2000 National Health Interview Survey (NHIS)
Public Use Data Release

NHIS Survey Description

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

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

March, 2002

2000 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 2000 NHIS, as well as important information about any modifications or updates to the data and/or documentation. Published reports from previous years' surveys are also available, as are updates about future surveys and datasets.

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

2000 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. Data 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), and (2) one or more sets of questions (called Supplements) on current health topics. Despite periodic revisions to the Core questionnaire, Supplements played an increasingly important role in the survey as a means of enhancing topic coverage in the Core. Eventually, certain Supplements, such as "Family Resources" and "Childhood Immunization", were incorporated in the NHIS Core on an annual basis.

However, the unintended result was an increasingly unwieldy survey instrument and longer interviewing sessions: recent questionnaires (Core and Supplements combined) ran almost 300 pages, while interviews averaged two hours. This imposed an unacceptable burden on NCHS staff, NHIS interviewers, the data collection budget, and, most importantly, on the NHIS respondents. Furthermore, the excessive length of NHIS interviews contributed to declines in both response rate and data quality. For all of these reasons, NCHS initiated a redesign of the NHIS questionnaire that was implemented in 1997.

NHIS Redesign: Questionnaire Changes

The redesigned NHIS has three parts or modules: a Basic Module; a Periodic Module; and a Topical Module. The Basic Module functions as the new Core questionnaire. It will remain largely unchanged from year to year and will allow for trends analysis and for data from more than one year to be pooled to increase the sample size for analytic purposes. The Basic Module contains three components: the Family Core, the Sample Adult Core, and the Sample Child Core. The Family Core component collects information on everyone in the family, and its sample also serves as a sampling frame for additional integrated surveys, as needed. Information collected on the Family Core for all family members includes: household composition and socio-demographic characteristics, tracking information, information for matches to administrative data bases, and basic indicators of health status and utilization of health care services. The Family Core yields three data files: the Household-Level file, the Family-Level file, and the Person-Level file.

From each family in the NHIS, one sample adult and one sample child (if any children under age 18 are present) are randomly selected, and information on each is collected with the Sample Adult

Core and the Sample Child Core questionnaires. Because some health issues are different for children and adults, these two questionnaires differ in some items, but both collect basic information on health status, health care services, and behavior. These sections of the survey yield the Sample Adult, Sample Child, and Child Immunization files.

Data Collection Procedures

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

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

The NHIS interviews traditionally were conducted using paper and pencil. The redesigned NHIS fielded since 1997 is conducted using computer-assisted personal interviewing (CAPI). The CAPI version of the NHIS questionnaire is administered using laptop computers, which allow interviewers to enter responses directly into the computer during the interviews. This computerized mode offers distinct advantages in terms of timeliness of the data and improved data quality.

Sample Design

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

black and Hispanic populations are now oversampled to allow for more precise estimation of health in these growing minority populations. In the previous design, only black Americans were oversampled.

Two other important features first implemented in the 1985-1994 design continue. NCHS survey integration and followback surveys are facilitated by an all-area frame with independent address lists; while the area frame is based on the preceding decennial Census, the address lists are not. Also, the NHIS sample is divided into four representative panels to further facilitate integration with other NCHS surveys. See NCHS Series 2, Number 130, for a description of the 1995-2004 survey design, the methods used in estimation, and general qualifications of the data obtained from the survey. This publication is available on-line at **http://www.cdc.gov/nchs/data/sr2_130.pdf** . (Users may also be interested in another Series 2 (number 126) report, *National Health Interview Survey: Research for the 1995-2004 Redesign*, which is available at **http://www.cdc.gov/nchs/data/sr2_126.pdf** .)

Weighting Information

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

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

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

Weight - Final Annual (WTFA) is based on design, ratio, non-response and post-stratification adjustments. This should be used in most analyses of the Family/Person data. National estimates of all person-level variables can be made using these weights.

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

The Sample Adult data file contains two sets of weights:

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

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

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

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

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

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

Weight - Final Annual (WTFA_IM) includes design, ratio, non-response and post-stratification adjustments for sample children under 18 years of age and additional children ages 12-35 months. This should be used in analyses for a full year of Immunization data.

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

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

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

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

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

NOTE: Analysts should be aware that 263 persons are on the Person-Level file who were active duty members of the Armed Forces at time of interview, despite the fact that NHIS covers only the civilian noninstitutionalized household population. The value of WTFA for these persons is zero, so they will not be counted when making national (i.e., weighted) prevalence 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 with a two-week recall, (variable)(26)(WTFA) = annual estimate; for a variable with a three-month recall, (variable)(4)(WTFA) = annual estimate. This assumes that the average rate of occurrence is the same over the last year as over the last two weeks (or three months). Analysts are cautioned to check the accompanying file documentation and the questionnaire in order to insure that annual estimates for these kinds of event variables are possible and have intrinsic meaning.

Variance Estimation

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

Weighted data used in standard software packages may provide unbiased estimators for commonly computed first-order statistics like means or regression coefficients, but the computed standard errors of the estimates may be too small. Also, standard packages may produce hypothesis test results (such as p values) that are incorrect. Hence, it is recommended that users of NHIS data utilize computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs. NCHS uses Taylor series linearization methods for NHIS variance estimation. Appendix III provides SUDAAN code and a description of its use to compute standard errors of means, percentages and totals with the 2000 NHIS database.

Analyses of large NHIS domains usually produce reliable estimates with stable variances, but analyses of small domains may yield unreliable estimates along with unstable variances. The analyst should pay particular attention to the coefficient of variation 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 variance estimate instability.

General Information About the 2000 Data

The interviewed sample for 2000 consisted of 38,633 households, which yielded 100,618 persons in 39,264 families. The interviewed sample for the Sample Adult component, which required self-response to all questions, was 32,374 persons 18 years or age and older. The interviewed sample for the Sample Child component, by proxy response from a knowledgeable adult in the family, was 13,376 children 0-17 years old. Lastly, the interviewed sample for the Immunization section, again, by proxy response from a knowledgeable adult in the family, was 14,618 children aged 17 and younger. Data were not collected on any infant who was born during the assignment week of the interview.

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

The conditional response rate for the Sample Adult component was 82.6% of persons identified as sample adults. The final response rate for the Adult Sample Person component is calculated as (Overall Family Response Rate)(Sample Adult Response Rate), or (87.3%)(82.6%) = 72.1%.

The conditional response rate for the Sample Child component was 90.9%, which was calculated by dividing the number of completed Sample Child interviews (13,376) by the total number of eligible sample children (14,711). The unconditional or final response rate for the Sample Child component was calculated by multiplying the conditional rate by the overall family response rate of 87.3%, yielding a rate of 79.4%.

The total noninterview rate for the Immunization (Child) section of the Sample Child component was 1.8% of persons asked to respond to the Immunization (Child) section. The final response rate for the Immunization (Child) file was 79.5%.

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

Information About the 2000 Data File Documentation

Along with the redesign of the NHIS questionnaire, other aspects of the data production process were also modified. As with the previous design, each data file has its own data documentation file (also known as a “codebook” or “data dictionary”). Beginning with the 1997 data (and continuing with subsequent years), the format and content of these files, henceforth referred to as Dataset Documentation, has changed. As a result, more specific information about

each variable is now available to users. For most variables, the Dataset Documentation now provides the actual question that generated the data, questionnaire location information, universe, values, value labels, and frequency counts. Additional specific information is provided under "Sources", "Recodes", "Keywords", and "Notes". These terms are defined as follows:

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

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

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

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

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

The universe refers to those respondents deemed eligible to answer a given question. For example, the universes for all Sample Adult variables are specified as "ASTATFLG = 1 and AGE = 18+", or "ASTATFLG eq 1 and AGE ge 18 ", followed by any other universe descriptors specific to the variable. ASTATFLG = 1 refers to a variable on the Person file and indicates that the respondent was selected as a sample adult and answered at least the first three sections of the

Sample Adult component (constituting a completed interview or an acceptable partial interview). Responses for persons who stopped answering key questions after the first three sections will appear as 8's (not ascertained) for the corresponding variables throughout the remaining Sample Adult variables.

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

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

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

Information About the 2000 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 2000 NHIS Field Representative's Manual (available on the NHIS web site, **http://www.cdc.gov/nchs/nhis.htm**). Every effort was made to insure that the variable names in the data are consistent with the question items in the instrument. In a few cases, this was not possible. When there is a question about variable names, matching the question number in the instrument to the variable number in the Dataset Documentation can resolve any discrepancies.

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

When errors were detected and diagnosed, and time permitted, the instrument was changed to correct for the errors. For example, there were two major versions of the instrument in 1998, fielded in quarters 1 and 3. In 2000, instrument changes were kept to a minimum, so that there was basically one version of the NHIS in the field across all four quarters of the survey year. Analysts are encouraged to read the notes in the Dataset Documentation for important information pertaining to specific variables.

Questionnaire Sections

The 2000 NHIS contained the annual Basic Module, which is broken into various sections that group questions into broad and specific categories. Each section is designated by a section title and corresponding three-digit acronym (or section code); questionnaire items are numbered sequentially (but not consecutively) within their respective sections, with the section acronym making up part of the item number. For example, the first item in the Household Composition section is identified as HHC.010; note that HHC.010 also has an associated variable name, RPNAME. The list on the following page details the various questionnaire sections, their acronyms and description titles.

Table 1. 2000 NHIS Core Questionnaire Sections and Topics

A. Household

Section No.	Section Code	Description
I	HHC	Household Composition

B. Family Core

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

C. Sample Adult Core

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

D. Sample Child Core

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

E. Recontact

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

In addition to the three Core sections comprising the Basic Module, the 2000 NHIS contains several other data files: the Household- and Family-level files, the Injury and Poison Episode file, and the Injury and Poison Verbatim file. The Household file is derived largely from the Household composition section of the Module and describes characteristics of each household. The variables contained in the Family-level file are reconstructions of the person-level data from the Basic Module sections at the family level. The Injury and Poison files are derived from the information obtained from the injury/poisoning questions in the Family Core section.

Changes/Additions/Deletions to 2000 Core

A number of changes were introduced to the Core sections of the 2000 NHIS, resulting in new, changed, or deleted variables (relative to 1999). Following is a brief summary. Users are strongly encouraged to check the notes in the Dataset Documentation, as well as the relevant sections in this document, for more information.

Family Core

Section	Variable	Brief Description of Change
HHC	ORIGIN_I*	Replaces ORIGIN
HHC	ORIGIMPT	Hispanic origin imputation flag
HHC	HISPAN_I*	Replaces HISPANCR
HHC	HISPIMPT	Hispanic origin imputation flag
HHC	RCDT1P_I*	Replaces RACDET_P
HHC	RC_SMP_I*	Replaces RC_SUM_P
HHC	RACERP_I*	Replaces RACER_P
HHC	RACEIMPT	Race imputation flag
HHC	MRACRP_I*	Replaces MRACER_P

Family Core, continued

Section	Variable	Brief Description of Change
HHC	MRACBP_I*	Replaces MRACBR_P
HHC	ERIMPFLG	Race imputation flag
HHC	RACREC_I*	Replaces RACERECR
HHC	HISCOD_I*	Replaces HISPCODR
Note: Users should refer to Appendix II for more information on the Race/Ethnicity variables.		
FIJ	All injury and poison variables on the Person file.	The 2000 NHIS Person file does not contain and injury or poisoning variables. See page 29 of this document for more information.
FAU	PHCPH2WR	Replaces PHCPH2W; question wording was modified
FAU	PHCPHN2W	Universe description modification due to question wording change in PHCPH2WR
FHI	HIKINDN	New variable for 2000 (no health insurance)
FHI	MCCHOICE	New variable for 2000 (indicates enrollment in Medicare Plus Choice)
FSD	REGIONBR	New region-of-birth recode
FIN	PINTRSTR	Replaces PINTRST; change in question wording

Sample Adult Core

Section	Variable	Brief Description of Change
ACN	DIBAGE	Changes in response categories
AHS	FLWALK-FLPUSH	Response category added ("do not do this activity")
AAU	AHCSYR2	Minor question wording modification
AAU	AHCNOYR2	Replaces AHCNOYR; minor question wording modification and response categories were expanded
AAU	APOX	New variable for 2000 (ever had chicken pox)
AAU	APOX12MO	New variable for 2000 (had chicken pox in past 12 months)
AAU	AHEP	New variable for 2000 (ever had hepatitis)
AAU	AHEPLIV	New variable for 2000 (ever live with someone with hepatitis)
AAU	SHTHEPB	New variable for 2000 (ever had hepatitis B vaccine)
AAU	SHEPDOS	New variable for 2000 (number of doses of hepatitis B vaccine)
ADS	HIVTST	Replaces AIDSTST (ever been tested for the AIDS virus infection?)
ADS	WHYTST_C	New labels (Main reason why not tested for HIV)
ADS	TST12M_M; TST12M_Y	Changed to month and Year field (month and year of last test for HIV)
ADS	TIMETST	New variable for 2000 (time period of most recent HIV test)

Sample Adult Core, continued

Section	Variable	Brief Description of Change
ADS	REATST_R	New labels with verbatim recodes (main reason for your last HIV test)
ADS	REASWHO	New variable for 2000 (who suggested you be tested?)
ADS	WHYREQ_C	New variable for 2000 (why were you required to get last HIV test?)
ADS	LASTST_C	Values and labels modified (where did you have your last HIV test?)
ADS	CLNTYP_C	New variable for 2000 (type of clinic for your last HIV test)
ADS	WHOADM	New variable for 2000 (was test admin. by nurse or used self sampling kit?)
ADS	GIVNAM	New variable for 2000 (last time tested, did you give your f/l name?)
ADS	STD	New variable for 2000 (had an STD other than HIV in past 5 yrs?)
ADS	STDDOC	New variable for 2000 (saw doctor the last time you had an STD?)
ADS	STDWHER	New variable for 2000 (where did you go to be checked?)
ADS	TBHRD	New variable for 2000 (have you ever heard of TB?)
ADS	TBKNOW	New variable for 2000 (know anyone who had TB?)
ADS	TB	New variable for 2000 (how much do you know about TB?)
ADS	TBSPRD_1 - TBSPRD_6	New variable for 2000 (how is TB spread?)
ADS	TBCURED	New variable for 2000 (can TB be cured?)
ADS	TBCHANC	New variable for 2000 (what are you chances of getting TB?)
ADS	HOMELESS	New variable for 2000 (ever spent 24+ hrs living on the streets/shelter/jail?)
ADS	TBSHAME	New variable for 2000 (feel ashamed if you/family member diagnosed with TB?)

Sample Child Core

Section	Variable	Brief Description of Change
CHS	ADD2	ADHD is added to the question on ADD
CAU	CDNLONGR	Children 1-17 years of age are asked this question (previously, information was only obtained for children 2-17)
CAU	CHERNOY2	Replaces CHERNOYR (expanded response categories)
CAU	CHCHNOY2	Replaces CHCHNOYR (expanded response categories)
CAU	CHCNOYR2	Replaces CHCNOYR (expanded response categories; also, dental visits are now specifically excluded from number of possible contacts)

Immunization

Section	Variable	Brief Description of Change
CIM	ICSTAT	Changed from a question into a check item
CIM	ROT	New variable for 2000; number of times had Rotavirus vaccine by mouth (for those with shot records)
CIM	ROTDT	New variable for 2000; shot dates of Rotavirus vaccine by mouth (for those with shot records)
CIM	PNEU	New variable for 2000; number of times had Pneumococcal vaccine (for those with shot records)

Note: Please refer to description of the CIM section in this document.

CIM	PNEDT	New variable for 2000; shot dates of Pneumococcal vaccine (for those with shot records)
CIM	ROTMOR	New variable for 2000; ever received additional Rotavirus vaccine by mouth (for those with shot records)
CIM	ROTMNO	New variable for 2000; how many additional Rotavirus vaccines received (for those with shot records)
CIM	PNEMOR	New variable for 2000; ever received an additional pneumonia shot (for those with shot records)
CIM	PNEMNO	New variable for 2000; how many additional Pneumococcal vaccines received (for those with shot records)
CIM	ROTEV	New variable for 2000; ever received a Rotavirus vaccine by mouth (for those without shot records)
CIM	ROTENO	New variable for 2000; how many Rotavirus vaccines ever received (for those without shot records)
CIM	PNEEV	New variable for 2000; ever received Pneumococcal vaccine for Qtrs.1-4 for the entire year (for those without shot records)
CIM	PNEENO	New variable for 2000; number of times ever received Pneumococcal vaccine for Qtrs.1-4 for the entire year (for those without shot records)

Injury and Poison Episode Files

Section	Variable	Brief Description of Change

The Poison Episode variables are now located on a combined Injury and Poisoning Episode file. See page 31 for more information.

	IJMED_2 - IJMED_7	New variables for 2000; where person received medical advice or treatment
	CAUSNEW	Replaces CAUS; new response categories
	ECAUSNEW	Response categories now include poisoning
	FALLNEW1	Replaces FALL_1; new response categories
	FALLNEW2	Replaces FALL_2; new response categories
	ANIMAL	New variable for 2000; type of animal/insect bite
	POITP	New response categories

WHERNEW1	Replaces WHER_1; new response categories
WHERNEW2	Replaces WHER_2; new response categories
WKLS	Changed order of categories
SCLS	Changed order of categories

* indicates a variable on multiple files (e.g., Person, Sample Adult, and/or Sample Child).

New NHIS Topics/Questions for 2000

In addition to the Core sections of the survey, the 2000 NHIS survey contains a Cancer Control Module that is similar to the Cancer Supplements in the 1987 and 1992 NHIS. The 2000 module consists of seven sections covering such topics as Hispanic acculturation, diet and nutrition, physical activity, tobacco, cancer screening, genetic testing, and family history. These sections are described in greater detail on pages 53-59. Users should note that the same respondents who served as sample adults for each household also participated in the Cancer Control Module.

2000 National Health Interview Survey
Household-Level File

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

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

The universe for the Household file is all eligible households, including both responding households and non-responding (Type A) households. The Household file contains information on 43,437 households: 38,633 households were interviewed, while 3,179 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 11.1% 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 88.9%.

2000 National Health Interview Survey
Family-Level File

The Family-Level file contains variables that describe characteristics of the 39,264 families living in households that participated in the 2000 NHIS. Thus, each record in the file represents a unique family. The universe for all variables in this file is limited to all responding families in those households participating in the 2000 survey; this is specified as FM = ALL in the Family file Dataset Documentation. Users wishing to determine the number of responding and non-responding families in each household are referred to ACPT_FAM and REJ_FAM in either the Family or Household files.

Most NHIS families consist of a group of two or more related persons who are living together in the same occupied housing unit (i.e., household) in the sample. All relationships are recorded relative to the household and, in cases with more than one family per household, the family reference person, who is the person, or one of the persons, who owns or rents the housing unit. (If more than one person owns or rents the housing unit, the oldest among them is designated the household reference person; if none of the household members owns or rents the unit, then the oldest person in the household is designated the reference person.) In some instances, unrelated persons sharing the same household may also be considered as one family, such as unmarried couples (same-sex or opposite-sex couples) who are living together. In contrast, an individual living alone or, alternatively, each member of a group of unrelated individuals living in the same household as roommates would be considered as a separate family relative to the household reference person. Additional groups of persons living in the household who are related to each other, but not to the reference person, are also considered to be separate families; for example, a lodger and his/her family, or a live-in household employee and his/her spouse, or a single boarder with no one related to him/her living in the household. Hence, there may be more than one family living in a single household. If this is the case, the various NHIS questionnaires (e.g., Family Core, Sample Adult Core, etc.) will be administered separately to *each* family within the sampled household.

Family size may vary considerably. Table 2 shows a breakdown of the 39,264 families by number of family members. Again, note that multiple families may share one household; this information can be determined from ACPT_FAM and REJ_FAM, or HHX and FMX, in the Family file.

The first part of the Family file contains the technical variables that identify or describe the record type (all observations in this file have a record type value of “60”), the survey year, the household and family numbers, the interview quarter and week, characteristics of the family’s housing unit, geographic information associated with the housing unit, variables used for variance estimation, and a family-level weight variable. The second part of the file consists of a series of recodes derived from the six family core sections of the NHIS that collapse the 98,000+ individual-level observations into information about their respective families. Generally, two types of recodes are possible. The first is a simple “yes-no” measure that indicates whether *any* family member falls into a particular category or exhibits a particular characteristic. These variables are equivalent to, but not directly derived from, the family-level questions in the instrument.

Table 2. Size of Family, 2000 National Health Interview Survey (unweighted counts)

Number of Members	Frequency	Percent
1	11,400	29.0
2	11,763	30.0
3	6,172	15.7
4	5,537	14.1
5	2,664	6.8
6	1,039	2.6
7	362	0.9
8	168	0.4
9	88	0.2
10	38	0.1
11	14	0.0
12	2	0.0
13	4	0.0
14	6	0.0
15	1	0.0
16	1	0.0
18	2	0.0
21	1	0.0

Every yes-no measure also has a corresponding counter that indicates the number of family members in that category or with that characteristic. Note that counters always consist of values from zero to 30; in addition, blanks are also possible if a family is not contained in the universe for a specific question. For example, FSALYN and FSALCT, two recodes from the Income and Assets section of the Family Core, are limited to families with at least one member aged 18 or older; families consisting solely of emancipated minor(s) are coded as blanks to indicate that they are out of the universe. The Family file also contains some counters that lack corresponding yes-no indicators. FHSTATEX, FHSTATVG, FHSTATG, FHSTATFR, and FHSTATPR (all derived from PHSTAT, FHS.310) provide counts of the number of family members in excellent, very good, good, fair, and poor health, respectively. Counters were also constructed to indicate the number of working adults in the family, the number of adults in the family looking for work, the number of adults working full-time, the number of children (under age 18) in the family, and the number of family members aged 65 and older.

Because most of the variables in the Family file are recodes of the person-level variables in the family core, the sum of the number of persons across all families in each family-level counter should be equivalent to the number of "yes" responses in its person-level source. Returning to our previous example, consider FSALCT: 15,204 families have one member receiving income from wages/salary, 11,676 families have two members (or 2(11,676) = 23,352 persons) with wage/salary income, 1,812 families have three members (or 3(1,812) = 5,436 persons), 430 families have four members (or 1,720 persons), 72 families have five members (or 360 persons), and 18 families have six members with wage/salary income in 1999 (108 persons). Thus the sum of persons across the 28,465 families answering "yes" to FSALYN, the associated yes-no indicator, is 46,180 (15,204 +

23,352 + 5,436 + 1,720 + 360 + 108), which is equivalent to the 46,180 "yes" responses to the person-level source variable, PSAL. Users are advised to check the Dataset Documentation for each Family file recode in order to determine its person-level source variable.

The 2000 NHIS Family file contains two variables describing family type and structure in both general and detailed terms. FMTYPE, which was first included in the 1997 Family file, consists of just four categories, and represents an initial classification of families according to the numbers of adults and children that are present. Based on FMTYPE, FMSTR2 further categorizes families according to familial relationships. If FMTYPE is equivalent to "1" (an adult and no children under age 18 present) then FMSTR2 is either "11" (living alone) or "12" (living with one or more roommates, all of which would be distinct families). If FMTYPE is equal to "2" (multiple adults, no children present), then FMSTR2 is equivalent to either "21" (a married couple), "22" (an unmarried couple), or "23" (all other adult-only families). Note that the last category would include married or unmarried couples living with a related adult (a parent or adult sibling, for example). If FMTYPE is equal to "3" (one adult and at least one child present), then FMSTR2 is equal to either "31" (a single mother and her biological, adoptive, step, or foster child), "32" (a single father and his biological, adoptive, step, or foster child), or "33" (a single adult and unrelated child). Lastly, if FMTYPE is equivalent to "4" (multiple adults and at least one child under 18 present), then FMSTR2 is classified as either "41" (married parents and their biological or adoptive child(ren)), "42" (cohabiting parents and child(ren)), "43" (parent, step-parent, and child(ren)), "44" (at least one biological/adoptive parent and child(ren), and a related adult), or "45" (other related or unrelated adults and child(ren) who is/are not biological/adoptive son or daughter of the adults present, such as a child living with his grandparents, or a child living with his grandmother and her unmarried partner). Families that could not be classified according to this schema are coded "99". Finally, please note that emancipated minors are treated as adults with respect to both FMTYPE and FMSTR2, despite the fact that they may be under 18 years of age.

Users familiar with the 1998 Family file will recall that the family structure variable in that file was called FMSTRCT2. This variable coded married *and unmarried* parents living with their biological child (or children) in one category, parent/step-parent/child families in a second category, and parent/cohabiting partner/child families in a third category. A change was made in the way the 1999 and 2000 NHIS obtained family relationship data, such that we are now unable to distinguish families with two cohabiting parents living with their biological child (or children) from families with a child and biological parent who are living with the parent's unmarried partner. As a result, the 1999-2000 family structure variable includes all cohabiting couple families in the same category (FMSTR2 = 42), regardless of the adults' relationships to the child(ren) in the family. The 1999-2000 family structure variable has a new name (FMSTR2) in order to distinguish it from the previous variable (FMSTRCT2).

The Family File Weight

The 2000 NHIS Family file can be thought of as a household-level file of sorts for all families; as such, it is very similar to a household-level file. The ideal situation for creating weights for such a family file would be to use independent estimates of the number of households or families

from a reliable source, such as the U.S. Census Bureau, to perform post-stratification adjustments in a manner similar to what is done for the NHIS Person file weight. Unfortunately, no suitable independent estimates exist.

Due to the lack of appropriate independent estimates, a variation of the "principal person" method is used to create the 2000 NHIS Family file weight (WTFA_FAM). This method is similar to that used in the Current Population Survey to create their household- and family-level weights. Briefly, a person-level ratio adjustment is used as a proxy for the NHIS family-level ratio adjustment. Use of the person weight with the *smallest* ratio adjustment within each family (that is, the smallest post-stratification factor between the interim and final person weights within each family) is believed to provide a more accurate estimate of the total number of U.S. families than either the use of other person weights in the family or the use of no ratio adjustments whatsoever.

Accordingly, the weight provided with the 2000 NHIS Family file, WTFA_FAM, corresponds to the 2000 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.

2000 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 2000 NHIS. The information in the Family Core questionnaire is collected for all household members. Any adult household members who are present at the time of the interview may take part; information regarding adults not participating in the interview, as well as all household members under age 18, is provided by a knowledgeable adult member of the household. The six sections comprising the Family Core are discussed in greater detail below.

Beginning in 1998, the NHIS introduced a number of new variables describing family relationships; these are continued in 2000 The family structure variables developed for the Family file (i.e., FMTYPE and FMSTR2; please refer to page 22 of this document) were added to each person's record, and are called FM_TYPE and FM_STRP. Also included for 2000 are variables identifying the person number of the respondent's mother or father (called MOTHER and FATHER, respectively), if a mother or father is present in the household, or guardian (called GUARD), as well as variables describing the degree of relationship (biological, adoptive, step, foster, or in-law) between parents and children (i.e., MOM_DEG and DAD_DEG), and adult siblings (SIB_DEG). Note that in the case of the last variable, one sibling must be the family reference person. Lastly, PARENTS, MOM_ED, and DAD_ED are included on the Person file. PARENTS indicates the number of parents present in the family (regardless of the respondent's age), while MOM_ED and DAD_ED provide information on completed education of mother or father, respectively, for all children under 18 years of age. (Note that in 1997, PARENTS, MOM_ED, and DAD_ED were available only on the Sample Child file.)

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

The Health Status and Limitation of Activity (FHS) section of the Family Core for the 2000 NHIS contains information addressing respondent-assessed disabilities, disability-associated conditions, and overall health status for all family members. Users should note that additional information on health status and disability is also included in other sections of the Sample Adult file, as well as the Sample Child file.

Limitation of Activity at the Person-Level

Information on activity limitations, including questions about work limitations, the need for personal assistance with personal care needs such as eating, bathing, dressing, getting around inside the home, and the need for personal assistance with handling routine needs such as everyday household chores, doing necessary business, shopping or running errands, is collected for each family member (with some exclusions for children and youth). If any limitations are identified, respondents are asked to specify the health condition(s) causing the limitation(s) and indicate how long they have had each condition. The questions in the 2000 NHIS Family Core regarding activity limitations due to physical, mental or emotional problems are comparable to the 1997-1999 NHIS (with minor

exceptions discussed below) and are substantively comparable to previous NHIS surveys (with noteworthy changes in question wording).

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

The 2000 data on activity limitations were processed in the same manner as similar data from the 1999 NHIS. The 1999 and 2000 data processing differed from the 1997 and 1998 protocols in minor ways. (See the Survey Description documents from these years for more information.)

Conditions

For each family member with a previously mentioned limitation, the respondent is asked about the condition or health problem associated with that limitation, as well as the amount of time he/she has had the condition. Respondents are then handed one of two flash card listing various condition categories. These categories are broad in scope, and vary according to age. Family members under age 18 were subject to the following fixed condition categories listed on the first flash card: "vision/problem seeing", "hearing problem", "speech problem", "asthma/breathing problem", "birth defect", "injury", "mental retardation", "other developmental problem (e.g., cerebral palsy)", "other mental, emotional, or behavioral problem", "bone, joint, or muscle problem", "epilepsy", and two instances of "other impairment problem". The fixed response categories in the instrument for adults were equally broad, and comprise the conditions listed on the second flash card: "vision/problem seeing", "hearing problem", "arthritis/rheumatism", "back or neck problem", "fractures, bone/joint injury", "other injury", "heart problem", "stroke problem", "hypertension/high blood pressure", "diabetes", "lung/breathing problem", "cancer", "birth defect", "mental retardation", "other developmental problem (e.g., cerebral palsy)", "senility", "depression/anxiety/emotional problem", "weight problem", and two instances of "other impairment problem". Respondents could supply a verbatim response for the "other impairment problem". Beginning in 1999 and continuing in 2000, verbatim responses could be entered in a longer, 50-character field. Also, as in 1999, respondents were no longer restricted to listing only five conditions.

The verbatim responses recorded by FRs in one or both of the 50-character fields indicating "other impairment problem" were subsequently analyzed during data processing. While most respondents named conditions that did not fall into the fixed response categories as originally specified in the instrument, some respondents named conditions that could be included in one of the fixed categories. In the latter case, these respondents were assigned codes corresponding to the original response categories (the first 11 for children, and the first 18 for adults). For adults, an

additional 16 *ad hoc* categories were created during coding to group responses that fell outside the original 18 condition categories included in the instrument: these ad hoc categories were assigned numbers 19-34. Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the new categories remained in the "other impairment problem" categories, were renumbered "90" and "91" for both children and adults. The resulting 36 categories for adults and 13 categories for children were generally informed by the International Classification of Diseases, Ninth Revision, Clinical Modification (see Table 3, below; note that the ICD-9-CM codes shown in this table are *not* included on the data file). These specific condition categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any activity (a "mention-not mention" format). Because the 16 adult *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions causing limitations will be underestimated. Therefore, these variables should be analyzed with care. Moreover, none of the FHS condition variables (the 11 child variables, LAHCC1 through LAHCC11, and the 34 adult variables, LAHCA1 through LAHCA34) should be used to estimate prevalence for the conditions they represent, because only those persons with a previously reported limitation were eligible for the condition questions that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult and Child Cores.

Recodes

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

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

Also, a series of age-group-specific recodes (e.g., under 18 versus 18 and over) regarding conditions limiting activity and duration of limiting conditions have been created. Because the questions about limitation of activity in the redesigned NHIS are asked in a different context, of differing age groups, and 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. However, the use of these new questions and recodes should enable researchers at NCHS to define new categories identifying the extent of limitation.

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 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 three months or longer to be considered chronic. An exception is made for children less than one year of age who have had a condition "since birth", as these conditions are considered chronic. Because the presence of a limitation determined whether persons were eligible for the condition questions and the chronicity recodes, we caution data users that these variables should *not* be used to produce estimates of the prevalence rates of chronic conditions. These data should be analyzed and interpreted with care.

Table 3. FHS Categories with Approximate ICD-9-CM Ranges

A. Codes for Adults, aged 18+

NHIS Category	**ICD-9-CM Codes**
1 - Vision or seeing problem	360-379
2 - Hearing problem	387-389
3 - Arthritis / rheumatism	711-712, 714-716, 720.0, 721, 729.0
4 - Back or neck problem	722-724, 732.0, 737
5 - Fractures, bone or joint injury *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

NHIS Category	ICD-9-CM Codes
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
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*	

NHIS Category	ICD-9-CM Codes
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 NEC *1st other-specify verbatim, not elsewhere classified*	
91 - Others NEC *2nd other-specify verbatim, not elsewhere classified*	
B. Codes for Children, aged < 18	
1 - Vision or seeing problem	360-379
2 - Hearing problem	387-389
3 - Speech problem	307.0, 307.9, 315.3, 784.3, 784.5
4 - Asthma or breathing problem	460- 461, 465-466, 470-471, 473, 477, 480-487, 490-496, 500-508, 510-519
5 - Birth defect *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 *Includes learning disabilities*	315, 343, 783.4
9 - Other mental, emotional, or behavioral problem *Includes ADD, ADHD, and hyperactivity*	290-314, 799.2, V15.4
10 - Bone, joint or muscle problem	710-739
11 - Epilepsy and seizures	345, 779.0, 780.3
90 - Others NEC *1st other-specify verbatim that does not fit in any other category*	
91 - Others NEC *2nd other-specify verbatim that does not fit in any other category*	

II. Injury Section (FIJ)

As with the 1997-1999 NHIS, the Family Core portion of the 2000 survey included questions about medically attended injuries and poisoning episodes that occurred to any member of the family within a three-month reference period. All injury and poisoning information was provided

by the family respondent. Two data files containing injury and poisoning information were derived from the 2000 NHIS: (A) the Injury/Poisoning Episode file and (B) the Verbatim Injury/Poisoning Episode file. Note that a separate Poisoning Episode file is no longer available. Moreover, the 2000 NHIS Person file does **not** contain injury and poisoning data.

Between 1999 and 2000, the Family Core Injury Section of the NHIS was extensively redesigned. These changes included combining the injury and poisoning questions, revising response categories for some existing questions, adding a question about where medical advice or treatment was received, adding a question about animal/insect bites, and deleting questions about drownings and/or water-related injuries and firearm-related injuries.

When changes to the section were made, a key question was inadvertently reworded. The question asking respondents "How many different times in the past three months were you injured or poisoned seriously enough to seek medical advice or treatment?" was mistakenly changed to "How many times in the past three months did you seek medical advice because you were injured or poisoned?". As a result, respondents gave the number of times they sought medical advice for an injury or poisoning during the three months prior to the interview, rather than the number of times they were injured or poisoned during the three months prior to the interview. This may account for the fact that there are approximately 4.5 million fewer injuries and 460,000 fewer poisonings in 2000 than in 1999. The wording for this question was changed back to the former version in the 2001 survey. The final results from the 2001 NHIS will enable us to assess whether the decline in injuries and poisonings observed in 2000 was real, or simply the result of the change in the question text. Users familiar with the NHIS injury data are no doubt aware of the continual decline in the overall number of injuries reported since the injury and poisoning section was added to the NHIS in 1997. Early evidence from *preliminary* 2001 data suggest that the number of episodes is continuing to decline, despite a return to the preferred question wording. Thus, the decline may not be solely related to the wording of the question, but other factors may be involved, such as declining response rates, other changes in the section, and a general reluctance on the part of respondents to answer detailed questions about injuries.

Consequently, NHIS staff made the decision not to include the injury and poisoning data on the 2000 Person file. However, injury and poisoning data will continue to be available to users on the Injury/Poisoning Episode file and the Verbatim Injury/Poisoning Episode file. **Users are strongly cautioned that any national prevalence estimates derived from the 2000 NHIS injury and poisoning episode data will be underestimated.** Despite the significant decline in the overall number of injury and poisoning episodes in 2000, the percentage distributions of episodes for many injury-related variables are comparable across 1997-2000, such as the external cause of the injury, the nature of the injury, the injured person's activity at the time of injury, where the injury happened (e.g., inside the home, outside the home, at work, etc.) and any limitation of activity resulting from the injury.

A. Injury/Poisoning Episode File

The Injury/Poisoning Episode file is an episode-based file: each injury and poisoning episode occurred during the three months prior to the interview, and resulted in one or more conditions. An injury episode refers to the traumatic event in which the person was injured one or more times from an external cause (e.g., a fall, a motor vehicle traffic accident). An injury condition is the acute condition or the physical harm caused by the traumatic event. Likewise, a poisoning episode refers to the traumatic event resulting from ingestion of or contact with harmful substances, as well as overdoses or wrong use of any drug or medication, while a poisoning is the acute condition or the physical harm caused by the traumatic event. A person may have up to ten injury and/or poisoning episodes and will appear in this file as many times as he/she has unique injury and/or poisoning episodes. Each episode must have at least one injury condition or poisoning classified according to the nature-of-injury codes 800-999 in the Ninth Revision of the International Classification of Diseases (ICD-9-CM). Other health conditions that were reported as occurring with the injury or poisoning, even if they are not classified according to the nature-of-injury code numbers 800-999 (e.g. mononeuritis of unspecified site (355.9), other symptoms referable to back (724.8)), are also included in the Injury/Poisoning Episode file. The resulting file contains information about the cause of the injury or poisoning episode, what the person was doing at the time of the injury or poisoning episode, the date and place of occurrence, the elapsed time between the date of the injury or poisoning episode and the date of the interview, whether the person was hospitalized, whether the person missed any days from work or school due to the injury or poisoning, whether the injury or poisoning episode caused any limitation of activity, ICD-9-CM diagnostic codes, and ICD-9-CM external cause codes. Information about how the injury or poisoning happened, the body part injured or poisoned, the type of injury or poisoning, along with responses to questions about specific types of injury or poisoning episodes, place of occurrence, and activity were used to assign ICD-9-CM diagnostic and external cause codes for all injury and poisoning episodes.

During the editing process, some injury and poisoning episodes were removed. These included episodes with no information, episodes that did not occur within the reference period, duplicate episodes, and injury episodes consisting solely of health conditions that could not be classified according to nature-of-injury codes 800-999 of the Ninth Revision of the International Classification of Diseases (ICD-9-CM).

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

This file only contains information about injury, poisoning, food poisoning, and allergic reaction episodes. Other person-level information can be obtained by linking the Injury/Poisoning Episode file to other 2000 NHIS data files using the household serial number (HHX), family serial number (FMX), and person number (PX).

Recall Period and Weights

Questions in the Injury section of the 2000 NHIS have a recall period of the "last 3 months". To calculate estimates of the number of injury or poisoning episodes occurring annually, each three-month estimate should be multiplied by 4. It is not possible to estimate the number of *people* injured or poisoned annually using the NHIS. Although the number of persons who were injured or poisoned during the three-month recall period is known, this number cannot be assumed to be uniform over a twelve-month period. On the other hand, it is appropriate to estimate the number of injuries or poisonings over the twelve-month period (by multiplying the 3-month estimate by 4) because that figure is the same whether or not individuals had multiple injuries or poisonings. Analysts are cautioned to check the Dataset Documentation and the specific item in the questionnaire in order to insure that annual estimates for these kinds of injury or poison episodes are possible and have intrinsic meaning.

Variance Estimation

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

Technical Notes

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

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

Additionally, in the case of some injury and poisoning episodes, the respondent was only able to provide a month and year of occurrence. In these cases, it was only possible to determine whether

the injury or poisoning episode occurred during the same month as the interview or within a certain number of months. For injury and poison episodes that occurred during the same month as the interview, this means that the amount of time between the date of the injury or poisoning episode and the date of the interview could be anywhere from one day to 30 days. For injury and poisoning episodes that occurred during the month before the interview, this means that the amount of time between the date of the injury or poisoning episode and the date of the interview could be anywhere from one day to 60 days. The amount of time between the date of the injury or poisoning episode and the date of the interview in months is only used when the day of the injury or poisoning episode was not given.

B. Verbatim Injury/Poisoning Episode File

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

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

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

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

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

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

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

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

Technical Notes

Due to the way in which IJBODY1, IJBODY2, IJBODY3, IJBODY4 (all referring to the body part injured or poisoned), IJKIND1, IJKIND2, IJKIND3, IJKIND4 (all referring to the kind of injury or poisoning), and IJHOW1, IJHOW2, IJHOW3, IJHOW4 (all referring to how the injury or poisoning happened) were recorded, the information contained in these variables may not

correspond exactly to the ICD-9-CM codes (ICD9_1, ICD9_2, ICD9_3, ICD9_4) and E codes (ECODE_1, ECODE_2, ECODE_3) associated with a given injury or poisoning episode. For example, the body part or body parts listed in IJBODY1 (question FIJ.050) may not be the one or ones listed in ICD9_1. When comparing verbatim injury/poisoning episode information and ICD-9-CM codes and E codes, it is better to look at all the information together rather than to try to match first body part field with first ICD-9-CM code field, second body part field with second ICD-9-CM code field, etc.

III. Health Care Access and Utilization Section (FAU)

The Health Care Access and Utilization (FAU) data from the Family Core of the 2000 NHIS contain information addressing access to health care and utilization services. The FAU section has remained largely unchanged since 1997 and consists of three parts: Part A, Access to Care; Part B, Hospital Utilization; and Part C, Health Care Contacts. The data items for Access to Care from 1997 onward differ from earlier years, with the exception of two questions: delay in receiving medical care (PDMED12M), and not receiving needed medical care due to cost (PNMED12M). The data items for Hospital Utilization are similar to those questions from the Hospital Probe and Hospital Page contained in the core questionnaire prior to 1997. In addition, the data items for Health Care Contacts are similar to the 2-week doctor visit probe questions from years prior to 1997 and include visits from medical doctors as well as other health care professionals.

In NHIS surveys prior to the 1997 redesign, questions about physician contacts, office visits, and home care included only contacts and visits to medical doctors or health care professionals working with or for a medical doctor. In addition, previous surveys included home care visits in the same category -- and thus in the same question -- as visits to or contacts with a doctor's office, hospital, etc. In contrast, the redesigned NHIS (1997-2000) distinguishes between home care and office visits, and includes separate questions for both. Moreover, beginning in 1997 the new instrument allowed respondents to consider an expanded list of health care professionals; 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, new for 1997 and continuing through 2000 is a question (P10DVYR) asking about 10 or more visits to doctors or other health care professionals in the last 12 months.

Health care utilization estimates based on the 2000 NHIS (as well as 1997 through 1999) may differ from those for earlier years of the NHIS due to changes in the questions and/or the context of the questions. Thus, estimates of health care utilization based on these data may not be comparable to estimates from previous years. For example, the estimated proportion of persons reporting one or more telephone contacts with a health professional in the past two weeks is higher than estimates from years prior to the 1997 NHIS. Users are advised to compare 1997-2000 NHIS questionnaire items pertaining to health care utilization to those used in previous NHIS surveys.

The question wording for the variable PHCPH2W, which asks about receiving medical advice over the telephone, was modified in the 2000 NHIS survey, such that billing questions and prescription refills were specifically excluded. Consequently, the name of the output variable was changed to PHCPH2WR. Because PHCPH2WR determines the universe for the next variable in the section, PHCPH2NW, the universe description for latter variable was modified accordingly. Beginning in 1999 and continuing in 2000, "excluding telephone calls" was added to the dependent person question (P10DVYR) "Who received care 10 or more times?"; however, it was not part of the family lead-in question. Analysts are advised to read the notes in the Dataset Documentation for further information pertaining to these changes.

IV. Health Insurance Section (FHI)

The Health Insurance section of the 2000 NHIS Family Core has a full range of data items addressing health insurance. The questions pertaining to health insurance programs covered by this section are similar to those asked on the 1993-96 NHIS Health Insurance Supplements and the 1997-1999 NHIS Family Core.

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

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

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

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

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

Types of single service plans;

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

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

The 2000 FHI data contain several modifications, as well as some new variables. The beginning of the FHI portion of the instrument was restructured to allow an active response to the question pertaining to health insurance coverage status for each individual family member. As a result, the HIKIND item now includes a category (HIKINDN) that allows the respondent to

indicate whether each family member has health insurance coverage. Health insurance coverage status is subsequently verified for all family members (instead of just persons deemed not to have coverage) with HICHANGE (FHI.075). (HICHANGE is not available on the public use data file, but HIKINDN is included.) Moreover, a new question was added to the Medicare detail (MCCHOICE), which provides information on whether the Medicare recipient is enrolled in a Medicare + Choice Plan. Lastly, the variables HICHECK and PHICOV have been removed from the 2000 instrument.

Technical Notes

During the course of data editing, it was determined that some respondents indicated plans (in response to the questions HIPNAM, NEXTPNM, NEXTPNM2, and NEXTPNM3) that were not private health insurance plans, or were single service plans that were excluded from the private health insurance coverage category. These respondents were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, CHIP, IHS, MILITARY, OTHERPUB, OTHERGOV and SINGLE. Similarly, in looking at the verbatim responses to the question STNAME asking respondents for the name of their CHIP, state sponsored or other government coverage, it was found that some respondents indicated plans and names of programs that were clearly private health insurance, Medicare, Medicaid, military coverage, Indian Health Service, single service plans or no coverage at all. Persons with these forms of coverage were reassigned to the appropriate enrollment recodes for MEDICARE, MEDICAID, PRIVATE, IHS, MILITARY and SINGLE. Additionally, respondents who answered "other state sponsored" or "other government coverage" who were subsequently determined through the STNAME field to be covered by the Children's Health Insurance Program were assigned to the CHIP recode. Also, some respondents offering an "other" response to the survey item (HISTOP@SPC) that inquired about the reason(s) their coverage stopped subsequently indicated in their verbatim responses that they did in fact have health insurance. These persons were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, CHIP, PRIVATE, IHS, MILITARY, OTHERPUB and OTHERGOV. Analysts are therefore strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, CHIP, IHS, MILITARY, OTHERPUB, OTHERGOV, and SINGLE for types of health care coverage, because these take into account the above-mentioned back edits. In contrast, the data contained in HIKINDA-HIKINDN were not back-edited and reflect the respondents' original replies. In addition, a recode (NOTCOV) is included in the data file that reflects the definition of non coverage as used in *Health, United States* (i.e., persons with *only* Indian Health Service coverage are considered uninsured).

As a result of the complicated editing process that takes place in the FHI section, the variables HILAST and HINOTYR, which reflect periods of non coverage, cannot be used to estimate the rate of uninsurance. Users should derive such estimates from NOTCOV (if they do not count IHS as coverage) or, alternatively, the health insurance recodes (MEDICARE, MEDICAID, PRIVATE, CHIP, IHS, MILITARY, OTHERPUB, and OTHERGOV). Using the most conservative estimate of the uninsured (which would exclude persons with IHS coverage only), a total of 1,433 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 1,061 respondents were not asked either the HILAST or the HINOTYR questions.

V. Socio-demographic Section (FSD)

The Socio-demographic (FSD) section of the Family Core in the 2000 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 1999, an estimate of their earnings from wages in 1999, and whether their employer provided health insurance.

New for 2000 is a "region of birth" recode (REGIONBR), which categorizes all respondents into one of 12 categories, depending on their country of origin. The CIA on-line World Factbook (**http://www.cia.gov/cia/publications/factbook/index.html**) was used to place countries into the regional categories shown below. Note that respondents born in Canada were included in the "Elsewhere" category in order to satisfy NCHS confidentiality requirements.

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

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

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

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

Technical Notes

Editing procedures have reconciled inconsistencies between DOINGLW and the relevant occupation and work-related variables in the Sample Adult Demographics section (ASD), but no such corrections are possible for non-sample adults. Moreover, data captured in DOINGLW have not been reconciled with other variables pertaining to the same subject matter within other sections of the data file. The "major activity" variable available in pre-1997 versions of the NHIS is now roughly approximated by a recode, MAJR_ACT, which sorts adults (persons aged 18 and older) into five categories: working at a job or business, keeping house, going to school, "something else" (e.g., retirement), and unknown. Users are advised that MAJR_ACT is derived in an entirely different fashion than the major activity variable available prior to 1997; comparable frequencies should not be expected. In particular, the frequency for those "keeping house" is low relative to data obtained prior to the 1997 NHIS redesign. The earlier instrument asked respondents for their primary activity very early in the interview, and well before the work/employment section; also, "keeping house" was an acceptable response category, along with "working". With the redesign, however, respondents could only "keep house" if they had previously said that they were not working. In effect, "keeping house" is a residual category of sorts: as a result of this change, the count of persons keeping house according to the redesigned NHIS is much lower.

Respondents who were not born in the United States were asked the year in which they came to the United States to stay. Respondents who could not recall or refused to answer were

subsequently asked to estimate the number of years they had been in the United States. This information was combined to create a recode that indicates how long these respondents have been living in the United States (YRSINUS). Regarding USBORN_P, respondents born in Puerto Rico, Guam, and other outlying territories of the United States are included in response category "2"; that is, they were not born in one of the fifty United States or the District of Columbia. The 2000 data also contain a citizenship recode that distinguishes between U.S. citizens and non-citizens.

VI. Income and Assets Section (FIN)

The Income and Assets (FIN) section of the Family Core contains information regarding a variety of income sources, as well as estimates of total combined family income and home tenure status. Many of the Income and Assets questions in this section have appeared in previous NHIS supplements (e.g., Family Resources). However, with the redesign of the 1997 NHIS, "Income and Assets" is now a permanent part of the Basic Module.

Respondents were first asked whether anyone in the family received income from a particular source. If a "yes" answer was obtained for any source, the respondent was 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 1999, and their home tenure status. The basic universe for most questions is "All families"; however, note that some universes for several questions (most importantly, PSAL, PSEINC, and PWIC) are further limited with respect to age (of family members). All variables in the Income and Assets section were converted from the family level to the person level during the editing process (i.e., the information in the respondent's record was transferred to the records for each family member).

Sources of Income

The first two questions in the section ask about income from wages and salary, and from self-employment (business or farm) for family members 18 years of age and older. Subsequent questions are not limited to adult family members. Respondents were asked about income from Social Security or Railroad Retirement (including that which was received as a disability benefit); other pensions; Supplemental Security Income (SSI); Welfare/Temporary Assistance to Needy Families (TANF); other kinds of government assistance (e.g., job training or placement, transportation assistance, or child care); interest from savings or other bank accounts; dividends from stocks, mutual funds, and/or net rental income from property, royalties, estates or trusts; child support payments; and other income sources (the question asked of respondents specifically mentioned alimony, contributions from family or friends, VA payments, Worker's Compensation, and Unemployment Compensation as possible sources of "other" income). Respondents are told at the start of the Income and Assets section that all questions are seeking information about possible income sources in the previous calendar year (i.e., 1999).

It should be noted that the wording for FIN.170 changed from 1999 to 2000. In 2000, the wording for the question was, "Did {you/any family members living here} receive income from saving accounts, money market funds, treasury notes, IRA's or certificates of deposit, interest bearing checking accounts, bonds or any other investment that earn interest?". However in 1999 the wording for FIN.170 was "Did {you/any family members living here} receive interest from savings

or other bank accounts?". As a result of this change in wording, the resulting output variable has a new name, PINTRSTR (it was previously called PINTRST).

Amounts and Home Ownership

In previous years, NHIS obtained information about the amount of income received from each financial source, but that was dropped in the redesigned NHIS in favor of a single overall estimate of combined family income. And, unlike previous NHIS surveys, the redesigned instrument contained three questions to elucidate the family's combined income from all sources during the previous calendar year, including a question (FIN.250) that allowed the respondent to supply a specific dollar amount (up to $999,995). Any family income responses greater than $999,995 were entered as $999,996. Respondents who did not know or refused to give a dollar amount to this question were then asked if their total combined family income for the previous year was $20,000 or more, OR less than $20,000 (FIN.260). If the respondent answered this question, he/she was then given a flash card and asked to indicate which income group listed on the card best represented the family's combined income during the previous year (FIN.270). Information from these three variables was combined into an income recode (INCGRP) that uses 13 categories to describe the family's income, as well as a second recode (AB_BL20K) that indicates all families at or above $20,000 or below $20,000.

Additionally, a more detailed indicator of poverty status was created by utilizing published information from the U.S. Census Bureau regarding 1999 poverty thresholds (see *Poverty in the United States, 1999*; U.S. Census Bureau). A ratio of the 1999 income value reported by respondents to the poverty threshold for the same year was constructed, given information on the family's overall size as well as the number of children aged 17 and under present in the family. The resulting ratio was subsequently ordered into a poverty gradient consisting of 14 categories (RAT_CAT). Users should note that the universe for this variable is considered to be all families, because the initial income question was asked of all families. However, the income-to-poverty ratios and resulting RAT_CAT values could not be calculated in two cases: when families simply did not supply adequate income information (e.g., those who would only indicate that their income was above or below $20,000, as well as those who declined to give any income information whatsoever), or those families where the number of children aged 17 or under *equaled* the overall number of family members (these observations are coded "99" and "96", respectively, on RAT_CAT). Respondents were also asked whether the family's house or apartment was owned or being bought, rented, or occupied by some other arrangement. If the family was renting the current residence, the question asked was whether the family was paying lower rent due to governmental rental assistance (FIN.282) .

Program Participation

Respondents were asked in the final part of the FIN section if any family members were authorized to receive food stamps in 1999, and if so, which members. In addition, respondents were asked whether any family member(s) had *ever* applied for Supplemental Security (SSI) or Social Security disability benefits (even if the claim(s) had been denied). It should be noted that most of the respondents who refused/didn't know if they were receiving SSI were not asked if they have ever applied for SSI (FIN.310). This was due to a CAPI initialization error that was not corrected until 2001. These individuals are included in the "not ascertained" response category.

Lastly, if one or more family members had received food stamps or Temporary Assistance to Needy Families (TANF), the respondent was asked, in two separate questions, for how many months during the last calendar year were food stamps and/or TANF provided.

It should be noted that the responses to FIN.385, PWIC, are only shown for interview quarters 3 and 4. This question was to be asked of all families who had a WIC age-eligible person (children aged 0-5 years or females aged 12-55 years). However, in quarters 1 and 2 this question was asked only if the family respondent was WIC age-eligible (in effect, females aged 12-55 years). Thus, quarter 1 and 2 responses for this variable were edited as being out of universe. Users wishing to derive national estimates based on PWIC should use the annual weight (WTFA) multiplied by 2. (Note that such an approach is based on the assumption of little or no seasonality in the degree to which Americans utilized WIC benefits in 1999.)

2000 National Health Interview Survey
Sample Adult File

The Sample Adult section of the 2000 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. More importantly, proxy responses are not acceptable in this section: each person chosen as the sample adult for a particular household must answer for himself/herself. The six sections comprising the Sample Adult section are discussed below.

I. Adult Conditions Section (ACN)

Prior to 1997, the NHIS covered 133 conditions across six condition lists and contained ICD-9 codes; each NHIS family was randomly assigned one of the six different condition lists. With the redesign, the six lists have been combined into a single, shorter list for adults and a single, shorter list for children, each consisting of several domains. Additionally, the current NHIS data files contain no ICD-9 codes. The domains for adults are now organized by organ system or health topic and include the following: cardiovascular disease, emphysema and asthma, ulcers, cancer, diabetes, other respiratory conditions, renal conditions, liver conditions, joint symptoms, sensory impairments, pain, hearing, vision, oral health, and mental health. Table 5 shows the specific health-related conditions covered in the redesigned NHIS, as well as the various reference periods covered by the questions. With the exception of head or chest colds, or stomach or intestinal illnesses, no question in the ACN section refers to a two-week reference period.

Given the previous NHIS questionnaire design, most condition analyses were carried out at the condition-level; person-level analyses were possible, but cumbersome. The redesigned NHIS makes person-level analyses much easier to carry out. Moreover, there are several notable differences in the way information on conditions is collected in the redesigned NHIS. As mentioned, all data in the redesigned Sample Adult component were required to be self-reported; proxy respondents were not allowed. In addition, most questions in the 1997-2000 NHIS now ask about conditions diagnosed by a doctor or health professional. Finally, while many of the condition questions are very similar to, if not identical to, those asked in the previous NHIS, questions are quite different for several conditions, notably asthma, hearing impairments, and vision impairments. All of these changes must be considered when attempting to compare condition prevalence estimates derived from either 1997, 1998, 1999, or 2000 data with those from earlier years. Moreover, users are advised that the condition data in the Person and Sample Adult files have not been compared for consistency of reported conditions.

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

Table 5. Sample Adult File: Conditions and Reference Periods

Reference Period in 2000 NHIS							
CRQ #	**Condition**	**Ever**	**12 months**	**3 months**	**30 days**	**2 weeks**	**Now**
ACN.010	High blood pressure	X					
ACN.031	Coronary heart disease	X					
ACN.031	Angina	X					
ACN.031	Heart attack	X					
ACN.031	Other heart condition	X					
ACN.031	Stroke	X					
ACN.031	Emphysema	X					
ACN.080-ACN.090	Asthma; Episode Attack	X	X				
ACN.110-ACN.120	Ulcer	X	X				
ACN.130	Cancer (Any + list max. 3 of 30 specific types)	X					
ACN.160	Diabetes	X					
ACN.201	Hay fever		X				
ACN.201	Sinusitis		X				
ACN.201	Chronic bronchitis		X				
ACN.201	Weak or failing kidneys		X				
ACN.201	Liver condition		X				
ACN.250	Joint pain		X				
ACN.300	Neck pain			X			
ACN.310-ACN.320	Low back pain, with or without leg pain			X			
ACN.331	Facial pain			X			
ACN.331	Severe headaches			X			
ACN.350	Head or chest cold					X	
ACN.360	Stomach or intestinal illness					X	

Reference Period in 2000 NHIS							
CRQ #	**Condition**	**Ever**	**12 months**	**3 months**	**30 days**	**2 weeks**	**Now**
ACN.370	Pregnancy						X
ACN.410-ACN.420	Hearing impairment	X					X
ACN.430-ACN.440	Vision impairment						X
ACN.451	Lost all upper/lower teeth						X
	Emotional Health:						
ACN.471	Sad				X		
ACN.471	Nervous				X		
ACN.471	Restless				X		
ACN.471	Hopeless				X		
ACN.471	Everything an effort				X		
ACN.471	Worthless				X		

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

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

The Adult Health Status and Limitation of Activity component of the Sample Adult file contains information from respondents on illness behavior, health status, special equipment, limitations in functional activities, and the conditions underlying such limitations. While the AHS section may seem similar to the FHS section in the Person file, the questions in these sections have a somewhat different focus. For example, both sections asked about the ability to walk without special equipment. However, the walking limitation question in the FHS section (FHS.210) only captured whether a person has difficulty walking without using special equipment. In contrast, the Sample Adult question on walking (AHS.091) asked about the degree of difficulty the respondent has walking a specified distance (a quarter mile, or about three city blocks) by him/herself and without using any special equipment.

The 2000 NHIS survey includes some minor modifications in the way information on functional limitations and the conditions that cause them are collected and processed. These revisions are highlighted below.

Health Indicators: Illness Behavior and Health Status

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

Limitation of Functional Activities

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

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

Conditions

Each sample adult indicating any functional limitation (regardless of the degree of the limitation) is asked about the condition(s) or health problem(s) associated with that limitation (up to five conditions may have been indicated), as well as the amount of time he/she has had the condition. Sample adults were given the following fixed response categories: "vision/problem seeing", "hearing problem", "arthritis/rheumatism", "back or neck problem", "fractures, bone/joint injury", "other injury", "heart problem", "stroke problem", "hypertension/high blood pressure", "diabetes", "lung/breathing problem", "cancer", "birth defect", "mental retardation", "other developmental problem (e.g., cerebral palsy)", "senility", "depression/anxiety/emotional problem", "weight

problem", and two instances of "other impairment problem". Respondents could supply a verbatim response (entered as up to an expanded 50-character field) for the "other impairment problem".

The AHS condition data were edited very much like the condition data in FHS. The verbatim responses recorded by FRs in one or both of the 50-character fields indicating "other impairment problem" were subsequently analyzed during data processing. While most respondents named conditions that did not fall into the fixed response categories as originally specified in the instrument, some respondents named conditions that could be included in one of the fixed categories. In the latter case, these respondents were assigned codes corresponding to the one of the original 18 response categories. An additional 16 *ad hoc* categories were created; these were assigned numbers 19-34. Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the new categories remained in the "other impairment" categories and renumbered "90" and "91". The resulting 36 categories were generally informed by the International Classification of Diseases, Ninth Revision, Clinical Modification (see Table 3 on page 27). These specific condition categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any functional activity (a "mention-not mention" format). Because the 16 *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions may be underestimated. Therefore, these variables should be analyzed with care. Moreover, none of the AHS condition variables (AFLHCA1 through AFLHCA34) should be used to estimate the prevalence rates for the conditions they represent, because only those sample adults with a previously reported functional limitation were eligible for the condition question that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult Conditions (ACN) section (above).

Recodes

The recode FLA1AR is a summary measure that indicates sample adults who reported *any* difficulty with *any* one or more of the functional activities discussed during the course of the AHS section of the interview. In other words, individuals who indicated *any* degree of difficulty to FLWALK, FLCLIMB, FLSTAND, FLSIT, FLSTOOP, FLREACH, FLGRASP, FLCARRY, FLPUSH, FLSHOP, FLSOCL, *or* FLRELAX are coded "1" for FLA1AR. This variable includes three response levels: "1" for limited, "2" for not limited and "3" for unknown if limited. ALCHRONR is based on FLA1AR but adds the additional criterion of whether at least one of the reported causal conditions is a chronic condition. The section also includes time recodes (ALANTR1-34, ALANTR90 and ALANTR91) and chronic recodes (ALCNDR1-ALCNDR34, ALCNDR90 and ALCNDR91) for each of the 36 categories. These recodes for conditions that cause functional limitation(s) are similar to those used in the FHS section and described above.

III. Adult Health Behaviors Section (AHB)

The NHIS Sample Adult questionnaire contains questions on four health-related behaviors: cigarette smoking, leisure-time physical activity, alcohol use, and height and weight. The health behavior questions have remained essentially unchanged since 1997.

Smoking

Smoking questions have been included in the NHIS periodically since 1965, although there has been some variation in question wording. Smokers continue to be defined as persons who have ever smoked 100 cigarettes and currently smoke every day or some days. In 1992, the NHIS basic smoking question changed from "Do you now smoke?" to "Do you now smoke every day, some days, or not at all?". This version of the question continues to the present. In addition to smoking status, data are collected on age of initiation, amount smoked, and quit attempts.

In 1997, the universes for most of the smoking variables were restricted to respondents with known smoking status. Beginning in 1998 and continuing to the present, the universes reflect the actual skip patterns in the questionnaire. A decision was made to make the data available in their original form, allowing the analyst to decide how to treat respondents who were asked the question but were not the primary target group for it. For example, CIGDAMO (number of days smoked in the past month) was primarily designed to be asked of some day smokers (SMKNOW = 2), but is also appropriate for persons who could not say how frequently they smoked (SMKNOW = 7-9). The universe for CIGDAMO includes SMKNOW = 2, 7-9. The overall number of cases affected by the inclusion of unknowns is quite small and neither inclusion nor exclusion of unknowns from the universes should affect estimates. Data users must decide whether they want to re-define these variables, limiting the universes to persons with known data in the lead-in question, or leave the universe as is, presenting data for all persons who were asked the questions.

Leisure time physical activity

The physical activity questions that have been included in the NHIS Sample Adult (core) Module since 1997 are substantially different from those included in pre-1997 NHIS questionnaires. Because of the large number of topic areas covered in the redesigned NHIS, only a brief set of leisure-time physical activity questions could be included in the Sample Adult Module. For this reason, the questions are general and lend themselves to broad classifications of activity levels. In 1998, the core physical activity questions were supplemented with additional physical activity questions in the Sample Adult *Prevention* Module. The Sample Adult *Prevention* Module contained a set of questions on 23 specific physical activities, and obtained information regarding frequency, duration, and intensity for each activity. In 1999 and 2000, the physical activity questions were once again limited to those that appeared in the Sample Adult Module. Due to space limitations, questions concerning occupation-related physical activity are not available on the NHIS Sample Adult Module. It is hoped that such questions can be included periodically in future surveys.

Data users who have used the 1997 NHIS data file should note that the variables from quarters 1-2 of 1997 have the same names as the variables included in annual 1998-2000 data files. However, the minimum duration for the two sets of questions are different. In 1997 quarters 1-2 variables, respondents were asked how often they did vigorous and light-moderate activities for *at least 20 minutes*. In 1998-2000, respondents were asked how often they did vigorous and light-moderate activities *for at least 10 minutes*.

Quarters 3-4 of 1997 were a transition period. The questionnaires contained leisure time physical activity questions with the new minimum duration (i.e., *at least 10 minutes*) but different variable names. Each of the 1997 quarters 3-4 physical activity variable names ended with the digit "2", to indicate a second version of the variables for that year. A recode was constructed for each of the physical activity variables that combined quarters 1-4 for activity with a minimum duration of 20 minutes. Data users interested in merging 1997 physical activity variables with later years should be mindful of the changes that were made to the physical activity questions mid-year in 1997 and should consult the 1997 Survey Description document and Dataset Documentation for additional details.

Alcohol use

The alcohol questions are a variation on questions that have appeared in the NHIS periodically since 1977. Extensive supplements on alcohol use were fielded in 1983 and 1988, with smaller sets of questions asked in other data years including 1985, 1990, 1991, and 1998. As with other sections of the questionnaire, the reference period was chosen to capture as much information as possible for as many people as possible. Although a twelve-month reference period is less than ideal, it allows for measurement of alcohol use among all adults, rather than just those who drink frequently. Since the alcohol questions had to be limited to a very short set of items, a shorter reference period, although undoubtedly obtaining more precise estimates for frequent drinkers, would have made it impossible to accurately measure consumption among infrequent drinkers.

The order and reference period of the questions differ from many other NHIS data years. The most notable change is in the order of the lifetime drinking questions. The question "Ever have 12 drinks in one year?" now precedes, rather than follows, "Ever have 12 drinks in entire lifetime?". Similarly, the definition of a lifetime abstainer has changed from "less than 12 drinks in one year" to "less than 12 drinks in entire lifetime". The current drinking status questions are now asked of all persons who previously said that they had ever had 12 drinks in their entire life, thus capturing infrequent drinkers who may never have had as many as 12 drinks in any one year, but did drink in the preceding 12 months.

Given the twelve-month reference period of the alcohol consumption questions, the respondent has the opportunity to answer in days per week, per month, or per year. It is assumed that persons who drink frequently will answer in days per week, while less frequent drinkers will answer in days per month or per year. Standardized variables, which convert the various time response options to a single standardized unit (i.e., days per week, days per month, days per year), are provided in the data file. Standardized variables assume a constant rate of occurrence across time periods. The question asking how often the respondent had five or more drinks in one day during the past year was asked of all adults who had anything to drink in the past year. The

responses were not edited for consistency with usual quantity or frequency. Notice that the quantity questions are phrased in terms of drinks *per day*, not drinks at a sitting.

Body weight and height

Height and weight have been asked for adults in the NHIS Core questionnaire for many years prior to the 1997 redesign, with proxy reporting allowed for family members not home at the time of the household interview. In addition, height and weight were asked in selected special topic NHIS questionnaires prior to 1997. These questions were always self-reported in special topic questionnaires. It should be kept in mind that estimates based on proxy reports may differ from those based on self-reports, and users of historical NHIS data files should be cognizant of the method that was used to collect the data for data years prior to 1997. No physical measurements are taken in the NHIS. Estimates based on physical measurements, such as those available from NCHS's National Health and Nutrition Examination Survey, are likely to differ from those available from the NHIS.

Beginning in 1998 and continuing to present, the CAPI instrument allows respondents to report height and weight in either metric or non-metric units. Metric responses for height were recoded from meters and centimeters to feet and inches, and metric responses for weight were recoded from kilograms to pounds. Only non-metric variables are available on the public use data file. However, beginning in 1999 and continuing to the present, flags (MET_FLG1 and MET_FLG2) were added to the data file to allow data users to identify which responses were volunteered in metric units. The metric variables themselves are not included in the public use data file but can be created using the formulas found in the Dataset Documentation for the BMI recode.

Beginning in 1997 and continuing thereafter, when a very large or very small value was reported for either height or weight, the data for both variables were changed to "96" or "996" ("Not available") on public use data files. This was done in order to protect the confidentiality of NHIS respondents who might be identifiable by their unusual physical characteristics.

Lastly, two recodes were created to make the height and weight data easier to use: Desirable Body Weight (DESIREWT) and Body Mass Index (BMI). Desirable Body Weight is based on the Metropolitan Life Insurance Company (MLIC) standards of desirable body weight (1983), which have been used in the NHIS since 1985. The MLIC standards are based on the mortality experience of life insurance policy holders. The Desirable Body Weight variable is restricted to persons whose height and weight are within the range provided by the MLIC chart of desirable weight for height. Because of this, the number of unknowns is substantially larger in the DESIREWT recode (n=1,816) than in the BMI recode (n=1,086).

The BMI is calculated from the formula: weight /(height)^2, where weight and height are in kilograms and meters, respectively. Unlike the Desirable Body Weight variable, the BMI recode includes all persons who provided height and weight information, including those for whom specific height and weight values were changed to "96" (not available) on the public use file. The values for the BMI recode are continuous; users should insert a decimal point in the appropriate place. Using the BMI, overweight is defined as values greater than or equal to 25 for both men and women, while BMI values greater than or equal to 30 indicate obesity. The BMI cut points are discussed in the Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans,

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

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

The core Adult Health Care Access and Utilization (AAU) section of the 2000 NHIS has remained largely unchanged since 1997 and consists of four parts: Access to Care, Dental Care, Health Care Provider Contacts, and Immunizations. Because three of the core questions in the 2000 AAU section featured expanded answer categories, the output variable names have been modified. Also, six new permanent core questions/variables that ask about chicken pox and hepatitis have been added to the Immunization Section. These new additions are noted in the Dataset Documentation.

The "Access to Care" section includes questions that ask respondents whether they have a usual place for sick care and have a usual place for routine/preventive care, and whether they have experienced any changes in their place of care, any delays in getting medical care, and instances when they were unable to afford medical care. While all of these topics were covered in the previous, pre-1997 design of the NHIS, the redesigned NHIS introduced changes in question wording as well as the order in which questions were asked. For example, in 1996 the question regarding the reason for delaying care asked "because of worry about the cost?". In 1997, and continuing through 2000, the question about the reason for delaying care focused on such access issues as transportation difficulties, getting an appointment, and waiting time prior to actually seeing the doctor.

The "Dental Care" core section includes only one question: length of time since last dental visit. This item has been asked in previous years, but beginning in 1997 the question was re-phrased.

As with the Person-level FAU section described previously, NHIS surveys prior to 1997 focused on physician contacts, visits, and home care that included only contacts with a medical doctor or a health care professional working with or for a medical doctor. In addition, home care visits were included in the question asking about visits or contacts with a doctor's office, hospital, etc. While the redesigned NHIS features questions in the "Health Care Provider Contacts" section similar to the two-week doctor visit probe questions contained in the previous design, the new version allows for visits not only from medical doctors but from a variety of other health care professionals, including chiropractors. Most significantly, there has been a change in the reference period. Surveys in 1996 and earlier asked about health care contacts in the two weeks prior to the interview; in contrast, beginning in 1997 the survey asked about contacts during the past 12 months. Additionally, the questions about home care in the redesigned NHIS were asked independently of visits to the doctor's office, the hospital, etc. Also new for 1997, and continuing through 2000, is a question asking about the number of visits to a hospital emergency room in the past 12 months.

The "Immunizations" core section includes two questions related to adult immunizations: flu shots and pneumonia vaccinations. Beginning in 2000, six new permanent questions were added to this section that inquire whether the respondent had ever had chickenpox or had chickenpox in the

past 12 months, had ever had hepatitis, had ever lived with someone with hepatitis, had ever received the hepatitis B vaccine, and if so, the number of vaccines received.

Technical notes

Due to the inclusion of expanded response categories and/or modifications to question wording in the item text in a few of the variables in 2000, modifications may have been made to the universe description, and the output variable names may have changed during the data editing process. Also, beginning in 2000, six new permanent core variables have been added to this section asking about chickenpox and hepatitis. It is suggested that analysts compare the 2000 Dataset Documentation to documentation from the 1999 (and earlier) NHIS for any changes that may have occurred to the variables in the AAU section.

V. Adult Demographics Section (ASD)

The Sample Adult Demographics (ASD) section provides information regarding the occupation, industry, workplace, and employment conditions of *employed* sample adults during the last week before the interview. Note that in previous years, NHIS asked about employment during the *two* weeks preceding the date of interview.

Industry and Occupation Coding

Sample adults aged 18 years and older who were "working at a job or business" or "with a job or business but not at work" during the week prior to their interview (DOINGLW = 1, 2) were then asked a series of questions about their employment and work status. First, verbatim responses were obtained from each respondent regarding his/her industry and occupation. These were subsequently recoded into two two-digit industry recodes that are consistent with the 1995 revisions to the Standard Industrial Classification (SIC) system. A detailed recode (INDSTRY1) indicates 42 possible industries, while a more simple recode (INDSTRY2) distinguishes 14 industries. A similar pair of recodes was created from the occupation information; again, this coding is consistent with the 1995 revisions to the Standard Occupational Classification (SOC) system. OCCUP1 distinguishes 41 separate occupations, while OCCUP2 indicates 13 occupations. These coding categories are provided in the Industry and Occupation Appendices (following the Dataset Documentation for the Sample Adult file).

Other Employment Questions

Sample adults were also asked to describe their employment situation (whether they were an employee of a private company or business, the federal government, a state or local government, self-employed in their own business or professional practice, or working without pay in a family business or farm), the number of full- and part-time employees at their workplace, how long they had worked at their current job or business, whether they were paid by the hour and received paid sick leave, and whether they were working more than one job. Sample adults who indicated that they had a second job were asked two additional questions: whether they were working for an employer or were self-employed, and if the latter, whether their business was incorporated.

VI. Adult AIDS Section (ADS)

This section contains a series of questions related to testing for the AIDS virus. Respondents were asked whether they had ever donated blood and whether they had had a blood test for HIV, the virus that causes AIDS, their main reasons for getting or not getting tested, when they had their last test, and where (the location/facility) the testing was done. The section also contains questions on respondents' plans for being tested in the future and their reasons for those plans, as well as their perceived personal risk for getting AIDS. With some modifications and additions, these questions are similar to those asked in the AIDS Knowledge and Attitudes Supplements that were included in the NHIS from 1987 to 1995.

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.

2000 National Health Interview Survey
Cancer Control Module

The Cancer Control Module consists of seven sections covering Hispanic acculturation, diet and nutrition, physical activity, tobacco, cancer screening, genetic testing, and family history. These sections are described in greater detail below. Those respondents who served as sample adults for each household also participated in the Cancer Control Module. As a result, NHIS staff decided to append the variables derived from Cancer Control Module to the 2000 Sample Adult file, rather than create a separate, stand-alone file for the cancer variables.

Although all sample adults were initially asked questions in the Cancer Control Module, some persons did not complete the module. Persons who did not give valid substantive responses to at least 50% of the questions in selected fields had a value of "not ascertained" inserted in the appropriate fields of the Cancer Control Module. These persons were then given a coded value of "1" for the Cancer Control Module dummy record flag variable (DUMMY_CA), which indicates that their record is a dummy record. In other words, these respondents (n = 1,152, or 3.6% of sample adults) are retained in the file, but they are coded as "8" in all relevant fields of the Cancer Control Module. All persons who met the criteria for completing the module were given a coded value of "0" for the DUMMY_CA flag, which indicates that their record is not a dummy record.

Background

The first cancer supplement to the National Health Interview Survey (NHIS) was fielded in 1987. It consisted of two supplement "booklets", or instruments, entitled, "Cancer Control" and the "Epidemiology Study". The topics in the "Cancer Control" section were acculturation; medical care; food knowledge; general knowledge and attitudes; cancer screening knowledge and practice; smoking habits; former smoker; current smoker; other tobacco use; occupational exposure; height and weight. Topics in the "Epidemiology Study" were acculturation; food frequency; vitamin and mineral intake; food knowledge; smoking habits; other tobacco use; reproduction and hormone use; family history of cancer; occupational exposure; height; weight; and relationships and social activities. Because these booklets both contained at least one very long section, each was administered to only a half-sample of adult respondents (except for those topics that were included in both booklets, which were asked of the full sample).

The cancer supplement was repeated in 1992 as the "Cancer Risk Factor Survey". The design was the same as in 1987, with a split sample and two booklets (i.e., instruments). The topics included in the "Cancer Control" booklet were acculturation; access to medical care; height and weight; cancer screening knowledge and practice; cancer survivorship; general knowledge and attitudes; smoking habits; current smoker; former smoker; other tobacco use; and workplace tobacco smoke. Topics in the "Epidemiology Study" were acculturation; food frequency; vitamin and mineral intake; height and weight; food knowledge; cancer survivorship; smoking habits; and occupational exposure. The 1992 cancer supplement was administered to about a half sample of adult respondents, due to budgetary constraints, which reduced the response in many sections to about a quarter of the sample.

In 2000, the NHIS again fielded a supplement, called the "Cancer Control Module", which again utilized adult respondents and covered many of the same topics as the previous cancer supplements. The split-sample design used in previous NHIS cancer supplements was dropped in order to increase the statistical power for population subgroups, and to allow the 2000 Cancer Control Module to be appended to the Sample Adult Core. The 2000 NHIS Cancer Control Module asked questions about Hispanic acculturation; diet and nutrition; physical activity; tobacco; cancer screening; genetic testing; and family history. Because the redesigned NHIS core included permanent sections on cigarette smoking, alcohol intake, and leisure-time physical activity, these topics were not covered by the 2000 supplement.

Hispanic Acculturation Section (NAB)

The questions in the Hispanic Acculturation section (NAB) of the 2000 Cancer Control Module were asked only of those sample adults who reported their ethnicity as Hispanic, and covered only the English and Spanish languages. This section contains a series of questions related to language usage, both in childhood and as an adult. Respondents were asked to name the language in which they generally speak, and the language that they are better able to read. They were also asked to name the language they usually use when speaking at home, when talking with friends, while thinking, when watching television, and when listening to radio programs. In a final question, respondents were asked to name the state or country of birth of their parents.

Users should note that the ASD section of the Sample Adult Core contains additional topics used to measure acculturation (i.e., whether the respondent was foreign or U.S. born, and, for the foreign born, length of residence in the U.S.).

Diet and Nutrition (NAC)

The Diet and Nutrition section (NAC) of the 2000 Cancer Control Module collects information about selected foods consumed by the sample adult during the past month. Data were collected on the number of times the sample adult ate or drank cold cereal, milk, bacon, hotdogs, whole grain bread, 100% fruit juice, fruit, full fat salad dressing or mayonnaise, salad, French fries, other white potatoes, beans, other vegetables, pasta, nuts, and regular fat chips. Responses about the number of times a particular item is eaten were reported in terms of daily, weekly, monthly, or yearly consumption. In addition, respondents were asked about their use, during the past 12 months, of multi-vitamins, vitamin A, vitamin C, vitamin E, calcium, and a variety of herbal or botanical supplements.

Because the time unit (day, week, month, year) could vary for each food item, it is possible that some recording errors may have occurred during the course of the interview. NHIS staff decided *not* to edit these data for reasonableness; hence, users will note that some respondents reported consuming what seems to be unusually large amounts of particular foods. For example, the data may indicate that a particular respondent ate cold cereal 30 times a day. In this instance, one data user might decide that "day" was mistakenly recorded instead of "month", and change the data accordingly; another data user might choose to disregard the record entirely. Each analyst must determine his/her own preferred method for handling these unusual cases.

Users should be aware that the underlying framework of the NAC section in the 2000 Cancer Control Module is very different from the corresponding Diet/Nutrition sections in the earlier supplements. The earlier supplements did not include any questions on herbal supplements. Moreover, the earlier supplements attempted to cover the diet of the respondent using a comprehensive "food frequency" list, whereas the focus of the 2000 NAC section is on the consumption of the main predictor foods for percent energy from fat, fruits and vegetables, and fiber. However, the questions on vitamin and mineral usage in the 2000 module are quite similar to those appearing in the previous supplements, in order to allow for analyses of time trends. Lastly, users should note that the AHB section of the Sample Adult Core contains numerous variables describing alcohol consumption for all sample adults.

Physical Activity (NAD)

Along with the questions on leisure-time physical activity in the AHB section of the Sample Adult Core, and the diet questions in the NAC section of the module, the questions on "daily" physical activities help to give a fuller picture of some of the behaviors that contribute to obesity. All sample adults who reported that they could walk a quarter of a mile (FLWALK in the AHS section of the Sample Adult Core), as well as those who refused to answer or did not know if they could walk a quarter of a mile were asked whether they walked or biked to work or school, or when performing errands. *All* sample adults were then asked the extent to which they "move around" during their usual daily activities (excluding leisure-time activities); the extent to which they lift or carry things while performing their usual daily activities (again, excluding leisure-time activities); the average number of hours per day that they spend in a sitting position (separate questions distinguish weekdays from weekends); and whether a doctor or other health professional had recommended in the past 12 months that they begin or continue exercise or physical activity.

Tobacco (NAE)

The Tobacco section of the 2000 Cancer Control Module collects smoking/tobacco-related information from every sample adult in the NHIS interviewed sample. All sample adults were asked if they had ever smoked a pipe, a cigar, a bidi, or used snuff or chewing tobacco. Those respondents who answered affirmatively were then asked if they currently used these products. All sample adults were also asked the number of days during the past week that anyone had smoked cigarettes, cigars, or pipes inside their home. Additionally, those sample adults who had seen or talked to a doctor or other health care professional in the past 12 months were asked whether that medical professional had asked about their consumption of any tobacco products. Lastly, all sample adults who were employed during the last week *and* indicated that they worked "mainly indoors" were asked, first, if anyone had smoked in their work area during the last week, and, second, whether their employer had an official policy that restricted smoking in any way.

In addition to the above questions, those sample adults who had already indicated in the AHB section of the Sample Adult Core that they were *former* smokers were asked whether they had ever used reduced tar and nicotine cigarettes. They were also asked about the method(s) they had used to quit smoking (when they stopped smoking completely). Those sample adults who had previously indicated that they were *current* smokers were asked whether they had ever used

reduced tar and nicotine cigarettes, whether they had ever tried to quit smoking, and the method(s) they used the last time they tried to stop smoking. Current smokers who indicated that they had seen a doctor or other health professional in the past 12 months were also asked whether that medical professional had advised them to quit smoking.

The final three questions in the section asked all sample adults for their opinions on several smoking-related issues: whether smoking should be allowed in indoor public places, whether second-hand smoke is harmful, and whether the price of cigarettes should be increased to curtail smoking among young people.

Cancer Screening (NAF)

The Cancer Screening section (NAF) of the 2000 Cancer Control Module collected information about selected cancer screening tests received by the sample adult, including skin exams, Pap smear tests, mammographies, clinical breast exams, Prostate Specific Antigen (PSA) tests, colorectal screening exams, and Fecal Occult Blood (FOB) tests (performed in a doctor's office and at home). The recommendations for having different screening exams differ by age and sex; these criteria were taken into account when asking about the different exams during the course of the interview. All sample adults were asked the questions about a previous skin exam and all female sample adults were asked the questions about a previous Pap smear test. Female sample adults 30 years of age and older were asked the questions regarding a prior mammography/clinical breast exam, and male sample adults 40 years of age and older were asked the questions regarding a prior PSA test. All sample adults 40 years of age and older were asked the question regarding a prior colorectal screening exam/FOB test. Respondents who indicated that they had had a particular cancer screening exam were subsequently asked if abnormal results were obtained, and if so, what additional tests and/or surgery were performed.

For each type of cancer screening exam, information was collected on when the last screening exam was received (month/year, number of days/weeks/months/years ago, or time interval grouping (if the respondent did not answer in any of the other formats)). Also, for each screening exam, two time-since-test recodes are provided: number of months since the last cancer screening exam, and a recoded time interval grouping. The recode for the number of months since the last cancer screening exam combines information contained in month/year, days/weeks/months/years ago, the original time interval grouping, and the interview date. Data analysts are advised to use the time-since-test recodes rather than the answers to original subquestions. The recodes include all respondents who were eligible for that test, while the raw answers applied only to the subset of respondents who gave the time in a particular format.

It should be noted that several respondents only provided a year, but not a month, for the date of the last cancer screening exam. To generate the recode for these individuals, the date of the cancer screening exam was set to June 30th (otherwise, the interview date and the test date were set to the 15th of the month that was provided). The following table gives a brief overview of the variable names for the various cancer screening exams and the recoded time variable associated with each exam. Within the Pap smear questions, women were also asked if they had had a hysterectomy, because a woman with a hysterectomy is less likely to receive a Pap smear test.

Ever had cancer screening test/exam (Variable name)	Months since last test/exam	Time since last test/exam (year groupings)
Skin exam (SKNX)	RSKX_MO2	RSKX3
Pap smear test (PAPHAD)	RPAP_MO2	RPAP3
Hysterectomy (HYST)	RHYS_MO2	RHYST3
Mammography (MAMHAD)	RMAM_MO2	RMAM3
Clinical Breast Exam (CBEHAD)	RCBE_MO2	RCBE3
PSA test (PSAHAD)	RPSA_MO2	RPSA3
Colorectal Exam (CREHAD)	RCRE_MO2	RCRE3
Home FOB test (HFOBHAD)	RHFO_MO2	RHFOB3
Dr. office FOB test (FOBHAD)	RFOB_MO2	RFOB3

The NAF section of the 2000 Cancer Control Module also contains information on live births and birth control pill use among female sample adults 18 years of age and older, in order to be able to use the Gail model to predict development of breast and ovarian cancer. Female sample adults 40 years of age and older were asked about their use of hormone replacement therapy, and their use of Tamoxifan and Raloxifan (for cancer prevention or therapy). In addition, appropriate respondents were asked their reasons for not having ever had particular screening exams (e.g., Pap smear tests, mammography, colorectal screening exams, and FOB tests), or for not having had them within a specified time period. Also, these respondents were asked whether a doctor or other healthcare professional had recommended (in the last 12 months) that they receive the screening exam in question. The following table highlights the variable names for the applicable cancer screening exams. The "X" in the table refers to the varying time periods for the different cancer screening tests/exams: Pap smear test (3 years), mammography (2 years), colorectal cancer screening (10 years), and home FOB (1 year).

Ever had cancer screening test/exam (Variable name)	Reason for no exam or not within the last "X" years	Exam/test recommended by doctor/healthcare professional in last 12 months?
Pap smear test (PAPHAD)	PAPNOT	MDRECPAP
Mammography (MAMHAD)	MAMNOT	MDRECMAM
Colorectal Exam (CREHAD)	CRENOT	CREREC
Home FOB test (HFOBHAD)	HFOBNOT	MDHFOB

Analysts should also note that due to an inconsistency in the survey instrument, MDRECPAP and MDRECMAM were not asked the same way as CREREC and MDHFOB. In CREREC and MDHFOB, sample adults who had previously indicated that they had not seen/talked to a doctor or other health care professional in the last 12 months (AMDLONG) were not asked the question. However, MDRECPAP and MDRECMAM did not have this consistency check in the instrument. Therefore, MDRECPAP and MDRECMAM were edited against AMDLONG (in order to be consistent with CREREC and MDHFOB).

Genetic Testing (NAG)

The Genetic Testing section of the 2000 Cancer Control Module collects information about genetic testing for cancer risk from every sample adult in the NHIS interviewed sample. Respondents were told at the outset of this section that genetic testing for cancer risk involved testing a person's blood to see if he/she carries genes that may predict a greater chance of developing cancer at some point in his/her life, and that such tests did not include diagnostic procedures to determine if the person currently had cancer. Respondents were then asked whether they had ever heard of this kind of genetic testing. Those respondents who answered affirmatively were then asked whether they had ever discussed the possibility of undergoing such a test with a doctor or other health professional, whether they had been advised by a doctor or health professional to have such a test performed, and whether they had taken a genetic test to determine their risk of cancer. If they indicated that they had undergone genetic testing, they were asked for details about the test, such as the type of genetic test taken, the date of the most recent genetic test, whether the test was done as part of a research study, what health professional had ordered the test, whether any genetic test results were received, and if genetic counseling about the test had been available. These respondents were also asked if they were confident that their test results would remain confidential, and whether they thought that having such a test might currently (or in the future) affect their health insurance coverage.

The final two questions in the section asked all sample adults whether they felt their risk of getting cancer in the future was "low", "medium", or "high", and whether they believed the amount of cancer among their blood relatives was "low", "medium", or "high".

Analysts are strongly cautioned regarding their use of the data in the Genetic Testing section. In examining the data, it was discovered that even though there was an explanation of genetic testing for cancer risk in the introduction to the section, some answers were invalid or unreliable, and a few respondents reported that they had taken genetic tests for cancer risk when they probably had not. At the time this survey was administered, genetic tests outside a research setting were available for detecting the risk for breast cancer, ovarian cancer, and colon/rectal cancer. Yet, of the 138 persons who said they had taken a genetic test for cancer risk, 52 said their test was for something other than breast cancer, ovarian cancer, or colon/rectal cancer and were not part of a research study. In addition, several respondents said they had taken a genetic test for risk of breast or ovarian cancer before 1994. However, neither of these tests was available until 1994. Likewise, some respondents reported having taken a genetic test for risk of colon/rectal cancer prior to 1991, the year this test became available. Since such a large percentage of respondents reported having

genetic tests for cancer risk that did not exist at the time of the survey, we strongly caution analysts when using the data in this section.

Family History (NAH)

The questions in the Family History section (NAH) of the 2000 Cancer Control Module were asked of all sample adults. Respondents were asked whether their biological father or mother, as well as any full biological brothers and sisters, and sons and daughters had ever had cancer of any kind. If a "yes" response was obtained for any family member, the respondent was then asked to specify the type of cancer and whether the family member in question was less than 50 years of age when the cancer was first diagnosed. Respondents could provide information on any three different kinds of cancer for their parents and siblings, and on any two different kinds of cancer for their children. In addition, a variable was created to indicate those respondents having family members with more than three kinds of cancer.

2000 National Health Interview Survey
Sample Child File

The Sample Child section of the 2000 NHIS covers additional subject areas not included in the Family Core. Moreover, the questions in the Sample Child section are more specific, and are intended to gather more detailed information, than those in the Family Core. Sample children do not speak for themselves; instead a knowledgeable adult (typically a parent or guardian) answers questions in the sample child's behalf. The sections comprising the Sample Child section are discussed below.

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

The Child Health Status (CHS) section of the 2000 NHIS consists of two parts: "Conditions, Limitations of Activity, and Health Status" and "Child Behavior". The section on Conditions, Limitations of Activity, and Health Status includes questions on the following health conditions: mental retardation, developmental delays, Attention Deficit Hyperactivity Disorder (ADHD) or Attention Deficit Disorder (ADD), Down's syndrome, cerebral palsy, muscular dystrophy, cystic fibrosis, sickle cell anemia, autism, diabetes, arthritis, congenital and other heart disease, asthma, various allergies, colitis, anemia, ear infections, seizures, headaches, stuttering, and stammering. 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.

It is important to note that the question about Attention Deficit Disorder (ADD) has been changed to include Attention Deficit Hyperactivity Disorder (ADHD).

The questions pertaining to child behavior were designed to serve as a global mental health indicator. The items were taken from the Child Behavior Checklist for Ages 2-3, and the Child Behavior Checklist for Ages 4-18 (Achenbach and Edelbrock 1983); these are standardized instruments for obtaining parents' reports of their children's problems. The items 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 each sex and age group.

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.

Technical Notes

Several questions pertaining to child behavior are used to create several 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 the Mental Health Index, an appendix following the Dataset Documentation for the Sample Child file.

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

The Child Health Care Access and Utilization (CAU) section of the 2000 NHIS consists of three parts: "Access to Care", "Dental Care", and "Health Care Provider Contacts". The questions pertaining to "Access to Care" include: having a usual place for sick care; having a usual place for routine/preventive care; change in place of care; reasons for a delay in getting medical care; and the inability to afford medical care. These topics were covered in previous years; however, there has been some change in every question, including minor word changes, changes in the order in which questions were asked, and rewriting an entire question. For example, 1996 NHIS participants were asked if they delayed getting medical care for the sample child "because of worry about the cost...". In contrast, with the redesign in 1997, questions focused on wider access issues, such as not having transportation, difficulty in getting appointments, and waiting time to see the doctor.

The section on "Dental Care" includes only one question: length of time since last dental visit. This topic has been covered in previous years, but the question was re-phrased in 1997. In 2000 this question was asked for sample children 1-17 years of age; previously 1 year old children were not included in questions about dental visits.

Questions regarding "Health Care Provider Contacts" are similar to the doctor visit questions from previous years, and include visits to or from medical doctors and other health care professionals (such as chiropractors) in the past 12 months. As with the FAU section discussed previously, the category of "health care professional" has been expanded to include additional occupational capacities (i.e., chiropractors, various types of therapists, psychiatrists, psychologists, and social workers); moreover, contacts or visits are no longer restricted to medical doctors or professionals working with/for a medical doctor. In addition, previous instruments included home care visits in the same question as visits to or contacts with a doctor's office, hospital, etc. From 1997 on, questions about home care were asked independently of these other visits. Most significantly, there has been a change in the reference period. Surveys in 1996 and earlier asked about health care contacts in the two weeks prior to the interview; in contrast, the redesigned survey asked about contacts during the past 12 months. Lastly, beginning in 1997 a question was asked about the number of visits to a hospital emergency room in the past 12 months.

Lastly, the question about the number of times the sample child has seen a doctor or other health professional in the past 12 months has been modified (CHCNOYR/CAU.320). Prior to 2000, this question specified that overnight hospitalizations, emergency room visits, home visits, and telephone calls should be excluded from the number of contacts. Starting in 2000, dental visits were added to list of contacts to be excluded.

III. Child Immunization File (CIM)

The Child Immunization file of the 2000 Sample Child Core involves questions on the vaccination status of children under 18 years of age and within two age groupings (under 7 years, and 7-17 years) for one randomly selected sample child per family in a household, along with any non-sample children aged 12-35 months within families of the household. The inclusion of additional children in the younger age ranges increases the precision of estimates of vaccination coverage for young children. The age split at 7 years reflects a differential focus on vaccinations by age. Among younger children, the focus is on the standard shots for which NHIS has previously obtained information (in earlier surveys). Among older children, vaccines such as hepatitis, measles, and diphtheria-tetanus booster are emphasized.

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

Several new variables pertaining to Rotavirus vaccinations were added to the questionnaire in 2000. In addition, the wording and placement of the pneumococcal questions changed in mid-2000: in quarters 1 and 2, they were included as part of a list of vaccines in the variables OTHEV, OTH2, OTH2DT (for those with shot records), while in quarters 3 and 4 they became separate questions in the instrument, called PNEU and PNEDT. (PNEU and PNEDT are for children with shot records; the corresponding variables for children without shot records are PNEEV and PNEENO.) During the course of data editing, information derived from these items was aggregated across all four quarters into the final variables, PNEU and PNEDT. However, users should keep in mind that the new questions were placed in a different order within the questionnaire. Also, a new question regarding an additional pneumonia shot, PNEMOR, was included for children with shot records.

Technical notes

The variable ICSTAT no longer results from a question in the instrument, but is now derived from a check item within the questionnaire. The variables IMRESPNO, ICRELTIV, and ICAGEMR were first included on the 1999 CIM file. IMRESPNO indicates the person who answered the questions on behalf of the child, while ICRELTIV defines the (proxy) respondent's relationship to the child whose immunization status is assessed. Lastly, ICAGEMR indicates the child's age in months.

Guidelines for Citation of Data

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 data should acknowledge NCHS as the original source. The suggested citation to appear at the bottom of all tables and graphs is as follow:

Data Source: National Center for Health Statistics (2002)

In a bibliography, the suggested citation should read:

National Center for Health Statistics (2002). Data File Documentation, National Health Interview Survey, 2000 (machine readable data file and documentation). National Center for Health Statistics, Hyattsville, Maryland.

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

References

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

Dept. of Commerce, U.S. Census Bureau. (2000). *Poverty in the United States, 1999*. Washington, D.C.: U.S. Dept. of Commerce, Census Bureau.

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

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

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

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

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

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

Appendix I

Calculation of Response Rates for the 2000 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 2000 NHIS.

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

Household Response Rate

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

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

Conditional Family Core Response Rate

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

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

Final Family Core Response Rate

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

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

Conditional Sample Adult Response Rate

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

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

Final Sample Adult Response Rate

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

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

Conditional Sample Child Response Rate

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

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

Final Sample Child Response Rate

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

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

Conditional Immunization Response Rate

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

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

Final Immunization Response Rate

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

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

Appendix I, Table 1. Response Rates for the 2000 NHIS

Household	88.9%
Family/Person (Core) - Conditional Family/Person (Core) - Final	98.2% 87.3%
Sample Adult - Conditional Sample Adult - Final	82.6% 72.1%
Sample Child - Conditional Sample Child - Final	90.9% 79.4%
Immunization - Conditional Immunization - Final	98.2% 79.5%

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, but the sum of the numerators for all combined data years is used for the combined numerator and the sum of the denominators for all combined data years is used for the combined denominator. The following examples are shown for two years of data. The same methods apply for multiple years of data with the same sample design used in 1997-2004.

Household Response Rate for Combined Data Years

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

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

Conditional Family Core Response Rate for Combined Data Years

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

This file was created from Family Core data collected from the respondent about all persons in the family. Because all eligible persons in the family are included, the response rates for the Person file are identical to the response rates for the Family file. The response rates for the Family Core can be calculated

in two ways. The conditional Family response rate is the rate only for those families identified as eligible and does not take into account household non-response. The conditional Family response rate for combined data years is calculated by dividing the number of responding families in Years 1 and 2 by the number of families that are eligible for the survey in Years 1 and 2, that is, from interviewed households in Year 1 and Year 2. Note that a household can have multiple families, and rejected families are families that were deleted from interviewed households because of insufficient data.

Final Family Core Response Rate for Combined Data Years

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

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

Conditional Sample Adult Response Rate for Combined Data Years

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

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

Final Sample Adult Response Rate for Combined Data Years

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

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

Conditional Sample Child Response Rate for Combined Data Years

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

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

Final Sample Child Response Rate for Combined Data Years

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

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

Conditional Immunization Response Rate for Combined Data Years

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

The response rates for the Immunization section can be calculated in two ways. The conditional immunization response rate is the rate only for the Immunization section and does not take into account household or family non-response. The conditional immunization response rate for combined data years is calculated by dividing the respondents to the immunization section for Year 1 and Year 2 by the number of eligible children from interviewed families for Year 1 and Year 2. Note that certainty children are children who were selected to participate in the Immunization section with certainty based on their age, regardless of whether they were selected as the sample child. Non-certainty sample children were selected, because all sample children (under 18 years) who responded to the Sample Child section were eligible for the Immunization section.

Final Immunization Response Rate for Combined Data Years

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

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

Appendix I, Table 2. Number Eligible/Interviewed 2000 NHIS

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

Appendix I, Table 3. Number Eligible/Interviewed 1999 NHIS

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

Appendix I, Table 4. Number Eligible/Interviewed 1998 NHIS

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

Appendix I, Table 5. Number Eligible/Interviewed 1997 NHIS

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

Appendix II

Race and Hispanic Origin in the 2000 NHIS

Background

For over 20 years, the National Health Interview Survey (NHIS) has collected information on the race and ethnicity of its respondents, following guidelines set forth by the Office of Management and Budget in a policy known as OMB Directive 15 (Office of Management and Budget, 1977). The NHIS has relied on respondents to provide self-identified race and ethnicity information (proxy information is reported for children and non-present household members), although interviewer-observed race was also recorded through 1996, the last year of the paper questionnaire. NHIS data are routinely tabulated by race and ethnicity in NCHS publications such as Current Estimates, Health U.S., Healthy People 2000 updates, and Advance Data reports.

In response to the changing demographics of the U.S. population, the OMB revised Directive 15 in 1997 after an extensive period of research and public commentary. The new race and ethnicity standards allow respondents to the Census and federal surveys to indicate more than one race group in answering questions on race. A complete description of the new OMB guidelines on the collection of racial and ethnic data, including descriptions of the new race categories, the ordering of race and ethnicity questions, and guidelines for the tabulation and publication of data under the new standards, can be found on the OMB web site: **http://www.whitehouse.gov/OMB/inforeg**. Although this policy is not expected to be fully implemented across the federal statistical system until 2003, surveys like the NHIS that are reviewed by OMB for renewal on a yearly basis are expected to implement changes to their survey instruments when they apply for their first OMB clearance after the policy's effective date. In accordance with this requirement, the NHIS became fully compliant with the new race and ethnicity standards with the fielding of the 2000 questionnaire, although the NHIS had been following some aspects of the new guidelines for many years.

Race and Hispanic Origin Questions in the National Health Interview Survey

The 2000 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; users should refer to the appropriate flashcard on the NHIS web site).

There were no changes in the wording of the 2000 Hispanic origin question, but some responses were imputed and the variable name was changed from HISPANCR to HISPAN_I to indicate this fact (see section on the imputation of race and ethnicity later in this document).

The 2000 NHIS included two race questions to obtain information on a respondent's race: "What race {does/do} {name/you} consider {himself/herself/yourself} to be?" (HHC.200),

and

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

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

Data users are strongly urged to read carefully the 2000 public use documentation, where details on the specific response categories for the race questions can be found. The following table summarizes the changes made to the Hispanic origin and race variables in the 2000 data file. Additional details on these variables can be found in the survey documentation, and users are urged to read the variable descriptions carefully to determine how and when the variables should be used in analysis. Data users are also encouraged to examine frequencies of the unweighted data for these variables before computing weighted estimates.

Appendix II, Table 1. Description of the 2000 NHIS Race and Ethnicity Variables

1999 Variable Name	2000 Variable Name	Status for 2000 (compared to 1999)	Description
ORIGIN	**ORIGIN_I**	Variable name changed to indicate imputed values for some records	Hispanic origin/ancestry
N/A	**ORIGIMPT**	New for 2000	Hispanic origin imputation flag
HISPANCR	**HISPAN_I**	Variable name changed to indicate imputed values for some records	Type of Hispanic origin/ancestry
N/A	**HISPIMPT**	New for 2000	Type of Hispanic origin imputation flag
RACDET_P	**RCDT1P_I**	Variable name changed to indicate imputed values for some records	Detailed race variable; multiple race persons in separate category

1999 Variable Name	2000 Variable Name	Status for 2000 (compared to 1999)	Description
RC_SUM_P	**RC_SMP_I**	Variable name changed to indicate imputed values for some records	Summary race variable (i.e., no detailed groups); contains 4 of 5 OMB race groups and "Other race"; multiple race persons in separate category
RACER_P	**RACERP_I**	Variable name changed to indicate imputed values for some records	Contains 4 of 5 OMB race groups and "Other race"; multiple race persons coded differently than in RCDT1P_I and RC_SMP_I (see documentation).
MRACER_P	**MRACRP_I**	Variable name changed to indicate imputed values for some records	Detailed race variable; only multiple race persons not selecting a primary race group in separate category
MRACBR_P	**MRACBP_I**	Variable name changed to indicate imputed values for some records	See section below on bridging.
RACERECR	**RACREC_I**	Variable name changed to indicate imputed values for some records	All persons not coded in a race category are imputed to a race category on this variable.
RCIMPFGR	**RACEIMPT**	Variable name changed to make it consistent with other imputation flags	Imputation flag for use in determining which cases were imputed for the race variables
HISPCODR	**HISCOD_I**	Variable name changed to indicate imputed values for some records	Same categories as RACREC_I, crossed with ORIGIN_I (Hispanic/non-Hispanic)
N/A	**ERIMPFLG**	New for 2000	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 2000 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 in the 2000 NHIS. Prior to the 2000 NHIS, a crude imputation method that assigned a race to persons with missing values on the variable MAINRACE for race recodes #1 and #2 was used. Under these procedures, if an observed race was recorded by the field representative, it was used to code a race value. If there were no observed race value, all persons who had a missing value for MAINRACE and were identified as Hispanic on the Hispanic origin question were coded as "white". In all other cases, non-Hispanic persons were coded as "other race".

The variables ORIGIN (whether or not the respondent is of Hispanic origin), HISPTY01-HISPTY10 (type of Hispanic origin), RACE1-RACE5 (each of 5 possible race mentions), and MAINRACE (primary race selection for persons reporting more than one race) with missing values were imputed (note that the pre-imputation variable names are used in this description because the names were

not changed until the imputation was completed). The imputation was carried out in two stages: within households at the first stage, and between households at the second stage. Hot-deck imputation procedures developed for the Decennial Census Dress Rehearsal (conducted in 1998) were adapted for use on the 2000 NHIS data. These specifications formed the basis of the first stage of the imputation (within households), although they were adapted to utilize NHIS family relationship variables. However, the specifications obtained from Census did not contain information on the imputation of race and ethnicity between households. Staff in DHIS and NCHS's Office of Research and Methodology developed the specifications for the between-household imputation, using the secondary sampling unit (SSU) as the geographic unit for selecting donors.

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

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

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

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

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

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

2. Stage 2 Imputation - this was done for households in which all persons had missing values.

A. The imputation was divided into three parts:

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

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

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

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

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

Bridging to the Old OMB Standards

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

There is one major change to the race and ethnicity data in the 2000 NHIS that occurs as a result of the creation of a bridge variable. NCHS confidentiality standards do not permit NCHS to release data that might lead to the inadvertent identification of individual respondents to the survey (e.g., a combination of demographic, geographic and other characteristics of persons in relatively small population groups could lead to identification of an individual respondent). Beginning with the 2000 survey, 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 2000 Race Variables

The variables RACEEC_I and MRACRP_I correspond to the old OMB guidelines for collecting racial and ethnic data (see the survey documentation for further descriptions of these variables). They were created in the same fashion as their previous NHIS counterparts (National Center for Health Statistics 1996), with two exceptions. First, since observed race is no longer collected in the NHIS (beginning in

1997), it was not used to help classify persons with "Unknown" race on the RACREC_I recode. Second, the recodes "White/Non-White" and "Black/Non-Black" were not created because they are no longer used in the weighting and tabulation of NHIS data. As in the past, smaller subgroups have been collapsed for confidentiality reasons.

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

SAS Code for Single and Multiple Race Groups

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

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. The following SAS code (in bold text) processes these variables to create a final variable coding all combinations of single and multiple race groups:

Step 1: Creates and initializes the following variables to 0 (intermediate processing step):

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

Step 2: Sets non-mutually exclusive conditions for recoding the 5 race variables, and sets each of the above variables to the number designated (intermediate processing step):

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

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

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

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

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

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

Step 3: Creates the final variable using the SAS SUM function:

RACEFULL=SUM(OF RACEW RACEB RACEAIAN RACEASIA RACENHPI RACEOTHR);

The variables RACEW, RACEB, RACEAIAN, RACEASIA, RACENHPI, and RACEOTHR thus take on the numbers 0,1,2,4,8,16, and 32 during the intermediate processing stage, which add up to a series of unique numbers corresponding to specific combinations of races. The value of the final output variable RACEFULL tells which races (RACEW through RACEOTHR) combined to give that number. For example, if RACEFULL=3, the only the sum of the values for RACEW=1 and RACEB=2 could have produced the number 3. Therefore anyone with the value RACEFULL=3 falls into the "White/Black" race category. If RACEFULL=1, then those persons fall into the "White" category. This scheme accurately allocates persons with multiple API and AIAN mentions. The full listing of categories and the numbers to which they correspond are included in the following table:

Appendix II, Table 2. Coding Scheme for SAS Algorithm

Coding Scheme for OMB Race Category Data (including single and multiple race mentions)		
# of Category (reported in SAS frequency distribution of RACEFULL)	**Sum of Coding (breakdown of RACEFULL, which is the sum of RACEW, RACEB, RACEAIAN, RACEASIA, RACENHPI, and 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
23	1+2+4+0+16+0	White/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 Coding (breakdown of RACEFULL, which is the sum of RACEW, RACEB, RACEAIAN, RACEASIA, RACENHPI, and RACEOTHR)**	**Resulting Category (used in the PROC FORMAT statement to label categories in SAS)**
24	0+0+0+8+16+0	Asian/NHOPI
25	1+0+0+8+16+0	White/Asian/NHOPI
26	0+2+0+8+16+0	Black/Asian/NHOPI
27	1+2+0+8+16+0	White/Black/Asian/NHOPI
28	0+0+4+8+16+0	AIAN/Asian/NHOPI
29	1+0+4+8+16+0	White/AIAN/Asian/NHOPI
30	0+2+4+8+16+0	Black/AIAN/Asian/NHOPI
31	1+2+4+8+16+0	White/Black/AIAN/Asian/NHOPI
32	0+0+0+0+0+32	***Other***
33	1+0+0+0+0+32	White/Other
34	0+2+0+0+0+32	Black/Other
35	1+2+0+0+0+32	White/Black/Other
36	0+0+4+0+0+32	AIAN/Other
37	1+0+4+0+0+32	White/AIAN/Other
38	0+2+4+0+0+32	Black/AIAN/Other
39	1+2+4+0+0+32	White/Black/AIAN/Other
40	0+0+0+8+0+32	Asian/Other
41	1+0+0+8+0+32	White/Asian/Other
42	0+2+0+8+0+32	Black/Asian/Other
43	1+2+0+8+0+32	White/Black/Asian/Other
44	0+0+4+8+0+32	AIAN/Asian/Other
45	1+0+4+8+0+32	White/AIAN/Asian/Other
46	0+2+4+8+0+32	Black/AIAN/Asian/Other
47	1+2+4+8+0+32	White/Black/AIAN/Asian/Other

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

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

Further Information

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

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** (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.

Appendix III

Variance Estimation Using the NHIS Public Use Data, 1997-2000

Introduction

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

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

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

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

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

Conceptual NHIS design for 1995-2004

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

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

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

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

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

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

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

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

Design Information Available on the NHIS Public Use Data

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

Appendix III, Table 1. Variables Used for Variance Estimation, 2000 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 1997-2000 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 2000 NHIS public use data, except the Injury Episode, Injury Verbatim, and Poison Episode files (when available).

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

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

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

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

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

Subsetted Data Analysis

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

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

NEST **STRATUM PSU / MISSUNIT ;**

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

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

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

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

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

Adult Mental Health Documentation

The questions regarding adult psychological problems (ACN.530.030 – ACN.530.650) were designed to assess mental illness diagnoses within the past 12 months among respondents who were eighteen years of age or older. Items from the Composite International Diagnostic Interview – Short Form (CIDI-SF; Kessler et al. 1997) were included in the 1999 NHIS to evaluate three specific diagnostic domains: 1) *major depressive episode*, 2) *generalized anxiety disorder*, and 3) *panic attacks.* Due to survey constraints, other CIDI-SF diagnostic domains, including phobias, social phobia, agoraphobia, alcohol dependence, and drug dependence, could not be measured by the NHIS. For a full review of the origins and scoring of these diagnostic domains please refer to the following CIDI web site: **http://www.who.int/msa/cidi/index.htm**.

The CIDI-SF provides psychiatric diagnoses based on definitions and criteria provided by the Diagnostic and Statistical Manual of Mental Disorders – IV (DSM-IV; APA, 1994). It was developed from the larger Composite International Diagnostic Interview developed by the World Health Organization (CIDI; WHO 1990). CIDI-SF items used in the 1999 NHIS were selected from the larger interview based on analyses performed by the U.S. National Comorbidity Survey (NCS; Kessler et al. 1994).

In general, the CIDI-SF is a fully standardized diagnostic instrument that can be administered in approximately ten minutes by trained interviewers *without* clinical backgrounds. The short-form interview evaluates diagnoses based on probability-of-CIDI caseness ranging from 0.0 to 1.0 for each diagnostic domain. Full, standardized diagnoses may be extracted using guidelines provided by the NCS (Nelson et al. 1998). The noted brevity of the interview may largely be attributed to its stem-branch questionnaire design, where initial step questions are relied upon to skip-out respondents who are least likely to be potential cases regarding specific diagnostic domains.

Some of the Periodic questions pertaining to the mental health form include a slightly modified index based on the Composite International Diagnostic Interview - Short Form (CIDI-SF; Kessler et al. 1997) that measures the probability of psychiatric illness within the past 12 months (i.e. major depression, generalized anxiety disorder, and panic attack). Users seeking more information on these mental health items should refer to the references cited below. The results from the diagnostic algorithms will be released in a supplement to these data. Use of these item-level questions in their current form should be undertaken with caution. It is suggested that these item-level data should not be combined in an effort to create summary scores on the probability of diagnosis. Users will be notified of this future release via the NHIS listserve and the NCHS web page.

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*. Washington DC: American Psychiatric Association.

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

Kessler, R.C., McGonagle, K.A., Zhao, S.Y., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U., Kendler, K.S. (1994). Lifetime and twelve-month prevalence of DSM-III-R psychiatric-disorders in the United States – Results from the National Comorbidity Survey. *Archives of General Psychiatry*, 51, 8-19.

Nelson, C.B., Kessler, R.C., and Mroczek, D. (1998). *Scoring the World Health Organization's Composite International Diagnostic Interview Short Form* (CIDI-SF; v1.0 NOV98).

World Health Organization (1990). *Composite International Diagnostic Interview, Version 1.0*. Geneva: World Health Organization.