondents. Consequently, much of the NHIS design information cannot be publicly released, and other data are either suppressed or recoded to insure confidentiality. In order to satisfy this disclosure constraint, many of the original design strata, substrata, PSUs and SSUs are masked for public release by applying techniques to cluster, collapse, mix, and partition the original design variables. Through this process the original NHIS design variables are transformed into public use design variables. The public use design structures perform reasonably well when compared to internal NCHS design structures (NCHS 2000). The sampling weights have not been changed in any way for the public data. Data users who want access to the internal NCHS data have the option of accessing internal data through the NCHS Research Data Center (for further information, refer to **<http://www.cdc.gov/nchs/r&d/rdc.htm>**).

## Design Information Available on the NHIS Public Use Data

The Person file public use design variables utilized for variance estimation are provided in Table 1. Users should check the Dataset Documentation for exact names and locations of these variables for each of the files.

**Appendix III, Table 1. Variables Used for Variance Estimation, 2001 NHIS Person File**

| Variable Name | Variable Label |
|---|---|
| **Stratum** | **Stratum for variance estimation** |
| **PSU** | **PSU for variance estimation** |
| **WTFA** | **Weight - Final, annual Person weight** |

As discussed above, in order to mask true geographical locations the STRATUM and PSU levels are pseudo-levels or simplified versions of the true NHIS sample design variables. Analysts are cautioned that these simplified design structures do not support geographical analyses below the regional level.

☹ **CAUTION**. Significant changes were made to the Stratum and PSU values beginning with the 1997 survey year. More strata have been provided (compared to the 1995 public release) to improve statistical efficiency in various statistical estimation procedures. The sample design variables provided on the 1997-2001 NHIS public use data files are *not* comparable to those of previous data years. Users are

cautioned that variance estimation structures discussed here are for individual survey years only, not for pooled analyses of multiple years of the NHIS.

### Variance Estimation Method for Public Use Data

The method described below is applicable to all 2001 NHIS public use data, except the Injury Episode, Injury Verbatim, and Poison Episode files (when available).

For this method of variance estimation, the NHIS sample is treated as having 339 strata, each containing two sampled PSUs. While in reality the PSUs were not duplicated, the limited public release design information requires a mathematical simplification that the PSUs be treated as if they were sampled with replacement (WR). This public use method provides slightly more conservative standard errors than the true variance estimation method that is applied internally by analysts at NCHS (NCHS 2000). Additionally, this public use method is applicable in many of the statistical packages for complex survey data that require exactly two sample PSUs per stratum. Moreover, this method is robust when analyzing subsetted or subgroup data (see the section "Subsetted Data Analysis" below).

When implementing this public use method, users should observe 678 PSUs when analyzing the full database. The simplified design structure can be specified with the following statements in SUDAAN:

***PROC*** **... DESIGN = WR ;**
***NEST*** **STRATUM PSU ;**
***WEIGHT*** **WTFA ;**

Note that SUDAAN requires that the input file be sorted by the variables listed on the nest statement (i.e., STRATUM and PSU). Design statements for other data files should use the appropriate weight variables.

☹ **CAUTION.** A rule of thumb to calculate the number of degrees of freedom to associate with a standard error is the quantity *number of PSUs - number of strata.* Typically, this rule is applied to a design with two-PSU per stratum and when the variance components by stratum are roughly the same magnitude. The applicability of this rule depends upon the variable of interest and its interaction with the design structure (for additional information, see Chapter 5 of Korn and Graubard 1999). Given this rule of thumb, the number of degrees of freedom for the public use method described above is 339. The number of degrees of freedom is used to determine the *t*-statistic, its associated percentage points, p-values, standard error, and confidence intervals. As the number of degrees of freedom becomes large, the distribution of the *t*-statistic approaches the standard normal distribution. For example, with 120 degrees of freedom, the 97.5 percentage point of the $t_{120}$ distribution is 1.980, while the 97.5 percentage point of the standard normal distribution is 1.960. If a variable of interest is distributed across most of the NHIS PSUs, a normal distribution assumption may be adequate for analysis since the number of degrees of freedom would be large. The user should consult a mathematical statistician for further discussion.

### Subsetted Data Analysis

Frequently, studies using NHIS data are restricted to specific population subgroups, e.g., persons

aged 65 and older. Some users delete all records outside of the domain of interest (e.g., persons aged less than 65 years) in order to work with smaller data files and run computer jobs more quickly. This procedure of keeping only select records (and list-wise deleting other records) is called subsetting the data. With a subsetted dataset, which is appropriately weighted, correct point estimates (e.g., estimates of population subgroup means) can be produced. However, most software packages that analyze complex survey data incorrectly compute standard errors for subsetted data. When complex survey data are subsetted, oftentimes the sample design structure is compromised because the complete design information is not available; subsetting data deletes important design information needed for variance estimation. Note that SUDAAN has a SUBPOPN option that allows the targeting of a subpopulation while using the full (unsubsetted) data file which has all sample design information. (See a SUDAAN manual for more information).

***Strategy 1*** Use the MISSUNIT option on the NEST statement with the method described above for subsetted data:

***NEST*** **STRATUM PSU / MISSUNIT ;**

In a WR design with exactly two PSUs per stratum, when some PSUs are removed from the database through the listwise deletion of records outside the population of interest, the MISSUNIT option in SUDAAN "fixes" the estimation to produce standard errors identical to that achieved when using a full dataset with a SUBPOPN statement (see Strategy 2, below). Note that other calculations for design effects, degrees of freedom, and standardization may need to be carried out differently. Users are responsible for verifying the correctness of their results based on subsetted data.

***Strategy 2*** Use the SUBPOPN statement with the method described above for the full dataset:

***PROC ...*** **DESIGN = WR ;**
***NEST*** **STRATUM PSU ;**
***WEIGHT*** **WTFA ;**
***SUBGROUP*** **(variable names);**
***LEVELS ...*** **;**
***SUBPOPN*** **RACE=2 & SEX=2 / NAME="Analysis of African American women";**

Using the full dataset with the SUBPOPN statement in this example would constrain analysis to African American women only (RACE = 2 for black and SEX = 2 for female). Use of the SUBPOPN statement is equivalent to subsetting the dataset, except that any resulting variance estimates are based on the full design structure for the complete dataset.

## Combining Data Years in the National Health Interview Survey

It is sometimes possible to combine data from successive years of the National Health Interview Survey (NHIS) to increase the number of responses to questions and thus increase the precision of estimates. This is possible when the questions remain essentially the same over the years being combined.

Note that if data from the 1996 and 1997 NHIS (or any data before 1997 and data from 1997 and beyond) are combined, it is possible to obtain point estimates but not variance estimates, because the coded PSU identifications are not the same. This was done for confidentiality reasons.

Weights will normally need to be adjusted when combining data years. For example, if two years of NHIS data are combined, the sum of the weights will be about twice the size of the civilian noninstitutionalized population of the United States, so to achieve annualized results, each weight should be divided by two before analyzing the data.

A description of how to combine NHIS data years using SAS software follows. In SAS terminology, the process of adding observations is called concatenating data, or joining data sets one after the other, as opposed to merging data sets. The purpose of merging data is to add more variables for the same number of respondents, and this is done when data files within an NHIS data year are combined. Analysts wishing to do both—combine years and use data from multiple files for the same years—will need to both concatenate and merge data.

Below is a short explanation of the SAS ***SET*** command used to combine multiple years of NHIS data and an example of a program that will complete the task. The program is written to combine the data from the Person files of the 1999 NHIS and the 2000 NHIS. Note that the ***SET*** command does not subset a data file. That is, it is not used to sort and partition an analytic file by race, age, or any other variable that can be used to group respondents. If the research question being studied requires subsetting, please refer to the procedures and caveats discussed in the "Subsetted Data Analysis" section of this Appendix.

SAS ***SET*** Command

The ***SET*** statement tells the SAS system to read observations from one or more SAS data sets. An analyst will use the ***SET*** statement when she or he wants to concatenate or join observations from existing SAS data sets into a new data set. By default, the ***SET*** statement reads all of the observations from the input SAS data set. For example:

***DATA*** **output-SAS-data-set;**
***SET*** **input-SAS-data-set;**
***RUN;***

/* This program concatenates the 1999 NHIS and the 2000 NHIS SAS data sets for the Person file. */

***DATA*** **A;** /* SAS Data set */
***SET*** **PERSONSX;** /* The SET statement reads data from an existing SAS data set, e.g., the 1999 Person file. */
***KEEP*** **HHX FMX PX RAT_CAT AGE_P WTFA STRATUM PSU IHS OTHERGOV OTHERPUB MILITARY MEDICARE MCPART MCHMO MEDICAID MACHMD**

```
MAPCMD MAREF PRIVATE HITYPE1 PLNMGD1 HITYPE2 PLNMGD2 HITYPE3
PLNMGD3 HITYPE4 PLNMGD4 HISCOD_I HISPAN_I INCGRP CHIP;
/* The KEEP statement retains only the listed variables for processing. */
PROC SORT DATA=A; BY HHX FMX PX;  /*  Sorting SAS Data set A */

DATA B;  /* SAS Data set */
SET PERSONSX;  /* The SET statement reads data from an existing SAS data set, e.g., the
2000 Person file. */
KEEP  HHX FMX PX RAT_CAT AGE_P WTFA STRATUM PSU IHS OTHERGOV
OTHERPUB MILITARY MEDICARE MCPART MCHMO MEDICAID MACHMD
MAPCMD MAREF PRIVATE HITYPE1 PLNMGD1 HITYPE2 PLNMGD2 HITYPE3
PLNMGD3 HITYPE4 PLNMGD4 HISPCODR HISPANCR INCGRP CHIP;
/* The KEEP statement retains only the listed variables for processing. */
PROC SORT DATA=B; BY HHX FMX PX;  /* Sorting SAS Data set B */

DATA COMBO;  /*  New, combined SAS Data set */
SET A B ;  /*  Merging selected variables from 1999 and 2000 Data sets */
```

## References


Cochran, W.G. (1977), *Sampling techniques* (3rd ed), John Wiley & Sons.

Korn, E.L., and Graubard, B.I. (1999), *Analysis of Health Surveys*, John Wiley & Sons.

National Center for Health Statistics (1999), *National Health Interview Survey: Research for the 1995-2004 redesign*, Vital and Health Statistics, Series 2, No. 126.

National Center for Health Statistics (2000), *Design and Estimation for the National Health Interview Survey, 1995-2004*, Vital and Health Statistics, Series 2, No. 130.

Shah, B.V., Barnwell, B.G. and Bieler, G.S. (1997), *SUDAAN User's Manual; Release 7.5*, Research Triangle Institute, Research Triangle Park, NC.

# Appendix IV

## A Preliminary Evaluation and Recommendations for use of the Mental Health Indicator (MHI) in the NHIS for children ages 2 to 3

This is based on from a report by Thomas M.Achenbach, Ph.D., which was submitted to the Division of Health Interview Statistics on May 10, 1999**.**

## Introduction

The NHIS mental health indicators, MHIBOY2 and MHIGRL2 are located in the Child Health Status (CHS) section of the survey, and are based on items from the Child Behavior Checklist (CBCL) that were identified by Dr. Thomas Achenbach as providing the best discrimination between demographically similar children referred for mental health services versus nonreferred. To take account of gender and age differences in the discriminative power of particular items, the items were selected separately for each gender and age group. From the original ten items identified in Dr. Achenbach's 1995 analyses, the 1997 NHIS elected to include only 4 items (per gender). These include whether male sample children (aged 2-3 years) had been uncooperative, had trouble sleeping, had speech problems, or been unhappy or depressed in the past 2 months, and whether female sample children (aged 2-3 years) had temper tantrums, had speech problems, had been nervous or high-strung, or been unhappy or depressed in the past 2 months. Response categories included "Not true", "Sometimes true", or "Often true" (as well as "Refused" and "Don't know"). These items are also located in the CHS section (see CHS.321 and CHS.361).

It is essential to note that such a small set of items cannot be used to evaluate individual children for clinical or other purposes. Even for use as a mental health indicator in large surveys such as the NHIS, very small sets of items can serve only as approximate indicators of needs for mental health services. Multiple items tapping each of several specific areas of functioning would be needed to identify specific disorders, such as Attention Deficit Hyperactivity Disorder (ADHD), Depression, Conduct Disorder, and Somatization Disorder. (Note: The items for children ages 4 to 17 were replaced in the 2001 NHIS with a different instrument, the Strengths and Difficulties Questionnaire (SDQ). The SDQ is described in Appendix V.)

It should also be noted that different cutpoints on the distributions of item scores may be needed for different purposes. For example, a very low cutpoint may be useful if the goal is to identify every possible case for which mental health services might be considered. However, very low cutpoints result in relatively high false positive rates, i.e., the inclusion of substantial numbers of healthy individuals among those identified as potentially needing services. Conversely, higher cutpoints may yield greater overall accuracy in classifying potential cases versus noncases, but at the cost of missing more cases potentially needing services.

## Data Analyses

Dr. Achenbach specified and reviewed data analyses that were done at NCHS. These included tabulations of specific responses to each behavioral/emotional problem item; tabulations of relations between total problem scores and classification of children as deviant versus nondeviant on the basis of external criteria (e.g., parents ever being told by health professionals that their child had ADHD, mental retardation, other developmental delay, autism, down syndrome, or a learning disability; parents having talked to mental health professionals about their child in the preceding 12 months; or parents needing mental health services for their child but unable to afford it); and Relative Operating Characteristic (ROC) analyses of cutpoints on the total problem scores. Because each behavioral/emotional problem item was scored "0" (not true of the child), "1" (somewhat or sometimes true), or "2" (very true or often true), total scores across the 4 items for each gender/age group could range from "0" to "8". Dr. Achenbach examined the results and recommended changes and additions to the analyses.

Based on the analyses to date, Dr. Achenbach makes the following recommendations for boys and girls ages 2-3. Total scores on the 8 problem items are useful for quantitative analyses in relation to other variables. However, categorical mental health indicators should not be derived from specific cutpoints on the total scores for the 4 behavioral/emotional problem items on the basis of 1997 NHIS data for ages 2-3 for the following reasons:

The total number of children, 44 boys and 27 girls, classified as deviant according to external criteria (e.g., parents being told their child had ADHD; talking to mental health professionals about their child) was too small to provide a sound basis for establishing cutpoints;

Many disorders relevant to defining criterion groups (e.g., ADHD) are not identified as early as age 2-3;

The rates of referral for mental health services and other possible indicators of deviance are much lower at ages 2-3 than at older ages.

## Appendix V

## The Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) was developed and copyrighted by Dr. Robert Goodman, Institute of Psychiatry, London, England. It is used in the NHIS with Dr. Goodman's permission. The parent respondent version of the SDQ was added as a mental health supplement for children ages 4-17 as a part of a collaborative agreement between NCHS and the National Institute of Mental Health (NIMH). The first part of the SDQ consists of 25 scale items; the source questions for these items are located in the Child Access and Utilization (CAU) section of the 2001 NHIS (please refer to items CAU.345.010 through CAU.345.050). These items can be divided into five subscales measuring the following psychological attributes or dimensions:

- emotional symptoms;
- conduct problems;
- hyperactive behavior;
- peer relationships;
- prosocial behavior.

The second part of the SDQ, the extended questions (items CAU.345.060 through CAU.345.100), obtains additional information about the duration and impact of symptoms, which can be useful for determining psychiatric caseness. Scoring of the SDQ is described below.

The SDQ has been validated in studies done in Britain, and studies are currently being planned to evaluate the SDQ among U. S. adolescents aged 15-17 years. Validation studies among U. S. children aged 4-14 years are not yet available. Further information on the SDQ, including information on the teacher and self-respondent versions of the SDQ, is available at the web site: **<www.sdqinfo.com>.**

## Scoring of the SDQ

The five subscales described above can be scored separately to look at specific psychological problems, or the items in the first four subscales can be added for an overall score of 0-40. The scoring used has been:

| **Emotional symptoms:** | **Not true** | **Somewhat true** | **Definitely true** |
|---|---|---|---|
| Often complains of headaches, stomach-aches, or sickness | 0 | 1 | 2 |
| Many worries, often seems worried | 0 | 1 | 2 |
| Often unhappy, depressed or tearful | 0 | 1 | 2 |
| Nervous or clingy in new situations | 0 | 1 | 2 |
| Many fears, easily scared | 0 | 1 | 2 |

| **Conduct problems:** | **Not true** | **Somewhat true** | **Definitely true** |
|---|---|---|---|
| Often has temper tantrums or a hot temper | 0 | 1 | 2 |
| Generally obedient, usually does what parents want | 2 | 1 | 0 |
| Often fights with other children or bullies | 0 | 1 | 2 |
| Often lies or cheats | 0 | 1 | 2 |
| Steals from home, school, or elsewhere | 0 | 1 | 2 |
| **Hyperactivity behavior:** | | | |
| Restless, overactive, cannot stay still for long | 0 | 1 | 2 |
| Constantly fidgeting or squirming | 0 | 1 | 2 |
| Easily distracted, concentration wanders | 0 | 1 | 2 |
| Thinks things out before acting | 2 | 1 | 0 |
| Sees task through to the end, good attention span | 2 | 1 | 0 |
| **Peer relationships:** | | | |
| Rather solitary, tends to play alone | 0 | 1 | 2 |
| Has at least one good friend | 2 | 1 | 0 |
| Generally liked by other children | 2 | 1 | 0 |
| Picked on or bullied by other children | 0 | 1 | 2 |
| Gets on better with adults than other children | 0 | 1 | 2 |

The fifth subscale, the Prosocial subscale, describes children's *positive* behaviors. These items are excluded from the overall SDQ scoring when using the SDQ to identify children with psychological problems. The Prosocial subscale can be scored as follows:

| **Prosocial behavior:** | **Not true** | **Somewhat true** | **Definitely true** |
|---|---|---|---|
| Considerate of other people's feelings | 0 | 1 | 2 |
| Shares readily with other children | 0 | 1 | 2 |
| Helpful if someone is hurt, upset, or feeling ill | 0 | 1 | 2 |
| Kind to younger children | 0 | 1 | 2 |
| Often volunteers to help others | 0 | 1 | 2 |

## Additional References on the SDQ

Goodman, R. (1997) The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*. 38: 581-586.

Goodman, R. (1999) The Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *Journal of Child Psychology and Psychiatry*. 40 (5): 791-799.

Goodman, R. and Scott S. (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*,. 27(1): 17-24.