2012 National Health Interview Survey (NHIS)
Public Use Data Release

NHIS Survey Description

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

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

June 2013

2012 National Health Interview Survey (NHIS)
Public Use Data Release

Table of Contents

The NCHS Web Pages and NHIS Electronic Mail List

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

http://www.cdc.gov/nchs/nhis.htm .

The website features downloadable public use data and documentation for the 2012 NHIS and previous years, as well as important information about any modifications or updates to the data and/or documentation. Published reports from previous years' surveys are also available, as are updates about future surveys and datasets.

Special topics websites on Adult Physical Activity, Adult Tobacco Use, Injury and Poisoning, and Race and Hispanic Origin are included to provide additional detail about the NHIS data, including historical context, editing, frequently asked questions, and references. Special topics can be found at:

http://www.cdc.gov/nchs/nhis/special_topics.htm .

The website also features the 2012 Paradata File, which contains data about the NHIS data collection process. It may be used as a stand-alone data file or linked to the NHIS 2012 health data files. The Paradata File and documentation can be found at:

http://www.cdc.gov/nchs/nhis/2012paradata.htm .

Data users are encouraged to join the NHIS Listserv, an electronic mail list. The Listserv is made up of over 4,000 NHIS data users located around the world who receive e-news about NHIS surveys (e.g., new releases of data or modifications to existing data), publications, workshops, and conferences. To join, click on "Listserv" on the NHIS Web page.

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

What's New in 2012?

- In 2012, the NHIS sample size augmentation that began in 2011 continued in 32 states and the District of Columbia, with some increases in the level of augmentation. The main goal of the augmentation was to increase the number of states for which reliable estimates can be made. The 2012 NHIS sample size is the largest sample size since the current sample design was implemented in 2006.

- Beginning in 2012, NCHS made the transition to weights derived from 2010 census-based population estimates.

- Family supplements in 2012 included questions to measure the impact of the Affordable Care Act and questions to measure family food security.

- Sample adult supplements in 2012 included questions to measure the impact of the Affordable Care Act; complementary and alternative medicine; functioning and disability; immunization; Internet and email usage; non-cigarette tobacco use; and voice, speech, and language.

- Sample child supplements in 2012 included questions to measure the impact of the Affordable Care Act; balance; complementary and alternative medicine; immunization; mental health; mental health services; and voice, speech, and language.

- Data on the impact of the Affordable Care Act; balance; food security; immunization; Internet and email usage; mental health; mental health services; non-cigarette tobacco use; and voice, speech, and language are released in the Core data files in 2012.

- Three stand-alone data files have been created: the Adult Functioning and Disability File; the Adult Complementary and Alternative Medicine File; and the Child Complementary and Alternative Medicine File.

- Questions from the Disability Questions Tests of 2010, which were asked at the end of the Family Core questionnaire, were retained in 2011 and 2012. This supplementary field test is one component of a larger effort to develop and adopt a standard set of disability questions for multiple surveys across multiple countries. These data are released as three stand-alone data files (Family Disability Questions Test, Adult Disability Questions Test, and Child Disability Questions Test) and can be linked to other NHIS 2012 data files.

2012 National Health Interview Survey (NHIS)
Public Use Data Release

Introduction

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

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

The redesigned NHIS introduced in 1997 consists of a Core as well as Supplements that vary from year to year. The Core, which remains largely unchanged from year to year, consists of four main components: the Household Composition Section, the Family Core, the Sample Child Core, and the Sample Adult Core. The Household Composition Section of the questionnaire collects some basic demographic and relationship information about all persons in the household. The Family Core questionnaire, which is administered separately for each family in the household, collects information on all persons in the family. Topics on the Family Core include socio-demographic characteristics, basic indicators of health status, activity limitations, injuries, health insurance coverage, and access to and utilization of health care services.

From each family in the NHIS, one sample child (if any children aged 17 years or younger are present) and one sample adult aged 18 years or older are randomly selected, and information on each is collected with the Sample Child Core and the Sample Adult Core questionnaires. Because some health issues are different for children and adults, these two questionnaires differ in some items, but both collect basic information on health status, health care services, and behavior. When fielded, supplementary questions about the sample child and sample adult provide additional information. These sections of the survey yield the Sample Child and Sample Adult data files.

The Family Core yields several data files, including the Household-Level file, the Family-Level file, the Person-Level file, and the Injury and Poisoning Episode file. Because these files contain the same or comparable variables from one survey year to the next, they are suitable for trend analysis; moreover, multiple years of these data may be easily pooled to increase the sample size for analytic purposes.

NCHS developed the Research Data Centers (RDC) to allow researchers access to restricted data. Although state identifiers and some other identifying variables are not publicly released due to confidentiality concerns, use of that information can be made through the NCHS Research Data Centers (RDCs). Contact the RDC for more information, or visit their Web page at: **http://www.cdc.gov/rdc/**.

Data Collection Procedures

The U.S. Census Bureau, under a contractual agreement, is the data collection agent for the National Health Interview Survey. NHIS data are collected continuously throughout the year by Census interviewers. Face-to-face interviews are conducted in respondents' homes, but follow-ups to complete interviews may be conducted over the telephone. Nationally, the NHIS uses about 750 interviewers, trained and directed by health survey supervisors in the U.S. Census Bureau Regional Offices. Interviewers are observed by supervisors periodically and their work is monitored by the PANDA system, a performance and data analysis program that provides monthly checks on response rates and completion rates. The supervisors responsible for the NHIS are career Civil Service employees who are selected through an examination and testing process. Interviewers (also referred to as Field Representatives, or "FRs") receive thorough training on an annual basis in basic interviewing procedures and in the concepts and procedures unique to the NHIS.

For the Household Composition Section of the questionnaire, one household member who is at least the age of legal majority for a given state is identified as the "household respondent." In most states this age is 18 years, but in Alabama and Nebraska it is 19 years, and in Mississippi it is 21 years. The household respondent provides basic demographic and relationship information about all household members; these relationships determine the number of families that comprise the household. Note that in a multi-family household, a single household respondent provides household information for all families.

For the Family Core questionnaire, a resident family member who is at least the age of legal majority is identified as the "family respondent." The family respondent serves as the primary respondent for the family, providing information for all children and adult family members. However, all members of the family aged 18 years or older who are at home at the time of the interview may respond for themselves. Note that in a multi-family household, a family respondent is identified for each family.

For the Sample Child questionnaire, one child (the 'sample child') is randomly selected. Information about the sample child is obtained from the sample child respondent who is an adult residing in the household who is knowledgeable about the child's health.

For the Sample Adult questionnaire, one adult per family (the 'sample adult') is randomly selected with enhanced chances of selection for black, Hispanic, or Asian persons aged 65 years or older. This individual responds for him/herself to the questions in that section unless he/she is physically or mentally unable to do so, in which event a knowledgeable proxy is allowed to answer for the sample adult (468 cases in 2012).

An emancipated minor is any person aged 14 years to one year less than the age of legal majority for his/her state of residence and is married, widowed, divorced, separated, or living with a partner. Emancipated minors are not eligible for Sample Child or Sample Adult selection and are not eligible to be the household or family respondent.

The NHIS is conducted using computer-assisted personal interviewing (CAPI). The CAPI data collection method employs computer software that presents the questionnaire on computer screens to each interviewer. The computer program guides the interviewer through the questionnaire, automatically routing the interviewer to appropriate questions based on answers to previous questions. Interviewers enter survey responses directly into the computer, and the CAPI program determines if the selected response is within an allowable range, checks it for consistency against some of the other data collected during the interview, and saves the responses into a survey data file. The computer contains help facilities to aid interviewers in administering the CAPI questionnaire. This data collection technology reduces the time required for transferring, processing, and releasing data, and it ensures the accurate flow of the questionnaire.

Further information about data collection procedures is available in the 2012 NHIS Field Representative's (FR) Manual available on the NHIS web site,
http://www.cdc.gov/nchs/nhis.htm

Sample Design

The National Health Interview Survey is a cross-sectional household interview survey. The target population for the NHIS is the civilian noninstitutionalized population of the United States. Excluded from the survey are persons in long-term care institutions (for example, nursing homes for the elderly, hospitals for the chronically ill or physically or intellectually disabled, and wards for abused or neglected children), correctional facilities (for example, prisons or jails, juvenile detention centers, and halfway houses), U.S. nationals living in foreign countries, and active-duty Armed Forces personnel. Although active-duty Armed Forces personnel are excluded from the survey, their civilian family members are included (for example, a child whose parents are both active-duty military would be included, but the parents would not).

Sampling and interviewing for the NHIS are continuous throughout each year. The sampling plan follows a multistage area probability design that permits the representative sampling of households and noninstitutional group quarters (e.g., college dormitories). The sampling plan is redesigned after every decennial census. The current sampling plan was implemented in 2006. It has many similarities to the previous sampling plan, which was in place from 1995 to 2005. The first stage of the current sampling plan consists of a sample of 428 primary sampling units (PSU's) drawn from approximately 1,900 geographically defined PSUs that cover the 50 states and the District of Columbia. A PSU consists of a county, a small group of contiguous counties, or a metropolitan statistical area.

Within a PSU, two types of second-stage units are used: area segments and permit segments. Area segments are defined geographically and contain an expected eight, twelve, or

sixteen addresses. Permit segments cover housing units built after the 2000 census. The permit segments are defined using updated lists of building permits issued in the PSU since 2000 and contain an expected four addresses. The NHIS sample frame consists of two non-overlapping parts: the area frame (all of the area segments) and the permit frame (all of the permit segments).

The current NHIS sample design continues the oversampling of both black persons and Hispanic persons that was a new feature of the previous sample design. A new feature of the current sample design is that Asian persons are oversampled, as well. In addition, the sample adult selection process has been revised so that when black, Hispanic, or Asian persons aged 65 years or older are present, they have an increased chance of being selected as the sample adult. See Appendix III for more details.

One of the two procedures used for oversampling is "screening". Prior to interviewing, the sample addresses in area segments are randomly separated into two parts. In one part, the sample addresses are assigned to be "screened". In this part, the NHIS interview proceeds through the collection of the household roster. The interview then continues only if the household roster contains one or more black, Asian, or Hispanic persons. Otherwise, the interview terminates and the household is said to be "screened out". In the other part of the NHIS sample, full interviews occur at all households. No screening occurs in permit segments.

The other oversampling procedure is applied when area segments are sampled within PSUs. Segments are grouped by 2000 census concentrations of black, Asian, and Hispanic persons, and groups with higher concentrations are sampled at a higher rate.

As with the previous sample design, the NHIS sample is drawn from each state and the District of Columbia. Although the NHIS sample is too small to provide state level data with acceptable precision for each state, selected estimates for most states may be obtained by combining data years. Although state identifiers are not publicly released due to confidentiality concerns, use of that information can be made through the NCHS Research Data Centers. Contact the RDC for more information, or visit their Web page at: **http://www.cdc.gov/rdc/**.

NCHS survey integration and follow-back surveys continue to be facilitated by an area frame (the collection of approximately 1,900 geographically defined PSUs based on Census 2000) to draw the survey sample, with address lists obtained by the Census Bureau in a separate listing activity, independent of Census 2000 and explicitly for the NHIS. Also, the NHIS sample continues to be divided into four individually representative panels to further facilitate integration with other NCHS surveys and to allow for sample size reductions while retaining representativeness. In addition, NHIS interviewed households serve as a sampling frame for the Medical Expenditure Panel Survey (MEPS), as needed by the Agency for Healthcare Research and Quality (AHRQ).

The NCHS report describing the 1995-2005 design provides much information that still applies to the new sample design (Botman, Moore, Moriarity, & Parsons, 2000). This publication is available on-line at: **http://www.cdc.gov/nchs/data/series/sr_02/sr02_130.pdf.** A new report providing a complete description of the 2006-2015 NHIS sample design is being developed.

Sample Size Changes in the National Health Interview Survey, 2002-2012

The NHIS sample size can vary from year to year. The sample size may be reduced because of budgetary reasons or may be augmented, if supplementary funding is available. Sample sizes were reduced in 2002-2004, and the sample redesign implemented in 2006 reduced the targeted annual sample size by approximately 13%. Further reductions were made to the sample sizes in 2006-2008 due to budget shortfalls. The 2009 NHIS sample was reduced by approximately 50% during January-March 2009. Newly available funding later in 2009 permitted an expansion during October-December to increase that quarter's normal sample size by approximately 50%. The net effect of the January-March cut and the October-December expansion was that the 2009 NHIS sample size was approximately the same as it would be if the sample had been maintained at a normal level during the entire calendar year.

In 2010, the NHIS sample size was augmented during January-March by approximately 25%. There were no augmentations or reductions in the remaining months of 2010. Thus, the 2010 NHIS sample size was slightly larger than the 2009 NHIS sample size.

In 2011-2012, the NHIS sample size was augmented in 32 states and the District of Columbia. The main goal of the augmentation was to increase the number of states for which reliable state-level estimates can be made. In 2011, the sample size was augmented by approximately 13%, and in 2012, by approximately 21%.

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 non-response and post-stratification, are reflected in the sample weights that are provided in the accompanying data files.

Since the NHIS uses a multistage sample designed to represent the civilian noninstitutionalized population of the United States, it is necessary to utilize the person's basic weight for proper analysis of person-record data. In addition to the design and ratio adjustments included in the calculation of the Person-Level file's basic weights, the person weights are further modified by adjusting them to Census control totals for sex, age, and race/ethnicity populations (post-stratification). Beginning in 2012, NCHS made the transition to weights derived from 2010 census-based population estimates.

Each file has 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 and ratio (including non-response and post-stratification) adjustments. This should be used in most analyses of the Family/Person data. National estimates of all person-level variables can be made using these weights.

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

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

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

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

The Sample Adult data file contains two sets of weights:

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

The non-response adjustment was added in 2010 to the Sample Adult weights and the Sample Child weights. The weights of all supplements that are derived from the Sample Adult file or the Sample Child file also include a non-response adjustment.

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, both responding and non-responding.

From 1997-2007, WTIA_HH was nonzero for all households, responding and non-responding. However, some non-responding households were assigned incorrect values for WTIA_HH during that period. This had no effect on WTFA_HH, which is assigned nonzero values only for responding households.

Beginning in 2008, changes were made to assign correct values for WTIA_HH to all non-responding households. Additionally, some non-responding households now have a zero weight in the WTIA_HH field because they would have been ineligible if the interview had reached the decision point for the household to be "screened out." See Appendix III for information about the NHIS "screening" process.

Lastly, the Family File contains one set of weights which is discussed in greater detail in the section of this document pertaining to the Family File.

NOTE: Analysts should be aware that data about 314 persons who were active duty members of the Armed Forces at time of interview are on the Person-Level file. Despite the fact that NHIS covers only the civilian noninstitutionalized household population, active duty members of the Armed Forces will be counted in the unweighted frequencies, because at least one other family member is a civilian eligible for the survey. The value of the final annual person weight (WTFA) for these military persons is zero, so they will not be counted when making national (i.e., weighted) estimates. Data for these Armed Forces members are included in all relevant files in order to aid any analyses pertaining to family structure or relationships. No active duty Armed Forces members were selected as sample adults.

Recall Period and Weights

Some questions for particular events have recall periods referring to, for example, the "last 2 weeks" or the "last 3 months." In general, annual estimates of events can be made using these types of variables. For example, using a variable that counts events experienced by a person within a two-week recall period, an annual estimate of the number of events is 26 times the weighted estimate of the total number of events experienced by all persons within the two-week recall period. Similarly, using a variable with a three-month recall period, an annual estimate of the number of events is four times the weighted estimate of the total number of events experienced by all persons within the two-week recall period. 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 evaluate if annual estimates for these kinds of event variables are possible and have intrinsic meaning. Annual estimates of **events** should not be interpreted as annualized **person** experiences.

Variance Estimation

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. Because of this complex design and adjusted sampling weights, the direct application of standard statistical analysis methods for estimation

and hypothesis testing to unweighted data may yield misleading results. If data are not weighted, severely biased estimates may result. For this reason, as indicated previously, it is necessary to use the weights that are included in the accompanying data file for analyses.

Weighted data used in standard software packages may provide unbiased point estimates for commonly computed first-order statistics like means or regression coefficients, but the computed standard errors of the estimates may be too small. Also, standard packages may produce hypothesis test results (such as p values) that are incorrect. Hence, it is recommended that users of NHIS data utilize computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs. NCHS uses SUDAAN software (Research Triangle Institute 2008) with 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 NHIS database. Appendix III also provides example code for SPSS, Stata, R, SAS survey procedures, and VPLX.

Analyses of large NHIS subgroups usually produce reliable estimates, but analyses of small subgroups may yield unreliable estimates, as indicated by their large variances. The analyst should pay particular attention to the coefficient of variation (relative standard error) for estimates of means, proportions and totals. In addition, small sample sizes, or small numbers of primary sampling units containing targeted data, may be an indication of estimates lacking precision.

General Information about the 2012 Data

The interviewed sample for 2012 consisted of 42,366 households, which yielded 108,131 persons in 43,345 families. The interviewed sample for the Sample Child component, by proxy response from a knowledgeable adult in the family, was 13,275 children under 18 years of age. The interviewed sample for the Sample Adult component, which required self-response to all questions unless the sample adult was physically or mentally unable to do so, was 34,525 persons 18 years of age and older. There were 468 cases where a knowledgeable proxy answered for the sample adult.

The total household response rate was 77.6%: 14.6 percentage points of the noninterview rate (22.4%) were the result of respondent refusal and unacceptable partial interviews. The remaining 7.8 percentage points were primarily the result of failure to locate an eligible respondent at home after repeated contact attempts.

The conditional response rate for the family component was 99.0%, which was calculated by dividing the number of completed family interviews (43,345) by the total number of eligible families (43,785). The unconditional or final response rate for the family component was calculated by multiplying the conditional rate by the household response rate of 77.6%, yielding a rate of 76.8%.

The conditional response rate for the Sample Child component was 90.7%, which was calculated by dividing the number of completed Sample Child interviews (13,275) by the total

number of eligible sample children (14,637). The unconditional or final response rate for the Sample Child component was calculated by multiplying the conditional rate by the final family response rate of 76.8%, yielding a rate of 69.7%.

The conditional response rate for the Sample Adult component was 79.7%, which was calculated by dividing the number of completed Sample Adult interviews (34,525) by the total number of eligible sample adults (43,323). The unconditional or final response rate for the Sample Adult component was calculated by multiplying the conditional rate by the final family response rate 76.8%, yielding a rate of 61.2%.

No compensation or other incentives were provided for participation in the NHIS. Additional information about NHIS response rates can be found in Appendix I.

Information about the 2012 Data File Documentation

As with previous data years, questionnaires, datasets, and related documentation for each data file are available on the NHIS website, **http://www.cdc.gov/nchs/nhis.htm**. The website provides the Survey Description document; the Questionnaire Reference Guide, survey questionnaires, the Field Representative Manual, and survey flowchart; information on co-sponsors and supplements; Summary Health Statistics reports (when available), and the data release. The data release page contains a Readme File including a summary of data access instructions; notices for data users including a log of release history, and if necessary, information about data problems or changes; the Household, Family, Person, Sample Adult, Sample Child, and Injury/Poisoning Episode Files; Imputed Income Files; the Paradata File; and any supplemental files released that year.

Each of the 2012 data release files include the following documents. A description of each type of document follows:

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

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

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

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

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

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

1) If the original questionnaire item was asked at the family level but resulted, after the editing process, in a person-level variable, this note is added: Family/person variable conversion.

2) If other questions in the instrument ask about the same topic, or if similar questions appear in other sections of the instrument, this note is added: Refer to {variable name and section number} for a {family/person/child} level question on the same topic.

The universe refers to those respondents deemed eligible to answer a given question. For example, the universes for most Sample Adult variables are specified as ASTATFLG = 1 and (AGE GE '018' and AGE not IN ('997', '999')), followed by any other universe descriptors specific to the variable. ASTATFLG = 1 refers to a variable on the Person File and indicates that the respondent was selected as a sample adult and answered at least the first three sections of the Sample Adult questionnaire, constituting a completed interview or an acceptable partial interview. Sample adults who are not eligible to answer a given question are considered to be not-in-universe. For example, a sample adult who reported that he did not have surgery in the past 12 months (ASRGYR=2) would not be eligible for a follow-up question (ASRGNOYR) about the number of times that he had surgery in the past 12 months. It is important to note that for all data files, persons who are not-in-universe are no longer listed in the Variable Layout

Report response categories as "Blank- Not-in-universe." If a respondent discontinued the interview any time after completing the first three sections of the Sample Adult component, his or her responses will appear as 8's (not ascertained) for the remaining variables in the Sample Adult File where the universes are applicable. In addition, each year, there generally are a few records (less than 10) where age is corrected due to data entry error. For the records where age is corrected, neither the universes nor the variables affected will be changed; however, a new variable, AGE_CHG, will indicate that a correction has been made on the record. Occasionally universe inconsistencies between variables may exist due to collection or processing errors.

The universes for most Sample Child File variables are specified as CSTATFLG = 1 and (AGE LE '017' and AGE NE ' '), followed by any other universe descriptors specific to the variable. CSTATFLG = 1 refers to a variable on the Person File that indicates a selected sample child with a completed interview or an acceptable partial interview, defined as the completion of the CHS section. Again, responses from acceptable partial interviews have a code of 8, meaning "not ascertained," throughout the remaining, unanswered Sample Child sections where the universes are applicable.

The Variable Frequency Report provides the frequencies, percents, and the frequency missing (not-in-universe) for each variable. **For the 2005 data year and beyond, all response categories are shown in the Variable Frequency Report, including those response categories with a zero count in the data files. This allows users to see a complete list of response categories with frequencies for each variable without referring to additional documentation. In addition, the "frequency missing" label will be shown if a variable has not-in-universe cases or cases whose values fall out of range.** For example, if all sample adults are asked about a usual place for medical care when sick (AUSUALPL), then the "frequency missing" label is not shown.

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

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

It is also important to note that for the 2005 data year and beyond, some frequently used variables are repeated on various data files; therefore, merging of files may be required less often than for the 2004 data year files. However, each data file contains

household, family, and person numbers that make merging the files possible, if needed. Appendix VI provides sample code for merging the files.

Information about the 2012 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 2012 NHIS Field Representative's (FR) Manual (both the questionnaire and FR Manual are available on the NHIS website, **http://www.cdc.gov/nchs/nhis.htm**). Every effort was made to insure that the variable names in the data are consistent with the question items in the instrument. In a few cases, this was not possible. Users should match the question number in the instrument to the variable number in the File Layout Report to resolve any discrepancies.

Because the questionnaire for the NHIS is administered by computer, the questionnaire exists as a long and complex computer program. While stringent quality control measures were applied, a few errors are known to have occurred in the program. Instrument problems were identified over the course of the year, and efforts were made to correct these errors. Some of these problems were resolved through correction of skip patterns, question wording changes, addition of questions, or other internal instrument corrections.

When errors were detected and diagnosed, and time permitted, the instrument was changed to correct for the errors. In 2012, instrument changes were kept to a minimum, so that there was basically one version of the NHIS in the field across all four quarters of the survey year. Analysts are encouraged to read the notes in the Variable Layout Report for important information pertaining to specific variables that may have changed across quarters.

Questionnaire Sections

The 2012 NHIS contained the annual Core, which is broken into various sections that group questions into broad and specific categories. Each section is designated by a section title and corresponding three-digit acronym (or section code); questionnaire items are numbered sequentially (but not consecutively) within their respective sections, with the section acronym making up part of the item number. Multiple-part questions have an extension added to their three-digit acronym. For example, the first item in the FHS section is identified as FHS.010_00.000; note that FHS.010_00.000 also has an associated variable name, PLAPLYLM. The following table lists the various questionnaire sections, their acronyms and description titles.

Table 1. 2012 NHIS Core Questionnaire Sections and Topics

A. Household

Section No.	Section Code	Description
I	HHC	Household Composition

B. Family Core

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

C. Sample Child Core

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

D. Sample Adult Core

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

E. Recontact

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

In addition to the three Core sections, the 2012 NHIS contains several other data files: the Household- and Family-level files, and the Injury/Poisoning Episode File. The Household File is derived largely from the Household composition section of the Core and describes characteristics of each household. The variables contained in the Family-level file are reconstructions of the person-level data from the Core sections at the family level. The Injury/Poisoning File is derived from information obtained from the injury/poisoning questions in the Family Core section.

Supplements, Supplement Co-Sponsoring Agencies, and Question Locations, 2012 NHIS

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

A supplement or one or more supplementary questions may be interwoven among Core questions, or may be placed at the end of a Core section. The existence of three extra digits (.xxx) at the end of the question number helps to identify supplementary questions in the Core questionnaires. In 2012, NHIS supplementary questions to measure the impact of the Affordable Care Act (family, adult and child), balance (child), complementary and alternative medicine (adult and child), disability (adult), food security (family), heart disease and stroke prevention (adult), immunization (adult and child), Internet access and email usage (adult), mental health (child), mental health services (child), non-cigarette tobacco use (adults), and voice, speech and language (adult and child) are found in the Core questionnaires. Data based on most supplementary questions are released in the Core data files in 2012. However, three stand-alone data files have been created: the Adult Functioning and Disability File, the Adult Complementary and Alternative Medicine File, and the Child Complementary and Alternative Medicine File.

In addition, questions from the Disability Questions Tests of 2011, which were asked in the Family, Sample Adult, and Sample Child Core questionnaires, were retained in 2012. This supplementary field test is one component of a larger effort to develop and adopt a standard set of disability questions for multiple surveys across multiple countries. These data are released as three separate files (Family Disability Questions Test, Adult Disability Questions Test, and Child Disability Questions Test) and can be linked to other NHIS 2012 data files. The files and documentation can be found at:
http://www.cdc.gov/nchs/nhis/disabilityquestionstests2012.htm .

A chart of all 2012 co-sponsored supplements and their question numbers is below. Users can obtain information about co-sponsored supplements from 1997-2011 on our website:
http://www.cdc.gov/nchs/nhis/supplements_cosponsors.htm .

Supplement Sponsoring Agencies and Question Locations, 2012 NHIS

Topic	Sponsoring Agency	Title	Survey Section/Number
Affordable Care Act Questions	Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS)	Adult and Child Affordable Care Act Questions	FHI.202_01.010 to FHI.208_01.010; FHI.235_01.010 to FHI.237_02.030; FHI.248_05.000; FHI.249_03.000; FHI.312_00.010 to FHI.317_00.010; FHI.325_00.010 to FHI.327_00.020; AAU.050_00.010 to AAU.059_00.010; AAU.111_05.010 to AAU.127_06.010; AAU.243_00.010 to AAU.248_08.080; AAU.306_00.010 to AAU.309_00.050; AAU.500_00.010 to AAU.601_00.080; CAU.050_00.010 to CAU.056_00.010; CAU.133_00.010 to CAU.133_00.020; CAU.135_05.010 to CAU.135_06.010; CAU.281_00.010 to CAU.283_08.080
Balance	National Institute on Deafness and Other Communication Disorders (NIDCD)[1]	Child Balance	CBL.010_00.000 to CBL.070_00.000
Complementary and Alternative Medicine	National Center for Complementary and Alternative Medicine (NCCAM)[1]	Adult and Child Complementary and Alternative Medicine	ACN.020_00.010; ACN.031_01.010; ACN.031_04.010; ACN.121_00.010; ACN.121_00.030; ACN.121_00.040 to ACN.121_00.180; ACN.125_00.010 to ACN.125_00.270; ACN.297_00.010; ACN.372_00.010 to ACN.372_00.040; ALT.001_00.010 to ALT.765_00.000 CHS.105_00.010 to CHS.105_00.030; CHS.106_00.010 to CHS.106_00.040; CHS.120_00.010 to CHS.120_00.100; CHS.125_00.010 to CHS.125_00.070; CHS.370_00.010 to CHS.370_00.030; CHS.375_00.010 to CHS.375_00.020; CAL.001_00.000 to CAL.765_00.000
Disability	National Center for Health Statistics[2]	Disability Test Questions; Adult Functioning and Disability	FDB.020_00.000 to FDB.120_00.000 CDB.020_00.000 to CDB.120_00.000 ADB.020_00.000 to ADB.120_00.000 AFD.090_00.000 to AFD.560_00.000
Food Security	United States Department of Agriculture (USDA)	Family Food Security	FFS.010_00.000 to FFS.100_00.000
Heart Disease and Stroke Prevention	Department of Health and Human Services (HHS)	Questions to Measure the Million Hearts® Initiative	ACN.020_00.010; ACN.121_00.010 to ACN.121_00.020; ACN.040_00.010 to ACN.040_00.040
Immunization	National Center for Immunization and Respiratory Diseases (NCIRD)[2]	Adult and Child Immunization	AAU.310_00.000 to AAU.470_00.010; CFI.005_00.010 to CFI.005_00.080
Internet Access and Email Usage	Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS)	Internet Access and Email Usage	AWB.010_00.000 to AWB.050_02.000
Mental Health	Center for Mental Health Services (CMHS)[3]	Child Mental Health Brief SDQ; Child and Adult Mental Health Questions	FHS.065_00.000; CAU.265_00.000; CMB.010_00.000 to CMB.030_00.000; CHS.321_01.000 to CHS.361_04.000; ACN.470_00.000 to ACN.530_00.000
Mental Health Services	Center for Mental Health Services (CMHS)[3]	Child Mental Health Services	CMS.001_00.000 to CMS.150_12.000
Tobacco	Center for Tobacco Products (CTP, FDA)	Non-cigarette Tobacco Use	AHB.085_00.010 to AHB.085_00.060
Voice, Speech, and Language	National Institute on Deafness and Other Communication Disorders (NIDCD)[1]	Adult and Child Voice, Speech, (Swallowing), and Language	CCD.010_00.000 to CCD.170_00.000; ACD.010_00.000 to ACD.320_00.000; ASD.215_00.000

[1] National Institutes of Health (NIH)
[2] Centers for Disease Control and Prevention (CDC)
[3] Substance Abuse & Mental Health Services Administration (SAMHSA)

2012 National Health Interview Survey
Household-Level File

The Household File is considered as the base file from which all other files are built. That is, the main sampling unit in the NHIS is the household, and each record on the Household File represents either a responding household or a "Type A" non-responding household. A "Type A" non-responding household was eligible for the NHIS interview but was not interviewed for a variety of reasons including refusal, language barrier, no one at home, and insufficient partial interview. Each record on the Household File represents a unique household included in the NHIS sample or sampling frame. Each household has a unique unit number (HHX). This unique unit number is needed for merging data files.

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

The universe for the Household File is all responding households and non-responding (Type A) households. The Household File contains information on 54,603 households: 42,366 households were interviewed, while 12,237 were not interviewed. The nature of non-interviews for Type A households, such as refusal or failure to locate an eligible respondent, is detailed in the variable NON_INTV.

The total non-interview rate for the Household File was 22.4% 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 + Type A non-responding households), or 77.6%. For more information about the eligibility of Type A non-responding households, see Appendix I of this document.

2012 National Health Interview Survey
Family-Level File

The Family-Level file contains variables that describe characteristics of the 43,345 families living in households that participated in the 2012 NHIS. A family is defined as an individual or a group of two or more related persons who are living together in the same occupied housing unit (i.e., household) in the sample. In some instances, unrelated persons sharing the same household may also be considered as one family, such as unmarried couples who are living together. Each record in the file represents a unique family. The universe for all variables in this file is limited to all responding families in those households participating in the 2012 survey; this is specified as FM = ALL in the Family-Level file Variable Layout. Note that multiple families may share one household. Users wishing to determine the number of responding and non-responding families in each household are referred to ACPT_FAM and REJ_FAM in the Household File or HHX and FMX in the Family File.

As Table 2 indicates, 98% of NHIS households consist of one family. All relationships in the household are recorded relative to a household reference person, who is generally the person who owns or rents the housing unit. Note that when there is only one family per household, all household and family relationships (as indicated by the Person File variables RRP and FRRP, respectively) are identical.

Table 2. Number of Families per Household, 2012 NHIS (unweighted counts)

Families per household	Frequency	Percent
1	41,569	98.1
2	666	1.6
3	96	0.2
4	25	0.1
5	7	*0.0
6	2	*0.0
9	1	*0.0

* 0.0 means more than zero but less than 0.05

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

Family size may vary considerably. Table 3 shows a breakdown of the 43,345 families by number of family members. Beginning in 2011, the number of family members includes any infant who was born and came home from the hospital or birthing center before the household roster was created. Previously, data were not collected on any infant who was born during the assignment week of the interview.

Table 3. Size of Family, 2012 NHIS (unweighted counts)

Number of Members	Frequency	Percent
1	13,244	30.6
2	13,286	30.7
3	6,468	14.9
4	5,736	13.2
5	2,876	6.6
6	1,048	2.4
7	406	0.9
8	180	0.4
9	51	0.1
10	25	0.1
11	13	*0.0
12	3	*0.0
13	3	*0.0
14	2	*0.0
15	2	*0.0
16	2	*0.0

*0.0 means more than zero but less than 0.05

The first part of the Family File contains the technical variables that identify or describe the record type (all observations in this file have a record type value of "60"), the survey year, the household and family numbers, the interview month and year, characteristics of the family's housing unit, geographic information associated with the housing unit, variables used for variance estimation, and a family-level weight variable.

The second part of the file consists of a series of recodes derived from five Family Core sections (FHS, FAU, FSD, FIN, and FHI) of the NHIS that collapse the individual-level observations into information about their respective families. In 2004, the FHI recodes were removed from the Family File and were reinstated on the 2011 Family File. Beginning in 2011, a Food Security Supplement (FSS) was added to the Family File that includes recodes based on family information.

Generally, the Family File consists of two types of recodes. The first is a simple "yes-no" measure that indicates whether *any* family member falls into a particular category or exhibits a particular characteristic. Every yes-no measure also has a corresponding counter that indicates the number of family members in that category or with that characteristic. Note that counters

23

always consist of values from zero to the maximum number of family members; in addition, no frequencies will be shown if a family is not contained in the universe for a specific question. For example, FSALYN and FSALCT, two recodes from the Income and Assets section of the Family Core, are limited to families with at least one member aged 18 or older; families consisting solely of emancipated minor(s) are coded as blanks to indicate that they are out of the universe, and thus, are not shown. The Family File also contains some counters that lack corresponding yes-no indicators. For example, FHSTATEX, FHSTATVG, FHSTATG, FHSTATFR, and FHSTATPR (all derived from PHSTAT, FHS.500) provide counts of the number of family members in excellent, very good, good, fair, and poor health, respectively. Counters were also constructed to indicate the number of working adults in the family, the number of adults in the family looking for work, the number of adults working full time, the number of children (under age 18) in the family, and the number of family members aged 65 and older.

Because most of the variables in the Family File are recodes of the person-level variables in the Family Core, the sum of the number of persons across all families in each family-level counter should be equal to the number of "yes" responses in its person-level source. Returning to our previous example, consider FSALCT: 17,374 families have one member receiving income from wages/salary, 11,868 families have two members (or 2(11,868)=23,736 persons) with wage/salary income, 1,879 families have three members (or 3(1,879)=5,637 persons), 465 families have four members (or 4(465)=1,860 persons), 83 families have five members (or 5(83)=415 persons), 11 families have six members (6(11)=66 persons), 2 families have seven members (7(2)=14 persons), 1 family has eight members (8(1)=8 persons), and 1 family has eleven members (11(1)=11 persons) with wage/salary income in 2012. Thus, the sum of persons across the 31,684 families answering "yes" to FSALYN, the associated yes-no indicator, is 49,121 (17,374 + 23,736 + 5,637 + 1,860 + 415 + 66 + 14 + 8 + 11), which is equal to the 49,121 "yes" responses to the person-level source variable, PSAL. Users are advised to check the Variable Layout Report for each Family File recode in order to determine its person-level source variable.

Family Structure Variables

The 2012 NHIS Family File contains several variables describing family type and structure in both general and detailed terms. FM_TYPE consists of just four categories and represents an initial classification of families according to the numbers of adults and children that are present. In addition, FM_STRP and FM_STRCP categorize families according to familial relationships and, when children are present, parental marital status. FM_STRP and FM_STRCP differ in how they categorize unmarried parents with children. FM_STRP includes all cohabiting couple families in the same category (FM_STRP = 42), regardless of the adults' relationships to the child(ren) in the family. FM_STRCP is identical to FMSTRCT2, a recode on the 1998, 2001-2003 NHIS Family Files, and distinguishes between families consisting of unmarried parents who are related biologically or by adoption to all children in the family (FM_STRCP = 41), and families consisting of a parent, his or her child(ren), and his or her partner, who is unrelated to the child(ren) present in the family (FM_STRCP = 43). In both recodes, families that could not be classified are coded "99." Emancipated minors are treated as adults with respect to FM_TYPE, FM_STRP, and FM_STRCP, despite the fact that they may be under 18 years of age.

The Family File Weight

The ideal situation for creating weights for the Family File would be to use independent estimates of the number of families from a reliable source, such as the U.S. Census Bureau, to perform post-stratification adjustments in a manner similar to what is done for the NHIS Person File weight. Unfortunately, no suitable independent estimates exist.

Due to the lack of appropriate independent estimates, a variation of the "principal person" method is used to create the 2012 NHIS Family File weight (WTFA_FAM). Briefly, a person-level ratio adjustment is used as a proxy for the NHIS family-level ratio adjustment. Use of the person weight with the *smallest* ratio adjustment within each family (that is, the smallest post-stratification factor between the interim and final person weights within the family) is believed to provide a more accurate estimate of the total number of U.S. families than either the use of other person weights in the family or the use of no ratio adjustments whatsoever.

Accordingly, the weight provided with the 2012 NHIS Family File, WTFA_FAM, corresponds to the 2012 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 non-response adjustment, and several ratio adjustment factors that are applied to all person weights.

2012 National Health Interview Survey
Person-Level File

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

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

The 2012 Health Status and Limitation of Activity (FHS) section of the Family Core contains information addressing respondent-assessed disabilities, disability-associated conditions, and overall health status for all family members. Users should note that additional information on health status and disability is also included in other sections of the Sample Adult File, as well as in the Sample Child File.

Limitation of Activity at the Person Level

Information on activity limitations, including questions about work limitations; the need for personal assistance with personal care needs such as eating, bathing, dressing, and getting around inside the home; and the need for personal assistance with handling routine needs such as everyday household chores, doing necessary business, and shopping or running errands, is collected for each family member (with some exclusions for children and youth). If any limitations are identified, the respondent is asked to specify the health condition(s) causing the limitation(s) and indicate how long thethe family member has had each such condition.

Since cognitive impairment is increasingly recognized as a source of activity limitations among older adults, the FHS section includes an indicator that identifies family members who are limited because of difficulty remembering or periods of confusion. Other indicators in this section identify family members who have difficulty walking without any special equipment or limitations related to specific personal care needs. In addition, the section contains information about children who receive special education or early intervention services. Information regarding limitations in play activities is also collected for young children.

In 2012, NHIS conducted a one-time split-ballot experiment involving two questions, FLIMANY (FHS.250) and PLIMANY (FHS.260), which try to capture information about limitations missed by earlier questions in FHS. As part of that experiment, a random half of families received FLIMANY and PLIMANY, if **no** limitation was reported for one or more family members in a series of previous questions. The other random half of families received FLIMANY and PLIMANY **regardless** of answers to the series of previous limitations questions.

The variable PLIMANY includes all persons in all families, both those with and without limitations; it is consistent with previous data years. The variable PLIMANY2 is based on responses for persons from the random half of families that received FLIMANY and PLIMANY **regardless** of responses to previous limitations questions. PLIMANY2 is only available for 2012, when the split-ballot experiment was conducted.

The 2012 FHS time variables and recodes, which indicate how long respondents have had the condition(s) causing their limitation(s), were processed using procedures similar to those used in 2002-2011. Substantively, the 2002-2012 variables and recodes are similar to those from previous years (1997-2001), but the detailed unknown categories that were included in the earlier data were collapsed into broader categories starting in 2002.

Beginning in 2011, the wording on the questionnaire and flash card changed from "mental retardation" to "intellectual disability, also known as mental retardation." Consequently, the resulting child and adult condition variables and the time variables and recodes associated with these condition variables all have new names. The child condition variable is now called LAHCC7A and the time variables and recodes associated with this variable are called LCTIME7A, LCUNIT17A, LCDURA7A, LCDURB7A, and LCCHRC7A. The adult condition variable is now called LAHCA14A and the time variables and recodes associated with this variable are called LTIME14A, LUNIT14A, LDURA14A, LDURB14A, and LCHRC14A.

Conditions

For each family member with a previously mentioned limitation, the respondent was asked about the condition or health problem associated with that limitation, as well as the length of time he/she has had the condition. Respondents were then handed one of two flash cards listing various condition categories. These categories are broad in scope and vary according to age. Information about family members under age 18 was solicited for the following fixed condition categories listed on the first flash card: "vision/problem seeing," "hearing problem," "speech problem," "asthma/breathing problem," "birth defect," "injury," "intellectual disability, also known as mental retardation," "other developmental problem (e.g., cerebral palsy)," "other mental, emotional, or behavioral problem," "bone, joint, or muscle problem," "epilepsy or seizures," "learning disability," "attention deficit/hyperactivity disorder (ADD/ADHD)," and two instances of "other impairment problem" (if the family member was limited by a condition not listed in one of the fixed categories). Respondents could supply a verbatim response of up to 50 characters for one or both of the "other impairment problem" categories.

The fixed response categories in the instrument for adults age 18 or older were equally broad, and comprised the conditions listed on the second flash card: "vision/problem seeing," "hearing problem," "arthritis/rheumatism," "back or neck problem," "fracture, bone/joint injury," "other injury," "heart problem," "stroke problem," "hypertension/high blood pressure," "diabetes," "lung/breathing problem (e.g., asthma and emphysema)," "cancer," "birth defect," "intellectual disability, also known as mental retardation," "other developmental problem (e.g., cerebral palsy)," "senility," "depression/anxiety/emotional problem," and "weight problem." Starting in 2001 and continuing in 2012, if an adult family member was limited by a condition not listed in one of these 18 fixed categories, the interviewer accessed a second screen containing

17 additional condition categories and two "other impairment problem" categories. These conditions were not read aloud to respondents or listed on a flash card, but if the respondent said a family member's condition was limited by one of these 17 conditions, the interviewer recorded this information. If the family member was limited by a condition not included in one of the 18 fixed categories or on the interviewer's computer screen, then the interviewer entered a verbatim response of up to 50 characters for one or both of the "other impairment problem" categories. Respondents could list any number of applicable conditions.

During data processing, the verbatim responses recorded by interviewers were reviewed to determine if any responses could be back-coded to one of the 13 fixed categories for respondents under age 18, or to one of the 18 fixed categories for adult respondents. If so, these "other" responses were assigned to the appropriate response categories (the first 13 for children, and the first 18 for adults). For adults, an additional 16 *ad hoc* categories were created during data processing to categorize responses that fell outside the fixed 18 condition categories included in the instrument: these *ad hoc* categories were assigned numbers 19_ through 34_. (Due to a naming convention error, in the 2002 and 2003 data files, these same *ad hoc* categories were assigned numbers 19 through 34 without an underscore.)   In addition, responses in the 17 "second screen" categories seen only by the interviewer were also back-coded and categorized into 8 of the *ad hoc* categories; Table 4 shows how the 17 additional adult condition categories on the second screen were coded. The resulting 34 output categories for adults and 13 output categories for children were based on the International Classification of Diseases, Ninth Revision, Clinical Modification. Table 5 shows the final FHS categories with approximate ICD-9-CM ranges.ICD-9-CM codes shown in this table are *not* included on the data file.

Any verbatim conditions that could not be back-coded to one of the original categories or recoded to one of the *ad hoc* categories (for adult respondents) remained in the "other impairment problem" categories, and were renumbered "90" and, if necessary, "91" for both children and adults. The specific condition categories as well as the "other impairment problem" categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any activity (a mention/not-mention format). Note that the verbatim responses associated with the "other impairment problem" categories are not included as a separate field on the public use file. In addition, because the 16 adult *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions causing limitations will be underestimates. Therefore, these variables should be analyzed with care. Moreover, none of the FHS condition variables (the 13 child variables, LAHCC1 through LAHCC13, and the 34 adult variables, LAHCA1 through LAHCA34_) should be used to estimate prevalence for the conditions they represent, because only those persons with a previously reported limitation were eligible for the condition questions that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult and Child Cores.

Table 4. Reassignment of "Second Screen" Adult Condition Categories

Screen item:		Is assigned to:	
LAHCA 19	Missing limbs (fingers, toes or digits), amputee	LAHCA19_	Missing limbs (fingers, toes or digits), amputee
LAHCA 20	Kidney, bladder or renal problems	LAHCA25_	Genitourinary system conditions
LAHCA 21	Circulation problems (including blood clots)	LAHCA21_	Other circulatory system conditions
LAHCA 22	Benign tumors, cysts	LAHCA28_	Tumors and cysts, benign and unspecified
LAHCA 23	Fibromyalgia, lupus	LAHCA20_	Other musculoskeletal system and connective tissue conditions
LAHCA 24	Osteoporosis, tendinitis	LAHCA20_	Other musculoskeletal system and connective tissue conditions
LAHCA 25	Epilepsy, seizures	LAHCA23_	Other nervous system conditions
LAHCA 26	Multiple Sclerosis (MS), Muscular Dystrophy (MD)	LAHCA23_	Other nervous system conditions
LAHCA 27	Polio(myelitis), paralysis, paraplegia/quadriplegia	LAHCA23_	Other nervous system conditions
LAHCA 28	Parkinson's disease, other tremors	LAHCA23_	Other nervous system conditions
LAHCA 29	Other nerve damage, including carpal tunnel syndrome	LAHCA23_	Other nervous system conditions
LAHCA 30	Hernia	LAHCA24_	Digestive system conditions
LAHCA 31	Ulcer	LAHCA24_	Digestive system conditions
LAHCA 32	Varicose veins, hemorrhoids	LAHCA21_	Other circulatory system conditions
LAHCA 33	Thyroid problems, Grave's disease, gout	LAHCA22_	Other endocrine, nutritional, metabolic and immunity conditions
LAHCA 34	Knee problems (not arthritis (03), not joint injury (05))	LAHCA20_	Other musculoskeletal system and connective tissue conditions
LAHCA 35	Migraine headaches (not just headaches)	LAHCA23_	Other nervous system conditions

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, when respondents answered "yes" to PLAPLYLM, PSPEDEIS, PLAADL, PLAIADL, PLAWKNOW, PLAWKLIM, PLAWALK, PLAREMEM, *or* PLIMANY, then LA1AR is coded "1". LA1AR includes three response levels: "1" for limited, "2" for not limited, and "3" for unknown if limited. (For comparability with 1997 and 1998 data, level 3 may be collapsed into level 2.) LACHRONR is based on LA1AR but adds the additional criterion of whether at least one of the reported causal conditions is a chronic condition. Users can utilize the information contained in LA1AR to control for "unknown if limited" cases with respect to LACHRONR (that is, when LACHRONR = 0).

Also, a series of age-group-specific recodes (for example, under 18 *versus* 18 and over) regarding conditions limiting activity, chronic condition status, and duration of limiting conditions has been created.

The questions about limitation of activity in the redesigned (1997 to present) NHIS are asked differently for different age groups. In some cases the questions are more general; in other cases the questions are more specific than in pre-1997 years. Lastly, since the questions in the

FHS section only allow for 'yes" or "no" responses, the degree to which a person is limited cannot be determined.

Chronic Conditions

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

Table 5. FHS Categories with Approximate ICD-9-CM Ranges

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

NHIS Category	ICD-9-CM Codes
1 - Vision or seeing problem	360-379.99
2 - Hearing problem	387-389.9
3 - Arthritis / rheumatism	711-712.9, 714-716.9, 720.0, 721-721.9, 729.0
4 - Back or neck problem	722-724.9, 732.0, 737-737.9
5 - Fractures, bone or joint injury *Injury with specific mention of bone or joints*	800-848.9 or 850-999.9 with mention of bone/joint
6 - Other injury *Injury without specific mention of bone or joints*	850-999.9 without mention of bone or joint
7 - Heart problem	410-417.9, 420-429.9, 745-746.9, 785.0-785.3
8 - Stroke problem	430-438.9
9 - Hypertension or high blood pressure	401-405.9
10 - Diabetes	250-250.9
11 - Lung or breathing problem	460, 461-461.9, 465-465.9, 466-466.19, 470, 471-477.9, 480-487.8, 490-496, 500-508.9, 510-519.9
12 – Cancer	140-208.9, 230-234.9
13 - Birth defect *Excludes Down syndrome and microcephalus*	740-759.9
14 - Intellectual disability, also known as mental retardation *Includes Down syndrome and microcephalus*	317-319, 742.1, 758.0
15 - Other developmental problem *Includes learning disabilities*	315.0-315.9, 343-343.9, 783.4
16 - Senility (and other cognitive problems)	290.0-290.9
17 - Depression, anxiety or emotional problem *Includes neurotic disorders, personality disorders, and other nonpsychotic mental disorders, excluding alcohol and drug related problems and developmental problems*	300.0-302.9, 306-313.9
18 - Weight problem *Indicates a problem with being overweight or obese as defined by the respondent*	
19 - Missing limbs (any part) / amputee *Indicates loss of a limb or digit*	
20 - Other musculoskeletal system conditions *Diseases of the musculoskeletal system and connective tissue not coded to 3, 4, 5*	710-739.9
21 - Other circulatory system conditions *Any diseases of the circulatory system not coded to 7, 8, 9*	390-459.9

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

NHIS Category	ICD-9-CM Codes
22 - Other endocrine system, etc. conditions *Any Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders not coded to 10 or 18*	240-279.9
23 - Other Nervous system conditions *Diseases of the nervous system and sense organs not coded to 1, 2, 15, 16*	320-389.9
24 - Digestive system conditions	520-579.9
25 - Genitourinary system conditions	580-629.9
26 - Skin & subcutaneous system conditions	680-709.9
27 - Blood & blood-forming organ conditions	280-289.9
28 - Tumors & cysts, benign & unspecified *Any mention of "tumor" without cancer, malignancy, etc.*	210-229.9, 235-239.9
29 - Alcohol & drug related problems *Any mention of "alcohol," "drugs" (or specific drug types), or substance abuse*	291-292.9, 303-305.9
30 - Other mental conditions *Any mental disorders not coded to 14 or 15 or 17*	290-290.9, 293-299.9, 314.00, 314.01
31 - After effects of surgery or other medical treatment *Any mention of "surgery" or "operation" or other treatment as the causal condition; includes ongoing or recent treatment (1 year or less) or specific and sole mention of surgery/medical procedure as specific cause of limitation.*	
32 - Old age *Any mention of age as the only specified cause*	
33 - Fatigue/Tiredness *Any mention of tiredness, stiffness, or weakness without referring to any specific part of the body*	
34 - Pregnancy related conditions *Any mention of "pregnancy" or "childbirth"*	
90 - Others Not Elsewhere Classified *1st other-specify verbatim, not elsewhere classified*	
91 - Others Not Elsewhere Classified *2nd other-specify verbatim, not elsewhere classified*	

B. Codes for Children (ages under 18 years)

NHIS Category	ICD-9-CM Codes
1 - Vision or seeing problem	360-379.99
2 - Hearing problem	387-389.9
3 - Speech problem	307.0, 307.9, 315.3, 784.3, 784.5
4 - Asthma or breathing problem	460- 461.9, 465-466.1, 470-471.9, 473, 477, 480- 487.8, 490-496, 500-508.9, 510-519.9

B. Codes for Children (ages under 18 years), continued

NHIS Category	ICD-9-CM Codes
5 - Birth defect *Excludes Down syndrome and microcephalus*	740-742.0, 742.2-757.9, 758.1-759.9
6 - Injury	800-999.9
7 - Intellectual disability, also known as mental retardation *Includes Down syndrome and microcephalus*	317-319, 742.1, 758.0
8 - Other developmental problem	343, 783.4
9 - Other mental, emotional, or behavioral problem	290-313.9, 799.2, V15.4
10 - Bone, joint or muscle problem	710-739.9
11 - Epilepsy and seizures	345, 779.0, 780.3
12 - Learning disability	315
13 - Attention Deficit/Hyperactive Disorder (ADD/ADHD)	314
90 - Others Not Elsewhere Classified *1st other-specify verbatim that does not fit in any other Category*	
91 - Others Not Elsewhere Classified *2nd other-specify verbatim that does not fit in any other Category*	

Technical Notes

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

II. Health Care Access and Utilization Section (FAU)

The Health Care Access and Utilization (FAU) section of the Family Core of the 2012 NHIS has remained largely unchanged since 1997. The FAU section contains information addressing access to health care, utilization services, and health care contacts.

Since 1997, questions that ask about delay of health care because of worry about the cost, overnight hospital stays, home care, calls to health professionals, and office visits have been included in the survey; there is also an expanded list of health care professionals, and respondents were instructed to consider "care from ALL types of medical doctors, such as dermatologists, psychiatrists, ophthalmologists, and general practitioners," as well as nurses,

physical therapists, and chiropractors. Lastly, a question asking about 10 or more visits to doctors or other health care professionals in the last 12 months has been included.

Technical Notes

A few large values were found for hospitalizations (HOSPNO) and hospital nights (HPNITE). In addition large numbers may exist for home care visits (PHCHMN2W), doctor visits (PHCDVN2W), and calls to health professionals (PHCPHN2W). Analysts should be aware that the above mentioned variables have not been edited for reasonableness.

III. Health Insurance Section (FHI)

The Health Insurance section of the 2012 NHIS Family Core has a full range of data items addressing health insurance. The flow of the questions pertaining to health insurance programs covered by this section is similar to the 1993-96 NHIS Health Insurance Supplements and the 1997-2011 NHIS Family Cores.

The FHI section covers several different topic areas:

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

Managed care arrangement and the need for referrals for those covered by Medicare, Medicaid, Children's Health Insurance Program, other state-sponsored health plans and other government programs;

Medicare managed care;

Enrollment in the Medicare Part D program;

Private insurance characteristics reported by the family respondent, including covered individuals relationships to the policyholder, coverage of individuals outside the household, HMO, PPO, and POS status, high deductible health plan (annual deductible of $1,200 for self-only coverage and $2,400 for family coverage), health savings account or health reimbursement agreement for high deductible plans, source of coverage, existence of employer subsidies for premiums, amount paid by employer, amount paid by individual/family, managed care detail information, need for a referral, primary care provider requirement, prescription drug benefit, dental coverage;

Confidence in purchasing affordable private health insurance on one's own;

Types of single service plans;

Type of TRICARE coverage;

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

Previous health insurance coverage;

Out-of-pocket costs in the past year for medical expenses (excluding health insurance premiums), difficulty paying medical bills, paying medical bills over time, and having medical bills that cannot be paid at all;

Enrollment in a flexible spending account (FSA) for medical expenses.

Beginning in 2011, the FHI section contains numerous new variables to address provisions of the Affordable Care Act of 2010 (ACA). The extended health insurance questions can be divided into two lines of questioning. One line of questioning is limited to those who have private health insurance plans. The other line of questioning is asked of all persons.

For private plans where the policyholder resides outside the household, information is now collected on the relationship of the covered individual to the policyholder. For private plans where the policyholder resides within the household, information is collected to find out if a private health insurance plan covers persons who do not reside in the household, and if so, the relationships of these individuals to the policyholder. These modifications to the FHI section address a major provision of the ACA that became effective September 23, 2010 which allows parents to keep their children on their health policies until they turn age 26.

Questions were also added in 2011 to the private plan detail section to examine employer contributions to private plan premiums and the private plan's requirement for seeing a primary care physician. In addition, information on how private plans were obtained was expanded to include two new categories: "Through parents" and "Through relative other than parents." The variables PLNWRKN1 and PLNWRKN2 were renamed PLNWRKR1 and PLNWRKR2 to reflect these additional categories.

For all persons, additional questions were added in 2011 to address confidence in obtaining affordable, directly purchased private coverage and difficulties paying medical bills. For persons who are either uninsured or who have had a change in insurance coverage within the past year, history of previous type of coverage was collected.

Beginning in 2011, FHI variables were created for the Family File. Many of these variables were on the Family File prior to 2004. The inclusion of the FHI section on the Family File will make it easier to analyze insurance data on a family basis.

Beginning in 2011, the variables HITYPEN1 and HITYPEN2, which ascertained if a private plan is an HMO, and HITYPE1 and HITYPE2, which provided information on health insurance model types, were discontinued. These variables are no longer on the public use file or available through the NCHS Research Data Centers (RDCs). These variables remain available for previous data years.

Beginning in 2010, two new variables, MCADVR and MCPREM, were added to the public use file. If a person is indicated as having a Medicare Advantage or Medicare HMO plan,

MCADVR ascertains if they are in that type of plan or in a Medicare supplemental plan. MCPREM ascertains if persons enrolled in a Medicare Advantage or Medicare HMO plan pay a premium. The variables MCNAME and MCPAYPRE were removed from the data set.

Beginning in 2004, FHI data contain several modifications, as well as some new variables. The HIKIND list was shortened from 14 categories to 11. Private health insurance was combined into one category, HIKINDNA, and military health coverage was combined into one category, HIKINDNF. To more accurately count single service coverage, SINCOV was added following the HIKIND question for persons for whom it had been indicated earlier that they have a single service plan. If a person was indicated as having single service coverage either through the HIKINDNJ or SINCOV questions, the family respondent received the single service detail questions about that person. Response categories were changed in the PLNWRK question to get better precision as to how a private health plan was obtained. This detail is contained in PLNWRKN1 and PLNWRKN2. An additional question was added to the private plan detail to monitor the impact of the Medicare prescription drug benefit on private plan drug benefits. This information is contained in PRRXCOV1 and PRRXCOV2. Detailed information concerning persons' third and fourth plans is not available on the public use data file to protect confidentiality. Persons with three or more plans have a "yes" response to the PRPLPLUS variable. Detailed information on persons' third and fourth plans is still available through the NCHS Research Data Centers.

Since 2004, details on type of military coverage are now contained in the variables MILSPC1, MILSPC2, MILSPC3, MILSPC4, and MILMAN. The wording on the MCCHOICE question was changed to address the new name for Medicare Plus Choice, which is Medicare Advantage. Follow-up questions were added regarding the Children's Health Insurance Program, state-sponsored and other public programs, and other government programs to obtain managed care information for all types of public coverage.

Beginning with quarter 3, 2004, two new questions were added to reduce potential errors in reporting Medicare and Medicaid status. For persons 65 years and over for whom Medicare coverage was not reported, the family respondent was explicitly asked about Medicare coverage for that person in MCAREPRB. For persons under 65 for whom there was no reported coverage, the family respondent was explicitly asked about Medicaid coverage for that person in MCAIDPRB. If a person was reclassified as covered by either of these additional questions, the family respondent received the appropriate follow-up questions.

Technical Notes

Analysts are strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, SCHIP, IHS, MILCARE, OTHPUB, OTHGOV, and SINGLE for types of health care coverage because these recodes take into account the complicated editing process that takes place in the FHI section. The variables HILAST and HINOTYR, which reflect periods of noncoverage, cannot be used to estimate the rate of uninsurance. Users may derive such estimates from NOTCOV, a recode included in the data file that reflects the definition of noncoverage as used in *Health, United States* (in which persons with *only* Indian Health Service coverage are considered uninsured). Alternatively, if users want to count IHS as coverage or create any other definition of uninsurance, they may use the health insurance recodes (MEDICARE, MEDICAID, PRIVATE,

SCHIP, IHS, MILCARE, OTHPUB, and OTHGOV) for this purpose. Using the most conservative estimate of the uninsured (which would exclude persons with IHS coverage only), a total of 595 persons do not have data for the HILAST question as during the course of the interview the instrument determined that they had health care coverage. It was subsequently established during the course of editing that they lacked coverage (given the information that the family respondent provided about their insurance plan(s)). NHIS staff elected not to edit these individuals out of the universe for HINOTYR. In addition, for a total of 997 persons, the family respondent was not asked either the HILAST or the HINOTYR questions.

In looking at the verbatim responses to HIPNAM1, HIPNAM2, HIPNAM3, HIPNAM4 (which ask respondents for the names of their family members' private health insurance plans), STNAME1, STNAME2, STNAME3 (the names of their family members' SCHIP, state sponsored or other government coverage respectively), MCANAME (the name of their family members' Medicare Advantage or Medicare HMO plan), and MACHMD_1, MACHMD_2 (names of Medicaid managed care plans), it was found that some of these respondents indicated plans or programs that were clearly private health insurance, Medicare, Medicaid, Children's Health Insurance Program, military coverage, Indian Health Service, single service plans, or no coverage at all. These persons were reassigned to the appropriate enrollment recodes for MEDICARE, MEDICAID, SCHIP, PRIVATE, IHS, MILCARE, and SINGLE.

In addition, some respondents offering an "other" response to the survey item (HISTOPOT) that inquired about the reason(s) coverage stopped subsequently indicated in their verbatim responses that the person did in fact have health insurance. These persons were reassigned to the appropriate response category with the enrollment recodes for MEDICARE, MEDICAID, SCHIP, PRIVATE, IHS, MILCARE, OTHPUB, and OTHGOV. Analysts are therefore strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, SCHIP, IHS, MILCARE, OTHPUB, OTHGOV, and SINGLE for types of health care coverage, because these take into account the above-mentioned back edits. In contrast, the data contained in HIKINDNA-HIKINDNK and MCAREPRB, MCAIDPRB, and SINCOV were not back-edited and reflect the respondents' original replies.

Note that the HIKINDA-HIKINDK variable names from 2004-2007 were renamed in 2008 to HIKINDNA-HIKINDNK, respectively. The OTHERPUB, OTHERGOV, and MILITARY variables from 2004-2007 were renamed in 2008 to OTHPUB, OTHGOV, and MILCARE, respectively. Although the variables are the same as in 2004-2007, these 2008 name changes were made because answer categories or the order of answer categories changed between 2003 and 2004, but were not appropriately renamed in 2004.

IV. Socio-demographic Section (FSD)

The Socio-demographic (FSD) section of the Family Core in the 2012 NHIS collects information on place of birth, citizenship status, and educational attainment for all family members, regardless of age. In addition, the family respondent was asked whether family members 18 years of age or older were working last week, and if not, the main reason they were not working. Additional questions inquired about the number of hours employed family members worked during the previous week, whether their employer offered health insurance, and, if they worked during the previous calendar year, how many months they worked and an

estimate of their earnings from wages. Analysts may also refer to the Adult Core Socio-demographic section (ASD) for additional occupational and employment data regarding those individuals selected as sample adults.

DOINGLWP (FSD.050) and WHYNOWKP (FSD.060) are the FSD equivalents of DOINGLWA (ASD.062) and WHYNOWKA (ASD.065) in the ASD section of the Sample Adult data file. For the majority of respondents, DOINGLWP and DOINGLWA will have identical values (and, likewise, WHYNOWKP and WHYNOWKA). However, it is nevertheless possible that DOINGLWP and DOINGLWA (and WHYNOWKP and WHYNOWKA) may have inconsistent values across the Person and Sample Adult data files. Users wishing to reconcile any discrepant values are advised to use the values of DOINGLWA and WHYNOWKA (rather than DOINGLWP and WHYNOWKP, respectively), since the information obtained from the family respondent during the FSD portion of the interview (and reflected in DOINGLWP and WHYNOWKP) was subsequently confirmed or corrected by the sample adult during his or her interview (as reflected in DOINGLWA and WHYNOWKA). Additionally, both DOINGLWP and WHYNOWKP are substantively equivalent to previous years' versions of these variables (i.e., DOINGLW1 and WHYNOWK1 in 2001-2003 and DOINGLW and WHYNOWRK in 1997-2000).

As in 2011, the 2012 FSD section contains a series of new questions on active duty service in the United States Armed Forces. These replace PMILTRY in the 2004-2010 NHIS (and before that, MILTRYDS in the 1997-2003 NHIS) that asked whether family members 18 years of age or older had ever been honorably discharged from active duty in the U.S. Army, Navy, Air Force, Marine Corps, or Coast Guard. The first question in this new series, ARMFEV (FSD.021), asks whether adult respondents in the family have ever served on active duty in the U.S. Armed Forces, military Reserves, or National Guard. A series of follow-up questions ask about specific time periods of service ("December, 1946 or earlier" through "September, 2001 or later"); respondents could indicate all periods that applied. A final question, ARMFDS (FSD.024), asks whether respondents served in the Persian Gulf during Operation Desert Shield or Operation Desert Storm between August 1990 and April 1991. Users should be aware that the value of the final annual person weight (WTFA) for respondents who are currently serving in the U.S. Armed Forces is zero, so these respondents will not be included in national (i.e., weighted) estimates of active duty service during any given time period. The NOTE section on page 12 provides more information about active duty members of the Armed Forces who were included in the 2012 NHIS sample.

The 2012 FSD section contains a variable called PLBORN (FSD.001), which is based on a question in the instrument that asked whether the respondent was born in the United States. If respondents replied affirmatively, they were asked the state in which they were born (PLBORN1, or FSD.002). If respondents said they were not born in the U.S., they were asked the country in which they were born (PLBORN2, or FSD.003). PLBORN1 and PLBORN2 are not included on the public use file for confidentiality reasons. However, the 2012 NHIS includes two public use recodes, GEOBRTH and REGIONBR, that are based on this restricted birthplace information (as well as the variable, PLBORN). GEOBRTH indicates geographic place of birth, and has three categories: born in one of the 50 United States or the District of Columbia; born in a U.S. territory; or not born in the U.S. or a U.S. territory. In order to make GEOBRTH comparable to previous recodes (for carrying out analyses on multiple years of NHIS data), users

should collapse those respondents in the last two categories of GEOBRTH into a single category. This will result in a recode that is comparable to USBRTH_P from the 2000-2001 NHIS or USBORN_P from the 1997-1999 NHIS. The second recode, REGIONBR, categorizes all respondents into one of 12 categories depending on their country of origin. The CIA on-line World Factbook was used to place countries into the regional categories shown below (for more information about the Factbook, users should refer to **https://www.cia.gov/library/publications/the-world-factbook/index.html**). Note that respondents born in Canada were included in the "Elsewhere" category of REGIONBR in order to satisfy NCHS confidentiality requirements. Users are cautioned that neither GEOBRTH nor REGIONBR indicate legal status or citizenship.

Category	Countries/regions included
United States	The 50 United States and the District of Columbia
Mexico, Central America, Caribbean Islands	Mexico, all countries in Central America and the Caribbean Island area, including Puerto Rico
South America	All countries on the South American continent
Europe	Albania, Austria, Azores Islands, Belgium, Bosnia and Herzegovina, Bulgaria, Corsica, Crete, Croatia, Czechoslovakia or the Czech Republic, Denmark, Finland, France, Germany, Great Britain, Greece, Holland, Hungary, Iceland, Ireland, Italy, Kosovo, Liechtenstein, Luxembourg, Macedonia, Majorca, Malta, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Scotland, Serbia, Sicily, Slovakia, Spain, Sweden, Switzerland, the area formerly known as Yugoslavia
Russia (and former USSR areas)	Russia, Lithuania, Latvia, Ukraine, Belarus, and all places formerly a part of the USSR
Africa	All countries on the African continent, plus the Canary Islands, Comoros, Madagascar, Madeira Islands
Middle East	Aden, Arab Palestine, Arabia, Armenia, Bahrain, Cyprus, Gaza Strip, Iran, Iraq, Israel, Jordan, Kuwait, Syria, Lebanon, "Middle East," Oman, Palestine, Persia, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, West Bank, Yemen
Indian Subcontinent	Afghanistan, Bangladesh, Bhutan, British Indian Ocean Territory, East Pakistan, India, Maldives, Nepal, Pakistan, Sri Lanka or Ceylon, Tibet, West Pakistan
Asia	Asia, Asia Minor, China, Japan, Mongolia, North Korea, South Korea

SE Asia	Borneo, Brunei, Burma or Myanmar, Cambodia, Christmas Island, Hong Kong, Indonesia, Laos, Malaysia, Philippines, Singapore, Taiwan, Thailand, Vietnam
Elsewhere	Guam, Bermuda, Canada, Greenland, Oceania, as well as "At sea," "High seas," "International waters," "North America"
Unknown	Places that could not be classified in the above categories.

Use of more detailed information on respondents' place of birth can be made through the NCHS Research Data Centers. Contact the RDC for more information, or visit their Web page at: **http://www.cdc.gov/rdc/**.

Regarding family members who were not born in one of the 50 United States or the District of Columbia, the family respondent was asked the year in which each person came to the United States to stay. Family respondents who could not recall or refused to answer were subsequently asked to estimate the number of years each person had been in the United States since he or she came to stay. This information was combined to create a recode that indicates how long these family members have been living in the United States (YRSINUS). The 2012 data also contain a citizenship recode (CITIZENP) that distinguishes between U.S. citizens and non-citizens.

Technical Notes

ERNYRFLG is a flag indicating that earnings data for a small number of respondents were collected differently in 2012. For more information on this flag, please refer to "2012 Verification Test" in the FIN section of this document.

V. Income and Assets Section (FIN)

2012 NHIS Imputed Income files will be available shortly!

Each year the missing data on family income and personal earnings in the NHIS are imputed using multiple imputation methodology. Multiple imputation is a technique that allows analysts to incorporate the extra variability due to imputation into their analyses. Imputed income values can then be merged with other data from the NHIS for analysis. NHIS imputed family income/personal earnings files are released annually approximately 60 to 90 days after the release of the initial NHIS data files, along with technical documentation that describes the contents of these files and provides instructions for how to use them. Users are strongly encouraged to refer to this technical documentation when using these files. Please refer to **http://www.cdc.gov/nchs/nhis/2011imputedincome.htm** for more information on these files.

The Income and Assets (FIN) section of the Family Core contains information regarding a variety of income sources, as well as estimates of total combined family income and home tenure status. Respondents are asked whether anyone in the family received income from a variety of sources; if so, the respondent is then asked to name the member(s) receiving income from that source. The section also includes questions about the family's total income from all

sources in 2011 and their home tenure status. The basic universe for most questions is "all families;" however, note that universes for several questions (most importantly, PSAL or FIN.040, PSEINC or FIN.060, and PWIC or FIN.385) are further limited with respect to age (of family members).

Sources of Income

Respondents are told at the start of the Income and Assets section that all questions are seeking information about possible income sources in the previous calendar year (2011). 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 are asked about income from Social Security or Railroad Retirement (including that which was received as a disability benefit); other pensions; Supplemental Security Income (SSI); Temporary Assistance for Needy Families (TANF); other kinds of government assistance (e.g., job training or placement, transportation assistance, or child care); interest from checking or savings accounts, Individual Retirement Accounts (IRAs) or certificates of deposit, money market funds, treasury notes, bonds, or any other accounts; dividends from stocks, mutual funds, and/or net rental income from property, royalties, estates or trusts; child support payments; and other income sources (the question specifically mentions alimony, contributions from family or friends, Veteran's Administration (VA) payments, Worker's Compensation, and Unemployment Compensation as possible sources of "other" income).

Income Amounts (1997-2006 NHIS)

In survey years prior to 1997, the NHIS obtained information about the amount of income received from each financial source, but that was dropped in the redesigned NHIS (1997 and beyond) in favor of a single overall estimate of combined family income. Furthermore, the 1997-2006 instruments contained three questions to identify the family's combined income from all sources during the previous calendar year: the first question (FAMINC) allowed the family respondent to supply a specific dollar amount (up to $999,995). Two subsequent questions (FINC20 and FINCCAT) attempted to gain additional information in those instances when the respondent did not know or refused to give a dollar amount to the first income question.

In the 2004 NHIS, 2,133 persons (2.25%) had a response of "$2" to the FAMINC question. It was subsequently determined that these were meant to be "Refused" or "Don't know" responses. In an attempt to reduce the amount of these types of responses, an edit was added to the NHIS instrument in Quarter 2 of 2005 that required interviewers to verify very high or very low income amounts reported by family respondents. In both 2004 and 2005, all of the "$2" responses to FAMINC were assigned the value of "not ascertained" and were subject to income imputation.

Additionally, the 1997-2006 NHIS contained a detailed indicator of poverty status (RAT_CAT) that utilized published information from the U.S. Census Bureau regarding poverty thresholds (for information on the 2011 poverty thresholds see DeNavas-Walt, Proctor, and Smith, 2012). A ratio of the previous calendar year's income value reported by respondents to the poverty threshold for the same year was calculated, 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 family-level ratio was subsequently ordered into a poverty gradient consisting of 14 categories (RAT_CAT). Users seeking more information about RAT_CAT should consult the 1997-2006 Survey Description Documents.

Starting in 2004, INCGRP and RAT_CAT (the public use recodes based on FAMINC, FINC20, and FINCCAT) and HOUSEOWN (FIN.280) were moved from the Person File to the Family File, replacing the 1997-2003 Family File variables FINCGRP, FRAT_CAT, and FHOUSE, respectively. In addition, FGAH (FIN.282), a Person File variable in 1998-2003, was moved to the Family File in 2004 (and thereafter), while AB_BL20K was deleted from the NHIS public use files starting in 2004 because it became redundant.

Income Amounts (2007-2012)

New income follow-up questions utilizing an unfolding bracket methodology were field-tested during the second quarter of the 2006 NHIS. The unfolding bracket method asked a series of closed-ended income range questions (e.g., "is it less than $50,000?") if the respondent did not provide an answer to the exact income amount question. The closed-ended income range questions were constructed so that each successive question established a smaller range for the amount of the family's income. A figure demonstrating the flow of the tested income follow-up questions is shown below.

FLOW DIAGRAM OF FIELD-TESTED INCOME FOLLOW-UP QUESTIONS (UNFOLDING BRACKETS), NHIS 2006, 2nd QUARTER
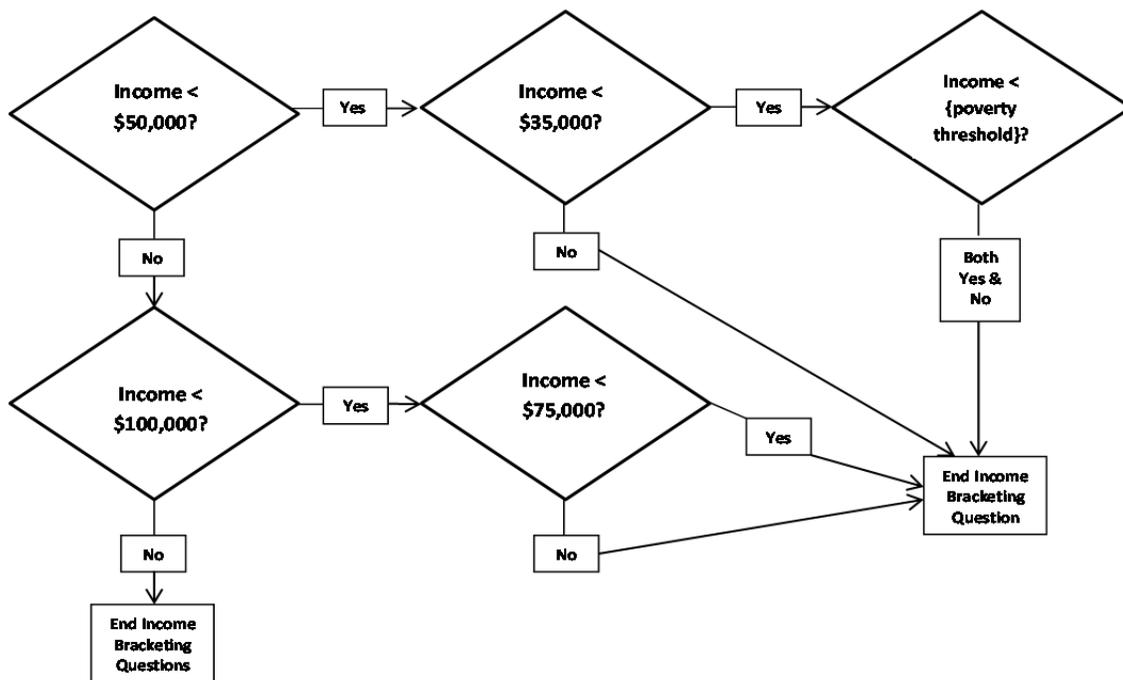
The unfolding income bracket follow-up questions used in the 2006 pilot test had much higher response rates than the original income follow-up questions (FINC20 and FINCCAT) that had been used since 1997. As a result, the poverty measure RAT_CAT from the 2006 pilot test had fewer responses with unknown values. The percentage of unknowns for RAT_CAT during the second quarter of the 2006 NHIS was 17.3% (unweighted) compared with 30.6% (unweighted) based on the original income follow-up questions. The positive results obtained by the unfolding bracket income follow-up questions led to their implementation in the 2007 NHIS, and FINC20 and FINCCAT were removed.

Because of these new income follow-up questions, new income variables were added to the NHIS starting in 2007. First, grouped income recodes INCGRP2 and INCGRP3 replaced the 1997-2006 income recode INCGRP, with INCGRP3 providing a bridge to the 1997-2006 variable INCGRP. In addition to the new grouped income recodes, two poverty status recodes (RAT_CAT2 and RAT_CAT3) were added to the NHIS in 2007 (and thereafter), replacing the 1997-2006 poverty status recode RAT_CAT. While the poverty status recodes in the 2007-2012 NHIS are somewhat similar to the 1997-2006 version, they have additional categories reflecting greater poverty ratio detail. The additional RAT_CAT2 categories allow data users to construct a three-category poverty ratio variable, while RAT_CAT3's additional categories allow data users to construct a four-category poverty ratio variable, based on the additional information provided by the income follow-up questions introduced in the 2007 NHIS.

Poverty Thresholds Used in FINCPOV (FIN.265)

As part of the new income amount questions that were introduced in 2007, families with incomes of less than $35,000 were asked a follow-up question (FIN.265) that used information on family size collected earlier in the interview and pre-defined NHIS poverty thresholds based on federal poverty thresholds for the previous calendar year. When the question is asked during the course of the interview, the appropriate poverty threshold relative to the family's size (in a dollar amount) is displayed on the interviewer's screen, so that the respondent is asked if the family's income was less than the applicable NHIS poverty threshold OR if the family's income was greater than or equal to the applicable NHIS poverty threshold in 2011.

The NHIS poverty threshold used in this follow-up question (called FINCPOV in the 2007-2011 NHIS; FINCPV01 in the 2012 NHIS) is a weighted poverty threshold derived from federal poverty thresholds that take into account family size and number of children under age 18. An example of the different federal poverty thresholds according to family size used in deriving the NHIS poverty thresholds for the 2012 instrument is shown below:

- Family size = 3:
    - 2011 Federal poverty thresholds by the number of children in the family:
        - No children: $17,595
        - One child: $18,106
        - Two children: $18,123
    - Weighted federal poverty threshold = $17,916

The poverty thresholds used in the 2012 NHIS survey instrument for families with three or more members were generally constructed as follows. First, the weighted federal poverty threshold was rounded to the nearest multiple of $500. Next, if the rounded weighted federal poverty

threshold was less than the original weighted federal poverty threshold OR was within $100 of the original weighted federal poverty threshold, the rounded weighted federal poverty threshold was used in the NHIS survey instrument. However, if the rounded weighted poverty threshold was at least $100 greater than the original weighted federal poverty threshold, the federal poverty threshold used in the NHIS survey instrument was the greatest multiple of $500 that was less than the original weighted federal poverty threshold. In the case of the poverty thresholds shown above for a family of three members, the original weighted federal poverty threshold was $17,916 and the rounded weighted average federal poverty threshold was $18,000. Therefore, because $18,000 is less than $100 greater than $17,916, the poverty threshold used in the 2012 NHIS survey instrument for a family with three members was $18,000.

For families of one or two members, the method was modified for the 2009-2012 NHIS when compared with the 2007-2008 NHIS. The U.S. Census Bureau publishes two separate age-based weighted federal poverty thresholds for families with one or two members. For a one-person family, the Census Bureau distinguishes between persons who are less than 65 years of age and those who are 65 years of age or older. For a two-person family, Census distinguishes between families where both members are less than 65 years of age and where one family member is 65 years of age or older. The 2007-2008 NHIS used the smaller of the two weighted age-based federal poverty thresholds for families of one or two members. In contrast, the 2009-2012 NHIS incorporated both weighted age-based poverty thresholds into the survey instrument.

Table 6. Weighted Federal Poverty Thresholds (calendar year 2011) and the NHIS Poverty Thresholds Used in the 2012 NHIS Instrument (FINCPV01, or family income in relation to the federal poverty threshold)

Family Size	2011 Weighted Federal Poverty Thresholds	NHIS Poverty Threshold used in 2012 Instrument (FINCPV01)
1[1]	$10,788 ($\geq$ 65 years of age)	$10,500
	$11,702 (< 65 years of age)	$11,500
2[1]	$13,609 (at least one member aged $\geq$ 65 years)	$13,500
	$15,139 (both members aged < 65 years)	$15,000
3	$17,916	$18,000
4	$23,021	$23,000
5	$27,251	$27,000
6	$30,847	$30,500
7	$35,085	$35,000
8	$39,064	$39,000
9	$46,572	$46,500

[1]Analysts should note that the methodology used to determine the poverty thresholds in the 2007-2008 NHIS instruments was different from that used to determine the thresholds in the 2009-2012 instruments. In the 2007-2008 NHIS, the lower weighted average federal poverty threshold for the appropriate family size was used for both age strata (< 65 years and $\geq$ 65 years). In the 2009-2012 NHIS, separate poverty thresholds were used for each age stratum for families of 1 or 2 members.
Source for column 2: U.S. Census Bureau; **http://www.census.gov/hhes/www/poverty/data/threshld/index.html**

Analysts should keep in mind that the reference period for income questions in the NHIS is the previous calendar year. Therefore, all income amounts in the 2012 NHIS are for calendar year 2011. Table 6 shows the calendar year 2011 weighted federal poverty thresholds and the

NHIS poverty thresholds that were used in the 2012 instrument that asked about the family's income in relation to the federal poverty threshold (FINCPOV). Data users should note that neither FINCPV01 nor the NHIS poverty thresholds are available on the public-use data file due to confidentiality concerns, but can be accessed through NCHS' Research Data Center.

Construction of Poverty Ratio Recodes RAT_CAT2 and RAT_CAT3

In general, poverty thresholds are used to construct the poverty ratio, which is a family's income in the last calendar year divided by the applicable poverty threshold. The variables RAT_CAT2 and RAT_CAT3 collapse the poverty ratio into gradients consisting of a finite number of categories. For both variables, categories "01" through "14" are equivalent, and are based on the U.S. Census Bureau federal poverty thresholds given the family's size and number of children. Categories "15" through "17" of RAT_CAT2 and categories "15" through "18" of RAT_CAT3 are based on the poverty thresholds used in the NHIS survey instrument (column 3 of Table 6), and reflect the additional poverty ratio information obtained from the income follow-up questions. Because RAT_CAT3 has one more category than RAT_CAT2, some families will not be coded equivalently on the two variables. In the 2012 NHIS, 2.8% of families were coded "17" on RAT_CAT2 and "18" on RAT_CAT3, while 3.2% of families were coded "17" on RAT_CAT2 and "99" on RAT_CAT3. Analysts should be aware of these discrepancies when analyzing data from RAT_CAT2 and RAT_CAT3.

Home Tenure Status

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

Program Participation

The 1997-2010 NHIS obtained information about receipt of food stamps from two questions, PFSTP (FIN.370) and FSTPMYR (FIN.380). In the 2011 and 2012 NHIS instruments, these questions were replaced by FSNAP (FIN.360) and FSNAPMYR (FIN.380), which ask respondents if any family members received food stamp benefits or participated in the Supplemental Nutrition Assistance Program (SNAP) at any time during the last calendar year, and if so, for how many months the food stamps or SNAP assistance were provided. Also, if one or more family members received Temporary Assistance for Needy Families (TANF) during the last calendar year, the respondent was asked for how many months during 2011 TANF was provided. In addition, respondents were asked whether any family member(s) had *ever* applied for Supplemental Security Income (SSI) or Social Security disability benefits (even if the claim(s) had been denied).

Finally, the NHIS contains three person-level variables relating to the Women, Infants, and Children (WIC) program. The first of these variables, ELIGPWIC or FIN.384, indicates if the person was in a family with at least one WIC age-eligible person (children 0-5 years of age or females 12-55 years of age). If there is at least one WIC age-eligible person in the family, the family respondent is asked if anyone in the family received WIC benefits in the previous

calendar year (PWIC or FIN.385). An additional recode, WIC_FLAG, is also included in the Person File. WIC_FLAG indicates if persons who received WIC benefits were age-eligible for the WIC program.

Technical Notes

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

2012 Verification Test

For most families, income (as reported in the FIN section) is expected to be equal to or greater than earnings (as reported in the FSD section) since total family income is based on salaries or self-employment earnings as well as other sources of income (e.g., pensions, retirement account disbursements, investments, entitlement programs, child support, unemployment compensation, etc.). NHIS data have not always been consistent with these expectations, however, so an Income-Earnings Verification Test was placed on the 2011 and 2012 NHIS to examine the problem of income-earnings inconsistencies. This test consisted of a series of questions that allowed interviewers to detect and correct large errors in income and earnings by comparing summed earnings across all employed family members to reported total family income. As part of the test, a random half of families received the verification questions if their reported earnings exceeded income by $5000 or more, or if their reported income exceeded earnings by $5000 or more *and* they did not report any other sources of income besides earnings. (Note that families with missing data for the income or earnings questions were ineligible for the verification questions.) The verification questions instructed interviewers to either make changes on their own or with the respondent. Interviewers also had the option to suppress the verification questions entirely and move on with the interview. Any information collected from these extra questions was stored as separate variables so as not to overwrite the original income and earnings entries. Data from the original income and earnings entries were used to create the earnings, income, and poverty status recodes on the 2012 NHIS Public Use data file.

During data processing, it was discovered that a small number of interviewers backed up and changed previously recorded income and/or earnings data to resolve or reduce income-earnings inconsistencies. In so doing, they overwrote the original income and earnings entries, which was not the intent of the verification test. Because income and earnings information was obtained differently for these families compared with the vast majority of families, two flags have been added to the Person (FSD) and Family (FIN) files that will allow data users to identify and remove these cases if they have concerns about data collection inconsistencies. Observations

(in this case, family members) are flagged "1" on ERNYRFLG in the Person file if the interviewer backed up in the instrument and changed the original value of ERNYR rather than completing the verification questions as intended. Likewise, observations (in this case, families) are flagged "1" on FINCFLG in the Family file if the interviewer backed up in the instrument and changed the original value of FINCTOT rather than completing the verification questions as intended. Cases that are "blank" on ERNYRFLG and/or FINCFLG are those for which data were collected per instrument instructions. The number of cases receiving a "1" on these flags is quite small (FINCFLG = 589 and ERNYRFLG = 275).

VI. Family Food Security Section (FFS)

Family food security refers to access by all persons in a family at all times to enough food for active, healthy lives. Beginning in 2011, a Family Food Security Supplement (FFS) that consisted of ten questions was added to the NHIS. The first three questions were asked about the family to determine if, in the last 30 days, the family was worried about food running out before there was money to buy more (FSRUNOUT), if the food purchased just didn't last until there was money to buy more (FSLAST), and if the family couldn't afford to eat balanced meals (FSBALANC). If any of these questions was answered affirmatively, that is, if the respondent answered "often true" or "sometimes true" to any of these three questions, additional questions were asked about the respondent and other adults in the family.

The next set of questions asked whether the adult respondent ever ate less than needed (FSLESS), was hungry but didn't eat (FSHUNGRY), or lost weight (FSWEIGHT) because there was not enough money for food, and whether any adults in the family ever cut or skipped meals (FSSKIP). If the respondent answered affirmatively to any of these four questions, the respondent was asked whether any adults in the family did not eat for a whole day because there was not enough money for food (FSNOTEAT). FSSKIP and FSNOTEAT were each followed by a question asking the number of days, in the past 30 days, this happened (FSSKDAYS and FSNEDAYS, respectively). The procedures for calculating raw food security scores and food security classification variables can be found in Appendix VII.

2012 National Health Interview Survey
Injury and Poisoning Episode File

The Family Core portion of the 2012 survey included questions about medically consulted injuries and poisonings that occurred for any member of the family within a three-month reference period. All injury and poisoning information was provided by the family respondent. The Injury/Poisoning Episode File was created from these data.

The inclusion criteria used beginning in 2004 were also used in 2012. In 1997-2003, the Injury/Poisoning Episode File contained episodes that were reported to occur within 104 days or four months of the interview and episodes where the date of the injury or poisoning was not reported. Beginning in 2004, the decision was made to retain all injury/poisoning episodes that reportedly occurred during the three months (91 days) prior to the date the injury/poisoning questions were asked based on responses to family level questions FIJ.010_01.000 to FIJ.028_00.000 about injury/poisoning, regardless of whether or not the date of the injury or poisoning episode subsequently reported by the family respondent in the family level questions was outside the 91 day reference period. Flags have been created to indicate which episodes may thus have occurred outside the 91 day reference period (ETFLG and BEIFLG).

The Injury/Poisoning Episode File is an episode-based file. Each episode is based on responses to family level questions (FIJ.010_01.000 to FIJ.028_00.000) about injury/poisoning episodes that resulted in one or more conditions. Each medically consulted injury/poisoning episode (for example, a call to a poison control center; visit to an emergency room, doctor's office or other health clinic; phone call to a doctor or other health care professional) reportedly occurred during the three months prior to the date the injury/poisoning questions were asked. An injury episode refers to the traumatic event in which the person was injured one or more times from an external cause (e.g., a fall, a motor vehicle traffic accident). An injury condition is the acute condition or the physical harm caused by the traumatic event. Likewise, a poisoning episode refers to the event resulting from ingestion of or contact with harmful substances, as well as overdoses or wrong use of any drug or medication, while a poisoning condition is the acute condition or the physical harm caused by the event. Up to a total of ten injury and/or poisoning episodes may be recorded, and each unique injury and/or poisoning episode is represented in this file... Each episode must have at least one injury condition or poisoning classified according to the nature of injury codes 800-909.2, 909.4, 909.9, 910-994.9, 995.5-995.59, and 995.80-995.85 in the Ninth Revision of the International Classification of Diseases (ICD-9-CM) and one external cause of injury code of E800-E848, E850-E869.9, E880-E929.9, or E950-E999. Other health conditions that were reported as occurring with the injury or poisoning, even if they are not classified according to the above mentioned nature of injury codes (e.g., mononeuritis of unspecified site (355.9), other symptoms referable to back (724.8)), are also included in the Injury/Poisoning Episode File.

The Injury/Poisoning Episode File contains information about the external cause and nature of the injury or poisoning episode, what the person was doing at the time of the injury or poisoning episode, the date and place of occurrence, the elapsed time between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked, where the person received medical advice, treatment, or follow-up care, whether the person was

hospitalized, whether the person missed any days from work or school due to the injury or poisoning, ICD-9-CM diagnostic codes, and ICD-9-CM external cause codes. ICD-9-CM diagnostic and external cause codes were assigned for all injury and poisoning episodes based on information about how the injury or poisoning happened, the body part injured or poisoned, and the type of injury or poisoning, along with responses to questions about specific types of injury or poisoning episodes, and activity.

Beginning in 2009 and continuing in 2012, for confidentiality reasons a decision was made to include only selected 4-digit external cause of injury codes in the public use file. All other 4-digit external cause of injury codes will be truncated to three digits. All the original 4-digit external cause of injury codes will be available through the RDC.

During the 2012 data editing process, 330 injury and poisoning episodes were removed out of an initial total of 3338. These included episodes with no information regarding cause, date and place of occurrence, duplicate episodes, etc. In addition, episodes were removed if they consisted *solely* of health conditions that could not be classified according to the nature of injury codes and external cause of injury codes listed above.

As in previous years, respondents reported episodes that they considered poisonings (e.g., food poisoning and allergic reactions) but that are not considered poisonings based on the ICD-9-CM. These types of episodes were included in the 1997-2003 data files. Beginning in 2004 and continuing in 2012, episodes that are not considered poisonings based on ICD-9-CM are no longer included in the Injury/Poisoning data file.

This file only contains information about injury and poisoning episodes. Other person-level information can be obtained by linking the Injury/Poisoning Episode File to other 2012 NHIS data files (Person, Sample Adult, and Sample Child) using the household identifier (HHX), family identifier (FMX), and person number (FPX). When using a linked Injury/Poisoning Episode File and Sample Adult File, analysis should be limited to those episodes for persons included in the Sample Adult File, and the Sample Adult weight should be applied. When using a linked Injury/Poisoning Episode File and Sample Child File, analysis should be limited to those episodes for persons included in the Sample Child File, and the Sample Child weight should be applied. See Appendix VI for additional information about merging data files.

Recall Period and Weights

Questions in the Injury/Poisoning section of the 2012 NHIS have a recall period of the "last 3 months." However, as the time between the injury/poisoning episode and the date the injury/poisoning questions were asked increases, the annualized number of injuries/poisonings reported decreases. For most analyses of the injury/poisoning data (e.g., estimates for all types of injury/poisoning episodes and estimates for less severe injuries/poisonings), limiting data to episodes with a reported five weeks or fewer between the injury/poisoning episode and the date the injury/poisoning questions were asked is recommended because analyses showed that respondents tend to forget less serious injuries (Warner, et al., 2005). For analysis of injury/poisoning episodes resulting in more serious outcomes (e.g., estimates for fractures and hospitalizations) that are unlikely to be forgotten, the data should not be limited to the five-week

period. The longer period of time between the injury/poisoning episode and the date the injury/poisoning questions were asked will increase the number of episodes reported and therefore increase the size of the sample and provide richer detail and greater stability in the estimate. It is suggested not to calculate two estimates, one for serious and one for non-serious injuries/poisonings and combine the two estimates.

Analysts may wish to use the recommended five-week reference period to maintain consistency with other studies using the five-week reference period with NHIS injury/poisoning data. However, because the number of days since the injury/poisoning occurred is now provided for each episode on the public use data file, analysts can choose the time period that is the most appropriate for their analysis.

To calculate an annual estimate of the number of injuries and poisonings, the weighted number of episodes reported during a time period is multiplied by the number of time periods in a year. For instance, to estimate the number of injury or poisoning episodes occurring annually using episodes with three months or less elapsing between the injury/poisoning and the date the injury/poisoning questions were asked, each three-month weighted count should be multiplied by 4 (i.e., by 12/3=4). If data are limited to episodes with five weeks or less between the injury/poisoning and the date the injury/poisoning questions were asked, each five-week weighted count should be multiplied by 10.4 (i.e., by 52/5=10.4).

Analysts are cautioned against estimating the number of different *people* injured or poisoned annually using the current NHIS questions. Estimating the number of persons injured using the annualizing method described in the above paragraph (i.e., multiplying the estimate by the number of time periods in a year) assumes that the same individuals experienced injuries at the same rate over the year. Analysts are cautioned to check the Dataset Documentation and the specific item in the questionnaire in order to insure that annual estimates for these kinds of injury or poisoning episodes have intrinsic meaning.

Variance Estimation

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

Technical Notes and Imputation Information

Two variables on the Injury/Poisoning Episode File, ICAUS and ECAUS, describe the external cause of the episode. ICAUS is the actual item found in the questionnaire. For each unique episode, the interviewer selected the category of ICAUS that he/she felt best described the episode based on the respondent's description of how the injury or poisoning happened (IPHOW). ECAUS is a recoded variable that describes the cause of the episode using categories based on ICD-9-CM external cause codes. The category into which an episode was placed was based entirely on the first ICD-9-CM external cause code listed for that episode. Appendix I in the Injury/Poisoning Episode Dataset Documentation contains a list of the ICD-9-CM external cause codes found in each category.

Analysts are cautioned regarding their use of the variable RPCKDMR, which indicates the elapsed time between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked. This variable is based on only the month, day, and year of the injury or poisoning episode provided by the respondent and the actual day the respondent was asked the injury/poisoning questions. No information from additional date questions that are currently in the survey were used in the creation of this variable. When possible, the elapsed time between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked is given in days. The time between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked is only given in months when the day of the injury or poisoning episode was not reported. Prior to 2004, the calculation of this variable was based on the last date when the interview was opened for examination or input of data, not necessarily on the date when the injury/poisoning questions were asked, which could be different. This could happen if the interviewer was unable to complete the interview in one visit and had to return at a later date, so the injury and poisoning questions may have been completed earlier than indicated by the date of the interview recorded by the CAPI instrument. If this occurred, the actual time between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked would be less than the elapsed time indicated by the variable RPCKDMR. Beginning in 2004, the actual date when the injury and poisoning questions were completed was recorded and used in the calculation of this variable.

Beginning in 2004 and continuing in 2012, imputation was implemented for episodes that did not have a valid month, day, and year of occurrence. Imputation was done so that it would be possible to calculate a specific elapsed time in days between the date of the injury/poisoning episode and the date the injury/poisoning questions were asked for all episodes in the Injury/Poisoning Episode File. Since all episodes in the file now have a specific elapsed time (RPD) between the date of the injury/poisoning episode and the date the injury/poisoning questions were asked, analysts will be able to calculate estimates based on the time period of their choice.

The variable RPD indicates the elapsed time in days between the date of the injury or poisoning episode and the date the injury/poisoning questions were asked. This variable is based on all date information that was given by the respondent, and when date information was missing, imputed information was used in the creation of this variable. For some injury and poisoning episodes, the respondent was only able to provide the month and year of occurrence; or a time period within the month (beginning, middle, or end) and year of occurrence; or the number of days, weeks, or months ago. For cases in which a month but no time period during the month was provided, a day was imputed between 1 and the last day of the month. For cases in which the month of the injury/poisoning episode and the time period within the month was provided, the day of the month within that time period was imputed. If the episode was reported as occurring during the beginning of the month, a day of 1-10 was imputed; for cases in the middle of the month, a day of 11-20 was imputed; and for cases at the end of the month, a day of 21 to either 28, 29, 30, or 31, depending on the month, was imputed. In other instances, the respondent was only able to provide a time period (i.e., number of days, weeks, or months) between the date the injury/poisoning occurred and the date the injury/poisoning questions were asked. For responses given in days ago, the corresponding value of RPD was calculated. For responses given in weeks ago or months ago, RPD was imputed from within, respectively, the interval 7(# weeks ago) ± 3 or the interval 30(# months ago) ± 15.

An elapsed time interval, with lower and upper bounds BIETD and EIETD, respectively, indicates the amount of uncertainty in the injury/poisoning episode date information that was provided by the respondent. If the specific day, month, and year of the episode were provided or could be deduced from information provided by the respondent, then BIETD = EIETD = RPD. Otherwise, BIETD and EIETD indicate the lowest and highest values of the elapsed time between the episode and the date the injury/poisoning questions were asked that were consistent with the reported episode date information, and RPD was imputed to be within that interval. In a few cases where insufficient information was provided to determine an elapsed time interval, values of BIETD, EIETD, and RPD were obtained from a random "donor" (another reported episode) using hot deck imputation.

There are several variables in the 2012 Injury/Poisoning Episode File that supply information about the imputed data and about the consistency of the episode date information provided by respondents. The variable IMPMETH indicates which episodes have a value for RPD that is based on a specific day, month, and year of the episode that was provided or was deduced from information provided by the respondent (i.e., no imputation was needed) and which episodes have a value for RPD that was imputed. Flag variables have been added to the file to indicate whether the elapsed time (RPD) or the elapsed time interval boundaries (BIETD and EIETD) fall within the 91-day reference period mentioned in family level questions FIJ.010_01.000 and FIJ.020_00.000. This was done because it is possible that the respondent provided inconsistent information (i.e., reported that the injury or poisoning occurred during the 91-day reference period mentioned in the family level questions, and then, in follow-up questions about the episode date, reported that the injury or poisoning occurred beyond the 91-day reference period mentioned in the family level questions). Also, the elapsed time interval boundaries and imputed values of the elapsed time were not constrained to be ≤ 91; they were only constrained to be consistent with the date information reported by the respondent. Variable ETFLG indicates whether the elapsed time (RPD) is ≤ 91 days. Variable BEIFLG indicates whether the boundaries (BIETD and EIETD) of the elapsed time interval are ≤ 91 days. These flags were created for convenience so that analysts can decide which version of inconsistently-reported date information to use. Analysts may also choose to re-impute values of RPD that are greater than 91, constraining them to be within the 91-day limit as well as within the elapsed time interval.

Additionally, the NHIS has a website devoted exclusively to the injury and poisoning data from the survey: **http://www.cdc.gov/nchs/nhis/rhoi.htm**. This site includes additional details on the NHIS injury and poisoning data, including historical context, links to documentation, questionnaires and other resources, more information on data editing, and references.

2012 National Health Interview Survey
Sample Child File

The Sample Child section of the 2012 NHIS covers additional subject areas not included in the Family Core. Moreover, the questions in the Sample Child section are more specific and are intended to gather more detailed information than those in the Family Core. Sample children do not self-report; instead a knowledgeable adult (typically a parent or guardian) answers questions about the sample child's health. A flag, QCCHILD, is included in the Person File to denote records that were removed from the Sample Child File for quality reasons.

In 2012, two new supplements were added to the Sample Child section for the first time. The Child Voice, Speech and Language Supplement: Child Communication Disorders (CCD) and the Child Balance Supplement (CBL) were asked for sample children aged 3 years and over. More details about the sections comprising the Sample Child File are provided below.

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

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

In 2012, there were 29 new condition questions embedded in the CHS section. The new questions about hypertension, high cholesterol, flu or pneumonia, constipation, sinusitis, strep throat or tonsillitis, headache, abdominal pain, joint pain, neck pain, lower back pain, other muscle or bone pain, severe sprains or strains, dental pain, other chronic pain, problems with overweight, other sore throat, fever, colds, nausea or vomiting, fatigue, excessive sleepiness, insomnia, depression, anxiety, stress, menstrual problems or PMS, and gynecologic problems were added to CHS in order to provide additional information for analysis of the Child Alternative Health supplement (CAL) data.

Beginning in 2011, NHIS terminology changed; "mental retardation" was replaced by "intellectual disability, also known as mental retardation." In addition, the term "autism" was expanded to "autism/autism spectrum disorder." Because of the changes in terminology, the variables with the original wording (ARM1, ARM2, CCONDL06) were replaced with new variables (ARM1R, ARM2R, CCONDL6R).

In 2010, there was a change to the universe for the question about whether the sample child had visited an emergency room because of asthma during the past 12 months (CASMERYR). Previously, this question was only asked for sample children who had an asthma attack during the past 12 months. Beginning in 2010, this question was asked for all sample children who were ever diagnosed with asthma.

In the NHIS, the Sample Child respondent (usually a parent) is asked to report the sample child's birth weight, current height, and current weight. Respondents have the option of reporting the child's height and weight in either U.S. Customary (lbs/oz; ft/in) or metric (kg; m/cm) format. Less than 1% of respondents reported in metric format. Metric responses on height and weight were converted into U.S. Customary format for inclusion on the micro data files. **No physical measurements were taken.** National estimates based on physical measurements, such as those available from NCHS' National Health and Nutrition Examination Survey (NHANES), may differ from those available from the NHIS, which are respondent-reported.

From 2004-2011, the Census editing program incorrectly read the two-digit birth weight in pounds variable, BWGT_LB, resulting in birth weights 10 pounds and over being changed to 1 pound. This error has been corrected for the 2004-2011 data years. Beginning in 2011, a new recode was developed to create a total birth weight in ounces variable, TOTOZ_P. This variable, which is top and bottom coded to protect the confidentiality of those who had extreme values in birth weight, replaces the separate birth weight in pounds and ounces variables, BWGTLB and BWGTOZ.

From 2004-2011, there was an additional error in the birth weight in grams variable, BWTGRM. The editing process did not allow data values above 5485 grams or below 500 grams, although data were collected up to 6900 grams. This error has been corrected for the 2004-2011 data files. In 2011, a new recode was developed to create a total birth weight in grams variable, BWTGRM_P. This new variable, which is top and bottom coded to protect the confidentiality of those who had extreme values in birth weight, replaces the birth weight in grams variable, BWTGRM.

Beginning in 2008, questions about children's current height and weight were limited to children aged 12-17 years. This limitation was introduced because of serious concerns about the reporting accuracy of height and weight information for younger children due to the rapid growth of children at younger ages. At the same time, an internal consistency check for the height and weight variables was added to the survey instrument to improve data quality. Extreme values for these height and weight variables triggered a request for interviewer verification of data entry and re-asking height and weight questions, if appropriate. In addition, body mass index (BMI) was calculated within the instrument, with extreme BMI values also triggering interviewer verification of height and weight. These consistency checks were solely within the survey instrument and are not reflected in the published questionnaire, documentation, or data file.

Variables with individual values for current height and weight for sample children aged 12-17 years are included in the public use Sample Child File. The current height and weight variables, CHGHT_TC, and CWGHT_TC, protect the confidentiality of sample children who might be identifiable by their unusual physical characteristics. Based on values from the

2004 – 2008 NHIS, the sex-specific height-for-age and weight-for-age values of the highest 1½ percent of records and the lowest 1½ percent of records were changed to "96" or "996" ("Not available"). For example, a 12 year old girl who weighed 60 lbs. or less (lowest 1½ percent of records) or 188 lbs. or more (highest 1½ percent of records) was coded "996." In cases where extreme values were reported for either current height or current weight, the data for both variables were changed to "96" or "996" ("Not available") on the public use data file.

Body Mass Index (BMI), a measure of body weight relative to height, was also added to the Sample Child File. BMI was calculated using the formula: BMI = kilograms /meters². Kilograms and meters were derived from (U.S. Customary) pounds and inches using the following factors: 1 kilogram=2.20462 pounds; 1 meter=39.37008 inches. BMI was calculated for all sample children aged 12-17 years with a reported current height and weight, including those for whom specific height and weight values were changed to "96" and "996" (not available) on the public use file for reasons of confidentiality. BMI variable values are released as 4 digit numbers with two decimal places implied. For example, a value of 2587 for the BMI variable indicates a 25.87 BMI.

Child mental health questions derived from the Child Behavior Checklist for children ages 2-3 years remain in the CHS section. The items in the checklist were chosen for their ability to discriminate between children who have not received mental health services in the preceding 12 months and those who have, by using demographically-matched normative and clinical samples for boys and girls. Each set of items can be viewed as comprising a scale with each item scored as either "0," "1," or "2." More information on the scale derived from the Child Behavior Checklist is included in Appendix IV of this document.

Beginning in 2008, the hearing status question, CHEARST1 (CHS.250), replaced the old core question, CHEARST. The 4 response categories in CHEARST (good, little trouble, lot of trouble, deaf) were expanded to 6 categories (excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, deaf) in CHEARST1.

In order to improve data collection about children with cerebral palsy (see Technical Notes below) the old core condition question, CONDL (CHS.060), was redesigned. In 2008, the question was divided into two separate questions, CONDL (CHS.060) and CONDL1 (CHS.061), in order to clarify the appropriate response codes. The child condition variables based on these questions were renamed CCONDL01-CCONDL10, replacing the variables CONDL1-CONDL10 used in previous years. As a result, in 2008 there were 20 cerebral palsy cases reported for sample children, and in 2009 there were 38 cases.

Technical Notes

From 2004-2007, there was a 10-fold increase in the number of sample children who were reported to have cerebral palsy (34 cases reported in 2003; 311-353 cases reported each year from 2004-2006). This increase was believed to be due to interviewer error arising from operational differences in the NHIS instruments used prior to 2004 and beginning in 2004. Beginning in Quarter 3 of 2007 an interviewer note was added to the instrument to address this difference, and the number of cerebral palsy cases decreased from 148 cases in the first 2

quarters of 2007 to 11 cases in the last 2 quarters of the year. Although this variable is included in the file for 2004-2007, it is suggested that it **not** be used for analysis from 2004-2007 Quarter 2.

Errors in past years complicate the ability to make comparisons over time. For example, in 1999, there was an incorrect skip pattern in the Sample Child questionnaire for question CHS.111 for children 2 years of age. As a result there are no data for 2 year olds for: HAYF1 (hay fever), RALLG1 (respiratory allergy), DALLG1 (food allergy), SALLG1 (skin allergy), DAIRH1 (frequent diarrhea), ANEMIA1 (anemia), EARINF1 (ear infection), and SEIZE1 (seizures).

Several questions pertaining to child behavior for children aged 2-3 years are used to create recodes regarding the child's mental health; only the recodes are included in the Public Use file. The background and usage of the mental health indicators can be found in Appendix IV.

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

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

The Child Health Care Access and Utilization Section (CAU) contains information on access to health care, dental care, and health care provider contacts. The questions pertaining to access to health care include: having a usual place for sick care; having a usual place for routine/preventive care; change in place of care; reasons for a delay in getting medical care; and the inability to afford medical care. A question on dental care asked about the length of time since last dental visit.

Questions regarding health care provider contacts include visits and other contacts with doctors and other health care professionals in the past 12 months. Doctors and other health care professionals include general doctors; specialists; dentists; orthodontists; oral surgeons; optometrists; ophthalmologists; foot doctors; chiropractors; physical, speech, respiratory or occupational therapists; audiologists; nurse practitioners; physician's assistants; midwives; obstetricians; gynecologists; psychiatrists; psychologists; psychiatric nurses; and clinical social workers; moreover, contacts or visits are not restricted to medical doctors or professionals working with/for a medical doctor. Note that questions about home care are asked independently of other types of health care visits. In addition, the reference period for all health care contacts is the past 12 months. Lastly, a separate question is asked about the number of visits to a hospital emergency room in the past 12 months.

Beginning in 2011, supplementary extended access and utilization questions were embedded throughout this section. Included were questions about reasons for not having a usual source of care (CNOUSLPL), difficulty finding providers (CPRVTRYR, CPRVTRFD, CDRNANP, CDRNAI), unmet medical need due to cost (CHCAFYRN, CHCAFYRF, CHCAFYR5, CHCAFYR6), timing and reasons for emergency room visits (CERVISND, CERHOS, CERREAS1–CERREAS8).

Starting in Quarter 2 of 2012, the universe was changed for eight CAU variables, CERREAS1–CERREAS8. The original universe for Quarter 1 included all sample children with at least one emergency room visit in the past year which did not result in a hospital admission. The new universe was expanded for Quarter 2-Quarter 4 and included all sample children with at least one emergency room visit in the past year (regardless of hospital admission). The variables with the new universe were renamed CERREA1R–CERREA8R. Both sets of variables are found on the data file. The original variables will have data for Quarter 1 and will be blank for Quarter 2-Quarter 4. The new variables will be blank for Quarter 1 and will have data for Quarter 2-Quarter 4. In order to get annualized estimates for this data, analysts will need to create a recode combining the variables. Similar changes were made to the AAU section for the sample adult.

III. Child Mental Health Brief Supplement (CMB)

As part of a collaborative agreement with the National Institute of Mental Health (NIMH), the Strengths and Difficulties Questionnaire (SDQ) was first used in 2001 in a Child Mental Health Supplement in the CAU section. The SDQ is a behavioral screening questionnaire for children aged 4 to 17 years with extended questions that provide information on the duration of a child's problem and the impact that the problem has on the child and his/her family. It is copyrighted by Dr. Robert Goodman, London, England and is used with his permission. More detailed information on the SDQ is provided in Appendix V of the 2004 Survey Description Document and/or the SDQ website at **http://www.sdqinfo.org**.

The SDQ was also fielded in 2003 in the CAU section and in 2004 in the CMH section. When the SDQ was transferred from the CAU section to the newly created Child Mental Health section (CMH) in 2004, the question/answer wording and the question order remained the same as in 2003. However, all variable names with the exception of CSCL7 were changed to accommodate the new editing system. The question numbers were also changed to reflect the new question numbering system and the new section name.

In 2002, the long version of the SDQ fielded in 2001 was deleted from the CAU section. A shorter version of the SDQ (a subset of the longer SDQ fielded in 2001) was added to the CHS section. This short SDQ was also fielded in 2005-2007 and 2010-2012 in the CMB section. More information about the short SDQ is provided in Appendix V of this document.

Because of the lack of funding in 2008 and 2009, most SDQ questions were dropped from the CMB Supplement. CMHDIFF (CMB.030_00.000), the question concerning the sample child's overall difficulties with emotions, concentration, behavior, or being able to get along with other people, was retained in the survey. In addition, the entire CMS Supplement about mental health services for sample children 4-17 years of age who have those difficulties was dropped from the survey.

More detail about the changes in the SDQ questions and variables can be found in the 2001-2011 Survey Description Documents on the NHIS web site: **http://www.cdc.gov/nchs/nhis.htm** .

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

<div style="border: 1px solid black; padding: 20px;">

Important Note

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

</div>

IV. Child Influenza Immunization Supplement (CFI)

The Child Influenza Immunization Supplement (CFI) was included for the first time in the 2005 NHIS. This supplement contains information on receipt of a flu vaccination in the past 12 months; month and year of the most recent flu vaccination; receipt of nasal flu spray vaccination in the past 12 months; and month and year of most recent nasal flu spray vaccination. These questions were also administered to all sample adults (see the AAU section).

There were no changes to the CFI section in 2012. The questions were the same as those used in 2011 and the latter part of 2010. However, analysts interested in trend analyses should note that, from January-August 2010, the NHIS included additional supplemental questions about the H1N1 flu vaccine, which was offered separately from the seasonal flu vaccine during the 2009-2010 flu season. Information regarding the 2010 flu vaccination data can be found in the CFI Appendix on the NHIS website, **ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/samchild_layout.pdf**.

V. Child Mental Health Services (CMS)

The purpose of the Child Mental Health Services Supplement (CMS) is to identify the interaction of sample children with health professionals or school staff personnel related to emotional or behavioral problems. In general, the questions are about the sample child's most recent consultations regarding emotional or behavioral problems and medication prescribed, if any. The questions are included in the Sample Child Questionnaire as supplemental questions for sample children aged 4 to 17 years.

The 2012 CMS supplement is a continuation of the 2011 CMS. Questions fielded include whether the sample child had difficulties with emotions, concentration, behavior, or getting along with others, life impact caused by the difficulties, medication prescribed for the difficulties, type of doctor prescribing medication, receipt of treatment or counseling services, attendance at a school specializing in children with difficulties, attendance at a regular school with program for children with difficulties, type of facility where child received treatment or counseling, type of counselor providing treatment, use of overnight treatment facility, type of overnight treatment facility used, use of case manager, payment method for treatment or counseling, unmet need for treatment or counseling because of difficulties, and reasons for unmet need.

The 2011 supplement was similar to the 2010 supplement, but some questions were revised and a few questions were added. Detailed information about question and variable changes to the 2011 CMS supplement can be found in the 2011 Survey Description Document on the NHIS web site: **http://www.cdc.gov/nchs/nhis.htm**.

A significantly shorter Child Mental Health Services Questionnaire was fielded on the 2005, 2006 and 2007 NHIS.

Technical Notes

In 2010, answer categories changed in the question OVERWHCH (CMS.060_00) where respondents were asked whether the sample child stayed overnight or longer in a facility to receive counseling or treatment for difficulties with emotions, concentration, behavior, or getting along with others. As a result, comparisons to 2010 may be complicated. Detailed information about question and variable changes in the 2010 CMS supplement can be found in the 2010 Survey Description Document on the NHIS web site: **http://www.cdc.gov/nchs/nhis.htm**.

VI. Child Voice, Speech and Language Supplement: Child Communication Disorders (CCD)

The Child Voice, Speech and Language Supplement: Child Communication Disorders (CCD) included in the 2012 NHIS is the first-ever large-scale communication supplement in the survey. The CCD Supplement encompasses the areas of voice, swallowing, speech and language problems for sample children aged 3 years and over. Data on the condition within the past 12 months, causes, age of onset, and severity are included for each CCD area. Further, detailed information on healthcare utilization as a result of the child's voice, swallowing, speech or language problem is available in CCD.

A companion Adult Voice, Speech and Language Supplement: Adult Communication Disorders (ACD) is available on the Sample Adult File.

VII. Child Balance Supplement (CBL)

The Child Balance Supplement (CBL) included in the 2012 NHIS is the first large-scale balance supplement in the survey. The CBL supplement encompasses the areas of vertigo, poor balance, motor coordination, falls and light-headedness/feeling faint for sample children aged 3 years and over. Data on the condition within the past 12 months are included for each of the specific areas of balance disorders. The causes, interference with the child's everyday activities, severity, and healthcare utilization as a result of the child's balance problems overall are also included. Age of first steps and general problems with standing, walking or using limbs are also included in this section.

2012 National Health Interview Survey
Child Complementary and Alternative Medicine File
(CAL)

In 2012 the Child Complementary and Alternative Medicine Supplement (CAL) collected information about all sample children aged 4 years or older on their use of 18 non-conventional health care practices. The modalities covered in the supplement include acupuncture, ayurveda, biofeedback, chelation therapy, chiropractic or osteopathic manipulation, craniosacral therapy, energy healing therapy, hypnosis, massage, naturopathy, traditional healers, movement therapies (Pilates/Trager psychophysical integration/ Feldenkrais), herbal and non-vitamin supplements, vitamins and minerals, homeopathy, special diets, yoga/tai chi/qi gong, and relaxation techniques (meditation/guided imagery/progressive relaxation). An earlier, similar version of the CAL was fielded in 2007, but some changes have been made for the 2012 supplement and not all content is comparable to the 2007 version. In particular, the 2007 CAL supplement asked questions of all sample children, regardless of age, while the 2012 CAL supplement excluded children under the age of 4. Analysts who are interested in comparing estimates from 2007 to 2012 should take this difference into account in their analysis.

For each health care practice covered on the CAL supplement, questions were asked about the following: whether the sample child ever used the modality, and, if so, whether the modality was used in the past 12 months; number of times the sample child had seen a practitioner for the modality; portion of the cost covered by insurance for the modality; amount paid out-of-pocket for the modality; and additional materials purchased to learn about the modality.

While some questions were asked of each health care practice, others were asked only for the top three modalities deemed by the respondent to be most important to the sample child's health. For the top 3 question series all modalities were included except for ayurveda, chelation therapy, and vitamins and minerals. These three therapies were excluded due to either very low or high prevalence. The topics covered in the top 3 series include: reasons for using the modality; whether the modality motivated the respondent to engage in other selected health behaviors; whether the modality was used to improve athletic/sports performance; outcomes associated with using the modality; whether the modality was used to treat a specific health problem or condition, and, if so, what health problems or conditions were treated and for which one of the health problems or conditions the modality was used the most; whether the respondent received any of a specified set of traditional medical treatments for the condition treated most by the modality, and, if so, when the conventional medical treatments were received in relation to when the modality was used; and disclosure of modality use to a specified set of conventional medical professionals. Given the number of modalities and questions, a separate stand-alone data file was created for the supplement rather than append it to the 2012 Sample Child file. The Sample Child weight is included on the CAL file and should be used for all analyses with variables on this file.

Although the questions in the CAL Supplement were asked about all sample children aged 4 years or older, some respondents did not complete the supplement. Respondents who did not give responses to any of the questions in the supplement had a value of "not ascertained" inserted in the appropriate fields in this file. In other words, these children (n=195, or 1.9 % of

sample children aged 4+) are retained in the file, but they are coded as "8" in all the relevant fields of the file.

Back Coding

In the CAL Supplement, respondents were given the opportunity to choose "Other specify" as a response to the questions about what health problems or conditions were treated with specific types of complementary and alternative medicine (CAM): CTP1CD56, CTP2CD56, CTP3CD56. If "Other specify" was selected, the respondent was then asked to give a verbatim response indicating the health problem or condition treated with the specified type of CAM (CTP1SPEC, CTP2SPEC, CTP3SPEC). As part of the editing process, verbatim responses were back coded (added) into existing categories (CTP1CD01-CTP1CD55; CTP2CD01-CTP2CD55; CTP3CD01-CTP3CD55) or when appropriate, into an *ad hoc* "not elsewhere classified" category (CTP1CD57, CTP2CD57, CTP3CD57). In all cases, the original verbatim text was retained by NCHS, but the original verbatim responses are excluded from the public use file in order to protect the confidentiality of the respondents.

The frequency counts for the questions about the condition treated most often with each CAM therapy have not been changed to reflect back coding the "other specify" responses. While back coding the condition verbatim responses it became apparent that some respondents were not using the selected CAM therapy to treat a specific condition (e.g., someone who received a massage at a spa as a gift). Therefore, in such cases, the values may have been changed for the following variables: CTP1TRET, CTP2TRET, and CTP3TRET.

Cost Variables

Responses to questions on the CAL Supplement about out-of-pocket cost of therapies and materials are reported exactly as reported by respondents, with the exception of upper limits placed on the cost variables (maximum value varies by question – see layout file for upper limit for each variable). Given the 12 month recall period for these questions, values may be subject to inherent memory limitations. Data users should be aware that these limitations may affect the data, although such effects have not been evaluated.

Technical notes

The sample child weight, WTFA_SC, provided on the CAL file is designed to produce annual-level estimates calculated based on data included in this file. In order to increase the number of variables available for analysis with the CAL file variables, the CAL file may be merged with one or more NHIS data files (e.g., Sample Child, Person, etc.) for a given individual. Information on merging data files can be found in Appendix VI of this document.

<div style="border: 2px solid black; padding: 10px; color: red;">
CAUTION: The CAM use variables should be analyzed with care. Many of the CAM use variables are asked of only a subset of the sample children and therefore should not be used to estimate prevalence for the health care practices they represent. In order to obtain accurate national prevalence estimates for use of health care practices, data must be weighted using WTFA_SC and recodes are required to appropriately account for each variable's universe. For specific information about who was asked each question, refer to the CAL Variable Layout file.
</div>

2012 National Health Interview Survey
Sample Adult File

The Sample Adult section of the NHIS covers many of the subject areas included in the Family Core. However, the questions in the Sample Adult section are more specific and are intended to gather more detailed information. In addition, sample adults generally respond for themselves, although in a small number of cases, proxy responses are allowed if the selected adult had a physical or mental condition prohibiting him/her from responding. The variable PROX1 indicates those cases where information was obtained from a proxy respondent. A flag, QCADULT, is included in the Person File to denote records that were removed from the Sample Adult File for quality reasons. The sections comprising the Sample Adult File are discussed below.

I. Adult Socio-Demographics Section (ASD)

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

Sample adults aged 18 years and older who were "working at a job or business," "with a job or business but not at work," or "working, but not for pay, at a job or business" during the week prior to their interview were asked a series of questions about their job and work status during the week prior to the interview. In addition, those sample adults who said that they were "looking for work" or "not working and not looking for work" during the week prior to the interview were asked if they had "ever held a job or worked at a business." Sample adults who responded affirmatively were then asked the occupation, industry and work status questions in the ASD section. Note that sample adults who had ever worked and were either retired or 65 years of age or older were asked about the job they had held the longest, whereas sample adults who had ever worked, were younger than 65 years of age, and were not retired were asked about their most recently held job. In a subsequent question (WRKLONGH or ASD.146), currently employed sample adults were asked if their current job was also the job they had held for the longest time. Likewise, sample adults who had ever worked and were not retired were asked if their most recently held job was also the job they had held for the longest.

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

Users should be aware that DOINGLWA (ASD.062) and WHYNOWKA (ASD.065) are the ASD equivalents of DOINGLWP (FSD.050) and WHYNOWKP (FSD.060) in the FSD section. For the majority of respondents, DOINGLWA and DOINGLWP will have identical

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

With the exception of BUSINC1A (ASD.112), WRKLONGH (ASD.146), and ONEJOB (ASD.170), the universe for all variables in the 2012 ASD section includes currently employed and ever employed sample adults. Variables with smaller universes (e.g., currently employed sample adults only) that were included in the 1997-2000 data files have been eliminated. Users wishing to replicate those variables are advised to use DOINGLWA to identify the subset of currently employed sample adults (i.e., DOINGLWA = 1, 2, or 4).

Industry and Occupation Coding

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

Changes in the U.S. economy led to changes in the SIC and SOC classifications. After an extensive period of review, the standard industry and occupation classifications – and the corresponding Census codes used by the NHIS and other non-economic Federal surveys – were replaced by the North American Industrial Classification System (NAICS) and a revamped (as of 2000) SOC. Accordingly, the Census Bureau has developed new industry and occupation codes to replace the obsolete codes informed by the old SIC and SOC.

The 2012 NHIS in-house data files contain 4-digit Census codes for industry and occupation consistent with the 2007 NAICS and 2010 SOC. As with the 2004-2011 NHIS public use data files, the 2012 NHIS public use data files contain 2-digit industry and occupation recodes based on the in-house Census codes. The 2004 NHIS public use data contained a second set of 2-digit industry and occupation recodes (OCCUP1A, OCCUP2A, INDSTR1A, and INDSTR2A) based on the 3-digit 1990 Census codes (and, in turn, the 1987 SIC and 1980 SOC); these were dropped in 2005 (and thereafter).

Users are advised that the previous coding scheme based on the 3-digit Census codes and the new coding scheme based on the new (as of 2004) 4-digit Census codes are entirely different classification systems that are not compatible with one another. Moreover, crosswalks showing how these systems compare to one another are not available at this time. However, the coding categories for these recodes are provided in the Industry and Occupation Appendices (following the Variable Layout documentation for the Sample Adult data file), and additional information is available on-line (see the final paragraph in this section).

While the 2012 NHIS Sample Adult public use file does not include the in-house Census codes, it does include a detailed occupation recode (OCCUPN1) with 94 distinct categories, while the associated simple recode (OCCUPN2) has 23 categories. These categories are derived from the 2010 SOC Occupation Subgroups and Major Occupation Groups, respectively, as determined by the U.S. Census Bureau and the Bureau of Labor Statistics. The detailed industry recode (INDSTRN1) informed by the 2007 NAICS has 79 distinct categories, while the associated simple recode (INDSTRN2) has 21 categories. These categories are derived from the NAICS Industry Subsectors and Sectors, respectively, as identified by Census.

Users should consult the Occupation and Industry Appendices at the end of the Sample Adult variable layout file for the response categories and labels of OCCUPN1, OCCUPN2, INDSTRN1, and INDSTRN2. Links to the complete lists of NAICS Industry Subsectors and Sectors and the SOC Occupation Subgroups and Major Groups are embedded within these appendices, and provide the classification framework for the NHIS public use recodes. These lists should <u>not</u> be used in place of the Occupation and Industry Appendices. For more information about the 2007 NAICS, please refer to **http://www.census.gov/epcd/www/naics.html**. For more information about the 2010 SOC, please refer to **http://www.bls.gov/soc/home.htm**.

Technical Notes

The 2012 ASD section contains a new variable (SCHOOLYR) based on a question that asked sample adults if they had attended any kind of school during the past 12 months. The question was added to obtain an appropriate universe for several questions in the 2012 Adult Complementary and Alternative Medicine and the Voice, Speech, and Language supplements.

II. Adult Conditions Section (ACN)

The ACN section of the 2012 NHIS obtains information from the sample adult as to whether he or she has, or has had, a selected number of medical conditions. In most instances, sample adults were asked whether a doctor or other health professional had told them that they had the condition in question (joint symptoms, pain, hearing, vision impairment, and tooth loss are the exceptions). Respondents are also asked about head colds and intestinal illness which began in the 2 weeks prior to the interview, and women age 18-49 are asked about current pregnancy status. In addition, the section contains information about the sample adult's current mental or emotional health (whether he or she experienced feelings of sadness, nervousness, restlessness, hopelessness, worthlessness, or that everything was an effort in the past 30 days), and the extent to which these feelings interfere with his or her life or daily activities (Kessler's "K6" screen for nonspecific psychological distress). For more information about Kessler's K6, please refer to

http://www.hcp.med.harvard.edu/ncs/k6_scales.php. In the 2012 NHIS, the ACN section
contained a number of supplemental questions that asked respondents about medical conditions, as
part of either the Adult Alternative and Complimentary Medicine Supplement, or the Million Hearts®
Initiative Supplement. These supplemental questions are detailed below. Table 9 shows the specific
health-related conditions in this section and the various reference periods covered by the questions.
Table 9 includes not only conditions that are asked as part of the NHIS core, but also those asked as
part of the Adult Alternative and Complimentary Medicine Supplement and Million Hearts
Initiative® Supplement.

Table 9. Sample Adult File: Conditions and Reference Periods

	Condition	Core or Supplement	Reference Period					
			Ever	12 mos.	3 mos.	30 days	2 weeks	Now
ACN.010	Hypertension	C	X					
ACN.020	Hypertension 2+ visits	C	X					
ACN.020	Hypertension recent	S		X				
ACN.031	Coronary heart disease	C	X					
ACN.031	Coronary heart disease recent	S		X				
ACN.031	Angina pectoris	C	X					
ACN.031	Heart attack (MI)	C	X					
ACN.031	Other heart condition or heart disease	C	X					
ACN.031	Other heart condition or heart disease recent	S		X				
ACN.031	Stroke	C	X					
ACN.031	Emphysema	C	X					
ACN.035	Chronic obstructive pulmonary disease (COPD)	C	X					
ACN.080	Asthma	C	X					
ACN.085	Asthma still have	C						X
ACN.090	Asthma episode / attack	C		X				
ACN.100	Asthma ER visit	C		X				
ACN.110	Ulcer ever told	C	X					
ACN.120	Ulcer recent	C		X				
ACN.121	High cholesterol	S	X					
ACN.121	High cholesterol	S		X				
ACN.121	Influenza or pneumonia	S	X					
ACN.121	Influenza or pneumonia recent	S		X				

	Condition	Core or Supplement	Reference Period					
			Ever	12 mos.	3 mos.	30 days	2 weeks	Now
ACN.121	Strep throat or tonsillitis	S	X					
ACN.121	Strep throat or tonsillitis recent	S		X				
ACN.121	Poor circulation in legs	S	X					
ACN.121	Poor circulation in legs recent	S		X				
ACN.121	Urinary problems	S	X					
ACN.121	Urinary problems recent	S		X				
ACN.121	Phobias or fears	S	X					
ACN.121	Phobias or fears recent	S		X				
ACN.121	Attention Deficit Hyperactivity Disorder (ADHD) or Attention Deficit Disorder (ADD)	S	X					
ACN.121	Bipolar Disorder	S	X					
ACN.121	Depression	S	X					
ACN.121	Depression recent	S		X				
ACN.121	Other mental health disorders	S	X					
ACN.121	Other mental health disorders recent	S		X				
ACN.125	Any respiratory allergy	S		X				
ACN.125	Any digestive allergy	S		X				
ACN.125	Eczema or skin allergy	S		X				
ACN.125	Other allergies (not including hay fever, respiratory, food, digestive, or skin)	S		X				
ACN.125	Acid reflux or heartburn	S		X				
ACN.125	Fever for more than one day	S		X				
ACN.125	Head or chest cold	S		X				
ACN.125	Nausea and/or vomiting	S		X				
ACN.125	Sore throat other than strep or tonsillitis	S		X				
ACN.125	Infectious diseases or problems of the immune system	S		X				
ACN.125	Recurring headache, other than migraine	S		X				

	Condition	Core or Supplement	Ever	12 mos.	3 mos.	30 days	2 weeks	Now
				Reference Period				
ACN.125	Memory loss or loss of other cognitive functions	S		X				
ACN.125	Neurological problems	S		X				
ACN.125	Abdominal pain	S		X				
ACN.125	Severe sprains or strains	S		X				
ACN.125	Dental pain	S		X				
ACN.125	Other muscle or bone pain	S		X				
ACN.125	Other chronic pain	S		X				
ACN.125	Excessive use of alcohol or tobacco	S		X				
ACN.125	Substance abuse (other than alcohol or tobacco)	S		X				
ACN.125	Problems with being overweight	S		X				
ACN.125	Skin problems (other than eczema or allergies)	S		X				
ACN.125	Fatigue or lack of energy for more than 3 days	S		X				
ACN.125	Regularly had excessive sleepiness during the day	S		X				
ACN.125	Regularly had insomnia or trouble sleeping	S		X				
ACN.125	Frequently felt anxious, nervous, or worried	S		X				
ACN.125	Frequently felt stressed	S		X				
ACN.130	Cancer any	C	X					
ACN.140	Cancer kind	C	X					
ACN.160	Diabetes	C	X					
ACN.165	Prediabetes	C	X					
ACN.180	Insulin	C						X
ACN.190	Oral agents/pills	C						X
ACN.201	Hay fever	C		X				
ACN.201	Sinusitis	C		X				
ACN.201	Chronic bronchitis	C		X				
ACN.201	Weak kidneys	C		X				
ACN.201	Liver condition	C		X				
ACN.250	Joint symptoms	C				X		
ACN.260	Joints affected	C				X		
ACN.270	Joint symptoms chronic	C			X			

	Condition	Core or Supplement	Ever	12 mos.	3 mos.	30 days	2 weeks	Now
					Reference Period			
ACN.280	Joints doctor consult	C	X					
ACN.290	Arthritis (arthritis, gout, fibromyalgia, rheumatoid arthritis, lupus) diagnosis	C	X					
ACN.295	Limited in activities due to arthritis/joint symptoms	C						X
ACN.297	Arthritis: arthritis	S	X					
ACN.297	Arthritis: rheumatoid arthritis	S	X					
ACN.297	Arthritis: gout	S	X					
ACN.297	Arthritis: lupus	S	X					
ACN.297	Arthritis: fibromyalgia	S	X					
ACN.297	Arthritis: other joint condition	S	X					
ACN.300	Neck pain	C			X			
ACN.310	Back pain	C			X			
ACN.320	Leg pain	C			X			
ACN.331	Jaw, face pain	C			X			
ACN.331	Migraine	C			X			
ACN.350	Head/chest cold	C					X	
ACN.360	Intestinal illness	C					X	
ACN.370	Pregnant	C						X
ACN.372	Menstrual problems	S		X				
ACN.372	Menopausal problems	S		X				
ACN.372	Gynecologic problems	S		X				
ACN.372	Men's health problems (including prostate trouble or impotence)	S		X				
ACN.400	Use hearing aid	C						X
ACN.410	Use hearing aid	C	X					
ACN.420	Hearing	C						X
ACN.430	Vision impairment	C						X
ACN.440	Blind	C						X
ACN.451	Lost all teeth	C						X
ACN.471	Sad	C				X		
ACN.471	Nervous	C				X		
ACN.471	Restless	C				X		

	Condition	Core or Supplement	Ever	12 mos.	3 mos.	30 days	2 weeks	Now
					Reference Period			
ACN.471	Hopeless	C				X		
ACN.471	Everything an effort	C				X		
ACN.471	Worthless	C				X		

The cancer questions were asked in a format that allowed a respondent who reported having had cancer to specify up to three types of cancer as well as to indicate if the sample adult had had more than three different cancers. The responses were recorded with the codes indicated in the questionnaire and were then transformed into "mentioned"/ "not-mentioned" variables during editing. These variables (CNKIND1-31) assign to every sample adult who reported having ever had cancer either a "mentioned," if he/she specified that particular cancer, a "not mentioned," if he/she did not specify that cancer, or a "refused," "don't know," or "not ascertained," if there was no information for any of the cancers. Thus, a sample adult may have a code in each of the cancer variables, but can have only up to three "mentions," with a fourth mention possible for the variable CNKIND31 ("More than 3 kinds").

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

In 2012, a new core question was added to the ACN section that asked the respondent if the sample adult had ever been told by a doctor or other health professional that he/she had chronic obstructive pulmonary disease (COPD).

The ACN section for sample adults included a series of supplemental questions pertaining to the Million Hearts® initiative. This series of questions pertained to aspirin use, hypertension (or high blood pressure), hyperlipidemia (or high cholesterol), and current smoking. These questions asked the respondent if the sample adult had ever been told by a doctor or other health professional that he/she had hypertension in the past 12 months (HYPYR), coronary heart disease in the past 12 months (CHDYR), any kind of heart disease or condition (other than coronary heart disease, angina pectoris, or myocardial infarction) in the past 12 months (HRTYR), ever had high cholesterol (CHLEV), and had high cholesterol in the past 12 months (CHLYR). In addition to these questions, four additional questions were included in the ACN section as part of this supplement, which asked about aspirin use among sample adults aged 40 years and over. These four questions included: has a doctor or other health professional ever told the sample adult to take low-dose aspirin each day (ASPMEDEV), is the sample adult following this advice (ASPMEDAD), has a doctor or other health professional advised the sample adult to stop taking low-dose aspirin (ASPMDMED), and if the sample adult is taking low-dose aspirin on his/her own (ASPONOWN).

The ACN section for sample adults included a series of questions pertaining to the Adult Alternative and Complimentary Medicine Supplement. These questions asked the respondent if sample adult ever had influenza/pneumonia (AFLUPNEV); had influenza/pneumonia in the past 12 months (AFLUPNYR); ever had strep throat/tonsillitis (ASTREPEV); had strep throat/tonsillitis in the past 12 months (ASTREPYR); ever had poor circulation in his/her legs (PRCIREV); had poor circulation in his/her legs in the past 12 months (PRCIRYR); ever had urinary problems (UREV); had urinary problems in the past 12 months (URYR); ever had phobias/fears (PHOBIAEV); had phobias/fears in the past 12 months (PHOBIAYR); ever had ADHD/ADD (ADDHYP1); ever had bipolar disorder (BIPDIS); ever had depression (ADEPRSEV); had depression in the past 12 months (ADEPRSYR); ever had other mental health disorders (MHDOTHEV); and had other mental health disorders in the past 12 months (MHDOTHYR).

Additional questions included in the ACN section pertaining to the Adult Alternative and Complimentary Medicine Supplement asked if the sample adult in the past 12 months had respiratory (RESPALYR), digestive (DGSTALYR), eczema/skin (SKNALYR), or other (OTHALYR) allergies; problems with acid reflux/heartburn (ACIDRYR); fever for more than one day (AFEVRYR); a head/chest cold (ACOLDYR); nausea/vomiting (ANAUSYR); sore throat (ATHOTHYR); infectious disease/problems with immune system (IMMOTHYR); recurring headache (AHEADYR); memory or other cognitive function loss (MEMLOSYR); neurological problems (NEUROYR); abdominal pain (AABDOMYR); any severe strains/sprains (SPNYR); dental pain (DENYR); other muscle/bone pain (AMUSCLYR); other chronic pain (APNOTHYR); excessive use of alcohol/tobacco (ALCTOBYR); substance abuse (SUBABYR); problems with being overweight (AOVRWTYR); skin problems other than eczema/allergies (SKNYR1); fatigue/lack of energy for more than three days (FATIGYR); excessive sleepiness during the day (FATYR); insomnia/trouble sleeping (INSYR); frequently felt anxious/nervous/worried (ANXNWYR); and felt stressed (ASTRESYR). In addition, supplement questions were also included that ask the respondent which form of arthritis the sample adult had (if the sample adult was identified as having arthritis), including arthritis (ARTHTYP1), rheumatoid arthritis (ARTHTYP2), gout (ARTHTYP3), lupus (ARTHTYP4), fibromyalgia (ARTHTYP5), or some other joint condition (ARTHTYP6). The final series of supplement questions asked the respondent if the sample adult had recently been pregnant (PREGFLYR); or if in the past 12 months the sample adult had menstrual problems (MENSYR), menopausal problems (MENOYR), gynecological problems (GYNYR), or men's health problems such as prostate trouble or impotence (PROSTYR).

Beginning in 2010, a change was made to the universe of the core question that asked whether or not a person had visited an emergency room in the past 12 months because of asthma (AASERYR1). Beginning in 2010, this question was asked of any sample adult who had ever been told he/she had asthma. Previously (1997-2009), this question was asked only of sample adults who had an asthma episode/attack in the past 12 months.

Major changes were made in 2002 to core questions about arthritis and joint symptoms, and those questions remained unchanged through 2012. Users are advised to read the 2002 Survey Description Document to learn about those changes. Because of those changes, any comparisons of 2002-2012 arthritis and joint symptom data with data prior to 2002 should be undertaken with caution.

In 2007, a new core diabetes question was added: DIBPRE1 ("Have you EVER been told by a doctor or other health professional that you have any of the following: prediabetes, impaired fasting glucose, impaired glucose tolerance, borderline diabetes, or high blood sugar?").

The hearing question AHEARST1 ("Is your hearing excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, or are you deaf?"), introduced in 2007 as a supplementary question, has been a core question since 2008, replacing the old core question AHEARST. AHEARST1 has more response categories than AHEARST. Two questions about hearing aid use, HRAIDNOW and HRAIDEV, were supplementary in 2007 and are now core questions, replacing the old core question HEARAID. In addition to the increase in response categories in the hearing question AHEARST1, researchers should note that the placement of the hearing aid questions relative to the hearing question changed in 2007, which may also result in differences in estimates relative to estimates prior to 2007.

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

The 2012 Adult Health Status and Limitation of Activity component of the Sample Adult File contains information from respondents on illness behavior, health status, use of special equipment, limitations in functional activities, and the conditions underlying such limitations. While the AHS section may seem similar to the FHS section in the Person File, the questions in these sections have a somewhat different focus. For example, both sections asked about the ability to walk without special equipment. However, the walking limitation question in the FHS section (PLAWALK or FHS.220) only captured whether a person has difficulty walking without using special equipment. In contrast, the Sample Adult question on walking (FLWALK or AHS.091_01) asked about the degree of difficulty the respondent has walking a specified distance (a quarter mile, or about three city blocks) by him/herself and without using any special equipment.

The 2012 AHS time variables and recodes, which indicate how long respondents have had the condition(s) causing their limitation(s), were processed using procedures similar to those used in 2002-2011. Substantively, the 2002-2012 variables and recodes are similar to those from previous years (1997-2001), but the detailed unknown categories that were included in the earlier data were collapsed into broader categories starting in 2002.

Beginning in 2011, NHIS terminology changed from "mental retardation" to "intellectual disability, also known as mental retardation." Consequently, the resulting adult condition variable and the time variables and recodes associated with this condition variable were renamed. The adult condition variable is now called ALHCA14A, and the time variables and recodes associated with this variable are called ATIME14A, AUNIT14A, ADURA14A, ADURB14A, and ACHRC14A.

Health Indicators: Illness Behavior and Health Status

The first questions in this section determined the number of days the respondent took off from work or spent in bed due to illness or injury during the 12 months prior to the interview. In

addition, respondents were asked to compare their health now (whether it is better, worse, or the same) to their health 12 months ago.

Limitation of Functional Activities

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

As in FHS, if the sample adult reported difficulty with any of these 12 activities, he/she was then asked what condition(s) cause the difficulty, as well as how long he/she has had the condition. The format of these condition data is similar to that found in the FHS section.

Conditions

Each sample adult indicating any functional limitation (regardless of the degree of the limitation) is asked about the condition(s) or health problem(s) associated with that limitation, as well as the amount of time he/she has had the condition. Sample adults were given the following fixed response categories: "vision/problem seeing," "hearing problem," "arthritis/rheumatism," "back or neck problem," "fracture, bone/joint injury," "other injury," "heart problem," "stroke problem," "hypertension/high blood pressure," "diabetes," "lung/breathing problem (e.g., asthma and emphysema)," "cancer," "birth defect," "intellectual disability, also known as mental retardation," "other developmental problem (e.g., cerebral palsy)," "senility," "depression/anxiety/emotional problem," and "weight problem. Starting in 2001 and continuing in 2012, if the sample adult was limited by a condition not listed in one of these 18 fixed categories, the interviewer accessed a second screen containing 17 additional condition categories and two "other impairment problem" categories. These conditions were not read aloud to respondents, but if the sample adult's condition was limited by one of these 17 conditions, the interviewer recorded this information. If the sample adult was limited by a condition not included in one of the 18 fixed categories or on the interviewer's computer screen, then the interviewer entered a 50-character verbatim response for one or both of the "other impairment problem" categories.

The AHS condition data were edited very much like the condition data in FHS. The verbatim responses recorded by interviewers in one or both of the 50-character fields indicating "other impairment problem," as well as those in the 17 additional "second screen" categories seen by the interviewers, were subsequently analyzed during data processing. While most respondents named "other" conditions that did not fall into the 18 fixed response categories as originally specified in the instrument, some respondents named conditions that should have been included in one of the fixed categories. In the latter case, these "other" responses were assigned codes during data processing corresponding to the appropriate category. An additional 16 *ad hoc*

categories were created, and were assigned numbers 19_ thru 34_. (Note: Due to a naming convention error, in 2002 and 2003, these same *ad hoc* categories were assigned numbers 19 thru 34 without an underscore.) Any verbatim conditions that could not be back-coded to one of the 18 fixed categories or recoded to one of the *ad hoc* categories remained in the "other impairment" categories, and were renumbered "90" and, if necessary, "91." In addition, responses in the 17 "second screen" categories seen only by the interviewer were also back-coded and categorized into 8 of the *ad hoc* categories (refer to Table 4 in the FHS section). The resulting 34 output categories were generally based on the International Classification of Diseases, Ninth Revision, Clinical Modification (see Table 5 in the FHS section).

These specific condition categories were subsequently transformed into variables indicating whether or not the condition was responsible for the respondent's difficulty with any functional activity (a mention/not-mention format). Because the 16 *ad hoc* categories were not included on the flash cards given to respondents during the course of the interview, it is possible that frequencies obtained for these conditions may be underestimates. Therefore, these variables should be analyzed with care. Moreover, none of the AHS condition variables (AFLHCA1 through AFLHCA34_) should be used to estimate prevalence rates for the conditions they represent, because only those sample adults with a previously reported functional limitation were eligible for the condition questions that followed. Analysts who are interested in estimating the prevalence of particular conditions are referred to the Sample Adult Conditions (ACN) section.

Recodes

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

Technical Notes

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

WRKLYR4 in the ASD section continues to be used to identify the universe for WKDAYR in the AHS section.

IV. Adult Health Behaviors Section (AHB)

The AHB section of the NHIS Sample Adult questionnaire contains questions related to cigarette smoking, leisure-time physical activity, alcohol use, height, weight, and sleep. The same questions have been included in the Sample Adult core every year since 1997, with the exception of sleep, which was added in 2004. A supplement consisting of six questions related to non-cigarette tobacco use was included in the AHB section in 2012.

Smoking

Lifetime and current cigarette smoking status were assessed with a series of questions about past and current smoking practices. All adults were asked if they had smoked at least 100 cigarettes in their entire life. Those who said "yes" were asked a series of questions about the age at which they began smoking and their current smoking practices (every day, some days, not at all). Every day and some day smokers were asked about the number of cigarettes smoked (on the days that they smoked) and whether they had stopped smoking cigarettes for more than one day in the past year because they were trying to quit. Those who no longer smoked were asked about the length of time since they had quit.

Current smokers are defined as persons who have ever smoked 100 cigarettes and who currently smoke every day or some days. Since 2004, there is only one smoking status recode summarizing lifetime and current cigarette smoking status of adults on the data file (SMKSTAT2), rather than three recodes during data years 1997-2003.

In 2012, six supplemental questions on the use of tobacco products other than cigarettes were included in the AHB section. These other tobacco questions, sponsored by the Food and Drug Administration's Center for Tobacco Products, were designed to estimate overall tobacco use in the population and to complement existing NHIS questions about cigarette smoking. The other tobacco questions explored both smoked tobacco (e.g. cigars, pipes, water pipes or hookahs, very small cigars that look like cigarettes, bidis, or cigarillos) and smokeless tobacco (e.g. chewing tobacco, snuff, dip, snus, or dissolvable tobacco). Respondents were asked if they had used either of these two types of non-cigarette tobacco products even one time. Those who said "yes" were asked if they currently used them every day, some days, rarely, or not at all. Interviewers were instructed to exclude electronic cigarettes ("e-cigarettes") and nicotine replacement therapy products (patch, gum, lozenge, spray) used for smoking cessation treatment. Finally, adults who had smoked at least 100 cigarettes or who had ever used other tobacco products were asked if they were using any kind of tobacco product 12 months earlier. Adults who were using tobacco products 12 months earlier and those who were current tobacco users were asked if they stopped using all kinds of tobacco for more than one day in the past year because they were trying to quit.

For additional information about the history and context of NHIS tobacco use information for adults, see: **http://www.cdc.gov/nchs/nhis/tobacco.htm**.

Leisure-time Physical Activity

The section on leisure-time physical activity is introduced with the following statement: "The next questions are about physical activities (exercise, sports, physically active hobbies...) that you may do in your leisure time." From 1997-2003, the term "leisuretime" was used only in this introductory statement. Beginning in 2004, "leisure-time" was inserted into each of the physical activity questions in the AHB section. In this section, respondents are asked to summarize their *usual* leisure-time physical activity – both in terms of frequency and duration. This requires some mental calculations by the respondent. Responses can be offered in terms of any time unit the respondent volunteers (times per day, per week, per month, or per year). A recode converting all responses into frequency in times per week is provided for each type of activity. The set of leisure-time physical activity questions included every year in the sample adult core module is: frequency and duration of vigorous activities, frequency and duration of light or moderate activities, and frequency of strengthening activities.

For additional information about the history and context of NHIS physical activity information for adults, see: **http://www.cdc.gov/nchs/nhis/physical_activity.htm**.

Alcohol Use

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

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

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

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

Since 2004, the category "former drinker, frequency unknown" (ALCSTAT=4) includes former drinkers for whom information is not available on whether or not they had 12 or more drinks in any one year. Previously, in ALC7STAT, this category of former drinker was combined with "unknown drinking status" (ALC7STAT=9). ALCSTAT can be created by the data user relatively easily for data years in which both ALCSTAT1 and ALC7STAT appear (2001-2003). Creating ALCSTAT for data years prior to 2001 can be done, but the coding is quite complex.

A documentation error that occurred for ALC7STAT (2001-2003) and ALCSTAT (2004-2008) has been corrected. Prior to 2009, the "notes" section of the documentation for these variables erroneously indicated that the definition of a "current drinker" included "12+ drinks in lifetime and 12+ drinks in 1 year." The correct definition of a current drinker is someone who had had 12+ drinks in their lifetime and at least one drink in the past year. Prior years' documentation will not be corrected.

Body Weight and Height

Sample adults were asked their current height and weight. **No physical measurements were taken.** National estimates based on physical measurements, such as those available from NCHS' National Health and Nutrition Examination Survey (NHANES), may differ from those available from the NHIS, which are self-reported.

Since 1997, the individual values of height for men were limited to 63-76 inches and the individual values of height for women were limited to 59-70 inches. Similarly since 1997, individual values of body weight were limited. However, because of the increasing weight of the US population, weight limits were revised in 2006. Since 2006, individual values of body weight for men were limited to 126-299 pounds and individual values of body weight for women were limited to 100-274 pounds. In cases where reported values were outside the limits for either height or weight, the data for both variables were changed to "96" or "996" ("Not available") on the public use data file. This was done in order to protect the confidentiality of NHIS respondents who might be identifiable by their unusual physical characteristics. For details concerning these and other changes, especially regarding body weight, please review the section entitled "Body Weight and Height" within the AHB section of the 2006 NHIS Survey Description Document and Appendix VIII.

Respondents have the option of reporting their height and weight in either U.S. Customary (lbs/oz; ft/in) or metric (kg; m/cm) format. Less than 1% of respondents reported in metric format. Metric responses on height and weight were converted into U.S. Customary format for inclusion on the data file. Since 2006, the factor used to convert metric values from centimeters to inches has been expanded to 2.54 for greater precision. The conversion factor was rounded to 2.5 during 1997-2005. For the earlier data years, estimates of height in feet and inches will be slight overestimates for respondents who initially reported their height in meters and/or centimeters (e.g., the number of such respondents was 216, less than 1% of adults in 2005).

Body Mass Index (BMI), a measure of body weight relative to height, was calculated using the formula: BMI = kilograms /meters². Kilograms and meters were derived from (U.S. Customary) pounds and inches using the following factors: 1 kilogram=2.20462 pounds; 1 meter=39.37008 inches. BMI was calculated for all persons who provided height and weight, including those for whom specific height and weight values were changed to "96" and "996" (not available) on the public use file for reasons of confidentiality. BMI variable values are released as 4 digit numbers with two decimal places implied. For example, a value of 2587 for the BMI variable indicates a 25.87 BMI.

The following classification of body weight status for both men and women, established by the World Health Organization, is used in the NHIS data files: underweight (BMI < 18.5); healthy weight (18.5 ≤ BMI < 25); overweight, but not obese (25 ≤ BMI < 30); overweight, including obese (BMI ≥ 25); and obese (BMI ≥ 30).

Although extreme values of height and weight are not publicly released due to confidentiality concerns, use of that information can be made through the NCHS Research Data Centers (RDCs). Contact the RDC for more information, or visit their Web page at: **http://www.cdc.gov/rdc/**.

Beginning in 2008, an internal consistency check for the height and weight variables was added to the survey instrument to improve data quality. Extreme values for these variables triggered a request for interviewer verification of data entry and re-asking height and weight questions, if appropriate. In addition, body mass index (BMI) was calculated within the instrument, with extreme values also triggering interviewer verification. These consistency checks were solely within the survey instrument and are not reflected in the published questionnaire, documentation or data file.

Sleep

A question asking about usual number of hours of sleep, first introduced in the Sample Adult Core in 2004, continues in 2012. Prior to 2004, a question about sleep was last asked in the NHIS in 1990 as part of the Health Promotion and Disease Prevention Supplement.

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

The core Adult Health Care Access and Utilization (AAU) section of the 2012 NHIS has remained largely unchanged since 1997 and contains information on access to health care, dental care, health care provider contacts, and immunizations**.**

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

Respondents were asked about health care provider contacts, including questions about doctor contacts during the past 12 months. Doctor visit probe questions allow for visits not only from medical doctors but from a variety of other health care professionals, including chiropractors. Questions about home care are included as well as a question asking about the number of visits to a hospital emergency room in the past 12 months. There is also a question that asks how long it has been since the respondent has seen or talked to a doctor.

There are several supplementary questions related to adult immunizations: flu shot and nasal spray flu vaccine, including month and year received; pneumonia vaccine; hepatitis B vaccine and hepatitis A vaccine, including number of doses; Zoster or Shingles vaccine; and tetanus shot, including if it was given in 2005 or later and whether it included the pertussis or whooping cough vaccine.

There were no changes to the flu vaccination questions in the AAU section in 2012. The questions were the same as those used in 2011 and the latter part of 2010. However, analysts interested in trend analyses should note that, from January-August 2010, the NHIS included additional supplemental questions about the H1N1 flu vaccine, which was offered separately from the seasonal flu vaccine during the 2009-2010 flu season. Information regarding the 2010 flu vaccination data can be found in the AAU Appendix on the NHIS website, **ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/samadult_layout.pdf** .

Additional supplementary questions inquire whether adult respondents ever had chickenpox and if it had been in the past 12 months; ever had hepatitis, ever lived with someone with hepatitis; ever told they had a chronic or long-term liver condition; and ever traveled outside the United States to countries other than Europe, Japan, Australia, New Zealand, or Canada since 1995.

Lastly, beginning in 2010, two supplemental questions were included that ask adult respondents if they currently volunteer or work in a hospital, clinic, doctor's office, dentist's office, nursing home or other health care facility and if they provide direct patient care.

Beginning in 2011, supplementary extended access and utilization questions were embedded throughout this section. Included were questions about reasons for not having a usual source of care, difficulty finding providers, monetary burden of care, emergency room visits and reasons for use, use of prescription medication, use of health information technology (HIT), and preventive services such as colon cancer screening, mammography, and cholesterol screening. In addition, questions were included that addressed long term care and experience with purchasing health insurance directly, not through an employer. A list of the supplemental question names and numbers can be found in the "Supplements, Supplement Co-sponsoring Agencies, and Question Locations" section within this document.

Also in 2011, four supplemental questions about the human papillomavirus (HPV) for adults aged 18-64 were moved from the 2010 NAF Cancer section to the AAU section. The variable name, section name and number for SHHPVHRD (NAF.228_03.000) changed to SHHPVHD1 (AAU.444_00.010) and SHTHPV (NAF.228_04.000) changed to SHTHPV1 (AAU.446_00.010). The section name and number for HPVHRD changed from NAF.228_01.000

to AAU.442_00.010 and SHHPVDOS changed from NAF.228_05.000 to AAU.448_00.010. In 2012, the HPV questions remained the same as in 2011.

In addition, a question that asked adult respondents about receiving an HIV test was moved from the AIDS section to this section in 2011. The variable name, section name and number were changed from HIVTST (AIDS.040) to HIVTST1 (AAU.700). In 2012 the HIV question remained the same as in 2011.

Beginning in 2011 and continuing in Quarter 1 of 2012, a series of questions were included that asked respondents about emergency room (ER) visits which did not result in a hospital admission. In Quarters 2-4 of 2012, the universe for these questions was modified to include hospital admissions resulting from an ER visit. Because of this change, the eight original variables, AERREAS1-AERREAS8, were renamed AERREA1R-AERREA8R. The final data file contains both the original and the new variables for all four quarters of 2012. However, the original variables contain data only for Quarter 1 and are blank for Quarters 2-4. Similarly, the new variables contain data only for Quarters 2-4 and are blank for Quarter 1. Similar changes were made to the CAU section for the sample child. For more information see the AAU and CAU Appendices which are located in the 2012 NHIS, Data Release, Sample Adult and Sample Child files, Variable layout found at:
http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm .

In 2012, the variable ALASTTYP was modified in two ways. An additional response category was added to the variable (Nurse practitioner /Physician assistant) and the variable now allows the respondent to give more than one answer. Therefore, the ALASTTYP variable has been exploded into four variables (ALASTYP1 through ALASTYP4) based on the four response categories and the source variable ALASTTYP does not appear on the file. In addition, a supplemental variable ALASTSPC was added which allowed for a verbatim response when the respondent answered "someone else" in ALASTTYP (ALASTYP4). The verbatim responses in ALASTSPC were back coded to an appropriate category ALASTYP1 – ALASTYP3 when possible. However, those that could not be back coded were cleaned for confidentiality and a recode (ALASTVRB) was created. ALASTSPC is not on the final data file.

Also in 2012, two new supplemental questions (FLUSHPG1 and FLUSHPG2) that asked female respondents 18-49 about receipt of a flu shot before, during, or after pregnancy based on month of interview were added to this section.

In addition, in 2012, the following variables were renamed: AVISAPTN changed to AVISAPN2 because a new instruction was added to the question allowing for an additional response category; AVISAPTT changed to AVISAPT2 because the universe changed; SHTTDAP changed to SHTTDAP2 because the universe changed and the question text was modified; and WRKHLTH changed to WRKHLT2 because the question text was modified. For more information, see the notes for the above mentioned variables in the AAU section of the 2012 NHIS, Data Release, Sample Adult File, Variable layout found at:
http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm .

Lastly, the 2011 supplemental questions LTCPRCH, LTCPAY, and LTC100M were deleted from the AAU section in 2012.

Technical Notes

Analysts are advised to read the notes in the Dataset Documentation for further information pertaining to any changes that may have occurred and to compare the 2012 Dataset Documentation to documentation from the 2011 (and earlier) NHIS for any other changes that may have occurred over time to the variables in this section.

VI. Adult Voice, Speech and Language Supplement: Adult Communication Disorders (ACD)

The Adult Voice, Speech and Language Supplement: Adult Communication Disorders (ACD), included in the 2012 NHIS, is the first-ever large scale communication supplement in the survey. The ACD Supplement encompasses the domains of voice, swallowing, speech, and language problems for sample adults age 18 and older. Data about the causes, age of onset, severity, and healthcare utilization are included for each domain. Further, detailed information on healthcare utilization as a result of the voice, swallowing, speech or language problem is available in ACD.

A companion Child Voice, Speech and Language Supplement: Child Communication Disorders (CCD) is available on the Sample Child File.

VII. Adult Internet and Email Usage Section (AWB)

The Adult Internet and Email Usage Section (AWB) is the first supplement on the NHIS to ascertain information on internet and email usage for sample adults aged 18 years and over. This section includes data on whether or not the adult uses the internet or email and the frequency at which they are used.

2012 National Health Interview Survey
Adult Functioning and Disability Supplement File (AFD)

The Adult Functioning and Disability Supplement (AFD) fielded with the 2012 NHIS Sample Adult module is part of an international project to develop and test improved measures of functioning. The project is part of a partnership with the Washington Group on Disability Statistics (WG) and the Budapest Initiative on the Measurement of Health State (BI). More information on the WG, the short and extended set of disability questions developed by this group, and testing efforts to date can be found on the NCHS website at **http://www.cdc.gov/nchs/washington_group.htm**. A set of questions developed by both the WG and BI was included as a supplement (the Disability Questions Tests) in the NHIS from October 2008 through December 2012. The questions included in the 2010, 2011, and 2012 NHIS reflect changes to that set as a result of cognitive testing and analysis of the 2008/2009 questions and data. In 2010, these questions were in the Quality of Life (QOL) Supplement file; in 2011 and 2012, they are in the Adult Functioning and Disability Supplement file. As development and testing of the questions is ongoing, modifications to the questions used in the NHIS will also continue. Check **http://www.cdc.gov/nchs/nhis/nhis_2012_data_release.htm** for additional information and updates.

Approximately one quarter of sample adults were randomly selected to receive the 2012 AFD Supplement. As a result of the selection process, different weights were generated for these respondents (WTFA_AFD). In addition, a separate, stand-alone public use file was created for the AFD variables, rather than appending these variables to the 2012 Sample Adult file. The file itself contains information about a respondent's functioning in various basic and complex activity domains: vision (difficulty seeing), hearing (difficulty hearing), mobility (difficulty walking, climbing steps, or moving around), communication (difficulty communicating), cognition (difficulty remembering or concentrating), upper body (difficulty with self-care), affect (feelings of being worried, nervous, or anxious; feelings of being depressed), pain, and fatigue (feelings of being very tired or exhausted). The file also contains a set of questions designed to capture an individual's ability to participate in society – a measure of quality of life. Follow-up questions on the degree of difficulty, use of assistive devices, and functioning with assistance were included for most domains. However, the 2012 AFD file does not contain as many follow-up questions as the 2010 QOL file, nor does it contain the two verbatim fields that were available in the 2010 QOL file.

Although approximately one quarter of all sample adults were to be asked the questions in the AFD supplement, some persons did not complete the section. Persons who did not give responses to any of the questions in the section were given a coded value of "1" on the record completion status variable in the data file (variable RCS_AFD), which indicates that their record only contains responses of "not ascertained." These respondents are retained in the file, but they are coded as "8" in all remaining relevant fields of the AFD file. Adults who met the criteria for completing the section were coded as either "2" (all answers were "refused" or "don't know"), "3" (all answers were "refused," "don't know," or "not ascertained"), or "4" (at least one valid answer) on RCS_AFD. Over 98% of AFD respondents are coded "4" on RCS_AFD.

Many NHIS analysts will want to produce estimates and perform comparisons within key subgroups such as age, sex, and race/ethnicity. This requires merging the Adult Functioning and

Disability File with one or more NHIS data files (e.g., Sample Adult, Person, etc.). Information on merging data files can be found in Appendix VI of this document. The weight WTFA_AFD provided with the AFD Supplement file is designed to produce annual-level estimates calculated based on data included in this file.

Technical Notes

Users interested in combining years of data should be aware that the 2012 AFD instrument is not identical to the 2011 AFD instrument. For the 2012 instrument, a new question (VIS_0) has been added to the vision sequence, and the order of questions in the hearing sequence has been changed. As a result, the vision questions and some of the hearing questions that were retained in 2012 have different contexts or universes relative to 2011, so these questions have been renamed. In addition, the wording for several of the questions in the mobility sequence has been changed, resulting in new variable names. Some of the changes in question wording are minor (e.g., "any" was removed from MOB_SS in the 2011 AFD section, and the resulting question renamed MOB_SS2 in 2012). Lastly, the universes for ANX_3 and DEP_3 in 2011 were modified in 2012 – hence, these questions have been renamed ANX_3R and DEP_3R, respectively. Users wishing to combine years of data should read the notes in the Variable Layout file in order to more completely understand the changes in 2012 variables relative to variables in the 2011 AFD file and in the 2010 QOL file.

2012 National Health Interview Survey
Adult Complementary and Alternative Medicine File (ALT)

In 2012 the Adult Complementary and Alternative Medicine Supplement (ALT) collected information from sample adults on their use of 18 non-conventional health care practices. The modalities covered in the supplement include acupuncture, ayurveda, biofeedback, chelation therapy, chiropractic or osteopathic manipulation, craniosacral therapy, energy healing therapy, hypnosis, massage, naturopathy, traditional healers, movement therapies (Pilates/Trager psychophysical integration/ Feldenkrais), herbal and non-vitamin supplements, vitamins and minerals, homeopathy, special diets, yoga/tai chi/qi gong, and relaxation techniques (meditation/guided imagery/progressive relaxation). Earlier, similar versions of the ALT were fielded in 2002 and 2007, but some changes have been made with each iteration of the supplement. Therefore, not all 2012 ALT supplement content is comparable to that of the earlier versions.

For each health care practice covered on the ALT supplement, questions were asked about the following: whether the respondent ever used the modality, and, if so, whether the modality was used in the past 12 months; number of times the respondent had seen a practitioner for the modality; portion of the cost covered by insurance for the modality; amount paid out-of-pocket for the modality; and additional materials purchased to learn about the modality.

While some questions were asked of each health care practice, others were asked only for the top 3 modalities deemed by the respondent to be most important to their health. For the top 3 question series all modalities were included except for ayurveda, chelation therapy, and vitamins and minerals. These three therapies were excluded due to either very low or high prevalence. The topics covered in the top 3 series include: reasons for using the modality; whether the modality motivated the respondent to engage in other selected health behaviors; outcomes associated with using the modality; whether the modality was used to treat a specific health problem or condition, and, if so, what health problems or conditions were treated and for which one of the health problems or conditions the modality was used the most; whether the respondent received any of a specified set of traditional medical treatments for the condition treated most by the modality, and, if so, when the conventional medical treatments were received in relation to when the modality was used; and disclosure of modality use to a specified set of conventional medical professionals. Given the number of modalities and questions, a separate stand-alone data file was created for the supplement rather than append it to the 2012 Sample Adult file. The Sample Adult weight is included on the ALT file and should be used for all analyses with the variables on this file.

Although all sample adults were to be asked the questions in the ALT supplement, some persons did not complete the supplement. Persons who did not give responses to any of the questions in the supplement had a value of "not ascertained" inserted in the appropriate fields in this file. In other words, these respondents (n=931 or 2.7% of sample adults) are retained in the file, but they are coded as "8" in all the relevant fields of the file.

Back Coding

In the ALT supplement, respondents were given the opportunity to choose "Other specify" as a response to the questions about what health problems or conditions were treated with specific types of complementary and alternative medicine (CAM): TP1CND86, TP2CND86, TP3CND86. If "Other specify" was selected, the respondent was then asked to give a verbatim response indicating the health problem or condition treated with the specified type of CAM (TP1_SPEC, TP2_SPEC, TP3_SPEC). As part of the editing process of the ALT file, verbatim responses were back coded into existing categories (TP1CND01-TP1CND85; TP2CND01-TP2CND85; TP3CND01-TP3CND85) or into an *ad hoc* "not elsewhere classified" category (TP1CND87, TP2CND87, TP3CND87). One additional *ad hoc* response category (any cardiovascular condition: TP1CND88, TP2CND88, TP3CND88) was added to the existing health problem or condition list for adults, for each top 3 therapy. In all cases, the original verbatim text was retained in the original variables. Note that the original verbatim responses are excluded from the public use file in order to protect the confidentiality of the respondents.

The frequency counts for the questions about the condition treated most often with each CAM therapy have not been changed to reflect back coding the "other specify" responses. While back coding the condition verbatim responses it became apparent that some respondents were not using the selected CAM therapy to treat a specific condition (e.g., someone who received a massage at a spa as a gift). Therefore, in such cases, the values may have been changed for the following variables: TP1_TRET, TP2_TRET, and TP3_TRET.

Cost Variables

Responses to questions on the ALT Supplement about out-of-pocket cost of therapies and materials are reported exactly as reported by respondents, with the exception of upper limits placed on the cost variables (maximum value varies by question – see layout file for upper limit for each variable). Given the 12 month recall period for these questions, values may be subject to inherent memory limitations. Data users should be aware that these limitations may affect the data, although such effects have not been evaluated.

Technical notes

The sample adult weight, WTFA_SA, provided on the ALT file is designed to produce annual-level estimates calculated based on data included in this file. In order to increase the number of variables available for analysis with the ALT file variables, the ALT file may be merged with one or more NHIS data files (e.g., Sample Adult, Person, etc.) for a given individual. Information on merging data files can be found in Appendix VI of this document.

CAUTION: The ALT use variables should be analyzed with care. Many of the ALT use variables are asked of only a subset of the sample adults and therefore should not be used to estimate prevalence for the health care practices they represent. In order to obtain accurate prevalence estimates for use of health care practices, data must be weighted using WTFA_SA and recodes are required to appropriately account for each variable's universe. For specific information about who was asked each question, refer to the ALT Variable Layout file.

2012 National Health Interview Survey
Coverage Section: Telephone Questions

The purpose of the cellular telephone questions is to track the use of wireless telephones in American families over time, allowing researchers to analyze the demographic characteristics of families who have substituted wireless service for landline home telephones. Having these data from a large population-based survey such as the NHIS provides useful information about potential bias from undercoverage in random-digit-dial telephone surveys that use only landline telephone numbers in their sampling frames.

In 2007, the cellular telephone questions were modified. In 2003-2006, all cellular telephone questions resided in the Recontact Section (REC) at the end of the survey; in 2007, all cellular telephone questions were moved to the Coverage Section (COV.330 – COV.337) at the beginning of the survey. Appendix VIII of the 2007 Survey Description document lists the cellular telephone variable changes.

From 2008-2012, no changes were made to the telephone questions.

Guidelines for Citation of Data Source

With the goal of mutual benefit, the National Center for Health Statistics (NCHS) requests that recipients of NHIS data files cooperate in certain actions related to their use. Any published material derived from the 2012 data should acknowledge "CDC/NCHS, National Health Interview Survey" as the original source. The suggested citation to appear at the bottom of all tables and graphs is as follows:

Data Source: CDC/NCHS, National Health Interview Survey, 2012

In a bibliography, the suggested citation should read:

National Center for Health Statistics. *Data File Documentation, National Health Interview Survey, 2012 (machine readable data file and documentation)*. National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland. 2013.

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.

NHIS questionnaires are in the public domain and no permission is required to use them. Citation as to source, however, is appreciated.

Information on how to cite electronic media is available at:
http://www.cdc.gov/nchs/products/citing_electronic_media.htm.

References

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

Botman SL, Moore TF, Moriarity CL, and Parsons VL. *Design and estimation for the National Health Interview Survey, 1995–2004*. National Center for Health Statistics. Vital Health Stat 2(130). 2000.

DeNavas-Walt C, Proctor BD, Smith JC. U.S. Census Bureau. Current Population Reports, P60-243, *Income, Poverty, and Health Insurance Coverage in the United States: 2011*. U.S. Government Printing Office, Washington, DC. 2012.

National Center for Health Statistics. *Health, United States, 2012: With Special Feature on Emergency Care*. Hyattsville, Maryland. 2013.

Research Triangle Institute. SUDAAN *Language Manual; Release 10.0*, Research Triangle Institute, Research Triangle Park, NC. 2008.

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

Warner M, Schenker N, Heinen MA, Fingerhut LA. The effects of recall on reporting injury and poisoning episodes in the National Health Interview Survey. *Injury Prevention*, 11(5):282-7. 2005.

Appendix I
Calculation of Response Rates for the 2012 NHIS

The response rates calculated here pertain to the Core questions in the 2012 NHIS. Note that the categories "interviewed households," "interviewed families," "interviewed sample children," and "interviewed sample adults" include those with completed interviews or acceptable partial interviews.

Household Response Rate

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

The Household Response Rate is calculated by dividing the number of interviewed households by the sum of the number of responding households and the number of Type A non-response households. Type A non-response households are households that were not interviewed for a variety of reasons: language problems, no one 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. NHIS includes all Type A non-response households in the Household Response Rate calculation, although a small number of Type A non-response households are ineligible for the survey because of the "screening" process. If the ineligible Type A households were omitted from the Household Response Rate calculation, the rate would increase slightly (less than one percent). See Appendix III for information about the NHIS screening process.

Conditional Family Response Rate

$$\frac{(\text{Interviewed Families})}{(\text{Eligible Families from Interviewed Households})}$$

Family Core data were collected from the respondent about all persons in the family. The response rates for the Family Core can be calculated in two ways: conditionally and finally. The Conditional Family Response Rate is the rate only for those families identified as eligible and <u>does not take into account household non-response</u>. The Conditional Family Response Rate is calculated by dividing the number of interviewed 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.

Final Family Response Rate

$$\frac{(\text{Interviewed Families})}{(\text{Eligible Families from Interviewed Households})}(\text{Household Response Rate})$$

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

Conditional Sample Child Response Rate

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

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

Final Sample Child Response Rate

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

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

Conditional Sample Adult Response Rate

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

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

Final Sample Adult Response Rate

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

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

Appendix I, Table 1A. Conditional Response Rates, NHIS 1997-2012

Survey year	Household module (same as below)	Family module	Sample Child module	Sample Adult module
1997	91.8	98.4	93.1	89.0
1998	90.0	98.0	93.3	83.8
1999	87.6	98.3	90.8	80.8
2000	88.9	98.2	90.9	82.6
2001	88.9	98.5	92.0	84.2
2002	89.6	98.3	92.3	84.4
2003	89.2	98.5	92.3	84.5
2004	86.9	99.6	91.8	83.8
2005	86.5	99.5	90.1	80.1
2006	87.3	99.6	90.6	81.4
2007	87.1	99.4	88.4	78.3
2008	84.9	99.5	85.6	74.2
2009	82.2	99.3	89.9	80.1
2010	79.5	99.1	89.8	77.3
2011	82.0	99.2	91.8	81.6
2012	77.6	99.0	90.7	79.7

Appendix I, Table 1B. Unconditional Response Rates, NHIS 1997-2012

Survey year	Household module (same as above)	Family module	Sample Child module	Sample Adult module
1997	91.8	90.3	84.1	80.4
1998	90.0	88.2	82.4	73.9
1999	87.6	86.1	78.2	69.6
2000	88.9	87.3	79.4	72.1
2001	88.9	87.6	80.6	73.8
2002	89.6	88.1	81.3	74.3
2003	89.2	87.9	81.1	74.2
2004	86.9	86.5	79.4	72.5
2005	86.5	86.1	77.5	69.0
2006	87.3	87.0	78.8	70.8
2007	87.1	86.6	76.5	67.8
2008	84.9	84.5	72.3	62.6
2009	82.2	81.6	73.4	65.4
2010	79.5	78.7	70.7	60.8
2011	82.0	81.3	74.6	66.3
2012	77.6	76.8	69.7	61.2

Calculation of Response Rates for Combined NHIS Data Years

The response rates for combined NHIS data years are calculated in the same basic way as for a single year. The following examples are shown for two years of data. Similar methods apply for multiple years of data in 1997 and beyond.

Household Response Rate for Combined Data Years

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

The Household Response Rate for Combined Data Years is calculated by dividing the number of interviewed households for Years 1 and 2 by the sum of the number of interviewed households and the number of Type A non-response households for the survey for Years 1 and 2. Type A non-response households are households that were not interviewed for a variety of reasons: language problems, no one 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. NHIS includes all Type A non-response households in the Household Response Rate calculation, although a small number of Type A non-response households are ineligible for the survey because of the "screening" process. If the ineligible Type A households were omitted from the Household Response Rate calculation, the rate would increase slightly (less than one percent). See Appendix III for information about the NHIS screening process.

Conditional Family Response Rate for Combined Data Years

$$\frac{(\text{Interviewed Families for Years 1 \& 2})}{(\text{Eligible Families from Interviewed Households for Years 1 \& 2})}$$

Family Core data were collected from the respondent about all persons in the family. The response rates for the Family Core can be calculated in two ways: conditionally and finally. The Conditional Family Response Rate is the rate only for those families identified as eligible and does not take into account household non-response. The Conditional Family Response Rate for Combined Data Years is calculated by dividing the number of interviewed families for Years 1 and 2 by the number of families that are eligible for the survey in Years 1 and 2, that is, from interviewed households for Years 1 and 2. Note that a household can have multiple families, and rejected families are families that were deleted from interviewed households because of insufficient data.

Final Family Response Rate for Combined Data Years

$$\frac{(\text{Interviewed Families for Years 1 \& 2})}{(\text{Eligible Families from Interviewed Households for Years 1 \& 2})}\left(\begin{array}{l}\text{Household Response}\\\text{Rate for Years 1 \& 2}\end{array}\right)$$

The Final Family Response Rate is the rate for those families identified as eligible that takes into account household non-response. The Final Family Response Rate for Combined Data Years is calculated by dividing the number of interviewed families for Years 1 and 2 by the number of families that are eligible for the survey in Years 1 and 2, that is, from interviewed households for Years 1 and 2, and then multiplying this quotient by the Household Response Rate for Combined Data Years.

Conditional Sample Child Response Rate for Combined Data Years

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

The response rates for the Sample Child section can be calculated in two ways: conditionally and finally. The Conditional Sample Child Response Rate is the rate only for sample children and does not take into account household or family non-response. The Conditional Sample Child Response Rate for Combined Data Years is calculated by dividing the number of interviewed sample children for Years 1 and 2 by the number of eligible sample children from interviewed families for Years 1 and 2.

Final Sample Child Response Rate for Combined Data Years

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

The Final Sample Child Response Rate is the rate for sample children that takes into account household and family non-response. The Final Sample Child Response Rate for Combined Data Years is calculated by dividing the number of interviewed sample children for Years 1 and 2 by the number of eligible sample children from interviewed families for Years 1 and 2, and then multiplying this quotient by the Final Family Response Rate for Combined Data Years.

Conditional Sample Adult Response Rate for Combined Data Years

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

The response rates for the Sample Adult section can be calculated in two ways: conditionally and finally. The Conditional Sample Adult Response Rate is the rate only for those sample adults identified as eligible and <u>does not take into account household or family non-response</u>. The Conditional Sample Adult Response Rate for Combined Data Years is calculated by dividing the number of interviewed sample adults for Years 1 and 2 by the number of eligible sample adults from interviewed families for Years 1 and 2.

Final Sample Adult Response Rate for Combined Data Years

$$\frac{(\text{Interviewed Sample Adults for Years 1 \& 2})}{(\text{Eligible Sample Adults from Interviewed Families for Years 1 \& 2})}\left(\frac{\text{Final Family Response}}{\text{Rate for Years 1 \& 2}}\right)$$

The Final Sample Adult Response Rate is the rate for those sample adults identified as eligible that <u>takes into account household and family non-response</u>. The Final Sample Adult Response Rate for Combined Data Years is calculated by dividing the number of interviewed sample adults for Years 1 and 2 by the number of eligible sample adults from interviewed families for Years 1 and 2, and then multiplying this quotient by the Final Family Response Rate for Combined Data Years.

Appendix I, Table 2. Number Eligible/Interviewed, 2012 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	54,603*	42,366
Family / families	43,785	43,345
Sample Child / persons	14,637	13,275
Sample Adult / persons	43,323	34,525

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix I, Table 3. Number Eligible/Interviewed, 2011 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	48,200*	39,509
Family / families	40,829	40,496
Sample Child / persons	13,998	12,850
Sample Adult / persons	40,470	33,014

Appendix I, Table 4. Number Eligible/Interviewed, 2010 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	43,208*	34,329
Family / families	35,509	35,177
Sample Child / persons	12,557	11,277
Sample Adult / persons	35,153	27,157

Appendix I, Table 5. Number Eligible/Interviewed, 2009 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	41,177*	33,856
Family / families	34,899	34,640
Sample Child / persons	12,404	11,156
Sample Adult / persons	34,616	27,731

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix I, Table 6. Number Eligible/Interviewed, 2008 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	33,911*	28,790
Family / families	29,569	29,421
Sample Child / persons	10,303	8,815
Sample Adult / persons	29,370	21,781

Appendix I, Table 7. Number Eligible/Interviewed, 2007 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	33,615*	29,266
Family / families	30,081	29,915
Sample Child / persons	10,658	9,417
Sample Adult / persons	29,875	23,393

Appendix I, Table 8. Number Eligible/Interviewed, 2006 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	33,468*	29,204
Family / families	29,974	29,868
Sample Child / persons	10,853	9,837
Sample Adult / persons	29,825	24,275

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix I, Table 9. Number Eligible/Interviewed, 2005 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	44,540*	38,509
Family / families	39,463	39,284
Sample Child / persons	13,906	12,523
Sample Adult / persons	39,227	31,428

Appendix I, Table 10. Number Eligible/Interviewed, 2004 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	42,089*	36,579
Family / families	37,653	37,466
Sample Child / persons	13,538	12,424
Sample Adult / persons	37,388	31,326

Appendix I, Table 11. Number Eligible/Interviewed, 2003 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	40,266*	35,921
Family / families	37,126	36,573
Sample Child / persons	13,275	12,249
Immunization /persons	13,275	11,665
Sample Adult / persons	36,524	30,852

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix I, Table 12. Number Eligible/Interviewed, 2002 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	40,377*	36,161
Family / families	37,458	36,831
Sample Child / persons	13,570	12,524
Immunization /persons	13,865	13,611
Sample Adult / persons	36,787	31,044

Appendix I, Table 13. Number Eligible/Interviewed, 2001 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	43,797*	38,932
Family / families	40,227	39,633
Sample Child / persons	14,766	13,579
Immunization /persons	15,000	14,709
Sample Adult / persons	39,564	33,326

Appendix I, Table 14. Number Eligible/Interviewed, 2000 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	43,437*	38,633
Family / families	39,998	39,264
Sample Child / persons	14,711	13,376
Immunization /persons	14,890	14,618
Sample Adult / persons	39,201	32,374

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix I, Table 15. Number Eligible/Interviewed, 1999 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	42,882*	37,573
Family / families	38,845	38,171
Sample Child / persons	14,217	12,910
Immunization /persons	14,178	13,881
Sample Adult / persons	38,117	30,801

Appendix I, Table 16. Number Eligible/Interviewed, 1998 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	42,440*	38,209
Family / families	39,559	38,773
Sample Child / persons	14,619	13,645
Prevention Sample Child /persons	13,645	13,610
Immunization /persons	15,041	14,775
Sample Adult / persons	38,729	32,440
Prevention Sample Adult	32,440	31,882

Appendix I, Table 17. Number Eligible/Interviewed, 1997 NHIS

File / Type of Records	Eligible	Interviewed
Household / households	43,370*	39,832
Family / families	41,291	40,623
Sample Child / persons	15,244	14,290
Immunization /persons	15,558	15,402
Sample Adult / persons	40,552	36,116

*Includes a small number of Type A non-response households that are ineligible for the survey. See description of Household Response rate earlier in this appendix.

Appendix II
Race and Hispanic Origin in the 2012 NHIS

Background

For nearly 30 years, the National Health Interview Survey (NHIS) had collected information on the race and Hispanic origin or 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 relied on respondents to provide self-identified race and ethnicity information (proxy information is reported for children and non-present household members), although interviewer-observed race was also recorded through 1996, the last year of the paper questionnaire. NHIS data are routinely tabulated by race and ethnicity in NCHS publications such as Summary Health Statistics, Health U.S., and National Health Statistics Reports.

In response to the changing demographics of the U.S. population, the OMB revised Directive 15 in 1997, after an extensive period of research and public commentary. The revised race and ethnicity standards allow respondents to the Census and federal surveys to indicate more than one race group in answering questions on race. A complete description of the revised OMB guidelines on the collection of racial and ethnic data, including descriptions of the race categories, the ordering of race and ethnicity questions, and guidelines for the tabulation and publication of data under the new standards can be found on the OMB website: **http://www.whitehouse.gov/omb/inforeg_statpolicy#dr.** In accordance with this requirement, the NHIS became fully compliant with the new race and ethnicity standards with the fielding of the 1999 questionnaire, although the NHIS had been following some aspects of the new guidelines for many years. This policy was expected to be fully implemented across the federal statistical system by the 2003 calendar year.

As noted previously, the U.S. Census Bureau is the data collection agent for the NHIS, as it is for a number of other federal surveys. The Census Bureau also provides the control totals for race/ethnicity (along with sex and age) that are used in the post-stratification adjustment of the person weights in the NHIS data file. In order to maintain consistency with the Census Bureau procedures for collecting and editing data on race and ethnicity, the NHIS made major changes to its editing procedures in the 2003 data year. Beginning in the 2003 NHIS, "Other race" was no longer available as a separate race response. This response category was treated as missing, and the race was imputed if this was the only race response. In cases where "Other race" was mentioned along with one or more OMB race groups (defined as White, Black, American Indian/Alaska Native (AIAN), Asian, and Native Hawaiian and Other Pacific Islander (NHOPI)), the "Other race" response was dropped, and the OMB race group information was retained. These editing changes are consistent with the procedures that the Census Bureau uses to create the Modified Race Data Summary File, which is the data file that provides the population control totals used in weighting the NHIS data. More information about the Modified Race Data Summary File and the editing procedures used to create it, can be found at the following Website: **http://www.census.gov/popest/research/modified.html** . These editing procedures remain in effect for the 2012 data file. Please refer to the 2012 Variable Layout Report for more information.

Race and Hispanic Origin Questions in the National Health Interview Survey

The 2012 NHIS included two questions about Hispanic Origin:

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

"Please give me the number of the group that represents {your/NAME's} Hispanic origin or ancestry. You may choose up to five (5) if applicable." [HHC.180; response categories shown to the respondent on a flashcard].

There were no changes in the wording of the 2012 Hispanic origin question, but some responses were imputed, and the variable name is labeled as HISPAN_I to indicate this fact (see section on the imputation of race and ethnicity later in this document).

The 2012 NHIS also included two questions to obtain information on a respondent's race:

"What race or races {do you/does NAME} consider {yourself/herself/himself} to be? Please select one or more of these categories." [HHC.200; response categories shown to the respondent on a flashcard], and

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

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

2012 Race and Hispanic Origin Variables

The following table (Appendix II, Table I) summarizes the Hispanic origin and race variables in the 2012 data file. Details on the specific response categories for the race questions and additional details on these variables can be found in the 2012 public use Variable Layout Report, and users are strongly urged to read these descriptions carefully to determine how and when the variables should be used in analysis. Data users are also encouraged to check the Variable Frequency Report to examine the unweighted data for these variables before computing weighted estimates.

Appendix II, Table 1. 2012 NHIS Race/Ethnicity Variable Names and Description

2012 Variable Name	Description
ORIGIN_I	Hispanic origin/ancestry with imputed values for some records
ORIGIMPT	Hispanic origin imputation flag
HISPAN_I	Type of Hispanic origin/ancestry with imputed values for some records
HISPIMPT	Type of Hispanic origin imputation flag
RACERPI2	Contains 4 of 5 OMB race groups; values imputed for some records. Does not include "Other race" category.
MRACRPI2	Detailed race variable; multiple race persons not selecting a primary race group in separate category. Values were imputed for some records. Does not include "Other race" category.
MRACBPI2	See section below on bridging; values were imputed for some records. "Other race" category included for bridging purposes.
RACRECI3	Variable that contains 4 race categories used in post-stratification and weighting. New category added to reflect changes in sample design. Values imputed for some records.
RACEIMP2	Imputation flag for use in determining which cases were imputed for the race variables. New categories added to account for new editing procedures.
HISCODI3	Same categories as RACRECI3, crossed with ORIGIN_I (Hispanic/non-Hispanic); values were imputed for some records.
ERIMPFLG	Summary race/ethnicity imputation flag – indicates that either race or ethnicity or both race and ethnicity were imputed.

Procedures for Imputation of Ethnicity and Race in the NHIS

Prior to the 2000 NHIS, race recodes #1 and #2 were created using a crude imputation method that assigned a race to persons with missing values for the variable MAINRACE. Under these procedures, in the 1996 and earlier NHIS, if an observed race was recorded by the interviewer, it was used to code a race value. If there were no observed race values, all persons with a missing value for MAINRACE who were identified as Hispanic (on the Hispanic origin question) were coded as "White," and those who were identified as non-Hispanic were coded as "Other race." Beginning with the 1997 NHIS, observed race was no longer collected. Therefore, the race imputation procedures for all persons with missing values for the variable MAINRACE in 1997-1999 matched the imputation procedures for earlier years when no observed race values were recorded.

In an effort to improve the quality of data on ethnicity and race in the NHIS, hot-deck imputation of selected race and ethnicity variables was done for the first time in the 2000 NHIS and continued to be used for the 2012 NHIS data. Changes implemented in the 2003 imputation procedures also remain in effect for 2012. Records for persons for whom "Other race" was the only race mentioned were treated as having missing data on race, and were added to the pool of records for which selected race and ethnicity variables were imputed.

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

Stage 1 imputation was used for households in which some persons had missing values and some persons had valid entries for ethnicity and race variables (imputation within households). Stage 1 imputation was based on the hot-deck imputation procedures developed for the 2000 Decennial Census Dress Rehearsal (conducted in 1998), which were adapted to utilize NHIS family relationship variables for imputation of the missing ethnicity and race data. Additional imputation procedures for "Other race" responses were adapted from the Census Bureau's Modified Race Data Summary File editing specifications for use with the NHIS race data.

Stage 2 imputation was used for households in which all persons had missing values for ethnicity and race variables (imputation between households). The specifications obtained from Census which were the basis of Stage 1 imputation did not contain information on the imputation of race and ethnicity between households. Therefore, NCHS developed the specifications for the between-household imputation, using the secondary sampling unit (SSU) as the geographic unit for selecting donors.

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

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

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

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

   Step 4.   Load donors' data to hot decks within each household, and conduct imputation for each donee in the same household. Donees are classified in twenty-six categories based on the relationship of the donees to the Reference Person in the household

(see following section). The allocation sequence of donors for each type of donee is different, depending on the type of the donee, and the relationship between the donor and the donee.

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

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

Step 1.   The imputation was divided into three parts:

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

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

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

Use of Imputation Flags

Since hot-deck imputation procedures have been implemented on the NHIS race and ethnicity data, imputation flags have also been added to the data file. These flags allow data users to keep track of the number of cases for which race and/or ethnicity was imputed by the type of original response. They also provide users with a means of accessing the data in their unimputed form. The flags also provide a mechanism for converting data back to the format in the data files prior to the implementation of imputation in 2000, which is critical for merging data files across survey years and maintaining trends in the data. There are four imputation flags on the 2012 public use data file: ORIGIMPT, HISPIMPT, RACEIMP2, and ERIMPFLG. These flags are described in Table 1 above.

Users who wish to merge across data years or create trend data must recreate the race variables RACERPI2, MRACRPI2, and MRACBPI2 in the format they had in previous years by using the flag <u>RACEIMP2</u>. Sample SAS code for using the imputation flags and merging across data years for the variable RACERPI2 (RACERP_I in 2000-2002 and RACER_P in 1999) is included below (the example uses 1999-2006 NHIS data, but other combinations of data years can be used with the appropriate adaptations to the code).

*** Merge 1999-2006 race variable using public use variables ***;

*** Recode 1999 data ***;

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

*** Code to add imputed responses for 2000-2002 RACERP_I ***;

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

*** Code to add imputed responses for 2003-2006 RACERPI2 ***;

```
        if RACEIMP2 in (1)      then RACP0306=7;    /* refused */
else    if RACEIMP2 in (2)      then RACP0306=8;    /* NA     */
else    if RACEIMP2 in (3)      then RACP0306=9;    /* DK     */
else    if RACEIMP2 in (4 5)    then RACP0306=5;    /* Other races[1] */
else                                 RACP0306=RACERPI2;
```

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

*** Combine 1999-2006 data into a single variable ***;

```
        if RACEPU99  ne .       then RACE9906=RACEPU99;
else    if RACP0002  ne .       then RACE9906=RACP0002;
else                                 RACE9906=RACP0306;
```

Bridging to the Old OMB Standards

The OMB tabulation guidelines for the race and ethnicity standards recognize that the complete transition from the old standards to the 1997 standards will take some time, and that many federal statistical systems have a primary mission to track data trends over time. During this transitional period, known as the "bridge," it has been recommended that data systems tabulate data for publication under the new standards, while also providing a means for data users to bridge the new data back to the old standards. This will allow data users to examine differences, if any, in tabulating the data under the old and new standards, assist in the maintenance of data trends, and allow users to become accustomed to data tabulated under the new standard before the transition is complete. In the NHIS, the second race question (commonly known as the "follow-up question") is used to create the bridge between data collected under the old standards and data collected under the new ones. The 2012 NHIS public use data file

contains one bridge race variable to allow comparisons of 2012 data with data from previous years, and to enable merging the 2012 data with 1997-2011 data.

There was one major change to the race and ethnicity data in the 1999 NHIS (which is also true for 2000-2012) that occurred as a result of the creation of a bridge variable. NCHS confidentiality standards do not permit NCHS to release data that might lead to the inadvertent identification of individual respondents to the survey. Beginning with the 1999 survey (and continuing in 2012), data on "Asian" persons and "Native Hawaiian and Other Pacific Islander (NHOPI)" persons were collected separately according to the OMB guidelines. Ideally, these two groups could be combined to recreate the old category "Asian and Pacific Islander (API)" as a bridge back to data collected under the old race standards. However, the NCHS Disclosure Review Board (DRB), consulting with DHIS analysts, determined that releasing data using an all-inclusive "Other Pacific Islander" category (which would include the Native Hawaiian, Samoan, Guamanian, and Other Pacific Islander groups) would pose a disclosure risk, especially when used in combination with other demographic and geographic information available on the file. For this reason, the decision was made to suppress the "Other Pacific Islander" category on all public use bridge variables. **This is important for data users to know because this change makes it impossible to bridge back to the old "Asian and Pacific Islander" category that existed in the 1998 and earlier NHIS surveys**. Data users who need this information for their analyses will have to contact the RDC to obtain controlled access to non-released data.

Creation and Editing of 2012 Race Variables

The variables RACRECI3 and MRACRPI2 correspond to the old OMB guidelines for collecting racial and ethnic data (see the Variable Layout Report for further descriptions of these variables). They were created in the same fashion as their previous NHIS counterparts (National Center for Health Statistics, 1996), with two exceptions. First, since observed race is no longer collected in the NHIS (beginning in 1997), it was not used to help classify persons with "Unknown" race on the RACRECI3 recode. Second, the recodes "White/Non-White" and "Black/Non-Black" were not created because they are no longer used in the weighting and tabulation of NHIS data. As in the past, smaller subgroups have been collapsed for confidentiality reasons.

Since the NHIS is now required to collect racial and ethnic data under the new OMB guidelines, variables have been created to allow users to tabulate NHIS data by race variables that correspond to the new OMB guidelines. These variables conform to the new OMB race standards; therefore they are created independently of the follow-up race question (see the section of this appendix on Race and Hispanic Origin Questions in the National Health Interview Survey). The variable RACERPI2 was created using an algorithm that first coded the five race mentions from the survey into the single and multiple race group combinations (shown in bold/italicized and regular font, respectively) included in Table 2, below. All of the multiple race categories in the table were then collapsed into a single "Multiple race" category, and along with 4 of the 5 OMB single race categories, the variable RACERP_I was created. The full algorithm is provided below so that data users can better understand how this variable is derived.

Algorithm used to Create Single and Multiple Race Groups

This algorithm (implemented using SAS) takes into account the OMB race categories. In the NHIS, data are collected in 15 race categories: White, Black/African American, Indian (American), Alaska Native, Native Hawaiian, Guamanian, Samoan, Other Pacific Islander (a verbatim mention that is back-coded to this category), Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian (a verbatim mention that is back-coded to this category). These can all be collapsed back to the OMB categories in the following fashion: *White*, *Black*, *AIAN* (includes Indian (American) and Alaska Native), *Asian* (includes Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese and Other Asian), and *NHOPI* (includes Native Hawaiian, Guamanian, Samoan and Other Pacific Islander).

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

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

> RACEW=0;
> RACEB=0;
> RACEAIAN=0;
> RACEASIA=0;
> RACENHPI=0;

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

> IF ((RACE1=1) or (RACE2=1) or (RACE3=1) or (RACE4=1) or RACE5=1))
> then RACEW=1;
> *This sets RACEW to 1 if there is any mention of the race "White" in any of the 5 race variables;*
>
> IF ((RACE1=2) or (RACE2=2) or (RACE3=2) or (RACE4=2) or RACE5=2))
> then RACEB=2;
> *This sets RACEB to 2 if there is any mention of the race "Black" in any of the 5 race variables;*
>
> IF ((RACE1=3) or (RACE2=3) or (RACE3=3) or (RACE4=3) or RACE5=3))
> then RACEAIAN=4;
> **This sets RACEAIAN to 4 if there is any mention of the race "AIAN" in any of the 5 race variables;*
>
> IF ((RACE1=4) or (RACE2=4) or (RACE3=4) or (RACE4=4) or RACE5=4))
> then RACEASIA=8;
> **This sets RACEASIA to 8 if there is any mention of the race "Asian" in any of the 5 race variables;*

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

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

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

Appendix II, Table 2. Algorithm Coding Scheme

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

Coding Scheme for OMB Race Category Data (including single and multiple race mentions)		
# of Category (reported in SAS frequency distribution of RACEFULL)	Sum of Codes (breakdown of RACEFULL= SUM (OF RACEW+RACEB+ RACEAIAN+RACEASIA+ RACENHPI+RACEOTHR))	Resulting Category (used in the PROC FORMAT statement to label the categories in SAS)
12	0+0+4+8+0+0	AIAN/Asian
13	1+0+4+8+0+0	White/AIAN/Asian
14	0+2+4+8+0+0	Black/AIAN/Asian
15	1+2+4+8+0+0	White/Black/AIAN/Asian
16	0+0+0+0+16+0	*NHOPI*
17	1+0+0+0+16+0	White/NHOPI
18	0+2+0+0+16+0	Black/NHOPI
19	1+2+0+0+16+0	White/Black/NHOPI
20	0+0+4+0+16+0	AIAN/NHOPI
21	1+0+4+0+16+0	White/AIAN/NHOPI
22	0+2+4+0+16+0	Black/AIAN/NHOPI
23	1+2+4+0+16+0	White/Black/AIAN/NHOPI
24	0+0+0+8+16+0	Asian/NHOPI
25	1+0+0+8+16+0	White/Asian/NHOPI
26	0+2+0+8+16+0	Black/Asian/NHOPI
27	1+2+0+8+16+0	White/Black/Asian/NHOPI
28	0+0+4+8+16+0	AIAN/Asian/NHOPI
29	1+0+4+8+16+0	White/AIAN/Asian/NHOPI
30	0+2+4+8+16+0	Black/AIAN/Asian/NHOPI
31	1+2+4+8+16+0	White/Black/AIAN/Asian/NHOPI

Data users should be aware that the variable RACEFULL and others derived from it are not available on public use data files for confidentiality reasons. However, the recode RACERPI2, based on RACEFULL, is available on the data file. Analysts who wish to use more detailed race data in their analyses should contact the RDC or visit their web page: **http://www.cdc.gov/rdc/.**

Further Information

Although the race variables included in the 2012 file have been edited and tested, analytic and methodological work with these variables continues. NCHS is also evaluating other recodes for possible public release at a later date. If these analyses should result in changes to the 2012 NHIS race data, information about this will be placed on the NHIS website: **http://www.cdc.gov/nchs/nhis.htm**.

Additionally, the NHIS has a website devoted exclusively to the race and Hispanic origin data from the survey: **http://www.cdc.gov/nchs/nhis/rhoi.htm**. This site includes additional details on the NHIS race and Hispanic origin data, including more information on editing and imputation of the data, and links to documentation, questionnaires and other resources.

References

Office of Management and Budget. *Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity, December 15, 2000 (Statistical Policy section).* **http://www.whitehouse.gov/omb/inforeg_statpolicy#dr**

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

National Center for Health Statistics. *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). 1996.

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

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

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

Appendix III
Variance Estimation and Other Analytic Issues, NHIS 2012

Introduction

The data collected in the NHIS are obtained through a complex, multistage sample design that involves stratification, clustering, and oversampling of specific population subgroups. The final weights provided for analytic purposes have been adjusted in several ways to permit calculation of valid estimates for the civilian, noninstitutionalized population of the United States. As with any variance estimation methodology, the techniques presented here involve several simplifying assumptions about the design and weighting scheme applied to the data. This appendix provides basic concepts of the NHIS sample design structure so that data users may compute standard error estimates.

Several software packages are available for analyzing complex samples. The website *Summary of Survey Analysis Software,* currently located at:

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

provides references for and a comparison of different software alternatives for the analysis of complex data. Analysts at NCHS generally use the software package SUDAAN® (Research Triangle Institute, 2008) to produce standard error estimates. In this appendix, examples of SUDAAN computer code for standard error calculation are provided for illustrative purposes. Examples also are provided for the Stata, SPSS, SAS, R, and VPLX software packages. However, the appropriate application of these procedures is the ultimate responsibility of data users, and the example command code is *not* "guaranteed." Both the computer command code and methods are subject to change without notification to the user. NCHS recommends that NHIS data be analyzed under the direction of or in consultation with a statistician who is cognizant of sampling methodologies and techniques for the analysis of complex survey data.

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

Conceptual NHIS design for 2012

A new sample design was implemented in 2006. The 2012 NHIS sample came from the seventh year of the current sample design. This appendix provides a brief outline of the current NHIS sample design. The current sample design is very similar to the previous sample design, which was in place from 1995 to 2005.

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

First, the target universe was partitioned into primary sampling units (PSUs), which are single counties, groups of adjacent counties (or equivalent jurisdictions), or metropolitan areas. These PSUs vary in population size and number of jurisdictions. Cost-effective field operations and efficient sampling result in those PSUs with the largest populations (e.g., the New York City metropolitan area) being sampled with certainty, and the smaller universe PSUs being represented by a sample. These smaller PSUs are called non-self-representing (NSR) or non-certainty PSUs. The universe of NSR PSUs is stratified geographically, for example by state, using multiple criteria consistent with NHIS objectives. Once these strata were defined, a sample of PSUs was selected; within most NSR strata, two PSUs were selected without replacement with probability proportional to population size, and the self-representing (SR) PSUs were selected with certainty. Within a few NSR strata with smaller population sizes, only one PSU was drawn.

The U.S. Census Bureau partitioned each selected NSR or SR PSU into substrata of Census blocks or combined blocks based on the concentrations of black, Asian, and Hispanic persons. These race and ethnicity density substrata were defined according to the population concentrations from the 2000 Decennial Census. New housing within a PSU was included as its own substratum in order to produce the most current sample of households. Each PSU could be partitioned into up to 21 substrata of dwelling units. Large metropolitan SR PSUs tend to have many substrata, while the NSR PSUs tend to have only a few.

Sampling within the PSU substrata is complex and involves clustering dwelling units within each substratum. These clusters form a universe of Secondary Sampling Units (SSUs). A systematic sample of SSUs is selected to represent each substratum.

Prior to interviewing, one part of the NHIS sample is assigned to be "screened." In this part of the sample, the NHIS interview proceeds through the collection of the household roster. The interview then continues only if the household roster contains one or more black, Asian, or Hispanic persons. Otherwise, the interview terminates and the household is said to be "screened out." In the other part of the NHIS sample, full interviews are attempted at all households. The proportion of the NHIS sample that is assigned to be "screened" varies across the 21 substrata. For selected dwelling units, the NHIS collects some information about all persons living in the unit, and additional information is obtained for randomly selected persons living in the unit. For

example, one adult per family is randomly selected for interview with the sample adult questionnaire.

In the previous NHIS sample design, all adults in a family had the same chance of being selected as the sample adult. In the current NHIS sample design, any black, Asian, or Hispanic adult aged 65+ years is given twice the chance of being selected as the sample adult as any other adult in the family. This new procedure was implemented to increase the proportion of sample adults who are black, Asian, or Hispanic, and aged 65+ years.

The hierarchy of sampling allows the creation of household- and person-level base weights. Each base weight is the product of the inverses of the probability 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, and if 100% response occurred, a base-weighted sample total will be an unbiased estimator for the true total in the target population. In practice, however, the base weights are adjusted for non-response and ratio-adjusted to create final sampling weights. The final person-level weights are adjusted according to a quarterly poststratification by age/sex/race/ethnicity classes based on population estimates produced by the U.S. Census Bureau. Most other weights receive some form of ratio adjustment as well.

Internally, NCHS uses the design and weighting information to formulate appropriate variance estimates for NHIS statistics. While recognizing the need to provide accurate information, NCHS also must adhere to the Public Health Service Act (Section 308(d)), which forbids the disclosure of any information that may compromise the confidentiality promised to its survey respondents. Consequently, much of the NHIS design information cannot be publicly released, and other data are either suppressed or recoded to insure confidentiality. In order to satisfy this disclosure constraint, many of the original design strata, substrata, PSUs, and SSUs are masked for public release by applying techniques to cluster, collapse, mix, and partition the original design variables. Through this process the original NHIS design variables are transformed into public use design variables (i.e., STRAT_P and PSU_P). Data users who want access to internal NCHS data have the option of accessing data through the NCHS Research Data Centers. For further information, refer to **http://www.cdc.gov/rdc/** .

Design Information Available on the NHIS Public Use Data Files

The 2012 Household, Person, Sample Adult, Sample Child, and supplement public use files contain the design variables necessary for variance estimation; Table 1 provides a summary of the Person File variables. The stratum and PSU variable names are the same in the other files, but the weight variable has a different name.

Appendix III, Table 1. Variables Used for Variance Estimation, 2012 NHIS Person File

Variable Name	Variable Label
STRAT_P	Stratum for variance estimation
PSU_P	PSU for variance estimation
WTFA	Weight - Final annual Person weight

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

CAUTION. The STRAT_P and PSU_P values for 2012 are based on the current NHIS sample design and have no connection with the STRATUM and PSU values for 2005 and earlier years. Refer to the final section of this appendix for variance estimation guidance for pooled analyses of adjacent years of the NHIS, including pooling 2006 - 2012 data with data for 2005 and earlier years.

Variance Estimation Method for Public Use Data

The method described below is applicable to the 2012 NHIS Household, Person, Sample Adult, Sample Child, and supplement public use data files.

The limited public release design information requires a mathematical simplification that the PSUs be treated as if they were sampled with replacement (WR). This public use method tends to provide slightly more conservative (larger) standard errors than the variance estimation method that is applied internally by analysts at NCHS. The public use method is robust when analyzing subsetted or subgroup data (see the section "Subsetted Data Analysis" below).

The simplified design structure can be specified with the following statements in SUDAAN for the Person File:

> *PROC <DESCRIPT, CROSSTAB, ...>* **...**   **DESIGN = WR ;**
> *NEST* **STRAT_P PSU_P ;**
> *WEIGHT*   **WTFA ;**

Note that SUDAAN requires that the input file be sorted by the variables listed on the NEST statement (i.e., STRAT_P and PSU_P). Design statements for other data files should use the appropriate weight variables found on these files.

Corresponding statements for other software packages are as follows:

Stata svy:

SVYSET **[PWEIGHT=WTFA]**,*STRATA***(STRAT_P)***PSU***(PSU_P)**
SVY: MEAN <name of variable to be analyzed for average>
Or
SVY: PROPORTION <name of variable to be analyzed for percentage/proportion>

SPSS csdescriptives (for averages) or cstabulate (for percentages/proportions):

One needs first to define a "plan file" with information about the weight and variance estimation, e.g.:

CSPLAN ANALYSIS
/PLAN FILE="< file name >"
*/PLANVARS ANALYSISWEIGHT=***WTFA**
*/DESIGN STRATA=***STRAT_P** *CLUSTER=***PSU_P**
*/ESTIMATOR TYPE=***WR**.

And then refer to the plan file when using csdescriptives or cstabulate, e.g.:

CSDESCRIPTIVES
/PLAN FILE="< file name >"
/SUMMARY VARIABLES =<name of variable to be analyzed>
/MEAN.

CSTABULATE
/PLAN FILE="< file name >"
/TABLES VARIABLES =<name of variable to be analyzed>
/CELLS TABLEPCT.

SAS proc surveymeans (for averages) or surveyfreq (for percentages/proportions) :

PROC SURVEYMEANS;
STRATA **STRAT_P**;
CLUSTER **PSU_P**;
WEIGHT **WTFA**;
VAR <name of variable to be analyzed>;
RUN;

PROC SURVEYFREQ;
STRATA **STRAT_P**;
CLUSTER **PSU_P**;
WEIGHT **WTFA**;

TABLES <name of variable to be analyzed>;
RUN;

R (including the "survey" add-on package):
(note: R syntax is case-sensitive)

 # load survey package
require(survey)
 # create data frame with NHIS design information, using existing data frame of NHIS data
*nhissvy <- svydesign(id=~***psu_p***, strata=~***strat_p***,*
                          nest = TRUE,
                          *weights=~***wtfa***,*
                          data=< existing data frame name>)
svymean(~<name of variable to be analyzed>,design=nhissvy)

note: svymean will produce proportions for "factor variables." For details consult the R
documentation at **http://cran.r-project.org/manuals.html** .

VPLX:

In the CREATE step, include the following statements:

STRATUM   **STRAT_P**
CLUSTER   **PSU_P**
WEIGHT   **WTFA**

Then specify the variable to be analyzed in the DISPLAY step:

LIST   *MEAN*(<name of variable to be analyzed>)

VPLX can produce percentages by including a CAT statement in the CREATE step. For details
consult the VPLX documentation at **http://www.census.gov/sdms/www/vdoc.html** .

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

Subsetted Data Analysis

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

Strategy 1 (recommended)   Use the SUBPOPN statement with the method described above for the full Person File dataset:

> *PROC ...*　　**DESIGN = WR ;**
> *NEST* **STRAT_P PSU_P ;**
> *WEIGHT*　　**WTFA ;**
> *SUBGROUP* **(variable names);**
> *LEVELS ...*　　**;**
> *SUBPOPN*　　**RACRECI3=2 & SEX=2 / NAME="Analysis of African American women;"**

Using the full dataset with the SUBPOPN statement in this example would constrain this analysis to African American women only (RACRECI3 = 2 for black and SEX = 2 for female). Use of the SUBPOPN statement is equivalent to subsetting the dataset, except that any resulting variance estimates are based on the full design structure for the complete dataset.

Strategy 2 (not recommended, except when Strategy 1 is infeasible)   Use the MISSUNIT option on the NEST statement with the method described above for subsetted data:

> *NEST*　　　　**STRAT_P PSU_P / MISSUNIT ;**

In a WR design, when some PSUs are removed from the database through the listwise deletion of records outside the population of interest, leaving only one PSU in one or more strata, the MISSUNIT option in SUDAAN "fixes" the estimation to avoid errors due to the presence of strata with only one PSU. In the special case of a WR design with exactly two PSUs per stratum, using the MISSUNIT option with subsetted data gives the same variance estimate as using

Strategy 1. However, except for this special case, there is no guarantee that the variance estimates obtained by this method are equivalent to those obtained using Strategy 1. Other calculations, such as those for design effects, degrees of freedom, standardization, etc., may need to be carried out differently. Users are responsible for verifying the correctness of their results based on subsetted data.

Implementing Strategy 1 in other software packages can be accomplished as follows:

Stata svy:

Add SUBPOP to the SVY statement, e.g.:

SVY,SUBPOP(**RACRECI3==2 & SEX==2**)*: MEAN* <name of variable to be analyzed>

SPSS csdescriptives or cstabulate:

One must first define an indicator variable, e.g.:

DO IF (RACRECI3 EQ 2 AND SEX EQ 2).
COMPUTE SUBGRP=1.
     ELSE.
COMPUTE SUBGRP=0.
     END IF.

And then refer to the indicator variable in csdescriptives or cstabulate, e.g.:

CSDESCRIPTIVES (or CSTABULATE)
/SUBPOP TABLE=SUBGRP

It is **very important** that the indicator variable be defined for all data records, otherwise an invalid result can occur.

SAS proc surveymeans or surveyfreq:

One must first define an indicator variable, e.g.:

IF RACRECI3=2 & SEX=2   *THEN SUBGRP=1;*
                               ELSE SUBGRP=0;

And then refer to the indicator variable in proc surveymeans using the DOMAIN statement, e.g.:

PROC SURVEYMEANS;
DOMAIN SUBGRP;

Proc surveyfreq does not have a DOMAIN statement. Instead, include the indicator variable in the TABLES specification:

PROC SURVEYFREQ;
TABLES SUBGRP*<name of variable to be analyzed>;

As with SPSS, it is **very important** that the indicator variable is defined for all data records; otherwise an invalid result can occur.

R (including the "survey" add-on package):

After applying the svydesign function to a data frame that contains the entire NHIS sample file being analyzed, specify the criteria that define the subgroup of interest in the subset function and apply the function to the R "object" created by the svydesign function to create a new R object. Note that R is very "feisty" when testing for equality, hence the syntax that follows specifies the subgroup of interest without using an equality test.

 # subset for racreci3=2 & sex=2 without using equal signs
subgrp <- subset(nhissvy,racreci3>1 & racreci3<3 & sex>1)
svymean(~<name of variable to be analyzed>,design=subgr

VPLX:

In the CREATE step, define one or more CLASS variables that can be used to specify the criteria that define the subgroup of interest.

COPY RACRECI3 INTO RACECAT
COPY SEX INTO SEXCAT
CLASS RACECAT (1/2/3-HIGH)
CLASS SEXCAT (1/2)

The second category of RACECAT, crossed with the second category of SEXCAT, defines the subgroup of interest.

Then, specify the variable to be analyzed in the DISPLAY step, and specify the subgroup of interest as well:

LIST   MEAN(<name of variable to be analyzed>) */CLASS* RACECAT(2)*SEXCAT(2)

Note that the specification of RACECAT(2) and SEXCAT(2) is to the second category of each variable, which happens to be the value "2" in both cases in this example. Specification of RACECAT(3) would include all values of RACRECI3 of 3 and higher ("3-HIGH").

Variance Estimation for Pooled Analyses of Adjacent Years of the NHIS

Adjacent years of NHIS data sometimes are combined for a pooled analysis, e.g., 2005 and 2006, or 2002-2004. A pooled analysis might be done, for example, to increase the sample size for some small population. An estimate from a pooled analysis can be interpreted to be an estimate for the midpoint of or the "average" over the time interval of the pooled data.

See Appendix VI, "Combining Years of Data" section, for an example SAS program that combines 2004 and 2005 NHIS data, and an example program that forms a combined 2004-2007 NHIS dataset.

The sampling weights for pooled data should be adjusted; otherwise, estimates of totals will be too high. For example, the estimated total U.S. civilian noninstitutionalized population from two years of pooled data, using unadjusted weights, would be about twice as large as it should be. A simple, valid weight adjustment procedure that NCHS recommends is to divide each sample weight in the pooled dataset by the number of years that are being pooled; e.g., divide by 2 when two years of data are combined, divide by 3 when three years of data are combined, etc. A sophisticated user may want to consider an alternative weight adjustment method that would minimize the variance of a particular estimate; however, in general, if the sample sizes are similar in the data years being combined, the simple procedure and the sophisticated alternative would give a similar adjustment.

Variance estimation for pooled analyses falls into one or more of the following three classifications:

#1. The years being pooled fall within the same sample design period with the same public use design variables, and no changes were made to the design variables within the years being pooled.

#2. The years being pooled fall into different sample design periods (e.g., design periods 1963-1972, 1973-1984, 1985-1994, 1995-2005, 2006 and later years).

#3. The years being pooled fall within the same sample design period, and there were changes to the public use design variables (e.g., from 1995-1996 to 1997-2005).

For #1, the sample has been drawn from the same geographic areas (same sample design), and the definitions of the variables used for public use variance estimation have not changed within the time period being analyzed. A valid method for variance estimation is to treat the pooled data like one year of data with a very large sample size. It is not correct to treat the different data years as being statistically independent, because the samples for the different years were drawn from the same geographic areas (i.e., same PSUs, nearby SSUs). Treating different data years as being statistically independent generally will lead to standard error estimates that are too small, and standard error estimates of contrasts (differences) between years would tend to be too large if the yearly estimates are positively correlated.

For #2, the different sample design periods should be treated as statistically independent. If there are multiple years of data being used for one or both design periods, each group should be treated in a similar manner as described in #1, assuming that the design variables within each group were unchanged. For example, if 1992-1995 NHIS data were pooled, the #1 procedure applies for the 1992-1994 data, and that aggregate is treated as being statistically independent from the 1995 data.

Note that it may be necessary to create new design variables to carry out this type of analysis. For example, consider an analysis of 1992-1995 NHIS data. The design variables have different names in the two sample design periods, and the stratum identifiers have different lengths. Referring to the first method described in "Variance Estimation for Person Data Using SUDAAN and the National Health Interview Survey (NHIS) Public-Use Person Data Files, 1985-94", currently available online at **http://www.cdc.gov/nchs/nhis/sudaan.htm**, the (Method 1) design variables for the 1992-1994 data are CSTRATUM (stratum), CPSU (PSU), and WTF (weight), while they are STRATUM, PSU, and WFTA, respectively, for the 1995 data. Suppose the names of the new design variables are NSTRATUM (stratum), NPSU (PSU), and NWT (weight). One method to create values for NSTRATUM that are of consistent length and take account of the different sample design periods is to do the following: for the 1992-1994 data, where the CSTRATUM values are 1, 2, ..., 62, first change these to 001, 002, ..., 062 (consistent length with STRATUM), and then do something to make them distinct from the STRATUM values, such as put a "1" in front: 1001, 1002, ..., 1062. For the 1995 data, where the STRATUM values are 1, 2,..., 339, first change these to 001, 002, ..., 339, and then do something to make them distinct from the CSTRATUM values, such as put a "2" in front: 2001, 2002, ..., 2339. NPSU can be set equal to CPSU for the 1992-1994 data, and equal to PSU for the 1995 data, as both CPSU and PSU are of length one. NWT can be set equal to WTF/4 for the 1992-1994 data, and to WFTA/4 for the 1995 data.

For #3, no entirely satisfactory approach is available. Grouping of years should be done over the periods where the same public use design variables are present (i.e., like #1). Then, for combining across years where there were changes to the public use design variables, the only option is to carry out an analysis as if the data years were statistically independent. For example, if 1995-1999 NHIS data were pooled, the #1 procedure applies for 1995-1996, and 1997-1999; then, the only alternative is to treat these two groups as statistically independent. The resulting standard error estimates may be too small, and standard error estimates of contrasts between years might be too large if the yearly estimates are positively correlated.

References

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

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

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

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

Research Triangle Institute. *SUDAAN Language Manual; Release 10.0*, Research Triangle Institute, Research Triangle Park, NC. 2008.

Appendix IV
A Preliminary Evaluation and Recommendations for use
of the Mental Health Indicator (MHI) in the NHIS
for Children Aged 2 to 3 years

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

Introduction

The NHIS mental health recodes MHIBOY2 and MHIGRL2 are located in the Child Health Status (CHS) section of the survey, and are based on items from the Child Behavior Checklist (CBCL) that were identified by Dr. Thomas Achenbach as providing the best discrimination between demographically similar children referred for mental health services versus nonreferred (Achenbach and Edelbrock, 1983). To take account of gender and age differences in the discriminative power of particular items, the items were selected separately for each gender and age group. From the original ten items identified in Dr. Achenbach's 1995 analyses, the NHIS elected to include only 4 items (per gender). These include whether male sample children (aged 2-3 years) had been uncooperative, had trouble sleeping, had speech problems, or had been unhappy or depressed in the past 2 months, and whether female sample children (aged 2-3 years) had temper tantrums, had speech problems, had been nervous or high-strung, or had been unhappy or depressed in the past 2 months. Response categories included "Not true," "Sometimes true," or "Often true" (as well as "Refused" and "Don't know"). These items are also located in the CHS section (see CHS.321_01-04.000 and CHS.361_01-04.000).

It is essential to note that such a small set of items cannot be used to evaluate individual children for clinical or other purposes. Even for use as a mental health indicator in large surveys such as the NHIS, very small sets of items can serve only as approximate indicators of needs for mental health services. Multiple items tapping each of several specific areas of functioning would be needed to identify specific disorders, such as Attention Deficit Hyperactivity Disorder (ADHD), Depression, Conduct Disorder, and Somatization Disorder. It should also be noted that different cut points on the distributions of item scores may be needed for different purposes. For example, a very low cut point may be useful if the goal is to identify every possible case for which mental health services might be considered. However, very low cut points result in relatively high false positive rates, i.e., the inclusion of substantial numbers of healthy individuals among those identified as potentially needing services. Conversely, higher cut points may yield greater overall accuracy in classifying potential cases versus noncases, but at the cost of missing more cases potentially needing services.

Data Analyses

Dr. Achenbach specified and reviewed data analyses that were done at NCHS. These included tabulations of specific responses to each behavioral/emotional problem item; tabulations of relations between total problem scores and classification of children as deviant versus nondeviant on the basis of external criteria (e.g., parents ever being told by health professionals that their child had ADHD, mental retardation, other developmental delay, autism,

down syndrome, or a learning disability; parents having talked to mental health professionals about their child in the preceding 12 months; or parents needing mental health services for their child but being unable to afford it); and Relative Operating Characteristic (ROC) analyses of cut points on the total problem scores. Because each behavioral/emotional problem item was scored "0" (not true of the child), "1" (somewhat or sometimes true), or "2" (very true or often true), total scores across the 4 items for each gender/age group could range from "0" to "8." Dr. Achenbach examined the results and recommended changes and additions to the analyses.

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

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

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

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

Appendix V
The Short Strengths and Difficulties Questionnaire (SDQ)

In the NHIS, questions CMHMF_1- CMHMF_5 (CMB.020_01.000 to CMB.020_05.000) and CMHDIFF (CMB.030_00.000) make up a brief version of the SDQ. The questions are derived from the parent version of the long Strengths and Difficulties Questionnaire Extended (SDQ), developed and copyrighted by Dr. Robert Goodman, Institute of Psychiatry, London, England. Questions from the SDQ are used in the NHIS with Dr. Goodman's permission. The short SDQ, constructed to save time and space in the questionnaire, was added for children aged 4-17 years as a part of a collaborative agreement between NCHS and the National Institute of Mental Health (NIMH) of the National Institutes of Health (NIH). Detailed information on the SDQ can be found in Appendix V of the Dataset Documentation for the 2004 NHIS and on the SDQ web site at: **http://www.sdqinfo.org** .

The items in the short SDQ correlate to the subscales in the long SDQ as follows:

CMB.020_01.000 *Generally obedient,* correlates 0.69 with the long SDQ conduct score.
CMB.020_02.000 *Many worries…,* correlates 0.71 with the long SDQ emotion score.
CMB.020_03.000 *Often unhappy…,* correlates 0.64 with the long SDQ emotion score.
CMB.020_04.000 *Gets along better…,* correlates 0.69 with the long SDQ peer problems score.
CMB.020_05.000 *Sees tasks through,* correlates 0.72 with the long SDQ hyperactivity-inattention score.

In order to score the short SDQ the response for each item in CMB.020 is assigned a value from 0 – 2 based on the scale below, then all values are summed to produce a total score. A total score correlates 0.84 with the long SDQ total difficulties score.

Scoring of the Short SDQ

Response	Not true	Somewhat true	Definitely true
Value	0	1	2

CMB.030 is taken from a set of SDQ extended or impact questions which measure the impact of the child's difficulty on various aspects of his/her life. CMB.030 correlates 0.62 with the SDQ impact score from the extended SDQ questions.

The Short Strengths and Difficulties Questionnaire Calibration Study

In 2006, the Substance Abuse and Mental Health Services Administration (SAMHSA) convened a technical group to provide advice on developing a reliable way to produce estimates of children with serious emotional disturbance (SED). The group recommended that SED estimates be developed from the NHIS, utilizing the short SDQ, which had been introduced in the NHIS in 2001 for the screening of mental health problems. Although the short SDQ items provided a total SDQ score, no cut-off point was established that could be used to determine SED. The group proposed a calibration study to develop a cut-off score by comparing short SDQ scores with the results of a standard clinical psychiatric assessment.

The advisory group recommended that the Child and Adolescent Psychiatric Assessment (CAPA) and the Preschool Age Psychiatric Assessment (PAPA) be used as the standard clinical psychiatric assessments for the calibration study. The CAPA and the PAPA are semi-structured interviews conducted by lay interviewers trained to assess the frequency, duration, and intensity of numerous mental health symptoms (Angold & Costello, 2000; Egger et al., 2004). The complete CAPA and PAPA assess the full range of child mental health disorders. For this study, five areas were selected for assessment: anxiety disorders, mood disorders, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD). Detailed information on the CAPA and the PAPA can be found at:
http://devepi.duhs.duke.edu/instruments.html .

For the calibration study, 1,187 cases were sampled from completed interviews about sample children aged 4-17 years from quarters 2-4 of the 2011 NHIS and quarter one of the 2012 NHIS. Clinical psychiatric assessments (CAPA, PAPA) were conducted over the telephone between February and August 2012 with a parent of children aged 4-17 years and with adolescents aged 12-17 years. There were 139 completed parent interviews for children aged 4-11 years. There were 78 completed sets of interviews of parent and child for adolescents aged 12-17 years.

The results of this study are currently being analyzed and will be available at a later date.

Additional References on the SDQ and/or its use in the NHIS

Angold, A., & Costello, E. J. (2000). The Child and Adolescent Psychiatric Assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 39-48.

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

Egger, H., Ascher, B. H., & Angold, A. (2004). *Preschool Age Psychiatric Assessment (PAPA): Core modules, Version 2.0.9.* Durham, NC: Duke University, Department of Psychiatry and Behavioral Sciences, Center for Developmental Epidemiology.

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

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

Goodman, R. and Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27(1): 7-24. 1999.

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

<div align="center">

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

</div>

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

<div align="center">

Merging Data Files

</div>

It is important to note that for the 2005 data year and beyond, some frequently used variables are repeated on various data files; therefore, merging of files may be required less often than for the 2004 data year files. However, each data file contains household, family, and person record identifiers that make merging the data files possible, if needed. Once the data files are sorted by record identifiers common to each file, merging is straightforward. Below is an example of a SAS program that will merge data files within an NHIS data year. Using the household, family, and person record identifiers (HHX, FMX and FPX, respectively), this program merges data from the 2012 Household, Family, Person, and Sample Child data files. **Variable names may change from one year to another. Users are advised to check variable names and update computer programs when changes occur.**

/* Merge the 2012 Household File and the 2012 Family File. */

/* Create a Household dataset with selected variables and sorted by HHX.*/
DATA HH (KEEP=HHX REGION); /* HH is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2012.HOUSEHLD; /*The SET statement reads data from the 2012 Household File. */
PROC SORT DATA=HH; /* Sort by HHX, the household identifier. */
BY HHX;
RUN;

/* Create a Family dataset with selected variables and sorted by HHX. */
DATA FM (KEEP=HHX FMX INCGRP2 RAT_CAT2 WTFA_FAM); /* FM is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2012.FAMILYXX; /*The SET statement reads data from the 2012 Family File. */
PROC SORT DATA=FM; /* Sort by HHX, the household identifier. */
BY HHX;
RUN;

DATA HHFM; /* New combined dataset called HHFM */
MERGE FM (IN=FROMFM) HH; /* Merge the newly created FM and HH Files, using an IN statement.*/
BY HHX;

IF FROMFM = 1; /* The combined dataset HHFM will contain only those records that are in the Family File; the Household File's REGION variable will be appended to these records. */
PROC SORT DATA=HHFM; /* Sort by HHX and FMX, the household and family identifiers. */
BY HHX FMX;
RUN;

       In the code above, the IN statement creates a temporary SAS variable (called FROMFM) that has a value of 1 if the dataset associated with the IN statement contributed to the current observation or a value of 0 if it did not. The subsequent statement, "IF FROMFM = 1" tells SAS to retain only those observations from the Family File (called FM), thereby eliminating Household File records corresponding to non-response cases (no family/person records are available for non-response cases). For more information on IN statements in SAS, consult Delwiche and Slaughter (1998).

/* Merge the 2012 Person File and the combined 2012 Family/Household File. */

/* Create a Person File with selected variables. */
DATA PR (KEEP=HHX FMX FPX SEX AGE_P WTFA STRAT_P PSU_P); /* PR is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2012.PERSONSX; /*The SET statement reads data from the 2012 Person File. */
PROC SORT DATA=PR; /* Sort by HHX and FMX, the household and family identifiers. */
BY HHX FMX;
RUN;

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

       The above code will create a person-level file, copying the family/household information to each matching person record.

/* Merge the 2012 Sample Child File and the combined 2012 Person/Family/Household File. */

/* Create a Sample Child File with selected variables. */
DATA CH (KEEP=FPX HHX FMX CASHMEV PROBRX WTFA_SC); /* CH is a SAS dataset; the KEEP statement retains only the listed variables for processing. */
SET NHIS2012.SAMCHILD; /*The SET statement reads data from the 2012 Sample Child File. */

PROC SORT DATA=CH; /* Sort by HHX, FMX, and FPX, the household, family, and person identifiers. */

BY HHX FMX FPX;
RUN;

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

     In the code above, no IN statement was used in the MERGE statement, so the resulting file will have records for all persons, sample child or not. The sample child data items will be missing for persons who do not have a matching sample child record.

Combining Years of Data

Important Note

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

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

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

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

If data from the period 1997-2005 are combined or the 2006-2012 data are combined, the combined data should be treated like a single year of data with a larger sample size for the purpose of variance estimation. If data from any year before 1997 are combined with data from 1997 and beyond, or data from 2005 or before are combined with data from 2006 and beyond, variance estimation is more complicated. Refer to Appendix III for more information about variance estimation methods when combining datasets from more than one year.

The following is an example of a SAS program that will combine data files across NHIS data years. The program is written to concatenate the data from the Person Files of the 2011 NHIS and the 2012 NHIS. This same program can be used to combine the 2004 and 2005 NHIS Person Files after minor modifications (e.g., change "2011" and "11" to "2004" and "04", change "2012" and "12" to "2005" and "05", and change STRAT_P PSU_P to STRATUM PSU).

Important Note

The person identifier was called PX in the 2003 (and earlier) NHIS and FPX in the 2004 (and later) NHIS; users may find it necessary to create an FPX variable in the 2003 and earlier datasets (or, alternatively, a PX variable in the 2004 and later datasets) in order to make the data compatible for analyses.

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

DATA PER_11; /* Create SAS dataset PER_11.*/
SET NHIS2011.PERSONSX /* The SET statement reads data from an existing SAS dataset, e.g., the 2011 Person File */ **(KEEP=HHX FMX FPX AGE_P SEX WTFA STRAT_P PSU_P);** /* The KEEP statement retains only the listed variables for processing. */
RUN;

PROC SORT DATA=PER_11; /* Sort SAS dataset PER_11. */
BY HHX FMX FPX;
RUN;

DATA PER_12; /* Create SAS dataset PER_12.*/
SET NHIS2012.PERSONSX /* The SET statement reads data from an existing SAS dataset, e.g., the 2012 Person File */ **(KEEP=HHX FMX FPX AGE_P SEX WTFA STRAT_P PSU_P);** /* The KEEP statement retains only the listed variables for processing. */
RUN;

PROC SORT DATA=PER_12; /* Sort SAS dataset PER_12. */
BY HHX FMX FPX;
RUN;

```
DATA COMB1112;   /* New, combined SAS dataset */
SET PER_11 PER_12;   /* Concatenate selected variables from 2011 and 2012 datasets. */
WTFA_2YR=WTFA/2;  /*Create a new weight by dividing the existing Person File weight
(WTFA) by 2, the number of Person data files combined to create the data file called COMBO.*/
RUN;
```

Now, suppose there exists a dataset "COMB0405" with the combined 2004 and 2005 Person Files, and there exists a dataset "COMB1112" with the combined 2011 and 2012 Person Files. As part of creating a dataset named "COMB0412" containing the combined 2004, 2005, 2011 and 2012 Person Files, two issues need to be addressed:

1. Adjustment of weights
2. Formation of new variance estimation variables, because this combination goes across sample design periods

The weights in COMB0405 and COMB1112 should be divided by 2, so that the original weights have been divided by 4 (four years of data being combined). To avoid the possibility of errors, NCHS recommends that new names be used for the new variance estimation variables, e.g., NSTRATUM (stratum), NPSU (PSU). The PSU and PSU_P values from COMB0405 and COMB1112 can be copied directly to NPSU. The NSTRATUM values need to be created in such a way to assure the values are distinct between 2004-2005 and 2011-2012. As STRATUM ranges from 1 to 339 and STRAT_P ranges from 1 to 300, an appropriate method for creating the NSTRATUM values would be to add 1000 to the STRATUM values and 2000 to the STRAT_P values.

```
 /*Combine 2004-2005 data file with 2011-2012 data file */

DATA COMB0405;
SET COMB0405;
DROP STRATUM PSU;
NSTRATUM=STRATUM+1000;
NPSU=PSU;
RUN;

DATA COMB1112;
SET COMB1112;
DROP STRAT_P PSU_P;
NSTRATUM=STRAT_P+2000;
NPSU=PSU_P;
RUN;

DATA COMB0412;
SET COMB0405 COMB1112;
DROP WTFA_2YR;
WTFA_4YR=WTFA_2YR/2;
RUN;
```

`

References

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

Appendix VII
Analysis of the Family Food Security Variables

Beginning in 2011, a Family Food Security Supplement (FFS) that consisted of ten questions was added to the NHIS. The following instructions for scoring the ten questions and determining the food security status of families are based on guidance from the US Department of Agriculture (USDA) Economic Research Service. See **http://www.fns.usda.gov/fsec/FILES/FSGuide.pdf** for more information about the USDA's food security research and standard procedures for measuring food insecurity and hunger in the United States. The protocol for analyzing the FFS data from NHIS is the same as is used for the food security supplement of the Current Population Survey and other national surveys that carry the food security module.

First, a raw food security score can be created to represent the number of affirmative responses (0-10) to the food security questions. Answers of "often true" or "sometimes true" to FSRUNOUT, FSLAST, and FSBALANC are considered affirmative. Responses to FSLESS, FSHUNGRY, FSWEIGHT, FSSKIP, and FSNOTEAT are considered affirmative if the respondent answered "Yes" to the question. Responses to FSSKDAYS and FSNEDAYS are considered affirmative if the respondent's answer was greater than or equal to 3 days.

If the respondent does not provide a valid response to any of the questions, the raw score should not be calculated. Few families (less than 1%) were screened out after the first three questions due to complete item nonresponse (e.g., don't know, refused).

If the respondent does not provide a valid response to some, but not all, of the questions, a raw food security score can still be calculated. When deciding how to deal with these missing responses, it is useful to note that the questions are administered in order of severity, so that families with any negative responses to early questions are considered unlikely to affirm later questions, and families with affirmative responses to later questions are considered likely to affirm earlier questions. Thus, it is possible to impute responses for missing items based on the nature of the answers—negative or affirmative—that the respondent gave to all the other scale items. USDA provides procedures for doing so at: **http://www.fns.usda.gov/fsec/FILES/FSGuide.pdf** . In practice, analyses with NHIS data have shown that imputing negative responses for all missing values produce scale scores that are highly similar to those produced using the imputation procedures, and scale scores based on this methodologically conservative procedure were included in the 2011 NHIS family data file. For 2012, scale scores are not included in the file, and analysts are encouraged to use their own judgment when deciding how to produce scale scores if the respondent did not provide a valid response to some, but not all, of the questions.

Creating food security classification variables

Two options for food security status classification variables are recommended: one with food security represented in a single "food secure" category, and one which distinguishes between families with high food security and families with marginal food security. The recommended classifications are given below:

Option 1
    1 Food secure (high or marginal food security, raw score 0-2)
    2 Low food security (raw score 3-5)
    3 Very low food security (raw score 6-10)

Option 2
    1 High food security (raw score 0)
    2 Marginal food security (raw score 1-2)
    3 Low food security (raw score 3-5)
    4 Very low food security (raw score 6-10)

For most analytic purposes, food security is best modeled categorically. If food security is an independent variable, it may be entered as a set of two or three dummy variables based on food security status. If food security is the dependent variable, a logistic, probit, or other discrete model based on dichotomous classification may be used.

Linear modeling based on raw score may be justified with some restrictions. Under assumptions of the Rasch measurement model, the raw score is an ordinal rather than interval measure of the severity of food security. However, with the exception of raw score 0, the raw score is so nearly collinear with the Rasch severity parameter that raw score can be used for linear modeling with negligible distortion. However, the size of the 0-1 interval is unknown. It is much larger than the other 1-unit intervals, and may differ among families. This should be accounted for if the raw score is used in a linear model. If food security is an independent variable in the model, raw score plus a dummy variable for raw score > 0 may be used. If food security is the dependent variable, either a model such as tobit may be used, or the analysis may be restricted to families with raw scores 1 and higher.

For more information on food security measurement, please refer to the USDA's website at: **http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us.aspx**.

Appendix VIII
Core Changes/Additions/Deletions in 2012

Sample Adult Core

Adult Condition Section (ACN)

Variable Name	Brief Variable Description	Brief Description of Change
COPDEV	Ever been told you had COPD	New variable