1997 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

February 2000

1997 National Health Interview Survey (NHIS)

Public Use Release

Table of Contents

The NCHS Web Page and NHIS Electronic Mail List

Users can find 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 1997 NHIS, as well as information and published data from previous years' surveys. Updates about future surveys and data sets are also available.

Researchers may also wish to join the NHIS electronic mail list. To do so, go to

http://www.cdc.gov/subscribe.html

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 list serve is made up of approximately 3,000 NHIS data users located around the world who receive e-news about NHIS surveys, publications, and conferences.

1997 National Health Interview Survey (NHIS)
Public Use 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 revision occurring in 1982. 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, U.S. Bureau of the Census 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 resulting data file is referred to as 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.

Both Periodic and Topical Modules are planned for the future. The content and timing of the Periodic Modules have not yet been determined and will depend in part on experience gained in fielding the Basic Module. The purpose of the Periodic Module is to collect more detailed information on some of the topics included in the Basic Module from the sample persons. This will provide greater depth in certain areas while retaining key measures in all areas for analysis. The Topical Module component is analogous to the Supplements of the 1982-1996 NHIS and will be used to respond to new public health data needs as they arise. As with the previous Supplements, the Topical Module questionnaires may be fielded only once or may be repeated as needed. These questionnaires may be used to provide additional detail on a subject already covered in the Basic or Periodic Modules or on a different topic not covered in other parts of the NHIS.

Data Collection Procedures

The U.S. Bureau of the Census, under a contractual agreement, is the data collection agent for the National Center for Health Statistics. 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 each of the 12 U.S. Bureau of the Census 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 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 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 sampling frame 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 locations from 198 to 358. This enhanced the capability of using the NHIS for state estimation and future dual-frame surveys at the state level. 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 that are not based on the preceding decennial census. Also, the NHIS sample is divided into four representative panels to further facilitate integration with other NCHS surveys. An NCHS Series 2 publication containing a description of the 1995-2004 survey design, the methods used in estimation, and general qualifications of the data obtained from this survey is in preparation. (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.)

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 the person record data. In addition to the design and ratio adjustments included in the person'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 (quarters 3 and 4) and 19-35 months (quarters 1 and 2). 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 311 persons on the family/person data file were active duty members of the Armed Forces at time of interview, despite the fact that NHIS covers only the civilian noninstitutionalized household population. These persons have a zero weight for their person weights and should not be counted when making national 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, a variable with a two-week recall, (two-week estimate)(26)(WTFA) = annual estimate; for a variable with a three-month recall, (three-month estimate)(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 1997 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 estimated 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 1997 Data

The interviewed sample for 1997 consisted of 39,832 households, which yielded 103,477 persons in 40,623 families. The interviewed sample for the Sample Adult component, which required self-response to all questions, was 36,116 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 14,290 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 15,402 children aged 17 and younger. Data were not collected on any infant who was born during the assignment week of the interview.

The *total* noninterview rate was approximately 8.2%: 5.0% was the result of respondent refusal, and the remainder was primarily the result of failure to locate an eligible respondent at home after repeated calls or unacceptable partial interviews.

The total noninterview rate for the Adult Sample Person component was 11.0% of persons identified as sample adults. The final response rate for the Adult Sample Person component is calculated as (Overall Family Response Rate)(Adult Sample Person Response Rate), or (90.3%)(89.0%) = 80.4%.

The conditional response rate for the Sample Child component was 93.1%, which was calculated by dividing the number of completed Sample Child interviews (14,290) by the total number of eligible sample children (15,351). 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 90.3%, yielding a rate of 84.1%.

The total noninterview rate for the Immunization (Child) section of the Sample Child component was 1.0% of persons asked to respond to the Immunization (Child) section. The final response rate for the Immunization (Child) file was 84.3%. For further information about response rates, see Appendix I.

Information About the 1997 Data File Documentation

Along with the redesign of the NHIS questionnaire, other aspects of the data production process were also modified. The format and content of the data file documentation has changed. As a result, more specific information about each variable on the file is now available. For most variables, the 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, it may be constructed from two or more variables, as the body mass index (BMI) included in the Sample Adult file. If a particular variable was used in making other recode variables, then those recode variables are listed.

Keywords - Descriptive words relevant to the topic of the variable; can be used for word searches.

Notes - 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. The universes for all Sample Adult variables are specified as ASTATFLG = 1 and AGE = 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, 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 new, 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 know, specifically, if this adult respondent answered the Child Immunization questions for the sample child, or for the additional children aged 19-35 months in the first half of 1997, and aged 12-35 months in the latter half of 1997.

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 1997 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 1997 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 data file documentation can resolve any discrepancies.

Because the questionnaire for the NHIS is computerized, 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 (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. These changes were kept to a minimum, but as a result, there were two major versions of the instrument fielded in 1997. The first version (referred to subsequently as version 3.1) was fielded during the first half of the year, while the second version (or version 3.5) was in the field during the latter half of the year. Some known errors in the 1997 NHIS CAPI

instrument were not corrected until 1998. In nearly all cases, the effects of these errors on the data are quite small; exceptions are noted in the "Notes" section of the documentation. Analysts are encouraged to read the notes in the documentation for important information pertaining to specific variables.

In October 1997, the CAPI instrument developed an error such that it was not selecting a sample adult within some families. This error was quickly remedied, but in order to make up for the lost sample members, more interviews were assigned and completed in November and December. Variables in the data that are sensitive to seasonality may be affected, and users should make note of this. In general, there were fewer interviews completed in October than usual, while more interviews were completed in November and December than in previous months. See the Household File documentation for frequency counts by interview month.

Questionnaire Sections

The 1997 NHIS contained only the annual Basic Module. This Module 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 below details the various questionnaire sections, their acronyms and description titles.

Table 1. 1997 NHIS Questionnaire Sections and Topics

A. Household

Section No.	Section Code	Description
I	HHC	Household Composition

B. Family Core

Section No.	Section Code	Description
I	FHS	Health Status and Limitation of Activity
II	FIJ	Injury
III	FAU	Health Care Access and Utilization
IV	FHI	Health Insurance
V	FSD	Socio-demographic
VI	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 1997 NHIS contains two other data files, the Household- and Family-level files. 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. Both of these files are described in greater detail below.

1997 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 43,370 households: 39,832 households were interviewed, while 3,538 refused participation. 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 8.0% of households. The response rate for the Household File is calculated as the number of responding households divided by the total number of households (responding + non-responding households), or 92.0%.

1997 National Health Interview Survey
Family-Level File

The Family-Level file contains variables that describe characteristics of the 40,623 families living in households that participated in the 1997 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 1997 survey; this is specified as FM = ALL in the Family file documentation. Users wishing to determine the number of responding and non-responding families in each household should consult the Household-level file.

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), with all relationships being relative to the household reference person. There are several exceptions to this, however. Unmarried couples (same-sex or opposite-sex couples) are also considered to be one family. In contrast, individuals 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. 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.

Family size may vary considerably. The table below shows a breakdown of the 40,623 families by number of family members. Again, note that multiple families may share one household; this information can be determined from the Household-level file (refer to the variable ACPT_FAM).

Table 2. Size of Family, 1997 National Health Interview Survey (unweighted counts)

Number of Members	Frequency	Percent
1	11,850	29.2
2	12,059	29.7
3	6,504	16.0
4	5,822	14.3
5	2,756	6.8
6	1,001	2.5
7	386	1.0
8	119	0.3
9	62	0.2
10	30	0.1
11	16	0.0
12	9	0.0
13	4	0.0
14	2	0.0
15	1	0.0
16	2	0.0

The first part of the Family file contains the technical variables that identify or describe the record type (all observations in this file have a record type value of “60"), the survey year, the household and family numbers, the interview quarter and week, characteristics of the family’s housing unit, geographic information associated with the housing unit, variables used for variance estimation, and a family-level weight variable. The second part of the file consists of a series of recodes derived from the six family core sections of the NHIS that collapse the 100,000-plus individual-level observations into their respective families. Generally, two types of recodes are possible. The first is a simple “yes-no” measure that indicates whether *any* family member falls into a particular category or exhibits a particular characteristic. These variables are equivalent to, but not directly derived from, the family-level questions in the instrument. Every yes-no measure also has a corresponding counter that indicates the number of family members in that category or with that characteristic. Note that counters always consist of values from zero to 30; in addition, blanks are also possible if a family is not contained in the universe for a specific question. For example, FSALYN and FSALCT, two recodes from the Income and Assets section of the Family Core, are limited to families with at least one member aged 18 or older; families consisting solely of emancipated minor(s) are coded as blanks to indicate that they are out of the universe. The Family file also contains some counters that lack corresponding yes-no indicators. 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: 15,669 families have one member receiving income from wages/salary, 12,226 families have two members (or 2(12,226) = 24,452 persons) with wage/salary income, 1,792 families have three members (or 3(1,792) = 5,376 persons), 454 families have four members (or 1,816 persons), 59 families have five members (or 295 persons), and 11 families have six members with wage/salary income in 1996 (66 persons). Thus the sum of persons across the 30,211 families answering “yes” to FSALYN, the associated yes-no indicator, is 47,674 (15,669 + 24,452 + 5,376 + 1,816 + 295 + 66), which is equivalent to the 47,674 “yes” responses to the person-level source variable, PSAL. Users are advised to check the documentation for each Family file recode in order to determine its person-level source variable.

The Family File Weight

The 1997 NHIS Family File can be thought of as a household-level file for all single family households; 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. Bureau of the Census, 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 1997 NHIS Family file weight [WTFA_FAM]. This methodology 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 1997 NHIS Family file, WTFA_FAM, corresponds to the 1997 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.

1997 National Health Interview Survey
Person-Level File

The Person-level variables are derived from the six sections making up the Family Core of the 1997 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.

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

The Health Status and Limitation of Activity (FHS) section of the Family Core for the 1997 NHIS contains information addressing respondent-assessed disabilities and associated conditions, and overall health status. Users should note that additional and more detailed information on health status is also collected in both the Sample Adult and Sample Child sections of the survey.

Limitation of Activity at the Person Level

Information on activity limitations, including questions about the need for assistance with personal care needs such as eating, bathing, dressing, getting around inside the home, or 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 the very young). 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 1997 NHIS Family Core regarding activity limitations due to physical, mental or emotional problems are substantively comparable to previous NHIS surveys, but users should take note of changes in question wording.

Since cognitive impairment is increasingly recognized as a source of activity limitations among older adults, a new question has been added to determine if anyone in the family is limited because of difficulty remembering or periods of confusion. Other new indicators in this section identify family members who have difficulty walking without any special equipment. In addition, the section also contains measures indicating children who receive Special Education or Early Intervention Services.

Conditions

Each family member with a previously mentioned limitation is asked about the condition or health problem associated with that limitation (up to five conditions may be indicated), as well as the amount of time he/she has had the condition. The original response categories in this section of the survey are broad in scope, and vary according to age. Family members under age 18 were subject to the following fixed condition categories in the instrument: “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" (to which the adult respondent could then supply a verbatim response no more than 20 characters in length). The fixed response categories in the instrument for adults were equally broad: "vision/problem seeing", "hearing problem", "arthritis/rheumatism", "back or neck problem", "fractures, bone/joint 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" (again, a verbatim response of no more than 20 characters was possible).

Many 1997 respondents provided conditions that did not fall into the fixed response categories specified in the instrument. These conditions were recorded verbatim by FRs in one of the 20-character fields indicating "other impairments" (or both, in the event that the respondent reported two such conditions). The verbatim responses were analyzed during data processing and, where appropriate, 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; these were assigned numbers 19-34. Any verbatim conditions that could not be back-coded to one of the original categories or, for adults, recoded to one of the new categories remained in the "other impairment" categories, which 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). 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 weighted frequencies obtained for these conditions may be underestimated. We therefore caution data users that these variables should be analyzed with care. Moreover, the FHS variables (the 11 child variables and the 34 adult variables) should not be used to estimate prevalence estimates for the conditions they represent, because only those respondents who first reported a limitation were asked the condition questions that followed. More information regarding conditions is obtained in 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. LACHRON 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.

Also for 1997, a series of age-group-specific recodes (e.g., under 18 versus 18 and over) regarding limitation of activity and duration of limitation have been created. Because the questions about limitation of activity in the redesigned NHIS are being asked in a different context, of differing age groups, and are more general (in some cases) or more specific (in other cases) than in previous years, the degree to which a respondent is limited cannot be determined. However, the use of these new questions and recodes should eventually enable researchers at NCHS to define new categories identifying the extent of limitation.

Chronic Conditions

Each condition reported in 1997 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 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 from the date of onset, 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;" these conditions are considered chronic. Again, because the presence of a limitation determined eligibility for the subsequent condition questions and, in turn, the chronicity recodes, we caution data users that these variables may *not* produce accurate estimates of the prevalence rates of chronic conditions, and should be analyzed with care.

Table 3. FHS Categories 1-34 and ICD-9 Ranges (Adults, aged 18+)

NHIS Category	ICD-9 Codes
1 - Vision / problem seeing *Disorders of the eye and adnexa*	360-379
2 - Hearing problem *Some level of deafness or a 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, 733.2, 737
5 - Fractures, bone or joint injury *Specific mention of bone or joints*	800-848, 850-999
6 - Other injury *W**ithout** 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

NHIS Category	ICD-9 Codes
9 - Hypertension or high blood pressure	401-405
10 - Diabetes	250.6
11 - Lung / 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	740-759
14 - Mental retardation	317-319, 742.1, 758.0
15 - Other developmental problem *Includes learning disabilities*	315, 343, 783.4
16- Senility (and other cognitive problems)	290
17 - Depression / anxiety / emotional problem *Includes neurotic disorders, personality disorders, and other nonpsychotic mental disorders, ...excluding alcohol & drug related problems and developmental problems*	300-302, 306-314
18 - Weight problem *Indicates a weight-related problem*	
19 - Missing limbs (or part) *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 & Metabolic Diseases & Immunity Disorder* ***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
28 - Tumors, benign & unspecified *Any mention of "tumor" without mention of cancer or "malignant", etc.*	
29 - Alcohol & drug related problems *Any mention of "alcohol", "drugs" (or specific drug types), or substance abuse*	

NHIS Category	ICD-9 Codes
30 - Other mental conditions *Any mental disorders **not coded to 14 or 15 or 17***	290-299.9
31 - After effects of surgery or other medical treatment *Any mention of "surgery", "operation", or otherwise indicates treatment as the causal condition; includes recent or ongoing 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 in the verbatim response.*	
33 - Fatigue/Tiredness *Any mention of tiredness, stiffness, weakness in the verbatim response without referring to any specific part of the body*	
34 - Pregnancy *Any mention of "pregnancy" or "childbirth"*	
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)

The data from the Family Core Injury Section of the 1997 NHIS contain information about medically attended injuries and poisonings that occurred to any member of the family within a three-month reference period. All injury and poisoning information is provided by the family respondent. Injury information is found in three files: (A) the Person file, (B) the Injury Episode file, and (C) the Injury Verbatim file. Poisoning information is found in two files: (A) the Person file and (B) the Poison Episode file.

A. Person File

The Person file contains the number of times a person was injured by specific causes, the number of times a person was injured while doing specific activities, the number of times a person was injured in specific places, and the number of injury episodes a person has in the Injury Episode file. During the editing process, some injury episodes were removed. These included episodes with no information, episodes that did not occur within the reference period, duplicate episodes, and episodes consisting solely of health conditions that could not be classified according to nature-of-injury codes 800-959 or 980-999 of the Ninth Revision of the International Classification of Diseases (ICD-9-CM). In two instances, the respondent reported a person as having more than four injury episodes. Since the NHIS only collects detailed information on the four most recent injury episodes, we do not have any information on these additional injury episodes.

The Person file also contains the number of poison episodes the individual has in the Poison Episode file. During the editing process, some poison episodes were removed from the Poison Episode file. These included episodes with no information, episodes that did not occur within the reference period, duplicate episodes, and episodes that involved illnesses such as poison ivy or food poisoning. The variable that indicates the number of poison episodes the

individual has in the Poison Episode file also contains episodes that did not meet the criteria for poisoning. In two instances, the respondent reported a person as having more than four poison episodes. Since the NHIS only collects detailed information on the four most recent poison episodes, we do not have any information on these additional poison episodes.

Recall Period and Weights

Questions in the Injury section of the NHIS have a recall period of the "last 3 months." In general, only 3-month estimates should be made from the variables in the Person file. However, annual estimates can be obtained for the variables INJCT and POICT by multiplying: (variable)(4)(WTFA).

B. Injury and Poison Episode Files

The Injury Episode file is an episode-based file, with each injury episode consisting of one or more injuries. A person may have up to four injury episodes and will appear in this file as many times as he/she has unique injury episodes. Each episode must have at least one injury classified according to the nature-of-injury codes 800-959 or 980-999 in the Ninth Revision of the International Classification of Diseases (ICD-9-CM). Other health conditions that were reported as occurring with the injury, even if they are not classified according to the nature-of-injury code numbers 800-959 or 980-999, are also included in the Injury Episode file. This resulting file contains information about the cause of the injury episode, what the person was doing at the time of the injury episode, the date and place of occurrence, the elapsed time between the date of the injury episode and the date of the interview, whether the person was hospitalized, whether the person missed any days from work or school, whether the injury episode caused any limitation of activity, ICD-9-CM diagnostic codes, and ICD-9-CM external cause codes. The ICD-9-CM diagnostic codes and external cause codes were assigned according to responses given to questions FIJ.050 to FIJ.220.

During the editing process, some injury episodes were removed. These included episodes with no information, episodes that did not occur within the reference period, duplicate episodes, and episodes consisting solely of health conditions that could not be classified according to nature-of-injury codes 800-959 or 980-999 of the Ninth Revision of the International Classification of Diseases (ICD-9-CM). In two instances, the respondent reported a person as having more than four injury episodes. Since the NHIS only collects detailed information on the four most recent injury episodes, we do not have any information on these additional injury episodes. This file only contains information about injury episodes. Other person level information can be obtained by linking the Injury Episode file to other 1997 NHIS data files using the household serial number (HHX), family serial number (FMX), and person number (PX).

The Poison Episode file is also an episode-based file. A person may have up to four poison episodes and will appear in this file as many times as he or she has unique poison episodes. Included in this file is the cause of the poisoning, whether a poison control center was contacted, the date of the poisoning, the elapsed time between the date of the poison episode and

the date of the interview, whether the person was hospitalized, and whether the person missed any days from work or school.

During the editing process, some poison episodes were removed from the Poison Episode file. These included episodes with no information, episodes that did not occur within the reference period, duplicate episodes, and episodes that involved illnesses such as poison ivy or food poisoning. In two instances, the respondent reported a person as having more than four poison episodes. Since the NHIS only collects detailed information on the four most recent poison episodes, we do not have any information on these additional poison episodes. This file only contains information about poison episodes. Other person level information can be obtained by linking the Poison Episode file to other 1997 NHIS data files using the household serial number (HHX), family serial number (FMX), and person number (PX).

After reviewing the data, it was discovered that 47 episodes coded “06" (Something else) for question FIJ.340 (POITP) did not meet the criteria for poisoning. Rather than remove these episodes, a new variable (POITPR2) was created which contained the original categories in variable POITP plus additional categories that could be used to classify the 47 episodes that may not be poisonings. These 47 episodes were recoded to values “07" (“Allergic/adverse reaction to medication or other substance”) or “08" (“Something else - NOT poisoning”). The latter value includes such things as spraying paint or hair spray into the eyes, chemotherapy, and sun poisoning. It is suggested that the 47 episodes that did not meet the criteria for poisoning be removed prior to calculating national estimates of poisoning.

Recall Period and Weights

Questions in the Injury section of the NHIS have a recall period of the “last 3 months.” In general, annual estimates of episodes (i.e., events) can be made from variables in this section: (variable)(4)(WTFA) = annual estimate of variable. Annual estimates of the number of *people* injured cannot be made from these files due to the limited three month reference period. Analysts are cautioned to check the 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

These files do not contain the design variables used in variance estimation. To obtain the design information, the Injury Episode file and the Poison Episode file must be linked to the Person file. Design information from the entire sample must be used in variance estimation.

Technical Notes

There are two variables on the Injury Episode file that describe the cause of the injury. These variables are CAUS and ECAUS. CAUS is the actual item found in the questionnaire. For each injury, the interviewer selected the category of CAUS that he/she felt best described the injury based on responses that were given to questions FIJ.050 (IJTYPE) and FIJ.070 (IJHOW).

ECAUS is a new variable that describes the cause of the injury using categories based on ICD-9-CM external cause codes. The category into which an injury was placed was based entirely on the first ICD-9-CM external cause code listed for that injury. Appendix 1 which is found in the Injury Episode file 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 episode and the date of the interview, and the variable RPCKDMP, which indicates the elapsed time between the date of the poison 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 poison questions may have been completed earlier than indicated by the date of the interview recorded by CAPI. If this occurred, the time elapsed between the date of the injury or poison episode and the date of the interview would actually be less than indicated by variables RPCKDM and RPCKDMP. It is for this reason that value “92" (92-99 days) was created and 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 poison 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 poison 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 poison episode and the date of the interview could be anywhere from one day to 30 days. For injury and poison episodes that occurred during the month before the interview, this means that the amount of time between the date of the injury or poison 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 poison episode and the date of the interview in months is only used when the day of the injury or poison episode was not given.

C. Injury Verbatim File

The Injury Verbatim file contains the edited narrative text descriptions of the injury provided by the respondent and includes the body part injured, the kind of injury, and a description of how the injury happened. (The pre-edited responses are “verbatim” only insofar as the interviewer could write them down and condense them to fit the field size.) Editing was done only to protect the injured person’s confidentiality. Text descriptions used to replace the original non-compliant text are surrounded by arrows (< >). Grammatical and/or spelling errors were not corrected. The following changes were made to the file in order to protect the injured person’s confidentiality:

- person names (first, middle, and/or surnames or initials) were replaced with <He>, <She>, or <Injured>;
- 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 which indicated unusual personal behavior or events was modified to make it less remarkable;
- any group or organization which 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), IJKIND1, IJKIND2, IJKIND3, IJKIND4 (all referring to the kind of injury), and IJHOW1, IJHOW2, IJHOW3, IJHOW4 (all referring to how the injury 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 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 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 1997 NHIS contain information addressing access to health care and utilization services. The FAU section 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 differ from earlier years, with the exception of two questions: delay in receiving medical care (FAU.020), and not receiving needed medical care due to cost (FAU.040). The data items for Hospital Utilization are similar to those questions from the Hospital Probe and Hospital Page contained in the core questionnaire from prior years. In addition, the data items for Health Care Contacts are similar to the 2-week doctor visit probe questions from previous years and include visits from medical doctors as well as other health care professionals.

In previous years' surveys, 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 1997 NHIS distinguishes between home care and office visits, and includes separate questions for both. Moreover, the 1997 instrument allows respondents to consider an expanded list of health care professionals; respondents are 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 is a question (FAU.210) 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 1997 NHIS may differ from those for earlier years of the NHIS due to changes in the questions and/or the context of the questions. Thus, 1997 estimates of health care utilization 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 previous years of the NHIS. Users are advised to compare 1997 NHIS questionnaire items pertaining to health care utilization to those used in previous NHIS surveys.

Technical Notes

Some changes in question wording were introduced in mid-1997 for the items asking about delays to medical care (FAU.010), hospitalizations in the last 12 months (FAU.050), 2-week telephone contacts with health care professionals (FAU.150), and number of 2-week telephone contacts (FAU.170). (Instrument version 3.1 was used during the first half of 1997 and thus contains the original questions, while version 3.5 was used in the latter half of 1997 and contains the modified questions.) These slight modifications in wording may affect respondents' answers to the questions. Analysts are therefore advised to read the notes in the file documentation for information pertaining to these changes.

IV. Health Insurance Section (FHI)

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

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

- Type of health care coverage (Medicare, Medicaid, military/VA, CHAMPUS/TRICARE/CHAMP-VA, State-sponsored health plan, other government programs, Indian Health Service, private insurance);
- Managed care arrangement for those covered by Medicare and Medicaid;
- Private insurance characteristics reported by the respondent (HMO or PPO status, source of coverage, existence of employer subsidies for premiums, amount paid by individual/family);
- Periods of time without health insurance and reasons for no health insurance;
- Out-of-pocket costs in past year (general categories).

Technical Notes

A major reordering of questions occurred in version 3.5 of the FHI section: several questions determining the knowledge and availability of family members regarding the family's health insurance were moved to the start of the section (they were the third, fourth, and fifth questions in the 3.1 version of FHI). Although these variables were not edited or included on the public use data file, this reordering nevertheless affected the responses to HIKIND, which asked respondents with health insurance to indicate the type of health insurance or coverage they had. In version 3.1, only one such person failed to indicate what type of health insurance he/she had in HIKIND. However, in version 3.5, approximately 900 persons with health insurance failed to indicate the type of insurance they had in response to HIKIND. These persons are coded "8" (or "Not ascertained") for the following variables and recodes: HIKINDA through HIKINDJ, MEDICARE, MEDICAID, PRIVATE, IHS, MILITARY, OTHERPUB, OTHERGOV. This inconsistency was corrected in the 1998 NHIS instrument.

In looking at the verbatim responses to the questions (HIPNAM, NEXTPNM, NEXTPNM2, and NEXTPNM3) asking respondents for the names of their private health insurance plans, some respondents indicated plans that were not private health insurance plans, or single service plans that were excluded from the private health insurance coverage category. These persons were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, IHS, MILITARY, OTHERPUB and OTHERGOV. Additionally, 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 their appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, PRIVATE, IHS, MILITARY, OTHERPUB and OTHERGOV. Analysts are therefore advised to use the recodes MEDICARE, MEDICAID, PRIVATE, IHS, MILITARY, OTHERPUB and OTHERGOV for types of health care coverage, because these take into account the above-mentioned back edits. In contrast, the data contained in PHICOV and HIKIND were not back-edited and reflect the respondents' original replies.

V. Socio-demographic Section (FSD)

The Socio-demographic (FSD) section of the Family Core for the 1997 NHIS collects information on place of birth 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 1996, an estimate of their earnings from wages in 1996, and whether their employer provided health insurance.

Analysts may also refer to the Adult Core socio-demographic section (ASD) for additional socio-demographic data regarding those individuals selected as Sample Adults.

Technical Notes

The "major activity" variable available in previous NHIS data files is now roughly approximated by the variables DOINGLW, which asks respondents what they were doing in the seven days prior to the interview week, and WHYNOWRK, which asks respondents what they were doing during that same time period if they were not working. Additional occupation and work-related variables can be found in the Sample Adult Demographics section (ASD). Editing procedures have reconciled inconsistencies between DOINGLW and variables in the Sample Adult section, but no such corrections are possible for non-sample adults. Moreover, data captured in DOINGLW have not been reconciled with other variables of the same subject matter within other sections of the data file.

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.

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 1996, 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 and PSEINC) 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; other pensions; Supplemental Security Income (SSI); Social Security Disability Insurance (SSDI); Welfare/Aid to Families with Dependent Children (AFDC)/General Assistance; 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). Several questions

specifically asked about income in “the previous calendar year” or “last year;” in other instances, however, the questions did not specify a time period. Thus, it is conceivable that some respondents may have been unclear about the reference period being covered in the question (this problem has been rectified in the 1998 instrument). Users are cautioned to take note of this potential problem with the 1997 data; see the “Notes:” section within the file documentation for these questions.

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 1997 in favor of a single overall estimate of combined family income. And, unlike previous NHIS surveys, the 1997 instrument contained three questions to elucidate the family’s combined income from all sources during 1996, including a question (FIN.250) that allowed the respondent to supply a specific dollar amount (up to $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 1996 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. Bureau of the Census regarding 1996 poverty thresholds (see *Poverty in the United States, 1996*; U.S. Bureau of the Census, Appendix A, Table A-2). A ratio of the 1996 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). Lastly, respondents were also asked whether the family’s house or apartment was owned, being purchased, rented, or occupied by some other arrangement. The last question did not include a reference to a specific year, so there is a possibility that some respondents were confused about the time period being covered.

Program Participation

Respondents were asked in the final part of the FIN section if any family members were authorized to receive food stamps in 1996, and if so, which members. In addition, respondents were asked if family members had received Supplemental Security Income (SSI) or (in a separate question) Social Security Disability Insurance (SSDI) because of a disability. Also asked was

whether any family member(s) had *ever* applied for SSI or SSDI (even if the claim(s) had been denied). Lastly, if one or more family members had received food stamps or Aid to Families with Dependent Children (AFDC), the respondent was asked, in two separate questions, for how many months during the last calendar year food stamps and/or AFDC had been provided.

1997 National Health Interview Survey
Sample Adult File

The Sample Adult section of the 1997 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. With the redesign, the six lists have been reduced to a single list for adults and a single 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, gastrointestinal conditions, cancer, diabetes, other respiratory conditions, renal conditions, joint symptoms, sensory impairments, pain, hearing, vision, oral health, and mental health. Table 4 shows the specific health-related conditions covered in the 1997 NHIS, as well as the various reference periods covered by the questions. Note that no question in the ACN section refers to a two week reference period.

In addition to being less cumbersome for users, the elimination of the six condition lists allows all conditions to be analyzed at the person-level; in contrast, the previous NHIS design required that condition analyses be carried out at the condition-level. Moreover, there are several notable differences in the way information on conditions is collected in the redesigned NHIS. As mentioned, all data in the 1997 Sample Adult component were required to be self-reported; proxy respondents were not allowed. In addition, most questions in the 1997 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 impairment, and vision impairment. All of these changes must be considered when attempting to compare 1997 condition prevalence estimates 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.

The release of the ACN data pertaining to colds and intestinal illnesses of sample adults has been postponed indefinitely. NCHS staff believe that the items used in the 1997 instrument are not adequately measuring the conditions in question. The relevant questions may be re-designed for future NHIS instruments. It is unlikely that the data obtained in 1997 will be released.

Table 4. Sample Adult File: Conditions and Reference Periods

Reference Period in 1997 NHIS					
CRQ #	**Condition**	**Ever**	**12 months**	**3 months**	**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.331	Ulcer	X	X		
ACN.331	Cancer (Any + list max. 3 of 30 specific types)	X			
ACN.350	Diabetes	X	X		X
ACN.360	Hay fever		X		
ACN.370	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.370	Pregnancy				X
ACN.410-ACN.420	Hearing impairment				X

Reference Period in 1997 NHIS					
CRQ #	**Condition**	**Ever**	**12 months**	**3 months**	**Now**
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 cancer to specify up to three kinds of cancer or to indicate that he/she 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 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 on respondent-assessed limitations regarding various physical tasks, as well as social activities, the underlying conditions associated with these disabilities, and information regarding the duration of the conditions. While the AHS section is substantively similar to the FHS section in the Person file, more detailed questions are generally found in the AHS section. For example, the question regarding special equipment in the FHS section (FHS.210) does not mention any specific items. In contrast, the Sample Adult "special equipment" question (AHS.070) specifically mentions canes, wheelchairs, special beds, and special telephones. Additionally, the limitation questions in the Sample Adult section are designed to assess the degree of physical limitation as well as the severity of reported health conditions.

Health Status/Health Indicators

The first few questions in this section determine the number of work-loss days and bed days reported during the 12 months prior to the interview. In addition, respondents were asked if their health is better, worse, or the same compared with 12 months ago.

Limitation of Activity

The limitation questions regarding routine tasks and activities in the AHS section ask 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 to a movie or sporting event, 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).

For three activities (i.e., shopping, participating in social activities, and relaxing at home), respondents had the opportunity to respond in the interview that they "do not do this activity." For other activities (related to walking, climbing, standing, sitting, stooping, reaching, grasping, carrying, and pushing), respondents were not permitted in the instrument to use this response category, but could be 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 twelve activities, he/she was then asked what condition(s) cause the difficulty as well as how long he/she has had the condition. This format is quite similar to that found in the FHS section.

Conditions

Each sample adult indicating any 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 be indicated), as well as the amount of time he/she has had the condition. Sample adults were given the following response categories: "vision/problem seeing," "hearing problem," "arthritis/rheumatism," "back or neck problem," "fractures, bone/joint 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" for which, (as in FHS, a verbatim response of no more than 20 characters was possible for these fields).

Many 1997 respondents provided conditions that did not fall into the fixed response categories specified in the instrument. These conditions were recorded verbatim by FRs in one of the 20-character fields indicating "other impairments" (or both, in the event that the respondent reported two such conditions). The verbatim responses were analyzed during data processing and, where appropriate, assigned codes corresponding to the original response categories (the first 18

for sample adults). Furthermore, 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, which were 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 20). 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 ad hoc categories were not included on the flash cards given to respondents during the course of the interview, it is possible that weighted frequencies obtained for these conditions may be underestimated. We therefore caution data users that these variables should *not* be used to estimate prevalence rates for the conditions they represent.

Recodes

The recode FLA1AR is a summary measure indicating that the sample adult respondent reported *any* difficulty with *any* one or more of the 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. The section also includes a time recode (ALANTR1-34; ALANTR90 and ALANTR91) for each of the 36 categories, which features a cut-point at three months so that users may easily identify chronic from non-chronic health problems based on the duration of the condition.

III. Adult Health Behaviors Section (AHB)

With the 1997 redesign of the NHIS, four health-related behaviors are included in the Core: cigarette smoking, physical activity, alcohol use, and body weight. Of these, only body weight was previously in the Core, and it was often based on proxy report by the family respondent.

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. 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 is used in the 1997 NHIS. In addition to smoking status, data are collected on age of initiation, amount smoked, and quit attempts.

Physical activity

The physical activity questions are substantially different from those in previous NHIS questionnaires. Because of the large number of topic areas covered in the new NHIS, only a brief set of physical activity questions could be included in the Sample Adult Core Module. For this

reason, the questions are general and lend themselves to broad classifications of activity levels. Special topical Modules will continue to provide information on specific physical activities.

Due to a technical oversight, the 1997 physical activity data provide a unique opportunity for methodological analyses. In quarters one and two of 1997, the items asking about frequency of vigorous and moderate activity specified a minimum duration of 20 minutes (e.g., "How often do you do vigorous/moderate activities for at least 20 minutes . . . ?"). In quarters three and four, the minimum duration was changed to 10 minutes and will remain 10 minutes for the foreseeable future (e.g., "How often do you do vigorous/moderate activities for at least 10 minutes . . . ?"). However, subsequent questions allowed those who engaged in such activity for a minimum of 10 minutes to indicate the number of minutes (beyond 10) that they generally exercised, so that it is possible to determine which respondents in quarters three and four were moderately active for at least 20 minutes. The 1997 data file includes variables based on both versions of the question (note the differences in the universes), as well as recodes that combine quarters 1-4, restricting responses in quarters three and four to those who gave a duration of at least 20 minutes. This permits annual estimates of persons who engaged in vigorous or moderate activity for a minimum of 20 minutes. After 1997, these recodes will not be applicable, and the data file will contain only one variable for each question, with a 10 minute minimum duration. The following table provides a guide to the variable names and their recodes for the two different versions of the 1997 NHIS.

Table 5. 1997 Physical Activity Variable Names: Guide to Identifying Versions

Question #	Question Content	Quarters 1-2 Variable	Quarters 3-4 Variable	Annual (Recode) Variable
AHB. 090	**Frequency of Vigorous Activity**			
	Number of units	VIGNO	VIGNO2	VIG20NO
	Time unit	VIGTP	VIGTP2	VIG20TP
	Standardized (times per wk)	VIGFREQW	VIGFRQW2	VIG20FRQ
AHB.100	**Duration of Vigorous Activity**			
	Number of units	VIGLNGNO	VIGLNGN2	VIG20LGN
	Time unit	VIGLNGTP	VIGLNGT2	VIG20LGT
	Standardized (# of minutes)	VIGMIN	VIGMIN2	VIG20MIN

AHB. 110	**Frequency of Moderate Activity**			
	Number of units	MODNO	MODNO2	MOD20NO
	Time unit	MODTP	MODTP2	MOD20TP
	Standardized (times per wk)	MODFREQW	MODFRQW2	MOD20FRQ
AHB.120	**Duration of Moderate activity**			
	Number of units	MODLNGNO	MODLNGN2	MOD20LGN
	Time unit	MODLNGTP	MODLNGT2	MOD20LGT
	Standardized (# of minutes)	MODMIN	MODMIN2	MOD20MIN

Alcohol

The alcohol questions are a variation on questions that have appeared in the NHIS periodically since the 1970's. Extensive supplements on alcohol use were fielded in 1983 and 1988, with smaller alcohol question batteries included in other data years, such as 1985, 1990, and 1991. As with other sections of the questionnaire, the reference period was chosen to capture as much information for as many people as possible. Although a 12-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, 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 years, however. The most notable change is in the order of the lifetime drinking questions ("Ever 12 drinks in one year?" now precedes "Ever 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 12-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, converting the various time response options to standardized units, are provided in the data file (i.e., days per week, days per month, days per year). (Again, these standardized variables assume constant rates of occurrence across time periods.) The same was also done for the question about how often the respondent had 5 or more drinks in one day during the past year. 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 basic Core questionnaire for many years, with proxy reporting allowed for family members not home at the time of the household interview. Prior to 1997, height and weight were self-reported only in selected special topic NHIS questionnaires. It should be kept in mind that estimates based on proxy reports may differ from those based on self-reports. They also will differ from estimates based on physical measurements, such as those available from NCHS's National Health and Nutrition Examination Survey.

Beginning in 1997, when a very large or very small value was reported for either height or weight, the data in 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.

Two recodes were created to make the height and weight data easier to use: Desirable Body Weight (DESIREWT) and Body Mass Index (BMI). These are based on the Metropolitan Life Insurance Company standards of desirable body weight (1983), which have been used in the NHIS since 1985. The Desirable Body Weight recode is restricted to persons whose height and weight are within the range provided by the Metropolitan Life chart of desirable weight for height. Because of this, the number of unknowns is substantially larger in the DESIREWT recode (1,816) than in the BMI recode (1,086). In contrast, the BMI is calculated from the formula: weight (kg) / height (meters)^2. Unlike the Desirable Body Weight recode, the BMI recode includes all persons who provided height and weight information, including those for whom individual height and weight values were changed to "96" on the public use file. For BMI, the values are continuous with two implied decimal places. Using the BMI, overweight is defined as values greater than or equal to 25.0 for both men and women, while BMI values greater than or equal to 30 indicate obesity. The BMI cutpoints 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 Adult Health Care Access and Utilization (AAU) section of the 1997 NHIS consists of four parts: Access to Care, Dental Care, Health Care Provider Contacts, and Immunizations.

The "Access to Care" section includes questions related to: having a usual place for sick care, having a usual place for routine/preventive care, change in place of care, delay in getting medical care, and the inability to afford medical care. While all of these topics have been covered in previous years, there have been some changes in question wording as well as the order in which questions were asked. For example, in 1996 the question about the reason for delaying care asked "because of worry about the cost?" In 1997 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" section includes only one question: length of time since last dental visit. This item has been asked in previous years, but the 1997 question has been re-phrased.

The section on "Health Care Provider Contacts" contains questions similar to the 2-week doctor visit probe questions from previous years, and includes visits to medical doctors and other health care professionals including chiropractors. As with the Person-level FAU section described previously, earlier NHIS surveys focused on physician contacts, visits, and home care that included only contact with a medical doctor or a health care professional working with or for a medical doctor. In addition, in prior years, home care visits were included in the question asking about visits or contacts with a doctor's office, hospital, etc. In 1997, questions about home care are asked independently of visits to the doctor's office, the hospital, etc. Also new for 1997 is a question asking about the number of visits to a hospital emergency room in the past 12 months.

Lastly, the "Immunizations" section includes two questions related to adult immunizations: flu shots and pneumonia vaccinations.

V. Adult Demographics Section (ASD)

The Sample Adult Demographics (ASD) section provides information on the year and decade that non-U.S.-born sample adults have been in the United States (permanently). Additionally, employed sample adults were asked several questions regarding their occupation, industry, workplace, and employment conditions 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.

Years in United States for Foreign-born Respondents

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.

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 Appendix (following the file 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. With the exception of the question determining employment situation, the remaining questions described in this paragraph are all new for 1997.

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 had a blood test for AIDS, their reasons for getting or not getting tested, and the location of any testing. 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. These questions are similar to those asked in the AIDS Knowledge and Attitudes Supplements that were included in the NHIS from 1987-1995.

Technical notes

There are a large number of "8" ("Not ascertained") codes in variable RSGVN (ADS.100) because of an incorrect skip pattern in quarters one and two. The variable was edited to reflect the correct skip pattern, while the error in the skip pattern was corrected for quarters three and four. It is not necessary to adjust the weight if data for the entire year are used for analysis. However, the weight should be multiplied by a value of 2 when analyses are based on only two quarters of data.

1997 National Health Interview Survey
Sample Child File

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

The Child Health Status (CHS) section of the 1997 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 disorder, 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.

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.

The release of the CHS data pertaining to colds and intestinal illnesses of sample children has been postponed indefinitely. NCHS staff believe that the items used in the 1997 instrument are not adequately measuring the conditions in question. The relevant questions may be re-designed for future NHIS instruments. It is unlikely that the data obtained in 1997 will be released.

Technical Notes

During quarters one and two, there was a CAPI error in the skip pattern for one of the items in Part A of the 1997 CHS section. As a result, the parents of 14 two-year-old children who had experienced an asthma attack in the past 12 months were asked several questions intended for the parents of children aged 3-17 years rather than for the 0-2 age group. Most of the questions were identical for both age groups, but there were two additional questions about the older group that were not relevant for the younger group. The usable data from the identical questions were moved to the appropriate fields, and the instrument's skip pattern changed for quarters three and four. However, there may be some further methodological consequences, because even though the questions were identical, one of the additional questions was placed at the end of the list while the other was placed in the middle of the list. No adjustment to the weighting is required for these fields.

All of the data from the Child Behavior Checklist questions for the age group 4-11 years (CHS.401 and CHS.441) were dropped for quarters one and two because there was an error in the questionnaire. The incorrect reference period for those quarters was given as "in the past two months." The corrected reference period, "in the past six months," was inserted in the instrument for quarters three and four. As a result, there are only six months of data for the following variables: CMHAGM22, CMHAGM23, CMHAGM24, CMHAGM25, CMHAGF22, CMHAGF23, CMHAGF24, and CMHAGF25. The variables CMHAGM22, CMHAGM23, CMHAGM24, CMHAGF22, CMHAGF23, and CMHAGF24 are used to create several mental health indicators; 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 file documentation for the Sample Child file. When using either CMHAGF25, CMHAGM25, or the recodes MHIGRL4 and MHIBOY4, the analyst must create a semi-annual weight that is the annual weight multiplied by 2.

There was an additional error affecting 58 cases in Child Behavior (Part B) where the CAPI system did not follow the correct gender-based skip pattern. Since most of the affected questionnaire items were identical for boys and girls, these data were transferred to the appropriate fields. However, some of the items were not identical; in these instances, the data were missing and coded as "not ascertained." No adjustment to the weighting is required for these fields.

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

The Child Health Care Access and Utilization (CAU) section of the 1997 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 questions were asked, and rewriting an entire question. For example, 1996 HIS participants were asked if they delayed getting medical care for the sample child "because of worry about the cost...." In contrast, the 1997 question 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 for 1997.

Lastly, 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 twelve 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); 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. In 1997, 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 1997 survey asked about contacts during the past twelve months. Lastly, new for 1997 was a question asking about the number of visits to a hospital emergency room in the past 12 months.

III. Child Immunization File (CIM)

The Child Immunization file of the 1997 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 a randomly selected sample child per family in a household, along with any non-sample children aged 12-35 months (in survey quarters three and four) and 19-35 months (in survey quarters one and two) 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 (although note that the 1997 Child Immunization section contains a new question about chickenpox, or varicella, shots). 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.

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 is as follow:

Data Source: National Center for Health Statistics (1997)

In a bibliography, the suggested citation should read:

National Center for Health Statistics (2000). Data File Documentation, National Health Interview Survey, 1997 (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. Bureau of the Census. (1997). *Poverty in the United States, 1996.* Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census.

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

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

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

U.S. Department of 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 1997 NHIS

The NHIS questionnaire was redesigned for the 1997 survey year, and the instrument was changed from a paper and pencil questionnaire to a new computer-assisted personal interviewing (CAPI) system. The new 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 and was the only module fielded in 1997. The response rates calculated here pertain only to the Basic Module.

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 because of: language problems; no one 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

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

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

Conditional Sample 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

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

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

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

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 selected as the sample child.

Final Immunization Response Rate

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

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 eligible for the survey from interviewed families.

Appendix I, Table 1. Response Rates for the 1997 NHIS

Household	91.8%
Family (Core) - Conditional	98.4%
Family (Core) - Final	90.3%
Sample Adult - Conditional	89.0%
Sample Adult - Final	80.4%
Sample Child - Conditional	93.1%
Sample Child - Final	84.1%
Immunization - Conditional	99.0%
Immunization - Final	84.3%

Appendix II

Race and Ethnicity in the 1997 NHIS[1]

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 (except in the case of children, where the information was provided by proxy), although interviewer-observed race was also recorded through 1996, the last year of the paper questionnaire. This information has proven critical to the understanding of health data collected in the NHIS, and has been widely used in the publication of NHIS data, including 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 and the growing need of persons with multiple racial heritages to indicate this on the Decennial Census and federal surveys, the OMB revised Directive 15 in 1997 after an extensive period of research and public commentary. The new Directive 15 allows respondents to the Census and federal surveys to indicate more than one group in answering questions on race. A complete description of the revised Directive 15, including descriptions of the new race categories and the ordering of race and ethnicity questions, can be found in the October 31, 1997 Federal Register notice (Office of Management and Budget 1997). Although this policy is not expected to be fully implemented 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. The NHIS has been fully compliant with the revised OMB Directive 15 since the fielding of the 1999 questionnaire.

The draft tabulation guidelines for the new Directive 15 (Interagency Committee for the Review of Standards for Data on Race and Ethnicity 1999) recognize that the complete transition from the use of race data collected under the old standard to race data collected under the new standard will take some time, given that the Census and many federal statistical systems have a primary mission to track trends over time. During this transitional period, known as the "bridge," it has been recommended that data systems tabulate data for publication under both the old and new standards. This would help to illustrate 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.

---

[1] Please note that while the title includes the word "ethnicity," the discussion in this section almost exclusively focuses on the race variables in the NHIS. Although the revised OMB Directive 15 contains changes to the collection of ethnicity data (primarily requiring Hispanic origin to be collected before race), these changes were already in effect in the NHIS and are therefore not discussed in any further detail here.

Race Questions in the National Health Interview Survey

In the 1997 NHIS, two questions were asked to obtain information on a respondent's race:

"What race {does/do} {name/you} consider {himself/herself/yourself} to be?" [HHC.360], and

"Which one of these groups, that is (FR: READ GROUPS) would you say BEST represents {your/name's} race?" [HHC.380].

With minor changes, these two questions are worded basically the same way they have been in the NHIS for several years. The first question is asked of all respondents, and the second question is asked only of respondents who give more than one race in the first question. For details on the specific response categories for the race questions, refer to the Questionnaire.

In preparation for revising the NHIS questions to meet the new OMB standards for collecting race data, these items were closely examined to determine their appropriateness under the new guidelines. After several discussions with the staff of the OMB, it was determined that the second of the two NHIS race questions did not meet the spirit of the new directive. As the new guidelines have been interpreted, the second NHIS question is thought to be reflective of the old standard, which required respondents to the Census and surveys like the NHIS to select only one race. As a result, the OMB has asked that the NHIS limit the use of the follow-up question to the presentation of data that show the bridge between data collected under the old standard and data collected under the new one.

Since the second question, commonly referred to as the "follow-up" question, has been integral to the production of NHIS data for many years, the OMB has allowed the NHIS to keep the item on the questionnaire for the next several years. However, although the NHIS has received this special clearance from the OMB, data tabulated under the new guidelines cannot contain information provided by this questionnaire item until sufficient research has been conducted to determine both respondents' feelings about the question and its analytic usefulness. As a result we have created two sets of race variables that represent the old and new OMB standards for race data collection.

Race Variables in the 1997 NHIS Public Use Data Files

There are two kinds of race variables on the 1997 NHIS data file: the first group corresponds very closely to the ones used on previous NHIS data files, and the second conforms to the new OMB Directive 15 standard as described above. The following table summarizes the 1997 NHIS race variables.

Appendix II, Table 1. Description of 1997 NHIS Race variables

1997 NHIS Variable	Variable Categories	Previous NHIS variable	Previous file location	Relationship to OMB Directive 15 (Old/New)	Example of Analytic Use
RACEREC	White, Black and Other	Race recode	PU: 43	Old	Used to define race in *Current Estimates* tables
MRACE_P	White, Black, and detailed groups for AIAN* and API**	Main racial background	PU: 41-42	Old	At the discretion of the analyst
RACE	4 OMB race groups with single multiple race category	NA	NA	New	IH: At the discretion of the analyst; for use in tabulations of NHIS data for publication

* AIAN represents American Indian and Alaska Native (Aleut and Eskimo); ** API represents Asian and Pacific Islander.

Creation and Editing of 1997 Race Variables

The variables RACEREC and MRACE_P were created in the same fashion as their previous NHIS counterparts (National Center for Health Statistics 1996), with two exceptions. First, since observed race was not collected in the 1997 NHIS, it was not used to help classify persons with "Unknown" race on the RACEREC 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.

The variable RACE was created using an algorithm created by NHIS staff that first coded the five race mentions from the survey into the single and multiple race group combinations (shown in bold 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 the 4 OMB race categories and the category "Other" (all in bold font), the variable RACE was created.

Appendix II, Table 2. Coding Scheme for 1997 NHIS Race Data (including single and multiple race mentions)

Number of Category	Sum of Coding (breakdown of RACEFULL sum)	Resulting Category
1	1+0+0+0+0	**White**
2	2+0+0+0+0	**Black**
3	1+2+0+0+0	White/Black
4	0+0+4+0+0	**AIAN**
5	1+0+4+0+0	White/AIAN
6	0+2+4+0+0	Black/AIAN
7	1+2+4+0+0	White/Black/AIAN
8	0+0+0+8+0	**API**
9	1+0+0+8+0	White/API
10	0+2+0+8+0	Black/API
11	1+2+0+8+0	White/Black/API
12	0+0+4+8+0	AIAN/API
13	1+0+4+8+0	White/AIAN/API
14	0+2+4+8+0	Black/AIAN/API
15	1+2+4+8+0	White/Black/AIAN/API
16	0+0+0+0+16	**Other**
17	1+0+0+0+16	White/Other
18	0+2+0+0+16	Black/Other
19	1+2+0+0+16	White/Black/Other
20	0+0+4+0+16	AIAN/Other
21	1+0+4+0+16	White/AIAN/Other
22	0+2+4+0+16	Black/AIAN/Other
23	1+2+4+0+16	White/Black/AIAN/Other
24	0+0+0+8+16	API/Other
25	1+0+0+8+16	White/API/Other
26	0+2+0+8+16	Black/API/Other
27	1+2+0+8+16	White/Black/API/Other
28	0+0+4+8+16	AIAN/API/Other
29	1+0+4+8+16	White/AIAN/API/Other

30	0+2+4+8+16	Black/AIAN/API/Other
31	1+2+4+8+16	White/Black/AIAN/API/Other

Since these variables conform to the revised OMB Directive 15, they were created independently of the MAINRACE variable (see the Background section of this document).

Further Information

Although the race variables included in the 1997 file have been throughly edited and tested, analytic and methodological work with these variables continues. If these analyses should result in changes to the 1997 NHIS race data, information about this can be found on the NCHS web site (see page 3).

References

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. (Users may access this documentation at: http://www.whitehouse.gov/WH/EOP/OMB/html/misc-doc.html)

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 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

Introduction

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling, and the final weights are subject to several adjustments. Any variance estimation methodology must involve numerous simplifying assumptions about the design and weighting. This document provides some simplified conceptual NHIS sample design structures that should allow users of these data to compute reasonably accurate standard errors.

There are several available software packages for analyzing complex samples. A comparison is beyond the scope of this document, but a website, *Summary of Survey Analysis Software*, currently located at

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

provides references and discussion. At NCHS the software package SUDAAN® (Shah, et al. 1997) is generally used to produce standard errors. In this document, SUDAAN computer code is provided, but without guarantees of any kind. The computer code and methods are subject to change without notification to the user. The entire risk as to the results is assumed by the user. NCHS recommends that any analysis of NHIS data be done under the supervision of a statistician who understands analytic methods for sample surveys with complex designs.

Conceptual NHIS design for 1995-2004

A thorough discussion of the NHIS design, the methods used for weighting data, and methods used for variance estimation is well beyond the scope of this document. The discussion presented below outlines some of the basic technical ideas (with some simplifications) that appear in two technical reports (NCHS 1999 and NCHS forthcoming).

To achieve sampling efficiency and to keep survey operations manageable and cost-effective, the NHIS survey planners used "multi-stage cluster sampling" techniques to select the NHIS sample. These methods partition the target universe into several nested levels of strata and clusters of target units. For operational reasons, the NHIS target universe is considered to be all dwelling units in the U.S. that contain members of the civilian noninstitutionalized population. Since the NHIS is a face-to-face survey, a simple random sample of dwelling units would be so dispersed throughout the Nation that the sampling costs of obtaining about 50,000 households would be prohibitive. Furthermore, target domains like black and Hispanic households would not achieve required sample sizes. To achieve survey objectives subject to resource constraints, the NHIS used the following methods of clustering, stratification and sampling.

First, the entire universe was partitioned into about 1900 primary clusters consisting of individual or combined state counties (or equivalent jurisdictions) along with metropolitan areas. These primary clusters are referred to as Primary Sampling Units (PSUs). These PSUs vary greatly in population size. The PSUs with the largest populations (e.g., the New York metropolitan area) easily support cost-effective sampling and are designated as self-representing (SR) PSUs, i.e., they are in the NHIS sample with certainty. Limited resources require that a "representative" sample of the smaller PSUs be selected for data collection. These smaller PSUs, which are sampled, are referred to as non-self-representing (NSR) or non-certainty PSUs. These NSR universe PSUs were stratified using criteria consistent with survey objectives. For the NHIS these NSR PSUs were first placed into strata defined at the state level according to metropolitan status (metro or non-metro). If an NSR stratum in a given state had a large total population, it was further substratified into several smaller strata by grouping PSUs with similar poverty rates together. Thus, the number of NSR strata varies from state to state, and the number of universe PSUs varies from stratum to stratum. To select a "representative" sample of NSR PSUs, two PSUs were selected without replacement within each NSR stratum. In some of the smaller population states only one PSU was drawn from some strata. In either case, a PSU's probability of selection was proportional to its population size relative to its NSR stratum population size.

Within each sampled NSR or SR PSU, the U.S. Bureau of the Census partitioned the universe of dwelling units into smaller geographical clusters called blocks and block groups. For the NHIS all of the PSUs' blocks were partitioned into minority-density substrata based upon black and Hispanic population concentrations as determined by the 1990 Decennial Census. Furthermore, to keep the PSU universe current and to adequately cover the target population, a new housing construction stratum is continually maintained to update yearly coverage for each PSU. In all, a PSU's dwelling units can be partitioned into up to 21 substrata. Large metropolitan SR PSUs tend to have many substrata, while the NSR PSUs tend to have only a few.

Clustering and sampling within these PSU substrata is a complicated process, but it can be conceptualized as partitioning the dwelling units of each substratum into clusters of dwelling units, with each internal cluster expected to have about the same number of dwelling units. This target number will vary by substratum. These clusters form a universe of Secondary Sampling Units (SSUs). A systematic sample of SSUs is selected to represent each substratum. Based upon survey objectives, each substratum has its own sampling rate.

Within each sample SSU all households containing black or Hispanic persons are selected for interview, while only a subsample of the other households are selected for interview. These non-black, non-Hispanic households are sampled at different rates within the 21 substrata. Some information is collected on all household members, and other information is obtained only for a sampled person. For example, the sample adult questionnaire requires that one sample adult per family be selected for interview.

This hierarchy of sampling just discussed 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. An important adjustment is a quarterly post-stratification to 88 age/sex/race/ethnicity Census control totals.

Internally, NCHS uses the design and weighting information just discussed to formulate appropriate variance estimators for NHIS statistics. While recognizing the need to provide accurate information, NCHS also has to 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 ensure confidentiality. In order to satisfy this disclosure constraint, many of the original design strata, substrata, PSUs and SSUs have been masked by using various cluster collapsing, mixing, and partitioning techniques. Through this process the original NHIS design-level variables have been transformed into public use design-level variables. The public use design structures perform reasonably well when compared to internal NCHS design structures. The sampling weights have not been changed in any way for the public data. Data users wanting access to the internal NCHS data have the option of working with the NCHS Research Data Center (for further information see http://www.cdc.gov/nchs/r&d/rdc.htm).

Design Information Available on the NHIS Public Use Data

The following public use design variables are used for variance estimation. The user should check the file documentation for exact locations of these variables for each of the files.

Appendix III, Table 1. Variables Used for Variance Estimation, 1997 NHIS

Variable Name	Variable Label
STRATUM	Stratum for variance estimation
PSU	PSU for variance estimation
WTFA	Weight - Final annual

As discussed earlier, to mask true geographical locations, the STRATUM and PSU levels are pseudo levels, and hence are the simplified conceptual NHIS sample design structures needed for variance estimation. These simplified design structures do not support geographical analyses below the regional level.

☹ **CAUTION**. Significant changes have been made to the Stratum and PSU values for 1997 NHIS Public Use data. Also, more strata have been provided (compared to the 1995 public release) to improve statistical efficiency in various statistical estimation procedures. The sample design structure variables provided on the 1997 NHIS Public Use data files are not comparable to those of previous data years. Users are cautioned that variance estimation structures discussed in this

and previous documents are for individual years only, not for multiple years of combined NHIS data. For example, if 1996 NHIS Public Use data are combined with 1997 NHIS Public Use data, point estimates are possible, but combined-year design structures for variance estimation are not publically available.

Variance Estimation Method for Public Use Data

The method described below is applicable to the following 1997 NHIS Public Use data sets: Household, Family, Person, Sample Adult, Sample Child, and Immunization.

For this method of variance estimation, the NHIS sample is treated as having 339 strata, each containing two sampled PSUs. While the PSUs are not duplicated, limited public design information and mathematical convenience suggest that the PSUs be treated as sampled with replacement (WR). This suggested method will provide somewhat more conservative standard errors on average compared with an internal NCHS variance estimation method. This method has the advantage of being applicable to the many complex survey sample design computer programs that require exactly two sample PSUs per stratum. Additionally, this method is robust when analyzing subsetted data (see the section "Subsetted Data Analysis" below).

When implementing this method, the user should observe 678 PSUs when using the full database. For the above simplification of the NHIS sample design structure, the following SUDAAN design statements may be used.

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

Note that the input file must first be sorted by STRATUM and PSU variables.

☹ **CAUTION**. A typically-used rule of thumb for number of degrees of freedom to associate with a standard error is the quantity *number of PSUs - number of strata* . This rule is typically applicable to a two-PSU per stratum design whenever the variance components by stratum are roughly of the same magnitude. Furthermore, the applicability of this rule is dependent upon the variable of interest and its interaction with the design structure (for additional information, see Chapter 5 of Korn and Graubard 1999). The rule of thumb number of degrees of freedom for the method given above is 339. Note, that 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 an NHIS 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 of NHIS variables are restricted to select sub-domains, e.g., persons aged 65 and older. To save on storage, some users delete all records outside of the domain of interest. This procedure of keeping only select records is called subsetting the data. With a subsetted data set one can produce correct point estimates, e.g., the sub-domain means, but for general complex survey designs, standard errors produced by most software packages are computed incorrectly when using a subsetted design structure. When data is collected using a complex survey design, and the data are then subsetted, it is likely that sample design structures would be compromised. Design structures are compromised with subsetted data because 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 sub-domain while using a complete (unsubsetted) data file which has all sample design information (see SUDAAN manual for details).

Subsetting the Public Use NHIS with SUDAAN

Strategy 1 Use Method above with the MISSUNIT option on the NEST statement:

NEST **STRATUM PSU / MISSUNIT ;**

This strategy can be justified since, if a WR design has exactly two PSUs per stratum, and some PSUs are removed from the database, then the SUDAAN MISSUNIT option performs a fix-up that produces a standard error identical to that achieved when using a full data set and SUBPOPN statement. Note that other calculations like design effects, degrees of freedom, and standardization may need to be carried out differently. The user is responsible for checking that subsetted input leads to correct results.

Strategy 2 Use Method above with the SUBPOPN statement with the full dataset:

PROC ... **DESIGN = WR ;**
NEST **STRATUM PSU ;**
WEIGHT **WTFA ;**
SUBGROUP **(variable names);**
LEVELS ... **;**
SUBPOPN **RACE = 2 & SEX = 2 / NAME “Analysis on black females only”;**

Using the full data set and the SUBPOPN statement in this example would constrain analysis to black females only. In using the SUBPOPN statement, this strategy is equivalent to subsetting the data set, but it has the advantage of using the full data set and thereby not compromising the design structure.

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 (forthcoming), *Design and Estimation for the National Health Interview Survey, 1995-2004*, Vital and Health Statistics, Series 2, No. to appear.

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.