2004 National Health Interview Survey (NHIS)
Public Use Data Release

NHIS Survey Description

Division of Health Interview Statistics
National Center for Health Statistics
Hyattsville, Maryland

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

July, 2005

2004 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 2004 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.

Data users are encouraged to join the NHIS Listserv, an electronic mail list. The listserv is made up of over 4,000 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. To join, scroll down to "Related Links" on the NHIS Web page, and then click on "NHIS Listserv."

The Division of Health Interview Statistics also provides support services to data users. Users may contact us at 301-458-4901, or send e-mail to us at nhislist@cdc.gov.

What's New in 2004

In an effort to release data as quickly as possible, DHIS began reengineering the NHIS instrument and data production process in 2001. The new NHIS instrument, which utilizes Windows-based computer-assisted interviewing software called Blaise, was implemented for the 2004 survey year. In addition to the new interviewing software, data production processes were streamlined: all unnecessary or redundant processing was eliminated, and new methods were used to create the documentation files. These changes shortened the release time by 5 months.

Users will find many changes in the organization and content of the 2004 NHIS data and documentation files. Each of the 2004 ASCII data files (i.e., the Household, Family, Person, Injury/Poisoning Episode, Injury/Poisoning Verbatim Episode, Sample Child, and Sample Adult files) is accompanied by an associated Variable Summary Report, a Variable Layout Report, a Variable Frequency Report, sample SAS statements, sample SPSS statements, and sample Stata statements for that file.

It is especially important that users read all relevant documentation because the reengineering process resulted in changes to the organization of the NHIS data, variable names, variable numbers, and variable locations. Some of the major changes are described below:

- Prior to 2004, persons were numbered within the household; beginning in 2004, persons are numbered within the family, and then families are numbered within the household. This change requires new programming code to select specific respondents.

- Unlike previous survey years, variables are not repeated on data files in the 2004 NHIS. As a result, users will find it necessary to merge the 2004 files in order to analyze the data. Each data file contains household, family, and person numbers that make merging the 2004 files possible. Appendix VII provides sample code for merging the files.

- Although the "In universe" descriptions are included in the Variable Layout Report, the "Blank- Not in universe" label is no longer included as a response category in data files.

- The count of "Not in universe" cases is shown in the Variable Frequency Reports and is labeled "Frequency missing;" it will be shown if a variable has cases that are "Not in universe" or cases whose values fall out of valid range.

- All response categories with a zero count will be omitted from the Variable Frequency Report. Users must refer to the questionnaire or the Variable Layout Report to see a complete list of response categories for each question.

- The Child Immunization Section (CIM) section has been dropped from the NHIS Core because of cessation of funding.

- The Injury/Poisoning Episode file, and the Injury/Poisoning Episode Verbatim file and associated documentation will be released separately at a later date.

2004 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 "Health Insurance," 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 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 Child Core and the Sample Adult 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 Child and Sample Adult data 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 and poisonings . Because these files contain the same or comparable variables from one survey year to the next, they are suitable for trend 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 responsible for the NHIS are career Civil Service employees who 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 members of the household 18 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. Information for the Sample Child questionnaire is obtained from a knowledgeable adult 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 that section unless he/she is physically or mentally unable to do so. In these cases (200-300 per year) a knowledgeable proxy is allowed to answer for the sample adult.

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 the screen and enter responses directly into the computer during the interviews.

Sample Design

Traditionally, the sample for the NHIS is redesigned and redrawn about 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 not currently 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 follow back 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 individually 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/products/pubs/pubd/series/sr02/130-121/130-121.htm**>. (Users
may also be interested in another Series 2 (Number 126) report, *National Health Interview
Survey: Research for the 1995-2004 Redesign*, which is also available at the above address.)

Sample Reductions in the 2004 National Health Interview Survey

As in 2002 and 2003, the 2004 National Health Interview Survey (NHIS) was faced with
a budget shortfall during Fiscal Year 2004 and possibly beyond. As a result, NCHS and the
Division of Health Interview Statistics (DHIS) decided to reduce the size of the 2004 NHIS
sample. The goal of the 2004 sample cuts was strictly monetary savings. Since grouping the
sample cuts into consecutive weeks yields the greatest monetary savings, the NHIS sample was
cut by three consecutive weeks in May 2004. Therefore, no new interviews were conducted
during 3 of the normal 50 weeks of interviewing.

Overall, about 6% of the households in the 2004 NHIS sample were deleted from
interviewers' assignments. The total number of households deleted was 4,504 out of the 77,019
households originally selected for interviewing. Note that these totals do not represent the
number of completed household interviews. The table below contains the dates and numbers of
households that were cut.

Date	Households Cut
May 3-9, 2004	1,515
May 10-16, 2004	1,518
May 17-23, 2004	1,471

NCHS studied the statistical impact of removing cases from the NHIS sample before actually
cutting the sample. The evaluation showed that the impact of the sample cuts on the variance of
key estimates was negligible.

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-Level file's basic weights, the person
weights are further modified by adjusting them to Census control totals for sex, age, and
race/ethnicity populations (post-stratification). Beginning in 2003, NCHS made the transition
to weights derived from the 2000-Census-based population estimates. See Appendix VI for
more detailed information.

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.

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.

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.

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, responding or nonresponding.

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

NOTE: Analysts should be aware that 345 persons who were active duty members of the Armed Forces at time of interview are on the Person-Level file, despite the fact that NHIS covers only the civilian noninstitutionalized household population. These active duty members of the Armed Forces are included in that file because at least one other family member is a civilian eligible for the survey. The value of the final annual person weight (WTFA) for these military 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 experienced by a person 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 to unweighted data 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 a software package of SUDAAN[R] (Shah et. al. 1997) for NHIS variance estimation. SUDAAN[R] utilizes 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 2004 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 2004 Data

The interviewed sample for 2004 consisted of 36,579 households, which yielded 94,460 persons in 37,466 families. The interviewed sample for the Sample Child component, by proxy response from a knowledgeable adult in the family, was 12,424 children under 18 years of age. Data were not collected on any infant who was born during the assignment week of the interview. The interviewed sample for the Sample Adult component, which required self-response to all questions unless the sample adult was physically or mentally unable to do so, was 31,326 persons 18 years of age and older. There were 342 cases where a knowledgeable proxy answered for the sample adult.

The total household response rate was approximately 86.9%: 8.0 percentage points of the noninterview rate (13.1%) were the result of respondent refusal and unacceptable partial interviews. The remaining 5.1 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 Child component was 91.8%, which was calculated by dividing the number of completed Sample Child interviews (12,424) by the total number of eligible sample children (13,538). 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 86.5%, yielding a rate of 79.4%.

The conditional response rate for the Sample Adult component was 83.8% of persons identified as sample adults. The final response rate for the Adult Sample Person component was calculated as (Overall Family Response Rate) (Sample Adult Response Rate), or (86.5%) (83.8%) = 72.5%.

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

Information about the 2004 Data File Documentation

As with previous data years, questionnaires, datasets, and related documentation for each data file is available on the NHIS Web site, **<http://www.cdc.gov/nchs/nhis.htm>**. The Web site provides the Survey Description Document; a Readme File containing a summary of data access instructions; Notices for Data Users; a log of release history and, if necessary, new notices about data problems or changes; Survey Questionnaires, Flashcards, the Field Representative Manual, and Survey Flowchart; Information on Co-Sponsors and Supplements; Race/Ethnicity Information; the Data Release with links to a page that contains the Family, Household, Person, Injury/Poisoning Episode, Injury/Poisoning Verbatim Episode, Sample Child, and Sample Adult Files; Imputed Income Files; and Summary Health Statistics reports (when available).

Beginning in 2004 (and continuing in subsequent years), the organization and format of these data files have changed. Each of the 2004 data release categories for Household, Family, Person, Injury/Poisoning Episode, Injury/Poisoning Verbatim Episode, Sample Child, and Sample Adult Files will include following documents. A description of each type of document follows:

- Variable Summary Report
- Variable Layout Report
- Variable Frequency Report,
- ASCII data
- Sample SAS statements
- Sample SPSS statements
- Sample STATA statements

The Variable Summary Report lists each variable, a brief description of the variable, the question number on which it was based, and variable location in the released ASCII file. For most variables, the Variable Layout Report provides the actual question that generated the data, questionnaire location information, instrument variable name, universe, response values, and response value labels. Additional specific information is provided under "Sources," "Recodes," "Keywords," and "Notes." These terms are defined below:

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, collapsing, or verbatim coding of another variable, such as the family income recode (INCGRP) found in the Family 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

recode variables, then those recode variables are listed as a cross reference. 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, followed by the type of time unit. During the course of data editing, this information is standardized into a single appropriate time unit. 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 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 the same topic.

The universe refers to those respondents deemed eligible to answer a given question. For example, the universes for most Sample Adult variables are specified as ASTATFLG = 1 and (AGE GE '018' and AGE not IN ('997','999')), 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 questionnaire (constituting a completed interview or an acceptable partial interview). Sample adults who are not eligible to answer a given question are considered to be not-in-universe. For example, a sample adult who reported that he did not have surgery in the past 12 months (ASRGYR=2), would not be eligible for a follow-up question (ASRGNOYR) about the number of times that he had surgery in the past 12 months. **It is important to note that for all data files, persons who are not-in-universe are no longer listed in the Variable Layout Report response categories as "Blank- Not-in-universe."** If a respondent discontinued the interview anytime after completing the first three sections of the Sample Adult component, his or her responses will appear as 8's (not ascertained) for the remaining variables in the Sample Adult file where the universes are applicable.

The universes for most Sample Child file variables are specified as CSTATFLG = 1 and (AGE LE '017' and AGE NE ' '), 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 of the CHS section, or about half the questions of the Sample Child Core). Again, responses from acceptable partial interviews have a code of 8, meaning "not ascertained," throughout the remaining, unanswered Sample Child sections where the universes are applicable.

The Variable Frequency Report provides the frequencies, percents, and the frequency missing (not-in-universe) for each variable. **It is important to note that all response categories with a zero count are omitted from the Variable Frequency Report. Users must refer to the questionnaire or the Variable Layout Report to see a complete list of response codes for each question. In addition, the "frequency missing" label will be shown if a variable has not-in-universe cases or cases whose values fall out of range.** For example, if all sample adults are asked about a usual place for medical care when sick (AUSUALPL), then the "frequency missing" label is not shown.

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).

In addition, statistical noise at both the variable level and record level may have been added to allow for the protection of respondent confidentiality, and, at the same time, allow for release of files with as many variables as possible.

It is also important to note that variables are generally not repeated on multiple data files as was done in previous years. As a result, users will find it necessary to merge the 2004 files in order to analyze the data. Appendix VII provides sample code for merging the files.

Information about the 2004 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 2004 NHIS Field Representative's Manual (both the questionnaire 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. Users should match the question number in the instrument to the variable number in the File Layout Report to 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. Instrument problems were identified over the course of the year, and efforts were made to correct these errors. Some of these problems were resolved through correction of skip patterns, question wording changes, addition of questions, or other internal instrument corrections.

    When errors were detected and diagnosed, and time permitted, the instrument was changed to correct for the errors. In 2004, 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 Variable Layout Report for important information pertaining to specific variables that may have changed across quarters.

Questionnaire Sections

    The 2004 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. Multiple-part questions have an extension added to their three-digit acronym. For example, the first item in the FHS section is identified as FHS.010_00.000; note that FHS.010_00.000 also has an associated variable name, PLAPLYLM. The following list details the various questionnaire sections, their acronyms and description titles.

Table 1. 2004 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/Poisoning
IV	FAU	Health Care Access and Utilization
V	FHI	Health Insurance
VI	FSD	Socio-demographic
VII	FIN	Income and Assets

C. 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	CMH	Child Mental Health

D. Sample Adult Core

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

E. Recontact

Section No.	Section Code	Description
I	REC	Recontact Information and Follow-up

In addition to the three Core sections comprising the Basic Module, the 2004 NHIS contains several other data files: the Household- and Family-level files, the Injury/Poisoning Episode file, and the Injury/Poisoning Verbatim Episode 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/Poisoning files are derived from information obtained from the injury/poisoning questions in the Family Core section.

Supplements, Supplement Co-Sponsoring Agencies, and Question Locations, 2004 NHIS

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

        The 2004 NHIS supplements are in the Core files. A supplement or one or more supplementary questions may be interwoven among core questions, or may be placed at the end of a core section. The existence of three extra digits (.xxx) at the end of the question number helps to identify supplementary questions. A chart of all 2004 co-sponsored supplements and their question numbers is below. A list of changes to existing questions in the NHIS Core is located in Appendix VIII. All 2004 supplements are released in the same files as other Core data.

        In addition, users can obtain information about co-sponsored supplements from 1997-2005 on our Web site: **< http://www.cdc.gov/nchs/about/major/nhis/co-sponsors.htm >**

Topic	Co-sponsoring Agency	Title	Survey Section/Number
Mental Health	National Institute of Mental Health (NIMH) [1]	Child Mental Health	FHS.065; CAU.265; CMH.010_01– CMH.100
Immunization	National Immunization Program (NIP)[2]	Adult Immunization	AAU.310 - AAU.380
Agencies Providing General Support			
n/a	Center for Mental Health Services (CMHS)	Collection and analysis of mental health data using NHIS	
n/a	Agency for Healthcare Research and Quality (AHRQ)	Use of 2 NHIS sample panels to support the MEPS	
n/a	National Center for Health Statistics[2]	Cell phone usage	

[1] National Institutes of Health (NIH)
[2] Centers for Disease Control and Prevention (CDC)

2004 National Health Interview Survey
Household-Level File

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. Each record on the Household file represents a unique household included in the NHIS sample or sampling frame. Each household has a unique unit number (HHX). This unique unit number is needed for merging data files.

Some of the variables found only in this file include: the nature/reason for "Type A" non-responses 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 42,089 households: 36,579 households were interviewed, while 5,510 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 13.1% 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 86.9%.

2004 National Health Interview Survey
Family-Level File

The Family-Level file contains variables that describe characteristics of the 37,466 families living in households that participated in the 2004 NHIS. A family is defined as an individual or 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. 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 2004 survey; this is specified as FM = ALL in the Family-Level file Variable Layout. Note that multiple families may share one household. Users wishing to determine the number of responding and non-responding families in each household are referred to ACPT_FAM and REJ_FAM in the Household files or HHX and FMX in the Family file.

In 2004, there were 24 ineligible persons (type B non-interview) in 20 multi-family households who were not excluded in the creation of the rejected person and rejected family variables (REJ_FAM, REJ_PER). The family response rate has been adjusted in the Survey Description Document to reflect the exclusion of these 24 ineligible persons, but they have not been excluded from the 2004 household data file. Users should exercise caution when using the above variables in their analyses.

As Table 2 indicates, 98% of NHIS households consist of one family. All relationships in the household are recorded relative to a household reference person, who is generally the person who owns or rents the housing unit. 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, 2004 NHIS (unweighted counts)

Families per household	Frequency	Percent
1	35,894	98.1
2	551	1.5
3	93	0.3
4	24	0.1
5	11	0.0
6	3	0.0
7	2	0.0
8	1	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 *and* one family reference person is designated 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 37,466 families by number of family members.

Table 3. Size of Family, 2004 NHIS (unweighted counts)

Number of Members	Frequency	Percent
1	11,189	29.9
2	11,287	30.1
3	5,796	15.5
4	5,143	13.7
5	2,545	6.8
6	959	2.6
7	301	0.8
8	145	0.4
9	65	0.2
10	15	0.0
11	10	0.0
12	4	0.0
13	3	0.0
14	2	0.0
15	1	0.0
16	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 month and year, 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 five Family Core sections of the NHIS that collapse the 90,000+ individual level observations into information about their respective families. (Starting in 2004, recodes from the FHI section of the Family Core are no longer included in the Family file.)

Generally, the Family file consists of two types of recodes. The first is a simple "yes-no" measure that indicates whether *any* family member falls into a particular category or exhibits a particular characteristic. 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, no frequencies will be shown 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, and thus, are not shown. 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.500) 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: 15,266 families have one member receiving income from wages/salary, 10,091 families have two members (or 2(10,091) = 20,182 persons) with wage/salary income, 1,632 families have three members (or 3(1,632) = 4,896 persons), 356 families have four members (or 1,424 persons), 66 families have five members (or 330 persons), 16 families have six members (96 persons), and 2 families have seven members (14 persons) with wage/salary income in 2004. Thus, the sum of persons across the 27,429 families answering "yes" to FSALYN, the associated yes-no indicator, is 42,208 (15,266 + 20,182 + 4,896 + 1,424 + 330 + 96 + 14), which is equal to the 42,208 "yes" responses to the person-level source variable, PSAL. Users are advised to check the Variable Layout Report for each Family file recode in order to determine its person-level source variable.

Family Structure Variables

The 2004 NHIS Family file contains two variables describing family type and structure in both general and detailed terms. FM_TYPE (formerly called FMTYPE) consists of just four categories, and represents an initial classification of families according to the numbers of adults and children that are present. In addition, FM_STRP (formerly called FMSTR2) categorizes families according to familial relationships and, when children are present, parental marital status. FM_STRP is identical to FMSTR2, which was included on the 1999-2003 NHIS Family files. Users will note that 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 FM_STRP, families that could not be classified are coded "99." Emancipated minors are treated as adults with respect to FM_TYPE and FM_STRP, despite the fact that they may be under 18 years of age. Also, note that FMSTRCT2 is not available in the 2004 NHIS.

The Family File Weight

The ideal situation for creating weights for the Family file would be to use independent estimates of the number of 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 2004 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 the 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 2004 NHIS Family file, WTFA_FAM, corresponds to the 2004 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.

2004 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 2004 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 about 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 the Family-Level File for more information.) The six sections comprising the Family Core are discussed in greater detail below.

Important Note

Users should be aware that the family structure variables FM_TYPE and FM_STRP that were previously located on the Person and Sample Child data files are now located exclusively on the Family file. The substantive meaning and response categories of these variables are unchanged, and they retain the same variable names (FM_TYPE and FM_STRP). Note that FM_STRCP is not available in the 2004 NHIS.

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

The Health Status and Limitation of Activity (FHS) section of the Family Core for the 2004 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 in 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, and getting around inside the home; and the need for personal assistance with handling routine needs such as everyday household chores, doing necessary business, and 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 such condition.

Since cognitive impairment is increasingly recognized as a source of activity limitations among older adults, the FHS section includes an indicator that identifies family members who are 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 or limitations related to specific personal care needs. In addition, the section contains information about children who receive special education or early intervention services. Information regarding limitations in play activities is also collected for young children.

The 2004 FHS time variables and recodes, which indicate how long respondents have had the condition(s) causing their limitation(s), were processed using procedures similar to those used in 2002 and 2003. Substantively, the 2002, 2003 and 2004 variables and recodes are similar to those from previous years (1997-2001), but the 2002-2004 data contain fewer unknown categories than previously, that is, detailed unknown categories were collapsed into broader categories.

Conditions

For each family member with a previously mentioned limitation, the respondent was 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 were then handed one of two flash cards listing various condition categories. These categories are broad in scope, and vary according to age. Information about family members under age 18 was solicited for 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" (if the family member was limited by a condition not listed in one of the fixed categories). Respondents could supply a 50-character verbatim response for one or both of the "other impairment problem" categories.

The fixed response categories in the instrument for adults age 18 or older were equally broad, and comprised 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," and "weight problem." Starting in 2001 and continuing in 2004, if an adult family member was limited by a condition not listed in one of these 18 fixed categories, the interviewer entered "M" for "More conditions," and a second screen containing 17 additional condition categories and two "other impairment problem" categories appeared on the interviewer's laptop computer screen. These conditions were not read aloud to respondents, but if the respondent said a family member's condition was limited by one of these 17 conditions, the interviewer recorded this information. If the family member was limited by a condition not listed on either the second flash card or on the interviewer's computer screen, then the interviewer entered a 50-character verbatim response for one or both of the "other impairment problem" categories. Respondents could list any number of applicable conditions.

During data processing, the verbatim responses recorded by interviewers were reviewed to determine if any responses could be back-coded to one of the 13 fixed categories for respondents under age 18, or to one of the 18 fixed categories for adult respondents. If so, these "other" responses were assigned to the appropriate response categories (the first 13 for children, and the first 18 for adults). For adults, an additional 16 *ad hoc* categories were created during data processing to categorize responses that fell outside the fixed 18 condition categories included in the instrument: these *ad hoc* categories were assigned numbers 19_ through 34_. (Note: Due to a

naming convention error in 2002 and 2003 these same *ad hoc* categories were assigned numbers 19 through 34 without an underscore.) In addition, responses in the 17 general categories seen only by the interviewer were also back-coded and categorized into 8 of the *ad hoc* categories. The resulting 36 categories for adults and 13 categories for children were based on 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).

Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the *ad hoc* categories (for adult respondents) remained in the "other impairment problem" categories, and were renumbered "90" and, if necessary, "91" for both children and adults. The specific condition categories as well as the "other impairment problem" 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). Note that the verbatim responses associated with the "other impairment problem" categories are not included as a separate field on the public use file. 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 underestimates. Therefore, these variables should be analyzed with care. Moreover, none of the FHS condition variables (the 13 child variables, LAHCC1 through LAHCC13, 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. Conditions are considered chronic for children less than one year of age who have had a condition "since birth." **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 prevalence rates of chronic conditions.**

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

A. Codes for Adults (ages 18 or more years)

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** *Injury with specific mention of bone or joints*	800-848, 850-999
6 - **Other injury** *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** *Excludes Down's syndrome and microcephalus*	740-742.0, 742.2-744, 747-757.9, 758.1-759
14 - **Mental retardation** *Includes Down's syndrome and microcephalus*	317-319, 742.1, 758.0
15 - **Other developmental problem** *Includes learning disabilities*	315, 343, 783.4
16 - **Senility (and other cognitive problems)**	290

17 - **Depression, anxiety or emotional problem**
    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** *Indicates a problem with being overweight or obese*	
19 - **Missing limbs (any part) / amputee** *Indicates loss of a limb or digit*	

20 - **Other musculoskeletal system conditions**
    Diseases of the musculoskeletal system and connective
    tissue not coded to 3, 4, 5
	710-739

A. Codes for Adults (ages 18 or more years)

NHIS Category	ICD-9-CM Codes
21 - Other circulatory system conditions *Any diseases of the circulatory system not coded to 7, 8, 9*	390-459
22 - Other endocrine system, etc. conditions *Any Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders not coded to 10 or 18*	240-279
23 - Other Nervous system conditions *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 *Any mention of "tumor" without cancer, malignancy, etc.*	210-239
29 - Alcohol & drug related problems *Any mention of "alcohol," "drugs" (or specific drug types), or substance abuse*	291-292, 303-305
30 - Other mental conditions *Any mental disorders not coded to 14 or 15 or 17*	293-299
31 - After effects of surgery or other medical treatment *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 *Any mention of age as the only specified cause*	
33 - Fatigue/Tiredness *Any mention of tiredness, stiffness, or weakness without referring to any specific part of the body*	
34 - Pregnancy related conditions *Any mention of "pregnancy" or "childbirth"*	
90 - Others Not Elsewhere Classified *1st other-specify verbatim, not elsewhere classified*	
91 - Others Not Elsewhere Classified *2nd other-specify verbatim, not elsewhere classified*	

B. Codes for Children (*ages under 18 years*)

NHIS Category	ICD-9-CM Codes
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 *Excludes Down's syndrome and microcephalus*	740-742.0, 742.2-757.9, 758.1-759
6 – Injury	800-999
7 - Mental retardation *Includes Down's syndrome and microcephalus*	317-319, 742.1, 758.0
8 - Other developmental problem	343, 783.4
9 - Other mental, emotional, or behavioral problem	290-313, 799.2, V15.4
10 - Bone, joint or muscle problem	710-739
11 - Epilepsy and seizures	345, 779.0, 780.3
12 - Learning disability	315
13 - Attention Deficit/Hyperactive Disorder (ADD/ADHD)	314
90 - Others Not Elsewhere Classified *1st other-specify verbatim that does not fit in any other category*	
91 - Others Not Elsewhere Classified *2nd other-specify verbatim that does not fit in any other category*	

Technical Notes

The condition variable LAHCA31_ includes any causal condition that specifically mentioned "surgery" or "operation," or otherwise indicates a medical treatment as the causal condition (either ongoing or occurring within the last year). The condition variable LAHCA33_ includes any causal condition that specifically and solely mentioned "fatigue," "weakness," "lack of strength," "tiredness," "exhaustion," etc. without reference to any particular part of the body. Lastly, the condition variable LAHCA34_ includes any causal condition that specifically and solely mentioned "pregnancy," "pregnant," or "childbirth."

II.   Health Care Access and Utilization Section (FAU)

The Health Care Access and Utilization (FAU) section of the Family Core of the 2004 NHIS has remained largely unchanged since 1997. The FAU section contains information addressing access to health care, utilization services, and health care contacts.

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.

 Technical Notes

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

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 2004 documentation to documentation from the 2003 (and earlier) NHIS for any other changes that may have occurred over time to the variables in this section.

III. Health Insurance Section (FHI) Health Insurance Section

The Health Insurance section of the 2004 NHIS Family Core has a full range of data items addressing health insurance. In 2004 this section was restructured to address changes in health care coverage and to simplify respondents' initial classification of health insurance coverage type. Questions were also added to address Medicare and private prescription drug benefits, to improve classification of military health coverage and to obtain managed care information for all types of public health programs. However, the flow of the questions pertaining to health insurance programs covered by this section is similar to the 1993-96 NHIS Health Insurance Supplements and the 1997-2003 NHIS Family Core.

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

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

Managed care arrangement and the need for referrals for those covered by Medicare, Medicaid, SCHIP, other State-sponsored health plans and other government programs;

Medicare managed care model types;

Enrollment in a Medicare Prescription Drug Discount Care Endorsement Program;

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, prescription drug benefit;

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

Types of single service plans;

Type of TRICARE coverage;

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 2004, FHI data contain several modifications, as well as some new variables. The HIKIND list was shortened from 14 categories to 11. Private health insurance was combined into one category, HIKINDA, and military health coverage was combined into one category, HIKINDF. To increase the counts of single service coverage, SINCOV was added following the HIKIND question for persons who had not indicated earlier that they have a single service plan. A person who responded to either HIKINDJ or SINCOV received the single service detail questions. Response categories were changed in the PLNWRK question to get better precision as to how a private health plan was obtained. This detail is contained in PLNWRKN1 and PLNWRKN2. An additional question was added to the private plan detail to monitor the impact of the Medicare prescription drug benefit on private plan drug benefits. This information is contained in PRRXCOV1 and PRRXCOV2. Detailed information concerning the third and fourth plans for a respondent is no longer available on the public use data file. Persons with three or more plans have a "yes" response to the PRPLPLUS variable. Detailed information on the third and fourth plan for a respondent is still available through the Research Data Center.

Details on type of military coverage are now contained in the new variables MILSPC1, MILSPC2, MILSPC3, MILSPC4, and MILMAN. An additional question, MCRXCARD, was added to the Medicare section to capture enrollment in the Medicare Prescription Drug Discount Care Endorsement Program. The wording on the MCCHOICE question was changed to address the new name for Medicare Plus Choice, which is Medicare Advantage. Follow-up questions were added regarding the State Children's Health Insurance Program (SCHIP), State-sponsored and other public programs (OTHERPUB), and other government programs (OTHERGOV) to obtain managed care information for all types of public coverage.

Beginning with quarter 3, 2004 two new questions were added to reduce potential errors in reporting Medicare and Medicaid status. Persons 65 years and over not reporting Medicare coverage were asked explicitly about Medicare coverage in MCAREPRB. Persons under 65 with

no reported coverage were asked explicitly about Medicaid coverage in MCAIDPRB. Respondents who were reclassified as covered by either of these additional questions received the appropriate follow-up questions.

Technical Notes

Analysts are strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, SCHIP, IHS, MILITARY, OTHERPUB, OTHERGOV, and SINGLE for types of health care coverage because these recodes take into account the complicated editing process that takes place in the FHI section. The variables HILAST and HINOTYR, which reflect periods of noncoverage, 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, SCHIP, IHS, MILITARY, OTHERPUB, and OTHERGOV). Using the most conservative estimate of the uninsured (which would exclude persons with IHS coverage only), a total of 672 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 1268 respondents were not asked either the HILAST or the HINOTYR questions.

It was determined that some respondents indicated plans (in response to the questions HIPNAM1, HIPNAM2, HIPNAM3, and HIPNAM4) 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, SCHIP, IHS, MILITARY, OTHERPUB, OTHERGOV, and SINGLE. Similarly, in looking at the verbatim responses to the questions STNAME1, STNAME2 or STNAME3 that asks respondents for the name of their SCHIP, state sponsored or other government coverage respectively, 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 STNAME2 or STNAME3 fields to be covered by the State Children's Health Insurance Program were assigned to the SCHIP recode. 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 plans or programs that 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. Likewise, 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 of these respondents indicated plans or programs that were clearly private health insurance, Medicare, SCHIP, military coverage, Indian Health Service, single service plans, or no coverage at all. These respondents were also reassigned to the appropriate enrollment recodes for MEDICARE, CHIP, PRIVATE, IHS, MILITARY, and SINGLE. In addition, due to reengineering changes, there are no longer "99" response codes for the variables HITYPE1 and HITYPE2.

In addition, some respondents offering an "other" response to the survey item (HISTOPOT) 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, SCHIP, PRIVATE, IHS, MILITARY, OTHERPUB, and OTHERGOV. Analysts are therefore strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, SCHIP, 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-HIKINDK and MCAREPRB, MCAIDPRB, and SINCOV 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 noncoverage as used in *Health, United State,2004* (in which persons with *only* Indian Health Service coverage are considered uninsured).

IV. Socio-demographic Section (FSD)

The Socio-demographic (FSD) section of the Family Core in the 2004 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 2003, an estimate of their earnings from wages in 2003, and whether their employer provided health insurance. 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.

Important Changes for 2004

DOINGLWP and WHYNOWKP are the FSD equivalents of DOINGLWA and WHYNOWKA in the ASD section of the Sample Adult data file. For the majority of respondents, DOINGLWP and DOINGLWA will have identical values (and, likewise, WHYNOWKP and WHYNOWKA). However, it is nevertheless possible that DOINGLWP and DOINGLWA (and WHYNOWKP and WHYNOWKA) may have inconsistent values across the Person and Sample Adult data files. Users wishing to reconcile any discrepant values are advised to use the values of DOINGLWA and WHYNOWKA (rather than DOINGLWP and WHYNOWKP, respectively), since the information obtained from the family respondent during the FSD portion of the interview (and reflected in DOINGLWP and WHYNOWKP) was subsequently confirmed or corrected by the sample adult during his or her interview (as reflected in DOINGLWA and WHYNOWKA). Additionally, both DOINGLWP and WHYNOWKP are substantively equivalent to previous years' versions of these variables (i.e., DOINGLW1 and WHYNOWK1).

The 2004 FSD section contains a new variable called PLBORN, which is based on a question in the instrument that asked whether the respondent was born in the United States. If respondents replied affirmatively, they were asked the state in which they were born (PLBORN1). If respondents said they were not born in the U.S., they were asked the country in which they were born (PLBORN2). PLBORN1 and PLBORN2 are not included on the public use file for confidentiality reasons. However, the 2004 NHIS includes two public use recodes, GEOBRTH

and REGIONBR, that are based on this restricted birthplace information (as well as the new variable, PLBORN). GEOBRTH indicates geographic place of birth, and has three categories: born in one of the 50 United States or the District of Columbia; born in a U.S. territory; or not born in the U.S. or a U.S. territory. In order to make GEOBRTH comparable to previous recodes (for carrying out analyses on multiple years of NHIS data), users should collapse those respondents in the last two categories of GEOBRTH into a single category. This will result in a recode that is comparable to USBRTH_P from the 2000-2001 NHIS or USBORN_P from the 1997-1999 NHIS. The second recode, REGIONBR, 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 (for more information about the Factbook, users should refer to <**http://www.cia.gov/cia/publications/factbook/index.html**>). Note that respondents born in Canada were included in the "Elsewhere" category of REGIONBR in order to satisfy NCHS confidentiality requirements. Users are cautioned that neither GEOBRTH nor REGIONBR indicate legal status or citizenship.

Category	**Countries included**
United States	All persons born in one of the 50 United States or the District of Columbia
Mexico, Central America, Caribbean Islands	Mexico, 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)	Russia, Lithuania, Latvia, 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	Guam, 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 through the NCHS Research Data Center. For more information, refer to the Research Data Center Web page: < **http://www.cdc.gov/nchs/r&d/rdc.htm** >.

Respondents who were not born in one of the 50 United States or the District of Columbia 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 since they came to stay. This information was combined to create a recode that indicates how long these respondents have been living in the United States (YRSINUS). The 2004 data also contain a citizenship recode (CITIZENP) that distinguishes between U.S. citizens and non-citizens.

MAJR_ACT and EDUC_R1, two recodes included in previous years' data files, have been discontinued in the 2004 NHIS. In addition, PMILTRY replaces (and is substantively equivalent to) MILTRYDS, the variable indicating honorably discharged veterans of active duty service in the U.S. Armed Forces. Also, WRKHRS2 replaces WRKHRS, and EDUC1 replaces EDUC. WRKHRS2 and WRKHRS are equivalent variables: the 2004 instrument allowed respondents to report that they had worked up to 168 hours in the previous week, but all responses greater than or equal to 95 hours are topcoded at 95 (as was the case with WRKHRS, the 1997-2003 variable). **Users should note that EDUC1 and EDUC, the education variable in the 1997-2003 NHIS, are *not* equivalent variables: GED recipients and High School Graduates are coded as "13" and "14," respectively, on EDUC1 (whereas they were coded "14" and "13," respectively, for EDUC).**

V. 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 2003, 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).

+---+
| **Important Note** |
| |
| **Users should be aware that the family income variables INCGRP, RAT_CAT, and HOUSEOWN, which were previously located on the Person and/or Sample Child data files, are now located exclusively on the Family file. The substantive meaning and response categories of these variables are unchanged, and they retain the same variable names (INCGRP, RAT_CAT, and HOUSEOWN). Note that AB_BL20K is no longer available in the 2004 NHIS.** |
+---+

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 checking or savings accounts, Individual Retirement Accounts (IRAs) or certificates of deposit, money market funds, treasury notes, bonds, or any other 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 specifically mentions alimony, contributions from family or friends, Veteran's Administration (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 (2003).

Amounts and Home Ownership

In years prior to 1997, the NHIS obtained information about the amount of income received from each financial source, but that was dropped in the redesigned NHIS (1997 and beyond) in favor of a single overall estimate of combined family income. Unlike previous NHIS instruments, the redesigned instrument contains three questions to identify the family's combined income from all sources during the previous calendar year, including a question (FIN.250) that allows the respondent to supply a specific dollar amount (up to $999,995). Any family income responses greater than $999,995 are entered as $999,996. Respondents who do not know or refuse to give a dollar amount to this question are 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 answers this question, he/she is then given one of two flash cards and asked to indicate which income group listed on the card best represents the family's combined income during the previous calendar year (FIN.270). One flash card lists incomes that are $20,000 or more, and the other flash card lists incomes that are less than $20,000.

Data analysts should be aware of an unanticipated issue in 2004 related to the collection of exact amount income data (FIN.250). Specifically, a much larger than expected proportion of respondents seemingly reported a family income in the last calendar year of "$2." In 2004, 2,133

persons (2.25%) have a response of "$2" to the exact amount of family income question (FIN.250). By comparison, in 2002 (the most recent data year without sample cuts), 136 persons (0.15%) had a response of "$2" to the exact amount of family income question (FIN.250). To determine what could have caused the increase in the "$2" responses, we examined interview records and consulted with the Census Bureau. Based on this review, we determined that a majority of these "$2" responses were ones that should have been entered as "Refused" or "Don't know" to the exact amount of family income question (FIN.250). However, because a "$2" response to FIN.250 is considered a valid response, these respondents did not receive the follow-up questions (FIN.260, FIN.270). Therefore, in 2004, these 2,133 persons with a "$2" response to FIN.250 were coded as "Not ascertained" and will have their family income imputed when the Multiple Imputed Income files are released at a later date. Although FIN.250 is not part of the public-use data release, it is used to construct the combined income recode INCGRP, which uses 13 categories to describe the family's income.

Additionally, a more detailed indicator of poverty status was created by utilizing published information from the U.S. Census Bureau regarding 2003 poverty thresholds (see *Income, Poverty, and Health Insurance Coverage in the United States, 2003*). A ratio of the 2003 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 situations: for families who 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), and for 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, for RAT_CAT). Analysts should also note that RAT_CAT will also be affected by the relatively higher numbers (compared with prior years) of respondents with a "$2" response to the exact amount of family income (FIN.250) described in the previous paragraph.

Respondents were also asked whether the family's house or apartment was owned or being bought, rented, or occupied by some other arrangement (FIN.280). If the family was renting the current residence, a follow-up question (FGAH or FIN.282) asked whether the family was paying lower rent due to governmental rental assistance.

As mentioned previously, INCGRP, RAT_CAT, and HOUSEOWN have been moved from the Person file to the Family file, replacing the 1997-2003 Family file variables FINCGRP, FRAT_CAT, and FHOUSE, respectively. Analysts should also note that a second income recode (AB_BL20K), which was included on the 1997-2003 Person file, has been deleted from the NHIS public use files for 2004 because it can be created from INCGRP and is redundant.

In addition, while FGAH is situated in the FIN portion of the Family Core, the actual variable can be found on the 2004 Family file. FGAH does not have a 1997-2003 counterpart in the Family file.

Program Participation

Respondents were asked in the final part of the FIN section if any family members were authorized to receive food stamps in 2003, 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). 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.

Finally, the NHIS contains three person-level variables 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 or females 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 the previous calendar year (PWIC). An additional variable, WIC_FLAG, is also included in the Person file. WIC_FLAG indicates if persons who received WIC benefits were age-eligible for the WIC program.

Technical notes

As previously mentioned, the majority of the questions in the FIN section are structured to ask first whether any family member received the applicable income source and, if yes, then to determine which family members received the income source. This format also applies to other items in the section such as TANF, food stamps, and WIC benefits. As mentioned in the 2003 Survey Description Document, the 1997–2003 NHIS only allowed six persons per family to be indicated as receiving the income and/or program source. However, this problem has been corrected for 2004. Analysts interested in using these program participation variables for 1997–2003 should refer to the 2003 Survey Description Document for guidance. Further, since qualification for these programs is usually based on a family's economic circumstances, these program participation variables may have limited analytic value at the person level. Therefore, analysts may find more utility in using the corresponding variables from the Family file.

2004 National Health Interview Survey
Injury/Poisoning Episode Files

The Family Core portion of the 2004 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 will be created from these data: the Injury/Poisoning Episode file, and the Injury/Poisoning Verbatim Episode file. These files and documentation will be released separately at a later date.

2004 National Health Interview Survey
Sample Child File

The Sample Child section of the 2004 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 self-report; instead a knowledgeable adult (typically a parent or guardian) answers questions on the sample child's behalf.

In 2004 a new section, the Child Mental Health Supplement (CMH), was created to provide a set location for the full Strengths and Difficulties Questionnaire (SDQ). In addition, the Child Immunization Section (CIM) was dropped from the survey. The sections comprising the Sample Child File are discussed below.

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

The Child Conditions, Limitation of Activity and Health Status Section (CHS) section of the 2004 NHIS contains information on conditions, limitations of activity, health status, and mental health. The CHS 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 question about whether the sample child still has asthma is included. 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 2004, a behavioral screening questionnaire for children aged 4 to 17 years, the full version of the Strengths and Difficulties Questionnaire (SDQ), was transferred from the CAU section to a new section, Child Mental Health (CMH). However, other child mental health questions derived from the Child Behavior Checklist for children ages 2-3 years 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. The Division of Health Interview Statistics also provides support services to data users. Users may contact us at 301-458-4901, or send e-mail to us at nhislist@cdc.gov.

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 2004 NHIS contains information on access to health care, dental care, and health care provider contacts. The questions pertaining to access to health 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. A question on dental care asked about the length of time since last dental visit.

Questions regarding health care provider contacts include visits and telephone contacts 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 2003, the CAU section included a behavioral screening questionnaire for children aged 4 to 17 years. In 2004, this questionnaire, the full version of the Strengths and Difficulties Questionnaire (SDQ), was removed from the CAU section and placed in a separate section, Child Mental Health (CMH). More detailed information on the SDQ is provided in the CMH section below and in Appendix V.

III. Child Mental Health Supplement (CMH)

As part of a collaborative agreement with the National Institute of Mental Health (NIMH), the Strengths and Difficulties Questionnaire (SDQ) was first used in 2001 in a Child Mental Health Supplement in the CAU section. The SDQ is a behavioral screening questionnaire for children ages 4 to 17 years with extended questions that provide information on the duration of a child's problem and the impact that the problem has on the child and his/her family. It is copyrighted by Dr. Robert Goodman, London, England and is used with his permission. In 2002 the full SDQ was deleted from the CAU section, and a brief version of the SDQ was added to the CHS core.

In 2003 the brief version of the SDQ was dropped from the CHS section, and the full SDQ was reinserted into the CAU section. The six items from the brief SDQ in 2002 reverted to their original names and question numbers in the full SDQ in 2003 as follows: CSCL2_C2 in 2003 (CMHMF12 in 2002), CSCL2_E2 in 2003 (CMHMF13 in 2002), CSCL3_E3 in 2003 (CMHMF14 in 2002), CSCL5_P5 in 2003 (CMHMF15 in 2002), CSCL5_H5 in 2003 (CMHMF16 in 2002), and CSCL6 in 2003 (CMHDIFF in 2002).

In 2004, the full SDQ was transferred from the CAU section to a newly created section, Child Mental Health (CMH). The question/answer wording and the question order remain the same as in 2003. Variable names (except CSCL7) have been changed to accommodate the new editing system. The question numbers have also been changed to reflect the new question numbering system and the new section name. More detailed information on the SDQ is provided in Appendix V.

Note that other child mental health questions derived from the Child Behavior Checklist for children ages 2-3 years remain in the CHS section.

Important Note

The original numbering system of the response categories in the instrument has been modified in the Variable Layout Report for all variables in the CMH section with the exception of CSCL7. In order to correspond with the SDQ scoring system detailed in Appendix V, all variables with original answer codes of 1, 2, 3 in the instrument were changed to 0, 1, 2 in the data file, Variable Layout Report, and Variable Frequency Report; all variables with original answer codes of 1, 2, 3, 4 in the instrument were changed to 0, 1, 2, 3 in the data file, Variable Layout Report, and Variable Frequency Report.

2004 National Health Interview Survey
Sample Adult File

The Sample Adult section of the 2004 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 PROX1 indicates those cases where information was obtained from a proxy respondent. The eight sections comprising the Sample Adult section are discussed below.

I. Adult Demographics Section (ASD)

The Adult Socio-Demographics (ASD) section 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 were asked a series of questions about their job and work status during the week prior to the interview. In addition, 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 were asked if they had "ever held a job or worked at a business." 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 were not retired were asked about their most recently held job. In a subsequent question, currently employed sample adults were asked if their current job was also the job they had held for the longest time. Likewise, sample adults who had ever worked and were not retired were asked if their most recently held job was also the job they had held for the longest.

Additional questions in the ASD section ask sample adults to describe their current/most recent/longest-held employment situation (whether they were employed by 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. Respondents who indicated that they were self-employed at their current/most recent/longest-held job were asked whether they had an incorporated business. Currently employed sample adults were asked whether they were working at more than one job.

Important Changes for 2004

DOINGLWA and WHYNOWKA are the ASD equivalents of DOINGLWP and WHYNOWKP in the FSD section. For the majority of respondents, DOINGLWA and DOINGLWP will have identical values (and, likewise, WHYNOWKA and WHYNOWKP).

However, it is nevertheless possible that DOINGLWA and DOINGLWP (and WHYNOWKA and WHYNOWKP) may have inconsistent values across the Sample Adult and Person data files. Users wishing to reconcile any discrepant values are advised to use the values of DOINGLWA and WHYNOWKA (rather than DOINGLWP and WHYNOWKP, respectively), since the information obtained from the family respondent during the FSD portion of the interview (and reflected in DOINGLWP and WHYNOWKP) was subsequently confirmed and corrected by the sample adult during his or her interview (as reflected in DOINGLWA and WHYNOWKA).

With the exception of BUSINC1A, WRKLONGH, and ONEJOB, the universe for all variables in the 2004 ASD section includes currently employed and ever employed sample adults. Variables with smaller universes (e.g., currently employed sample adults only) that were included in previous years' data files have been eliminated in the 2004 NHIS. Users wishing to replicate those variables are advised to use DOINGLWA to identify the subset of currently employed sample adults (i.e., DOINGLWA = 1, 2, or 4).

Industry and Occupation Coding

During the course of the interview, verbatim responses were obtained from each eligible respondent regarding his/her industry and occupation. This information was subsequently reviewed by U.S. Census Bureau coding specialists, who assigned appropriate industry and occupation codes. These 3-digit codes, developed by U.S. Census Bureau staff for use in Federal surveys, are consistent with the structures of the Standard Industrial Classification (SIC) and Standard Occupation Classification (SOC) (but are not actual SIC and SOC codes). Prior to the 1997 NHIS, the codes were included on all NHIS public use data files. However, a review of NHIS data suggested that the level of detail contained in the codes could compromise respondent confidentiality. Consequently, beginning in 1997, the 3-digit codes were restricted to in-house NHIS data files, and DHIS staff created several 2-digit industry and occupation recodes that could be included on the public use data files. The latter recodes were based on occupation and industry groups and subgroups consistent with the existent SIC and SOC structures.

Changes in the U.S. economy led to changes in the SIC and SOC classifications. After an extensive period of review, the standard industry and occupation classifications – and the corresponding 3-digit Census codes used by the NHIS and other Federal surveys – were replaced by the North American Industrial Classification System (NAICS) and a revamped SOC. Accordingly, the Census Bureau has developed new 4-digit industry and occupation codes to replace the obsolete 3-digit codes; the former are consistent with the 1997 NAICS and 2000 SOC. Because the 2000 SOC is considered compatible with the 1997 NAICS, it is referred to below as "New SOC."

Federal surveys are switching from the earlier classification systems to NAICS/New SOC. As a result, the 2004 NHIS in-house data files contain *both* the 3- and 4-digit Census codes for industry and occupation: the 3-digit codes are consistent with the 1987 SIC and SOC systems, while the 4-digit codes are consistent with the 1997 NAICS and New SOC. Likewise, the 2004 NHIS public use data files contain two sets of 2-digit industry and occupation recodes: one set is based upon the 3-digit Census codes, while the other set is based on 4-digit Census codes.

Because the 1987 SIC/SOC and the 1997 NAICS/New SOC are entirely different classification schemes, the resulting occupation recodes (and, likewise, the industry recodes) in the 2004 NHIS will not be compatible with one another. Moreover, a crosswalk showing how the 1987 SIC/SOC compares to the 1997 NAICS/New SOC is not available at this time. However, the coding categories for these recodes are provided in the Industry and Occupation Appendices (following the Variable Layout Report for the Sample Adult file), and additional information is available on-line (see the final paragraph in this section). **Users are advised that the 1987 SIC/SOC recodes will be discontinued in the 2005 NHIS** (the 1997 NAICS/New SOC recodes will be retained)**.**

The recodes OCCUP1A, OCCUP2A, INDSTR1A, and INDSTR2A are identical to previous years' recodes with the same names – they were created in the same manner and have the same substantive meaning and response categories as in previous data files. The detailed occupation recode (OCCUPN1) based upon the New SOC has 94 distinct categories, while the associated simple recode (OCCUPN2) has 23 categories. These categories correspond to 2000 SOC Occupation Subgroups and Major Occupation Groups, respectively, as determined by the U.S. Census Bureau and the Bureau of Labor Statistics. The detailed industry recode (INDSTRN1) informed by the 1997 NAICS has 79 distinct categories, while the associated simple recode (INDSTRN2) has 21 categories. These categories correspond to NAICS Industry Subsectors and Sectors, respectively, as identified by Census.

For more information about the 1997 NAICS, please refer to <http://www.census.gov/epcd/www/naics.html>. For more information about the 2000 SOC, please refer to <http://www.bls.gov/soc/home.htm>. To obtain a complete list of 1997 NAICS Sectors and Subsectors and 2000 SOC Major Groups and Subgroups, please see <http://www.cdc.gov/nchs/nhis.htm>. Lastly, the table below provides a summary of the industry and occupation recodes included in the 2004 NHIS.

Table 5. Industry and Occupation Recodes, 2004 NHIS

NHIS Recode Name	Which NHIS data years?	Name of source coding scheme	Source year	Number of categories in NHIS recode	After 2004?
Occupation:					
OCCUP1A	1997-2003	SOC	1987	41	Not available
OCCUP2A	1997-2003	SOC	1987	13	Not available
OCCUPN1	New in 2004	New SOC	2000	94	Available
OCCUPN2	New in 2004	New SOC	2000	23	Available
Industry:					
INDSTR1A	1997-2003	SIC	1987	42	Not available
INDSTR2A	1997-2003	SIC	1987	14	Not available
INDSTRN1	New in 2004	NAICS	1997	79	Available
INDSTRN2	New in 2004	NAICS	1997	21	Available

II. Adult Conditions Section (ACN)

The ACN section of the 2004 NHIS obtains information from the sample adult as to whether he/she has, or has had, a selected number of medical conditions. 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 impairment, and tooth loss are the exceptions). Respondents are also asked about head colds and intestinal illness which began in the 2 weeks prior to the interview, and women age 18-49 are asked about current pregnancy status. In addition, the section contains information about the sample adult's current mental or emotional 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 his/her life or daily activities. Table 6 shows the specific health-related conditions covered in this section, as well as the various reference periods covered by the questions.

Table 6. Sample Adult File: Conditions and Reference Periods

	Condition	Ever	12 mos.	3 mos.	30 days	2 weeks	Now	Other
ACN.010	Hypertension Hypertension 2+	X						
ACN.020	visits							twice
ACN.031	Coronary heart disease Angina	X						
ACN.031	pectoris	X						
ACN.031	Heart attack (MI) Other heart	X						
ACN.031	condition or heart disease Stroke	X						
ACN.031	Emphysema	X						
ACN.031		X						
ACN.080	Asthma	X						
ACN.085	Asthma still have						X	
ACN.090	Asthma episode / attack		X					
ACN.100	Asthma ER visit		X					
ACN.110	Ulcer ever told	X						
ACN.120	Ulcer recent		X					
ACN.130	Cancer any	X						
ACN.140	Cancer kind	X						
ACN.150	Cancer when							Age
ACN.160	Diabetes	X						
ACN.170	Diabetes when							Age
ACN.180	Insulin						X	
ACN.190	Oral agents/pills						X	
ACN.201	Hay fever		X					
ACN.201	Sinusitis		X					
ACN.201	Chronic bronchitis Weak		X					
ACN.201	kidneys		X					
ACN.201	Liver condition		X					
ACN.250	Joint symptoms				X			
ACN.260	Joints affected				X			
ACN.270	Joint symptoms chronic			X				
CAN.280	Joints doctor consult	X						
CAN.290	Arthritis (arthritis, gout, fibromyalgia, rheumatoid arthritis, lupus) diagnosis	X						
CAN.295	Limited in activities due to arthritis/joint symptoms						X	
CAN.300	Neck pain			X				
CAN.310	Back pain			X				
CAN.320	Leg pain			X				
CAN.331	Jaw, face pain			X				

	Condition	Ever	12 mos.	3 mos.	30 days	2 weeks	Now	Other
CAN.331	Migraine			X				
CAN.350	Head/chest cold					X		
CAN.360	Intestinal illness					X		
CAN.370	Pregnant						X	
CAN.410	Hearing aid	X						
CAN.420	Hearing						X	
CAN.430	Vision impairment						X	
CAN.440	Blind						X	
CAN.451	Lost all teeth						X	
CAN.471	Sad				X			
ACN.471	Nervous				X			
ACN.471	Restless				X			
ACN.471	Hopeless				X			
ACN.471	Everything an effort				X			
CAN.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 types of cancer as well as to indicate that he/she had had more than three different cancers. The responses were recorded with the codes indicated in the questionnaire and were then transformed into "mentioned"/ "not-mentioned" variables during editing. These variables (CNKIND1-31) assign to every sample adult who reported having ever had cancer either a "mentioned," if he/she specified that particular cancer, a "not mentioned," if he/she did not specify that 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 CNKIND31 ("More than 3 kinds").

In 2002 there were major changes to core questions about arthritis and joint symptoms. Those questions remained unchanged in 2003 and 2004. Users are advised to read the 2002 Survey Description document to learn about those changes. Because of those changes, any comparisons of 2002-2004 arthritis and joint symptom data with data prior to 2002 should be undertaken with caution.

Sample adults who had joint symptoms were asked which joint(s) were involved, and were given the option to identify up to 16 joints. The format for this question is also a "mentioned"/ "not-mentioned" format. The resulting variables for analysis (JMTHP1-17) have a value of "mentioned" or "not mentioned" (or "refused," "don't know," or "not ascertained") for each of the 16 joints (including a 17th "other joint not listed") for each sample adult who was asked the question.

Age questions CANAGE1-30 and DIBAGE ("How old were you when you were diagnosed [with this condition]?") are "top coded" to 85+ years to insure confidentiality among the oldest respondents. The recode DIFAGE2 ("How long have you had diabetes" [AGE minus DIBAGE]) is calculated prior to top coding AGE and DIBAGE, but is itself top coded to 83+ years to insure confidentiality. The answers to the age questions were not edited for

reasonableness, and some respondents appear to have given the length of time since they were diagnosed rather than their age at diagnosis.

III. 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 2004 AHS time variables and recodes, which indicate how long respondents have had the condition(s) causing their limitation(s), were processed using procedures similar to those used in 2002 and 2003. Substantively, the 2002, 2003 and 2004 variables and recodes are similar to those from previous years (1997-2001), but the 2002-2004 data contain fewer unknown categories than previously, that is, detailed unknown categories were collapsed into broader categories.

Health Indicators: Illness Behavior and Health Status

The first questions in this section determined the number of days the respondent took off from work or 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 (whether it is 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).

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 these condition data is similar to that 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," and "weight problem. Starting in 2001 and continuing in 2004, if the sample adult was limited by a condition not listed in one of these 18 fixed categories, the interviewer entered "M" for "More conditions," and a second screen containing 17 additional condition categories and two "other impairment problem" categories appeared on the interviewer's laptop computer screen. These conditions were not read aloud to respondents, but if the sample adult's condition was limited by one of these 17 conditions, the interviewer recorded this information. If the sample adult was limited by a condition not included in one of the 18 fixed categories or on the interviewer's computer screen, then the interviewer entered a 50-character verbatim response for one or both of the "other impairment problem" categories.

The AHS condition data were edited very much like the condition data in FHS. The verbatim responses recorded by interviewers 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 interviewers, 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 appropriate category. An additional 16 *ad hoc* categories were created, and were assigned numbers 19_ thru 34_. (Note: Due to a naming convention error in 2002 and 2003 these same *ad hoc* categories were assigned numbers 19 thru 34 without an underscore.) Any verbatim conditions that could not be back-coded to one of the 18 fixed categories or recoded to one of the *ad hoc* 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 interviewer were also back-coded and categorized 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 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 underestimates. Therefore, these variables should be analyzed with care. Moreover, none of the AHS condition variables (AFLHCA1 through AFLHCA34_) should be used to estimate prevalence rates for the conditions they represent, because only those sample adults with a previously reported functional 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 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 in 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 and chronic recodes for each of the 36 categories, which are very similar to those used in the FHS section described above.

Technical Notes

The condition variable AFLHCA31_ includes any causal condition that specifically mentioned "surgery" or "operation," or otherwise indicates a medical treatment as the causal condition (either ongoing or occurring within the last year). The condition variable AFLHCA33_ includes any causal condition that specifically and solely mentioned "fatigue," "weakness," "lack of strength," "tiredness," "exhaustion," etc. without reference to any particular part of the body. Lastly, the condition variable AFLHCA34_ includes any causal condition that specifically and solely mentioned "pregnancy," "pregnant," or "childbirth."

IV. 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, weight, and sleep. With the exception of a new (2004) question on sleep, all health behavior questions have been in the NHIS Sample Adult core questionnaire since 1997. For details concerning the history of the Adult Health Behavior section, refer to the Survey Description documents for 1997-2001. Although the basic questions remain unchanged, several changes did occur in the AHB section of the questionnaire and in the data file in 2004: (1) selected follow-up questions were dropped; (2) universes for some questions were restricted to known responses to lead-in questions; and (3) selected recodes were dropped. See Appendix VIII for detail about these changes.

Smoking

Current smokers are defined as persons who have ever smoked 100 cigarettes and who currently smoke every day or some days. For data years 1997-2003, three smoking status recodes were provided. Beginning in 2004, only one of the three, SMKSTAT2, remains on the data file.

Leisure-time physical activity

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 included every year in the sample adult core module is: frequency and duration of vigorous activities, frequency and duration of light or moderate activities, and frequency of strengthening activities. The questions on leisure-time physical activity are used for tracking Healthy People 2010 Objectives 22.1-22.4 and in NHIS Early Release and in *Health, United States* (beginning in 2005).

Alcohol use

Lifetime drinking status was assessed for all sample adults. Questions related to current drinking behavior were asked of all respondents who had had at least 12 drinks in their lifetime. Respondents were permitted to answer in terms of the number of days they drank per week, per month, or per year. Standardized variables that convert the various time unit responses to standardized units-- that is, days per week (ALC12MWK) and days per year (ALC12MYR)- are provided.

A question asking how often the respondent had five or more drinks in one day during the past year was asked of all adults who drank at least once in the past year. The responses were not edited for consistency with the respondent's *usual* quantity or frequency of alcohol consumption because there was no basis for evaluating which one might be the more accurate. Note that the questions related to quantity of alcohol consumption are phrased in terms of the number of drinks consumed in a *day* and not the number of drinks consumed *at a sitting*.

ALCSTAT, a new recode in 2004, classifies lifetime and current drinking status for all sample adults. It replaces ALCSTAT1 (1997-2003) and ALC7STAT (2001-2003) and captures, in a single variable, all of the information contained in these two earlier recodes.

ALCSTAT is consistent with the classification of lifetime and current drinking status shown annually in *Health, United States, 2004*. The category "current drinker, level unknown" is slightly different from the category of the same name in the earlier variable, ALC7STAT. Beginning in 2004, adults who said they did not know how often they drank were not asked the question about usual number of drinks (ALCAMT) and are classified as "drinking status unknown" in ALCSTAT. In contrast, in the earlier variable (ALC7STAT), adults who said they did not know the frequency of their alcohol consumption <u>were</u> asked the question about number of drinks (ALCAMT); those few (less than 0.5% of sample adults) who answered the second question without having answered the first were classified as "current drinkers, level unknown" in ALC7STAT.

The category "former drinker, frequency unknown" (ALCSTAT=4) is new and includes former drinkers for whom information is not available on whether or not they had 12 or more drinks in any one year. Previously, in ALC7STAT, this category of former drinker was combined with "unknown drinking status" (ALC7STAT=9). ALCSTAT can be created by the data user relatively easily for data years in which both ALCSTAT1 and ALC7STAT appear (2001-2003). Creating ALCSTAT for data years prior to 2001 can be done, but the coding is quite complex.

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 the public use data file. 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. National estimates based on physical measurements, such as those available from NCHS' National Health and Nutrition Examination Survey may differ from those available from the NHIS, which are self-reported.

Body Mass Index (BMI) is a recode of the height and weight data, calculated using the formula: BMI = kilograms /meters². Although respondents had the option of answering in metric units, less than 1% of respondents 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" and "996" (not available) on the public use file for reasons of 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, but not obese (25 <= BMI < 30); overweight, including obese (BMI >= 25); and obese (BMI >= 30).

Sleep

A question asking about usual number of hours of sleep was added in 2004. A question on sleep was previously asked in the NHIS in 1990 as part of the Health Promotion and Disease Prevention Supplement.

V. Adult Health Care Access and Utilization Section (AAU)

The core Adult Health Care Access and Utilization (AAU) section of the 2004 NHIS has remained largely unchanged since 1997 and contains information on access to health care, dental care, health care provider contacts, and immunizations.

Questions regarding access to health care include having a usual place for sick care, having a usual place for routine/preventive care, change in the place of care, any delays in getting medical care, and instances of being unable to afford medical care. The question about the reason for delaying care focused on such access issues as transportation, getting an appointment, and waiting time prior to actually seeing the doctor. A question on dental care asked about the length of time since last dental visit.

Respondents were asked about health care provider contacts, including questions 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. There is also a question that asks how long it has been since the respondent has seen or talked to a doctor.

There are five questions related to adult immunizations: flu shot, flu mist vaccine, pneumonia vaccination, 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.

Technical notes

Beginning in Quarter 4 of 2003, in addition to the question asking about the flu shot, respondents were asked about flu vaccine administered by a nasal spray. Since data were collected for only one quarter of that year, the nasal spray flu vaccination variable (SPRFLUYR) was not included on the 2003 Public Use file. The data are expected to be released in the future. In 2004, this question was asked throughout the entire data year and is included on the 2004 public use file.

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 2004 Dataset Documentation to documentation from the 2003 (and earlier) NHIS for any other changes that may have occurred over time to the variables in this section.

VI. Adult AIDS Section (ADS)

This section contains a series of questions related to testing for HIV, the virus that causes AIDS. Respondents were asked whether they had ever donated blood and whether they had a blood test for HIV, their main reasons for getting or not getting tested, when they had their last test, the number of times they had been tested, 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.

Beginning in 2000, questions on sexually transmitted diseases (STDs) and tuberculosis (TB) are included in this section. These questions asked respondents whether they had an STD other than HIV or AIDS, whether they saw a doctor or health professional, and the location/facility to which they went to be checked. In addition to STD questions, respondents were also asked about TB, whether they had heard of it, how much they knew about it, and if they knew anyone personally with the disease. In addition, respondents were asked about their perceived personal risk of getting TB, and if, in their opinion, TB could be cured.

2004 National Health Interview Survey
Recontact Section

The Recontact Section includes questions about cellular phone use. The purpose of the cellular phone questions is to track the prevalence of wireless telephones in American families over time, allowing researchers to analyze the demographic characteristics of families who may have substituted wireless service for a home telephone (land-line). Having these data from a large population-based survey such as the NHIS could aid in improving the quality of random digit dial telephone surveys, which are dependent on accurate phone listings. These questions (REC.131 – REC.141) from the 2004 survey instrument can be found on the DHIS Web site at: <**http://www.cdc.gov/nchs/nhis.htm**>.

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 2004 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 (2004)

In a bibliography, the suggested citation should read:

National Center for Health Statistics (2005). *Data File Documentation, National Health Interview Survey, 2004 (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.

Information on how to cite electronic media is available at: <**http://www.cdc.gov/nchs/howto/citelec.htm**>.

References

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

DeNavas-Walt C, Proctor BD, Mills RJ. U.S. Census Bureau. Current Populaton Reports, P60-226. *Income, Poverty, and Health Insurance Coverage in the United States: 2003*. U.S. Government Printing Office. Washington, DC. 2004.

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

National Center for Health Statistics. (2004) *Health, United States, 2004*. Hyattsville, Maryland.

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.

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 2004 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 2004 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, and one sample child, if any.

Household Response Rate

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

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 Response Rate

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

Family Core data were collected from the respondent about all persons in the family. The response rates for the Family Core can be calculated in two ways: conditionally and finally. 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.

> **Important Note**
>
> **In 2004, there were 24 ineligible persons (type B non-interview) in 20 multi-family households who were not excluded in the creation of the rejected person and rejected family variables (REJ_FAM, REJ_PER). The family response rate has been adjusted in the Survey Description Document to reflect the exclusion of these 24 ineligible persons, but they have not been excluded from the 2004 household data file. Users should exercise caution when using the above variables in their analyses.**

Final Family Core Response Rate

$$\frac{(\text{Interviewed Families})}{(\text{Interviewed} + \text{Rejected Families (from Interviewed HHs))}}(\text{Household Response Rate})$$

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 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: conditionally and finally. 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

$$\frac{(\text{Interviewed Sample Children})}{(\text{Eligible Sample Children from Interviewed Families})}(\text{Final Family Response Rate})$$

The final Sample Child response rate is the rate for sample children that <u>takes into account household and family non-response</u>. 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 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: conditionally and finally. The conditional Sample Adult response rate is the rate only for those sample adults identified as eligible and <u>does not take into account household or family non-response</u>. 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

$$\frac{(\text{Interviewed Sample Adults})}{(\text{Eligible Sample Adults from Interviewed Families})}(\text{Final Family Response Rate})$$

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

Appendix I, Table 2. Response Rates for the 2004 NHIS

Household	86.9%
Family - Conditional Family - Final	99.6% 86.5%
Sample Child - Conditional Sample Child – Final	91.8% 79.4%
Sample Adult - Conditional Sample Adult – Final	83.8% 72.5%

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. The following examples are shown for two years of data. Similar 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 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 Response Rate for Combined Data Years

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

Family Core data were collected from the respondent about all persons in the family. The response rates for the Family Core can be calculated in two ways: conditionally and finally. 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 Years 1 and 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 Response Rate for Combined Data Years

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

The final Family response rate is the rate for those families identified as eligible that <u>takes into account household non-response</u>. 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 Years 1 and 2, and then multiplying this quotient by the Household 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: conditionally and finally. The conditional Sample Child response rate is the rate only for sample children and <u>does not take into account household or family non-response</u>. 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

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

The final Sample Child response rate is the rate for sample children that <u>takes into account household and family non-response</u>. 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 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: conditionally and finally. The conditional Sample Adult response rate is the rate only for those sample adults identified as eligible and <u>does not take into account household or family non-response</u>. 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 <u>takes into account household and family non-response</u>. 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 Years 1 and 2, and then multiplying this quotient by the final Family response rate for the combined data years.

Appendix I, Table 3. Number Eligible/Interviewed, 2004 NHIS

File	Eligible	Interviewed
Household	42,089	36,579
Family/Person	37,653	37,466
Sample Child	13,538	12,424
Sample Adult	37,388	31,326

Appendix I, Table 4. Number Eligible/Interviewed, 2003 NHIS

File	Eligible	Interviewed
Household	40,266	35,921
Family/Person	37,126	36,573
Sample Child	13,275	12,249
Immunization	13,275	11,665
Sample Adult	36,524	30,852

Appendix I, Table 5. Number Eligible/Interviewed, 2002 NHIS

File	Eligible	Interviewed
Household	40,377	36,161
Family/Person	37,458	36,831
Sample Child	13,570	12,524
Immunization	13,865	13,611
Sample Adult	36,787	31,044

Appendix I, Table 6. Number Eligible/Interviewed, 2001 NHIS

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

Appendix I, Table 7. Number Eligible/Interviewed, 2000 NHIS

File	Eligible	Interviewed
Household	43,437	38,633
Family/Person	39,998	39,264
Sample Child	14,711	13,376
Immunization	14,890	14,618
Sample Adult	39,201	32,374

Appendix I, Table 8. Number Eligible/Interviewed, 1999 NHIS

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

Appendix I, Table 9. Number Eligible/Interviewed, 1998 NHIS

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

Appendix I, Table 10. Number Eligible/Interviewed, 1997 NHIS

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

Appendix II
Race and Hispanic Origin in the 2004 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 Summary Health Statistics, 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/statpolicy.html >. 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. This policy was expected to be fully implemented across the federal statistical system beginning with the 2003 calendar year.

As noted previously, the U.S. Census Bureau is the data collection agent for the NHIS, as it is for a number of other federal surveys. The Census Bureau also provides the control totals for race/ethnicity (along with sex and age) that are used in the post-stratification adjustment of the person weights in the NHIS data file. In order to maintain consistency with the Census Bureau procedures for collecting and editing data on race and ethnicity, the NHIS made major changes to its editing procedures in the 2003 data year. Beginning in the 2003 NHIS, "Other race" is no longer available as a separate race response. This response category is treated as missing, and the race is imputed if this was the only race response. In cases where "Other race" was mentioned along with one or more OMB race groups, the "Other race" response is dropped, and the OMB race group information is retained. These editing changes are consistent with the procedures that the Census Bureau uses to create the Modified Race Data Summary File, which is the data file that provides the population control controls used in weighting the NHIS data. More information about the Modified Race Data Summary File and the editing procedures used to create it can be found at the following web site:

http://www.census.gov/popest/topics/methodology/2003_nat_char_meth.html. These editing changes remain in effect for the 2004 data file. Please refer to the 2004 Variable Layout Report for more information.

Race and Hispanic Origin Questions in the National Health Interview Survey

The 2004 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; response categories shown to the respondent on a flashcard].

There were no changes in the wording of the 2004 Hispanic origin question, but some responses were imputed, and the variable name is labeled as HISPAN_I to indicate this fact (see section on the imputation of race and ethnicity later in this document).

The 2004 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? Please select one or more of the following groups." [HHC.200; response categories shown to the respondent on a flashcard], and

"Which one of these groups, that is (FR: READ GROUPS) would you say BEST represents {your/name's} race?" [HHC.220; response categories given are read back to the respondent by the interviewer].

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 to the response categories effective 1999. In compliance with the new race and ethnicity data collection standards, the category "Asian and Pacific Islander" is now split into two categories, "Asian" and "Native Hawaiian and Other Pacific Islander." 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, public use 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.

The following table summarizes the Hispanic origin and race variables in the 2004 data file. Details on the specific response categories for the race questions and additional details on these variables can be found in the 2004 public use Variable Layout Report, and users are strongly urged to read these descriptions carefully to determine how and when the variables should be used in analysis. Data users are also encouraged to check the Variable Frequency Report to examine the unweighted data for these variables before computing weighted estimates.

Appendix II, Table 1. 2004 NHIS Race/Ethnicity Variable Names and Description

2004 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
RACERPI2	Contains 4 of 5 OMB race groups; values imputed for some records. Does not include "Other race" category.
MRACRPI2	Detailed race variable; multiple race persons not selecting a primary race group in separate category. Values were imputed for some records. Does not include "Other race" category.
MRACBPI2	See section below on bridging; values were imputed for some records. "Other race" category included for bridging purposes.
RACRECI2	All persons not previously coded in a race category are imputed to a race category on this variable.
RACEIMP2	Imputation flag for use in determining which cases were imputed for the race variables. New categories added to account for new editing procedures.
HISCODI2	Same categories as RACRECI2, 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 2004 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 and continued to be used for the 2004 NHIS data. Changes implemented in the 2003 imputation procedures remain in effect for 2004. Records for persons for whom "Other race" was the only race mentioned were treated as having missing data on race, and were added to the pool of records for which selected race and ethnicity variables are imputed. Prior to the 2000 NHIS, a crude imputation method that assigned a race to persons with missing values for the variable MAINRACE for race recodes #1 and #2 was used. Under these procedures, if an observed race was recorded by the interviewer, it was used to code a race value. If there was 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 2000-2004 NHIS data. In addition, imputation procedures for "Other race" were adapted from the Census Bureau's Modified Race Data Summary File editing specifications for use in the 2004 NHIS data imputation. These specifications formed the basis of the first stage of the imputation (within households), and 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' 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** - 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.   Re-classify 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** - for households in which all persons had missing values.

    Step 1.   The imputation was divided into three parts:

    A) Imputation among Hispanic households (ORIGIN=1).
    B) Imputation among Non-Hispanic households (ORIGIN=2).
    C) Imputation for households with unknown Hispanic origin (ORIGIN=7, 8, 9).

Step 2.   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.

Step 3.   After all imputations were completed, datasets from Stage 1 and Stage 2 were combined, records that were imputed were flagged for the in-house and public use data files, and comparisons of the distributions of the variables before and after imputation were examined.

Use of imputation flags

Since hot-deck imputation procedures have been implemented on the NHIS race and ethnicity data, imputation flags have also been added to the data file. These flags allow data users to keep track of the number of cases for which race and/or ethnicity was imputed by the type of original response. They also provide users with a means of accessing the data in their unimputed form. The flags also provide a mechanism for converting data back to the format in the data files prior to the implementation of imputation in 2000, which is critical for merging data files across survey years. There are four imputation flags on the 2004 public use data file: ORIGIMPT, HISPIMPT, RACEIMP2, and ERIMPFLG. These flags are described in Table 1 above.

Users who wish to merge across data years must recreate the race variables RACERPI2, MRACRPI2, and MRACBPI2 in the format they had in previous years by using the flag <u>RACEIMP2</u>. Sample SAS code for using the imputation flags and merging across data years for the variable RACERPI2 (RACERP_I in 2000-2002 and RACER_P in 1999) is included below.

```
************************************************************************
***  merge 1999-2004 race variables using public use variables   ****
***********************************************************************;
*** recode 1999 data ***;

     if RACER_P in (97)      then RACEPU99=7;      /* refused */
else if RACER_P in (98)      then RACEPU99=8;      /* NA      */
else if RACER_P in (99)      then RACEPU99=9;      /* DK      */
else                              RACEPU99=RACER_P;

*****   code to add in imputation flags for 2000-2002 RACERP_I   ******;

     if RACEIMPT in (1)      then RACP0002=7;      /* refused */
else if RACEIMPT in (2)      then RACP0002=8;      /* NA      */
else if RACEIMPT in (3)      then RACP0002=9;      /* DK      */
else                              RACP0002=RACERP_I;

******   code to add imputation flags for 2003-2004 RACERPI2   ********;

     if RACEIMP2 in (1)      then RACP0304=7;      /* refused */
else if RACEIMP2 in (2)      then RACP0304=8;      /* NA      */
else if RACEIMP2 in (3)      then RACP0304=9;      /* DK      */
else if RACEIMP2 in (4 5)    then RACP0304=5;      /* Other races[1] */
else                              RACP0304=racerpi2;
```

1 Note that this category contains "Other race only," "Unspecified Multiple race" and NHOPI persons.

```
*** combine 1999-2004 data into single variable *****;

     if RACEPU99 ne .          then RACE9904=RACEPU99;
else if RACP0002 ne .          then RACE9904=RACP0002;
else                               RACE9904=RACP0304;
```

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 2004 NHIS public use data file contains one bridge race variable to allow comparisons of 2004 data with data from previous years, and to enable merging the 2004 data with 1997-2003 data.

There was one major change to the race and ethnicity data in the 2000 NHIS (which is also true for 2001-2004) that occurred 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. Beginning with the 2000 survey (and now in 2004), 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 2004 Race Variables

The variables RACRECI2 and MRACRPI2 correspond to the old OMB guidelines for collecting racial and ethnic data (see the Variable Layout Report 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 RACRECI2 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 RACERPI2 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, 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.

Algorithm used to Create 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 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 that is back-coded to this category), Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian (a verbatim mention that is back-coded to this 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), and *NHOPI* (includes Native Hawaiian, Guamanian, Samoan and Other Pacific Islander.

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;

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;

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

The variables RACEW, RACEB, RACEAIAN, RACEASIA, and RACENHPI, are thus assigned the numbers 1, 2, 4, 8, and 16, which add up to a series of unique numbers corresponding to specific combinations of races. The value of RACEFULL tells which races (RACEW through RACENHPI) combined to give that number. For example, if RACEFULL=3, then 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 Asian, AIAN, and NHOPI 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 Codes (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
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

Coding Scheme for OMB Race Category Data (including single and multiple race mentions)		
# of Category (reported in SAS frequency distribution of RACEFULL)	**Sum of Codes (breakdown of RACEFULL= SUM (OF RACEW+RACEB+ RACEAIAN+RACEASIA+ RACENHPI+RACEOTHR))**	**Resulting Category (used in the PROC FORMAT statement to label categories in SAS)**
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

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 RACERPI2 is a recode based on RACEFULL. Analysts who wish to use more detailed race data in their analyses should contact the NCHS Research Data Center or visit their web page: < **http://www.cdc.gov/nchs/r&d/rdc.htm** >.

Further Information

        Although the race variables included in the 2004 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 2004 NHIS race data, information about this will be placed on the NCHS web site: < **http://www.cdc.gov/nchs/nhis.htm** >.

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://www.whitehouse.gov/omb/inforeg/statpolicy.html >.** (Statistical Policy 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.

 U.S Bureau of the Census. *Census 2000 Modified Race Data* [MR(31)-CO.txt]. 2002.

Appendix III
Variance Estimation and Other Analytic Issues, NHIS 2004

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, the techniques 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 generally 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 be 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 that 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 excessive Type I error.

Conceptual NHIS design for 1995-2005

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, cost-effective, and timely, 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 40,000 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), which are single counties or groups of adjacent counties (or equivalent jurisdictions) and/or metropolitan areas. These PSUs vary in population size and number of jurisdictions. Cost-effective field operations and efficient sampling result in those PSUs with the largest populations (e.g., the New York metropolitan area) being sampled with certainty, and the smaller universe PSUs being represented by a sample. 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 strata with smaller population sizes, 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 2000 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 (i.e., STRATUM and PSU). 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 2004 Household and Person public use files contain the design variables necessary for variance estimation; Table 1 provides a summary of the Person file variables. Users should check the Variable Layouts of the Household and Person files for any additional information about these variables. Note that, beginning with the 2004 NHIS data, the design variables (STRATUM and PSU) will no longer be included on the Family, Sample Child and Sample Adult files.

Appendix III, Table 1. Variables Used for Variance Estimation, 2004 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 1997-2004 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 2004 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 no PSU was sampled more than once, 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 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 found on these files.

⊗ **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 PSUs 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. This 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 selected records (and list-wise deleting other records) is called subsetting the data. With a subsetted dataset that 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 containing the design information for the entire sample. (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 those 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 Person file dataset:

> *PROC ...*   **DESIGN = WR ;**
> *NEST* **STRATUM PSU ;**
> *WEIGHT*   **WTFA ;**
> *SUBGROUP* **(variable names);**
> *LEVELS ...*   **;**
> *SUBPOPN*   **RACRECI2=2 & SEX=2 / NAME="Analysis of African American women;"**

Using the full dataset with the SUBPOPN statement in this example would constrain this analysis to African American women only (RACRECI2= 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.

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 Aged 2 to 3 years

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 (Achenbach and Edelbrock, 1983). 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 had 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 had 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_01-03.000 and CHS.361_01-03.000).

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 being 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 Mental Health (CMH) section of the 2004 NHIS (please refer to items CMH.010._01.00 through CMH.050_05.00). 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 CMH.060_00.000 through CMH.100_00.000), 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 in the field to validate the SDQ among U.S. adolescents aged 15-17 years. Results from the validation studies in the U.S. are not yet available. Evaluation studies of the SDQ in the 2001 NHIS include a factor analysis of the SDQ symptom items by Dickey and Blumberg (2004) and an evaluation of the psychometric properties by Bourdon et al. (2005). The NCHS Advance Data Report Number 360 also utilizes a question from SDQ data from the 2001, 2002, and 2003 NHIS. Further information on the SDQ, including information on the teacher and self-respondent versions of the SDQ, is available at the Web site: **<http://www.sdqinfo.com>.**

Appropriate analytic use of the SDQ requires thorough understanding of the development and use of the SDQ in the United States. Users may contact the Division of Health Interview Statistics at 301-458-4901, or send e-mail to us at nhislist@cdc.gov for help with these analyses.

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 who have or are at increased risk of having psychological disorders. 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

Bourdon KG, Goodman R, Rae Donald S, Simpson G, and Koretz D. *The Strengths and Difficulties Questionnarie: U. S. Normative Data and Psychometric Properties.* Journal of the American Academy of Child and Adolescent Psychiatry. 44(6): 557-564. 2005.

Dickey WC and Blumberg SJ. *Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire.* Journal of the American Academy of Child and Adolescent Psychiatry. 43(9): 1159-1167. 2004.

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

Goodman R. *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. 1999.

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

Simpson GA, Bloom B, Cohen RA, Blumberg S, and Bourdon K. *U. S Children with Emotional and Behavioral Difficulties: Data from the 2001, 2002, and 2003 NHIS.* Advance Data from vital and health statistics; no. 360. Hyattsville, MD. National Center for Health Statistics. 2005.

Appendix VI
Transition to the 2000-Census-Based Weights

Between 1995 and 2002, the weights for the NHIS data were derived from 1990-Census-based postcensal population estimates. Beginning with the 2003 data, the NHIS has made the transition to weights derived from the 2000-Census-based population estimates. The new population estimates no longer contain any adjustment for under-enumeration of the population. The NHIS sample weights were calibrated to 2000-Census-based totals for sex, age, and race/ethnicity of the U.S. civilian noninstitutionalized population.

During the preparation of the 2003 Summary Health Statistics reports, the impact of this transition was assessed for the 2002 NHIS by comparing estimates for selected health characteristics using the 1990-Census-based weights with those using the 2000-Census-based weights. The effect of new population controls on survey estimates differed by type of health characteristic. The person health estimates and sample adult health estimates were more affected than sample child estimates. The percent of health estimates expressed as percentages and rates with significant differences were 0.27% for person estimates, 0.27% for sample adult estimates, and 0.0% percent for sample child estimates. The percent of health estimates expressed as frequencies with significant differences were 13% for person estimates, 16% for sample adult estimates, and 1% for sample child estimates (Lynch and Parsons, 2004).

The impact of this transition was also assessed for the Early Release estimates from the 2000-2002 NHIS by comparing estimates using the 1990-Census-based weights with those using the 2000- Census-based weights. The changes for all selected measures are at most 1 percentage point. Results of these findings are presented in tables II and III at:
< **http://www.cdc.gov/nchs/data/nhis/earlyrelease/200409_app.pdf** >. Separate data files containing the 2000-Census-based weights for NHIS 2000–2002 will be made available at a later date on the NHIS Web site: < **http://www.cdc.gov/nchs/nhis.htm >**.

References

Lynch, C. and Parsons, V. (2004), *"The Impact of 2000 Census Based Population Controls on Health Estimates in the National Health Interview Survey,"* 2004 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM]*,* Alexandria, VA: American Statistical Association.

Appendix VII
Merging Data Files and Combining Years of Data in the NHIS

Merging Data Files

NHIS data files can be merged within years as well as combined across years. The purpose of merging data *within* a particular data year is to incorporate variables from different data files when respondents are common to both files, thereby increasing the number of variables available for analysis for a given individual. In contrast, the purpose behind combining NHIS data files *across* survey years is to combine respondents from different data years while retaining variables common to both files, thereby increasing the number of respondents (as long as the same variables are found in both files) and the precision of estimates. See the following subsection on Combining Years of Data for more information.

Unlike previous survey years, variables are not generally repeated on multiple data files in the 2004 NHIS. **As a result, users may find it necessary to perform additional merging of the 2004 files in order to analyze the data.** Each data file contains household, family, and person record identifiers that make merging within the 2004 files possible. Once the data files are sorted by record identifiers common to each file, merging is straightforward. Below is an example of a program that will merge data files **within** an NHIS data year. Using the household, family, and person record identifiers (HHX, FMX and FPX, respectively), this program merges data from the 2004 Household, Family, Person, and Sample Child data files.

```
/* Merge the Household file and the Family file. */

/* Create a Household dataset with selected variables and sorted by HHX.*/
DATA HH (KEEP=HHX REGION);  /* HH is a SAS dataset; the KEEP statement retains only
the listed variables for processing. */
SET NHIS2004.HOUSEHLD; /*The SET statement reads data from the 2004 Household file. */
PROC SORT DATA=HH; /* Sort by HHX, the household identifier. */
BY HHX;
RUN;
/* Create a Family dataset with selected variables and sorted by HHX. */
DATA FM (KEEP=HHX FMX INCGRP RAT_CAT WTFA_FAM); /* FM is a SAS dataset;
the KEEP statement retains only the listed variables for processing. */
SET NHIS2004.FAMILYXX; /*The SET statement reads data from the 2004 Family file. */
PROC SORT DATA=FM; /* Sort by HHX, the household identifier. */
BY HHX;
RUN;


DATA HHFM; /* New combined dataset called HHFM */
MERGE FM (IN=FROMFM) HH ; /* Merge the newly created FM and HH files, using an IN
statement.*/
BY HHX;
IF FROMFM = 1; /* The combined dataset HHFM will contain only those records that are in the
Family file; the Household file's REGION variable will be appended to these records. */
PROC SORT DATA=HHFM; /* Sort by HHX and FMX, the household and family identifiers. */
```

BY HHX FMX;
RUN;

In the code above, the IN statement creates a temporary SAS variable (called FROMFM) that has a value of 1 if the dataset associated with the IN statement contributed to the current observation, or a value of 0 if it did not. The subsequent statement, "IF FROMFM = 1" tells SAS to retain only those observations from the Family file (called FM). For more information on IN statements in SAS, consult Delwiche and Slaughter (1998).

/* Merge the Person file and the combined Family/Household file. */

/* Create a Person file with selected variables. */
DATA PR (KEEP=HHX FMX FPX SEX AGE_P WTFA STRATUM PSU); /* PR is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2004.PERSONSX; /*The SET statement reads data from the 2004 Person file. */
PROC SORT DATA=PR; /* Sort by HHX and FMX, the household and family identifiers. */
BY HHX FMX;
RUN;

DATA PRHHFM; /* Combined Person, Family, and Household dataset called PRHHFM*/
MERGE PR HHFM (DROP=WTFA_FAM); /* Merge the newly created PR file and HHFM, the combined Family/Household file, by the identifiers common to both files. At this point, users may drop the Family file weight and retain only the Person file weight for person-level analyses.*/
BY HHX FMX;
PROC SORT DATA=PRHHFM; /* Sort by HHX, FMX, and FPX, the household, family, and person identifiers. */
BY HHX FMX FPX;
RUN;

/* Merge the Sample Child file and the combined Person/Family/Household file. */

/* Create a Sample Child file with selected variables. */
DATA CH (KEEP=FPX HHX FMX CASHMEV PROBRX WTFA_SC); /* CH is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2004.SAMCHILD; /*The SET statement reads data from the 2004 Sample Child file. */
PROC SORT DATA=CH; /* Sort by HHX, FMX, and FPX, the household, family, and person identifiers. */
BY HHX FMX FPX;
RUN;

DATA CHPRHHFM; /* Combined Sample Child, Person, Family, and Household dataset called CHPRHHFM*/
MERGE PRHHFM CH; /* Merge CH, the newly created Sample Child file, and PRHHFM, the combined Person/Family/Household file, by the identifiers common to both files. */
BY HHX FMX FPX;
RUN;

Combining Years of Data

Important Note

Variable names may change from one year to another. Users are advised to check variable names and where names differ, make certain it is appropriate to combine years of data for a given variable.

As previously mentioned, the purpose of combining or concatenating years of data (in SAS terminology) is to increase the number of observations or respondents for the same number of variables, and thus increase the precision of estimates. It is possible to combine data from successive years of the National Health Interview Survey (NHIS) when the questions remain essentially the same over the years being combined.

Combining datasets from more than one year joins them one after the other (concatenates), as opposed to merging datasets. Analysts wishing to do both – merge data from multiple files within years and combine years of data – will need to first merge the data within each single year and then concatenate the files for the selected years of data (see the preceding section on Merging Data Files).

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.

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.

The following is an example of a program that will combine data files across NHIS data years. The program is written to concatenate the data from the Person files of the 2003 NHIS and the 2004 NHIS. Note that the person identifier was called PX in the 2003 (and earlier) NHIS and FPX in the 2004 NHIS; users may find it necessary to create an FPX variable in the 2003 dataset (or, alternatively, a PX variable in the 2004 dataset) in order to make the data compatible for analyses.

/*Combine data files from 2 different years. */

DATA PER_03; /* Create SAS dataset PER_03.*/
SET NHIS2003.PERSONSX /* The SET statement reads data from an existing SAS dataset, e.g.,

the 2003 Person file */ **(KEEP=HHX FMX PX AGE_P SEX WTFA STRATUM PSU);** /* The KEEP statement retains only the listed variables for processing. */
FPX=PX; /* Create a variable called FPX from the existing PX variable.*/
RUN;

PROC SORT DATA=PER_03; /* Sort SAS dataset PER_03. */
BY HHX FMX FPX;
RUN;

DATA PER_04; /* Create SAS dataset PER_04.*/
SET NHIS2004.PERSONSX /* The SET statement reads data from an existing SAS dataset, e.g., the 2004 Person file */ **(KEEP=HHX FMX FPX AGE_P SEX WTFA STRATUM PSU);** /* The KEEP statement retains only the listed variables for processing. */
RUN;

PROC SORT DATA=PER_04; /* Sort SAS dataset PER_04. */
BY HHX FMX FPX;
RUN;

DATA COMBO; /* New, combined SAS dataset */
SET PER_03 PER_04; /* Concatenate selected variables from 2003 and 2004 datasets. */
WTFA_2YR=WTFA/2; /*Create a new weight by dividing the existing Person file weight (WTFA) by 2, the number of Person data files combined to create the data file called COMBO.*/
RUN;

References

Delwiche, LD and SJ Slaughter (1998), *The Little SAS Book: A Primer* (2nd ed), SAS Institute: Cary, NC.

Appendix VIII
Changes/Additions/Deletions in 2004

A number of changes were introduced to the Core sections of the 2004 NHIS, resulting in new, changed, or deleted variables (relative to 2003). Following is a brief summary. Users are *strongly* encouraged to check the notes in the Variable Layout Report, as well as the relevant sections in this document, for more information.

Household File

Variable Name	Brief Variable Description	Brief Description of Change
NOWAF_HH	Armed Forces member in household	Deleted

Family File

Variable Name	Brief Variable Description	Brief Description of Change
FINT_M_P	Family interview month	New variable similar to INT_M_P on 2003 Household file
FINT_Y_P	Family interview year	New variable similar to INT_Y_P on 2003 Household file
FLNGINTV	Language of interview for each Family	New variable similar to LNG_INTV on 2003 Household file
INSIDE	Working telephone inside home	Moved from Household file
NOSERV	No telephone service past 12 months	Moved from Household File
HOWLNGWK	Number of weeks without telephone Service	Moved from Household File
TELN_FLG	Supplied telephone number (yes/no)	TELENUM moved from Household file and Renamed
TELH_FLG	Telephone number valid (yes/no)	New name for TELHOME
FM_TYPE	Family type	New name for FMTYPE
FM_STRP	Family structure	New name for FMSTR2
FMSTRCT2	Family structure	Deleted
FWICYN2	Receipt of WIC benefits	Deleted
FM_RESP	Family respondent person number	Deleted
ACPT_FAM	Number of responding families	Dropped/Family file; Retained/Household file
REJ_FAM	Number of non-responding families	Dropped/Family file; Retained/Household file
LIVQRT	Type of living quarters	Dropped/Family file; Retained/Household file
ASSIGNWK	Assignment week	Dropped/Family file; Retained/Household file
INTV_QRT	Interview quarter	Dropped/Family file; Retained/Household file
PSU	PSU	Dropped/Family file; Retained/Household file
STRATUM	Stratum	Dropped/Family file; Retained/Household file
REGION	Region	Dropped/Family file; Retained/Household file

Person File

Variable Name	Brief Variable Description	Brief Description of Change
FPX	Person number in family	Replaces PX, person number in Household

Family Core

Health Status and Limitation of Activity Section (FHS)

Variable Name	Brief Variable Description	Brief Description of Change
LAHCA19_ thru LAHCA34_	Conditions associated with any functional limitation	Name change only

Health Insurance Section (FHI)

Variable Name	Brief Variable Description	Brief Description of Change
HIKINDA	Initial response "Private Health Insurance"	Initial health insurance list shortened
HIKINDB	Initial response "Medicare"	Initial health insurance list shortened, note new meaning
HIKINDC	Initial response "Medigap"	Initial health insurance list shortened, note new meaning
HIKINDD	Initial response "Medicaid"	Initial health insurance list shortened, note new meaning
HIKINDE	Initial response "SCHIP"	Initial health insurance list shortened, note new meaning
HIKINDF	Initial response "Military health care"	Initial health insurance list shortened, note new meaning
HIKINDG	Initial response "Indian Health Service"	Initial health insurance list shortened, note new meaning
HIKINDH	Initial response "State-sponsored health plan"	Initial health insurance list shortened, note new meaning
HIKINDI	Initial response "Other government program"	Initial health insurance list shortened, note new meaning
HIKINDJ	Initial response "Single service plan"	Initial health insurance list shortened, note new meaning
HIKINDK	Initial response "No coverage of any type"	Initial health insurance list shortened, note new meaning
MCAREPRB	Medicare probe question	New variable, beginning quarter 3
MCAIDPRB	Medicaid probe question	New variable, beginning quarter 3
SINCOV	Single service plan probe question	New variable
MCNAMEN	Medicare Advantage plan model type	New name for MCNAME
MCCHOICE	Medicare Advantage plan enrollment	Change in question wording
MCRXCARD	Medicare RX Discount Card	New variable
PLNWRKN1	How private health plan was obtained (Plan)	New name for PLNWRK1; New answer categories
PRRXCOV1	Private plan has drug benefit (Plan 1)	New variable
PLNWRKN2	How private health plan was obtained	New name for PLNWRK2; New answer

Variable Name	Brief Variable Description	Brief Description of Change
	(Plan)	categories
PRRXCOV2	Private plan has drug benefit (Plan 2)	New variable
HITYPE3	Model type of plan (Plan 3)	Variable is no longer available on public use file
WHONAM3	Plan in whose name (Plan 3)	Variable is no longer available on public use file
PLNWRKN3	How private health plan was obtained (Plan3)	New name for PLNWRK3; New answer categories, Variable is not available on public use file
PLNPAY13	Paid for by self or family (Plan 3)	Variable is no longer available on public use file
PLNPAY23	Paid for by employer or union (Plan 3)	Variable is no longer available on public use file
PLNPAY33	Paid by someone outside household (Plan 3)	Variable is no longer available on public use file
PLNPAY43	Paid for by Medicare (Plan 3)	Variable is no longer available on public use file
PLNPAY53	Paid for by Medicaid (Plan 3)	Variable is no longer available on public use file
PLNPAY63	Paid for by SCHIP (Plan 3)	Variable is no longer available on public use file
PLNPAY73	Paid for by government program (Plan 3)	Variable is no longer available on public use file
HICOSTR3	Out-of-pocket premium cost (Plan 3)	Variable is no longer available on public use file
PLNMGD3	Type of plan (Plan 3)	Variable is no longer available on public use file
MGCHMD3	Doctor choice (Plan 3)	Variable is no longer available on public use file
MGPRMD3	Preferred list (Plan 3)	Variable is no longer available on public use file
MGPYMD3	Out of plan use (Plan 3)	Variable is no longer available on public use file
MGPREF3	Private referral (Plan 3)	Variable is no longer available on public use file
PRRXCOV3	Private plan has drug benefit (Plan 3)	New variable, variable is not available on public use file
HITYPE4	Model type of plan (Plan 4)	Variable is no longer available on public use file
WHONAM4	Plan in whose name (Plan 4)	Variable is no longer available on public use file
PLNWRKN4	How private health plan was obtained (Plan4)	New name for PLNWRK4 because of new answer categories, variable is not available on public use file
PLNPAY14	Paid for by self or family (Plan 4)	Variable is no longer available on public use file
PLNPAY24	Paid for by employer or union (Plan 4)	Variable is no longer available on public use file
PLNPAY34	Paid by someone outside household (Plan 4)	Variable is no longer available on public use file

Variable Name	Brief Variable Description	Brief Description of Change
PLNPAY44	Paid for by Medicare (Plan 4)	Variable is no longer available on public use file
PLNPAY54	Paid for by Medicaid (Plan 4)	Variable is no longer available on public use file
PLNPAY64	Paid for by SCHIP (Plan 4)	Variable is no longer available on public use file
PLNPAY74	Paid for by government program (Plan 4)	Variable is no longer available on public use file
HICOSTR4	Out-of-pocket premium cost (Plan 4)	Variable is no longer available on public use file
PLNMGD4	Type of plan (Plan 4)	Variable is no longer available on public use file
MGCHMD4	Doctor choice (Plan 4)	Variable is no longer available on public use file
MGPRMD4	Preferred list (Plan 4)	Variable is no longer available on public use file
MGPYMD4	Out of plan use (Plan 4)	Variable is no longer available on public use file
MGPREF4	Private referral (Plan 4)	Variable is no longer available on public use file
PRRXCOV4	Private plan has drug benefit (Plan 4)	New variable, this variable is not available on public use file
PRPLPLUS	Person has more than two private plans	New variable
SCHIP	State Children's Health Insurance Program	New name for CHIP to reflect changes in how program is referred; New categories
STDOC1	Choose a doctor SCHIP	New variable
STPCMD1	Must go to certain doc or clinic SCHIP	New variable
STREF1	Referral needed SCHIP	New variable
OTHERPUB	Other state-sponsored health plan	New categories
STDOC2	Choose a doctor other state plan	New variable
STPCMD2	Must go to certain doc or clinic other state plan	New variable
STREF2	Referral needed other state plan	New variable
OTHERGOV	Other government programs	New categories
STDOC3	Choose a doctor other government program	New variable
STPCMD3	Must go to certain doc or clinic other government program	New variable
STREF3	Referral needed other government program	New variable
MILITARY	Military Health Coverage	New categories
MILSPC1	Covered by TRICARE	New variable
MILSPC2	Covered by VA	New variable
MILSPC3	Covered by CHAMP-VA	New variable
MILSPC4	Covered by Other military coverage	New variable
MILMAN	Type of TRICARE coverage	New variable

Family Sociodemographics (FSD)

Variable Name	Brief Variable Description	Brief Description of Change
PLBORN	Born in the United States	New variable
EDUC1	Highest level of school completed	Replaces EDUC
PMILTRY	Honorable discharge status	Replaces MILTRYDS
DOINGLWP	Employment status in the previous week	Replaces DOINGLW1
WHYNOWKP	Main reason for not working last week	Replaces WHYNOWK1
WRKHRS2	Total hours worked last week/usually	Replaces WRKHRS
MAJR_ACT	Major activity	Dropped
EDUC_R1	Education recode	Dropped

Income and Assets Section (FIN)

Variable Name	Brief Variable Description	Brief Description of Change
AB_BL20K	Total combined family income (grouped)	Dropped
INCGRP	Total combined family income (grouped)	Moved to the Family file
RAT_CAT	Ratio of family income to the Poverty Threshold	Moved to the Family file
HOUSEOWN	Home tenure status	Moved to the Family file
FGAH	Pay lower rent due to government program	Moved to the Family file

Sample Child Core

Child Health Care Access and Utilization Section (CAU)

Variable Name	Brief Variable Description	Brief Description of Change
CSCL1_S1	Considerate of others feelings	Moved to CMH and renamed RSCL1_S1
CSCL1_H1	Restless/overactive	Moved to CMH and renamed RSCL1_H1
CSCL1_E1	Often complains of headache, Stomach aches or sickness	Moved to CMH and renamed RSCL1_E1
CSCL1_S2	Shares with other children/youth	Moved to CMH and renamed RSCL1_S2
CSCL1_C1	Often loses temper	Moved to CMH and renamed RSCL1_C1

Variable Name	Brief Variable Description	Brief Description of Change
CSCL2_P1	Solitary/prefers to play/be alone	Moved to CMH and renamed RSCL2_P1
CSCL2_C2	Well behaved/does what requested	Moved to CMH and renamed RSCL2_C2
CSCL2_E2	Many worries/often seems worried	Moved to CMH and renamed RSCL2_E2
CSCL2_S3	Helpful if someone hurt/ill	Moved to CMH and renamed RSCL2_S3
CSCL2_H2	Constantly fidgeting/squirming	Moved to CMH and renamed RSCL2_H2
CSCL3_P2	At least one good friend	Moved to CMH and renamed RSCL3_P2
CSCL3_C3	Fights with other children/or bullies them (youth)	Moved to CMH and renamed RSCL3_C3
CSCL3_E3	Unhappy/depressed/tearful	Moved to CMH and renamed RSCL3_E3
CSCL3_P3	Liked by other children/youth	Moved to CMH and renamed RSCL3_P3
CSCL3_H3	Distracted/wandering concentration	Moved to CMH and renamed RSCL3_H3
CSCL4_E4	Nervous/clingy/loses confidence	Moved to CMH and renamed RSCL4_E4
CSCL4_S4	Kind to younger children	Moved to CMH and renamed RSCL4_S4
CSCL4_C4	Often lies OR cheats	Moved to CMH and renamed RSCL4_C4
CSCL4_P4	Picked on or bullied by other children/youth	Moved to CMH and renamed RSCL4_P4
CSCL4_S5	Offers to help others	Moved to CMH and renamed RSCL4_S5
CSCL5_H4	Thinks things out before acting	Moved to CMH and renamed RSCL5_H4
CSCL5_C5	Steals from home/school	Moved to CMH and renamed RSCL5_C5
CSCL5_P5	Gets along better w/adults than children/youth	Moved to CMH and renamed RSCL5_P5
CSCL5_E5	Has many fears/easily scared	Moved to CMH and renamed RSCL5_E5
CSCL5_H5	Good attention/completes chores, homework	Moved to CMH and renamed RSCL5_H5
CSCL6	Difficulties w/ emotions/concentration/ behavior/getting along	Moved to CMH and renamed RSCL6
CSCL7	Length difficulties have been present	Moved to CMH
CSCL8	Difficulties upset/distress child	Moved to CMH and renamed RSCL8
CSCL9_HL	Difficulties interfere with home life	Moved to CMH and renamed RSCL9_HL
CSCL9_FR	Difficulties interfere with friendships	Moved to CMH and renamed RSCL9_FR
CSCL9_CL	Difficulties interfere with classroom Learning	Moved to CMH and renamed RSCL9_CL
CSCL9_LA	Difficulties interfere with leisure activities	Moved to CMH and renamed RSCL9_LA
CSCL10	Difficulties put burden on respondent/family	Moved to CMH and renamed RSCL10

Child Immunization Section (CIM)

The entire CIM section has been discontinued in the NHIS.

Child Mental Health (CMH)

In 2004, questions on children's emotional and behavioral difficulties from the Strengths and Difficulties Questionnaire (SDQ) that were included in the 2003 NHIS have been moved from the CAU section to a new section, Child Mental Health (CMH). The questions are identical to the 2003 questions; only the variable names and question numbers have changed. See the CAU section changes above for the old and new names of these variables.

Sample Adult Core

Adult Sociodemographics (ASD)

Variable Name	Brief Variable Description	Brief Description of Change
DOINGLWA	Employment status in the previous week	New variable
WHYNOWKA	Main reason for not working last week	New variable
INDSTRYN1	Detailed industry classification (NAICS)	New variable
INDSTRYN2	Simple industry classification (NAICS)	New variable
OCCUPN1	Detailed occupation classification (SOC)	New variable
OCCUPN2	Simple occupation classification (SOC)	New variable

Adult Health Status and Limitation of Activity Section (AHS)

Variable Name	Brief Variable Description	Brief Description of Change
AFLHCA19_ thru AFLHCA34_	Conditions associated with any functional limitation	New names for AFLHCA19 thru AFLHCA34

Adult Health Behaviors Section (AHB)

Variable Name	Brief Variable Description	Brief Description of Change
SMKQTD	Follow-up for former smokers who did not know how long ago they quit.	Dropped
CIGDAMO	Number of days smoked in past 30 days.	Universe restricted to known someday smokers.
CIGQTYR	Tried to quit smoking in past year	Universe restricted to known current smokers.

Variable Name	Brief Variable Description	Brief Description of Change
SMKQTD	Follow-up for former smokers who did not know how long ago they quit.	Dropped
CIGDAMO	Number of days smoked in past 30 days.	Universe restricted to <u>known</u> someday smokers.
CIGQTYR	Tried to quit smoking in past year	Universe restricted to <u>known</u> current smokers.
SMKSTAT1	4 category smoking status recode	Dropped
SMKSTAT3	7 category smoking status recode	Dropped
VIGLONGD	Follow-up for adults who did vigorous activity but did not know how long they spent.	Dropped
MODLONGD	Follow-up for adults who did moderate activity but did not know how long they spent	Dropped
ALC12MMO	Frequency of drinking (times per month).	Dropped
ALCAMT	Usual number of drinks	New universe restricted to <u>known</u> current drinkers.
ALC5UPNO, ALC5UPTP	5 or more drinks in one day	New universe restricted to <u>known</u> current drinkers.
ALCSTAT	10 category alcohol drinking status	Replaces ALCSTAT1 and ALC7STAT.
ALCSTAT1	3 category alcohol drinking status	Dropped
ALC7STAT	8 category alcohol drinking status	Dropped
MET_FLG1	Flag indicating metric response (cm) for height.	Dropped
MET_FLG2	Flag indicating metric (kg) response for body weight.	Dropped
DESIREWT	Metropolitan Life Insurance Company (MLIC) 1983 Standard of Desirable Body Weight	Dropped
SLEEP	Hours of sleep	New variable

Adult Health Care Access and Utilization Section (AAU)

Variable Name	Brief Variable Description	Brief Description of Change
SPRFLUYR	Nasal spray flu vaccination	New variable