

---

**2015 National Health Interview Survey (NHIS)  
Public Use Data Release**



## **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 2016**

## Table of Contents

NCHS Website and NHIS Electronic Mail List .....	5
Guidelines for Citation of Data Source .....	6
What's New in 2015? .....	7
Introduction.....	8
Sample Design .....	9
2015 Questionnaire.....	11
Interviewing Procedures .....	14
Sample Sizes and Response Rates.....	16
Overview of the 2015 Data Files and Data File Documentation .....	17
Editing the Data During and After the Interview .....	17
Edits to Protect Confidentiality.....	17
Data File Documentation .....	18
Household File .....	21
Family File.....	22
Family Structure Variables.....	24
Person File .....	25
I. Coverage Section: Telephone (COV).....	25
II. Household Composition Section (HHC) .....	25
III. Family Identification Section (FID).....	27
IV. Health Status and Limitation of Activity Section (FHS).....	28
V. Family Food Security Section (FFS) .....	35
VI. Health Care Access and Utilization Section (FAU) .....	35
VII. Health Insurance Section (FHI) .....	36
VIII. Socio-demographic Section (FSD).....	40
IX. Income and Assets Section (FIN).....	42
X. English Language Proficiency Section (FLG).....	45
Injury and Poisoning Episode File.....	46
Family Disability Questions File (FDB) .....	51
Background .....	51
Sample Child File .....	53
I. Child Identification Section (CID) .....	53
II. Child Conditions, Limitation of Activity and Health Status Section (CHS) .....	53
III. Child Health Care Access and Utilization Section (CAU) .....	55
IV. Child Mental Health Brief Supplement (CMB).....	55
V. Child Influenza Immunization Supplement (CFI) .....	56

Sample Adult File.....	57
I. Adult Identification Section (AID) .....	57
II. Adult Socio-Demographics Section (ASD) .....	57
III. Adult Conditions Section (ACN) .....	59
IV. Adult Health Status and Limitation of Activity Section (AHS).....	62
V. Adult Health Behaviors Section (AHB).....	65
VI. Adult Health Care Access and Utilization Section (AAU) .....	68
VII. Adult Selected Items Section (ASI).....	70
VIII. Adult Internet and Email Usage Section (AWB).....	71
Adult Functioning and Disability Supplement File (AFD).....	72
Cancer Control Supplement File.....	73
I. Diet and Nutrition Section (NAC) .....	73
II. Physical Activity Section (NAD) .....	73
III. Tobacco Section (NAE) .....	74
IV. Cancer Screening Section (NAF) .....	74
V. Genetic Testing Section (NAG).....	75
VI. Family History Section (NAH).....	76
Weighting .....	77
Variance Estimation.....	80
Merging Data Files and Combining Years of Data in the NHIS .....	81
NCHS Record Linkage Program.....	82
Guidelines for Data Use.....	83
References.....	84
Appendix I. Advance Letter .....	85
Appendix II. Calculation of Response Rates .....	87
Appendix III. Variance Estimation Method for Public Use Data.....	91
Variables Available for Variance Estimation .....	91
Variance Estimation for Analyses of Single Years of the NHIS.....	92
Degrees of Freedom.....	94
Subsetted Data Analysis.....	94
Appendix IV. Merging Data Files and Combining Years of Data in the NHIS.....	98
Merging Data Files .....	98
Combining Years of Data.....	100
Variance Estimation for Pooled Analyses of Adjacent Years .....	102
Appendix V. Analysis of the Family Food Security Variables.....	104
Creating Food Security Scores .....	104
Creating Food Security Classification Variables.....	104

---

Appendix VI. Mental Health Indicator (MHI) for Children Aged 2–3 Years.....	106
Introduction .....	106
Data Analyses.....	106
Appendix VII. The Short Strengths and Difficulties Questionnaire (SDQ) .....	108
Creating SDQ Scores .....	108
The Short Strengths and Difficulties Questionnaire Calibration Study.....	108
Appendix VIII. Core Changes/Additions/Deletions in 2015.....	110

---

## NCHS Website and NHIS Electronic Mail List

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

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

The website features downloadable data and documentation for the 2015 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.

The website also features the 2015 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 2015 health data files. The Paradata File and documentation can be found at:

- <http://www.cdc.gov/nchs/nhis/2015paradata.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 website.

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](mailto:nhislist@cdc.gov).

---

## 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 2015 data should acknowledge “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: NCHS, National Health Interview Survey, 2015

In a bibliography, the suggested citation for this document should read:

- National Center for Health Statistics. Survey Description, National Health Interview Survey, 2015. Hyattsville, Maryland. 2016.

The suggested citation for 2015 NHIS survey data and other documentation should read:

- National Center for Health Statistics. National Health Interview Survey, 2015. Public-use data file and documentation. [http://www.cdc.gov/nchs/nhis/quest\\_data\\_related\\_1997\\_forward.htm](http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm). 2016.

The published material should also include a disclaimer that credits the author’s analyses, interpretations, and conclusions 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 ensure 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/citations.htm>.

---

## What's New in 2015?

- In 2015, the NHIS sample size augmentation that began in 2011 continued in 32 states and the District of Columbia. The main goal of the augmentation was to increase the number of states for which reliable estimates can be made.
- Supplements to the Family Core in 2015 included expanded content on health care access and utilization, and functioning and disability, along with questions to measure family food security.
- Supplements to the Sample Child Core in 2015 included expanded content on health care access and utilization, immunization, and mental health.
- Supplements to the Sample Adult Core in 2015 included expanded content on health care access and utilization, cancer control, and functioning and disability, along with questions about Crohn's Disease/colitis, epilepsy, heart disease and stroke prevention, hepatitis B/C screening, immunization, internet and email usage, mental health, occupational health, and infrequent tobacco use.

---

## Introduction

The National Health Interview Survey (NHIS) is the principal source of information on the health of the civilian noninstitutionalized population of the United States and is one of the major data collection programs of the National Center for Health Statistics (NCHS). The National Health Survey Act of 1956 provided for a continuing survey and special studies to secure accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services rendered for or because of such conditions. The survey referred to in the Act, now called the National Health Interview Survey, was initiated in July 1957. Since 1960, the survey has been conducted by NCHS, which was formed when the National Health Survey and the National Vital Statistics Division were combined.

The main objective of the NHIS is to monitor the health of the United States population through the collection and analysis of data on a broad range of health topics. A major strength of this survey lies in the ability to categorize these health characteristics by many demographic and socioeconomic characteristics.

NHIS data are used widely throughout the Department of Health and Human Services (HHS) to monitor trends in illness and disability and to track progress toward achieving national health objectives. The data are also used by the public health research community for epidemiologic and policy analysis of such timely issues as characterizing those with various health problems, determining barriers to accessing and using appropriate health care, and evaluating Federal health programs.

While the NHIS has been conducted continuously since 1957, the content of the survey has been updated periodically. In 1996, a substantially revised NHIS questionnaire began field testing. The resulting revised questionnaire, described in detail below, was implemented in 1997 and has been used since then, with relatively small modifications made over time when needed.



---

## 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 residing in the United States at the time of the interview. 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), and U.S. nationals living in foreign countries. Active-duty Armed Forces personnel are also excluded from the survey, unless at least one other family member is a civilian eligible for the survey (for example, a child whose parents are both active-duty military). In that case, data for these Armed Forces members (269 persons in 2015) are collected and included in all relevant files in order to aid any analyses pertaining to the family (e.g., family structure, relationships, or income), but these persons are given a final weight of zero so their individual characteristics will not be counted when making national (i.e., weighted) estimates. Weighted estimates cover only the civilian noninstitutionalized household population.

Sampling and interviewing for the NHIS are continuous throughout each year. The sample design follows a multistage area probability design that permits the representative sampling of households and noninstitutional group quarters (e.g., college dormitories). The sample design is redesigned after every decennial census. The current sample design was implemented in 2006, and 2015 is the final year of its implementation. An NCHS report describing the 2006–2015 sample design (Parsons, Moriarity, Jonas, et al., 2014) is available from: [http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_165.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_165.pdf). Below is a summary of the design described in detail in that report.

Because the NHIS is conducted in a face-to-face interview format, the costs of interviewing a large simple random sample of households and noninstitutional group quarters would be prohibitive; randomly sampled dwelling units would be too dispersed throughout the nation. 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 first stage of the current sample design consists of a sample of 428 primary sampling units (PSUs) drawn from approximately 1,900 geographically defined PSUs, with some PSUs in each of 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.

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. These larger PSUs are called self-representing (SR) PSUs. The PSUs with smaller populations are called nonself-representing (NSR) or noncertainty PSUs. The set of NSR PSUs is stratified geographically (for example, by state). Once these strata were defined, a sample of PSUs was selected. Within most strata, two NSR PSUs were selected without replacement with probability proportional to population size, and the SR PSUs were selected with certainty. Because of this approach, nearly all states have at least two PSUs selected for the sample, and most have substantially more.

Within a PSU, two types of second-stage units are used: area segments and permit segments. Area segments are defined geographically. 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. The NHIS sampling frame consists of two non-overlapping parts: the area frame (all of the area segments) and the permit frame (all of the permit segments).

---

In order to increase the precision of estimates of the black, Hispanic, and Asian populations, the current NHIS sample design oversamples black persons, Hispanic persons, and Asian persons. In addition, the sample adult selection process provides 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.

Prior to interviewing, some of the sample addresses in area segments are assigned to be “screened,” and those interviews proceed through the collection of the household roster (i.e., list of household members and selected demographic characteristics), after which the interview 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 rest of the area sample, full interviews are attempted at all households. No screening occurs in permit segments.

Another oversampling procedure is applied when area segments are sampled within PSUs. Area segments that have high concentrations of black, Asian, and Hispanic persons, based on information from the 2000 Census, have an increased chance of being selected.

As noted earlier, the NHIS sample design is stratified by state; sample is drawn from each state and the District of Columbia. Although the NHIS sample is normally too small to provide state level estimates with acceptable precision for every state, selected estimates for many states may be reliable, especially if multiple data years are combined. State identifiers are not publicly released due to confidentiality concerns. However, state identifiers can be accessed through the NCHS Research Data Center (RDC), a secure data enclave described later in the section on “Edits to Protect Confidentiality.”

The total NHIS sample is subdivided into four separate panels such that each panel is representative of the U.S. civilian noninstitutionalized population (as is any combination of the four panels). This design feature has a number of advantages, including flexibility for the total sample size. For example, it allows for sample size reductions while retaining representativeness. However, since 2011, the NHIS has received sufficient funding to augment the normal sample size. (See the later section on “Sample Sizes and Response Rates” for details on the magnitudes of these increases.) Starting in July 2013, the PSU augmentation was facilitated by an expansion of the number of PSUs selected for the sample. This expansion of PSUs continued in 2015. The methods for selecting these additional PSUs are described elsewhere (Moriarity and Parsons, 2013).

NHIS address lists are generally obtained for NCHS by the U.S. Census Bureau in an address-listing activity that is independent of decennial census operations and explicitly for the NHIS. However, given the high cost of address listing, for the 2013 NHIS, half the addresses in the new PSUs were obtained from commercial address listings. NHIS continued to use commercial address listings for 2015, albeit for only a small portion of the sample.

---

## 2015 Questionnaire

The content of the NHIS questionnaire is revised periodically, with the last major revisions occurring in 1982 and 1997. The redesigned NHIS questionnaire introduced in 1997 consists of a Core that remains largely unchanged from year to year, plus an assortment of Supplements that may be sponsored by agencies other than NCHS, with the assortment varying from year to year. The Core 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 (defined as an occupied housing unit). The Family Core questionnaire, which is administered separately for each family in the household, collects information on all persons in the family. A family is defined as an individual or a group of two or more related persons who are living together in the same household. In some instances, unrelated persons sharing the same household may also be considered as one family, such as unmarried couples who are living together. 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” aged 17 years or less is randomly selected (if any children are in the family) and one “sample adult” aged 18 years or more is randomly selected. Information about the sample child is collected from a knowledgeable adult, and information about the sample adult is collected from the sample adult him/herself, using the Sample Child Core questionnaire and the Sample Adult Core questionnaire, respectively. 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 health-related behaviors. When fielded, supplementary questions about the sample child and sample adult provide additional information.

### Questionnaire Sections

The NHIS Core is divided into various sections that group questions into broad and specific categories. Each section is designated by a section title and a corresponding three-digit acronym (or section code); questionnaire items are numbered sequentially in ascending order 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. Table 1 lists the various questionnaire sections for the 2015 NHIS, their acronyms and description titles.

**Table 1. Core questionnaire sections and topics, 2015 National Health Interview Survey**

Questionnaire and section code	Description
Household	
HHC	Household Composition
Family Core	
FID	Family Identification and Verification
FHS	Health Status and Limitation of Activity
FIJ	Injury/Poisoning
FAU	Health Care Access and Utilization
FHI	Health Insurance
FSD	Socio-demographic
FIN	Income and Assets
Sample Child Core	
CID	Child Identification and Verification
CHS	Conditions, Limitation of Activity and Health Status
CAU	Health Care Access and Utilization
Sample Adult Core	
AID	Adult Identification and Verification
ASD	Demographics
ACN	Conditions
AHS	Health Status and Limitation of Activity
AHB	Health Behaviors
AAU	Health Care Access and Utilization
ASI	Adult Selected Items
Recontact	
REC	Recontact Information and Follow-up

## Supplements, Supplement-Sponsoring Agencies, and Question Locations, 2015 NHIS

The term “supplement” refers to one or more questions sponsored by NCHS and/or other agencies and added to the Core questionnaire to address current data needs. A supplement may be embedded in an appropriate place within a Core section, placed at the end of a Core section, or placed in a separate section of its own. The existence of three extra digits (.xxx) at the end of the question number helps to identify supplementary questions in the NHIS questionnaire.

In 2015, NHIS supplementary questions included expanded content on health care access and utilization (family, adult, and child), cancer control (adult), Crohn’s Disease/colitis (adult), epilepsy (adult), food security (family), functioning and disability (family and adult), heart disease and stroke prevention (adult), hepatitis B/C screening (adult), immunization (adult and child), Internet access and email usage (adult), mental health (adult and child), occupational health (adult), and infrequent tobacco use (adult).

Table 2 presents a list of all 2015 supplements and their question numbers. Information about supplements from 1997–2015 is available on our website: [http://www.cdc.gov/nchs/nhis/supplements\\_cosponsors.htm](http://www.cdc.gov/nchs/nhis/supplements_cosponsors.htm).

**Table 2. Supplement-sponsoring agencies and question locations, 2015 National Health Interview Survey**

Topic	Sponsoring Agency	Title	Survey Section/Number
Cancer Control	National Cancer Institute (NCI) <sup>1</sup> ; National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) <sup>2</sup> ; Center for Tobacco Products (CTP) <sup>4</sup>	Adult Cancer Control Module	NAC.005_00.000 to NAC.300_00.000; NAD.010_00.000 to NAD.070_00.000; NAE.010_00.000 to NAE.240_00.000; NAF.010_00.000 to NAF.770_01.000; NAG.001_00.000 to NAG.090_00.000; NAH.010_00.000 to NAH.210_00.000
Occupational Health	National Institute of Occupational Safety and Health (NIOSH) <sup>2</sup>	Adult Occupational Health	ASD.220_00.010 to ASD.220_00.240; ACN.296_00.010 to ACN.297_00.020; ACN.325_00.010 to ACN.325_00.080
Expanded Content on Health Care Access and Utilization	Office of the Assistant Secretary for Planning and Evaluation (ASPE) <sup>5</sup>	Family, Adult, and Child Affordable Care Act Questions	FHI.135_00.010 to FHI.137_00.030; FHI.202_01.010 to FHI.204_01.010; FHI.215_01.010; FHI.225_00.000; FHI.248_05.000; FHI.249_03.000; FHI.250_00.010 to FHI.250_00.030; FHI.257_00.010 to FHI.257_00.030; FHI.264_00.010 to FHI.264_00.030; FHI.312_00.010 to FHI.317_00.010; FHI.325_00.010 to FHI.327_00.020; CAU.052_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; AAU.051_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.309_00.010 to AAU.309_00.050; AAU.500_00.010 to AAU.605_00.010
Disability	National Center for Health Statistics (NCHS) <sup>2</sup>	Disability Questions Test; Adult Functioning and Disability	FDB.020_00.000 to FDB.120_00.000 AFD.090_00.000 to AFD.560_00.000
Epilepsy	National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) <sup>2</sup>	Adult Epilepsy	ACN.192_00.010 to ACN.192_00.050
Crohn's Disease/ Ulcerative Colitis	National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) <sup>2</sup>	Adult Crohn's Disease/ Ulcerative Colitis	ACN.120_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	Centers for Disease Control and Prevention (CDC) <sup>5</sup>	Questions to Measure the Million Hearts® initiative	ACN.020_00.010; ACN.022_02.020 to ACN.023_00.020; ACN.023_03.030 to ACN.023_04.040; ACN.040_00.010 to ACN.040_00.040
Hepatitis	National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) <sup>2</sup>	Adult Hepatitis B/C Screening	AAU.365_00.010; AAU.405_00.010 to AAU.405_00.020
Immunization	National Center for Immunization and Respiratory Diseases (NCIRD) <sup>2</sup>	Adult and Child Immunization	AAU.310_00.000 to AAU.360_00.000; AAU.370_00.000 to AAU.400_00.010; AAU.410_00.010 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) <sup>5</sup>	Adult Internet Access and Email Usage	AWB.010_00.000 to AWB.050_02.000
Mental Health	Center for Mental Health Services (CMHS) <sup>3</sup>	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; ASI.390_00.000 to ASI.400_00.000
Tobacco	Centers for Disease Control and Prevention (CDC)	Infrequent Tobacco Use	AHB.081_00.000-AHB.085_00.000

<sup>1</sup>National Institutes of Health (NIH)<sup>2</sup>Centers for Disease Control and Prevention (CDC)<sup>3</sup>Substance Abuse and Mental Health Services Administration (SAMHSA)<sup>4</sup>Food and Drug Administration (FDA)<sup>5</sup>US Department of Health and Human Services (HHS)

---

## Interviewing 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. A telephone interview may also be conducted when the respondent requests a telephone interview or when road conditions or travel distances would make it difficult to schedule a personal visit before the required completion date.

Nationally, the NHIS uses about 750 interviewers (also called "Field Representatives" or "FRs") who are 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 Census Bureau's PANDA system, a performance and data analysis program that provides monthly checks on response rates, completion rates, item response times, item nonresponse, telephone usage rates, and other data quality indicators. The supervisors responsible for the NHIS are career Civil Service employees who are selected through an examination and testing process. Interviewers receive thorough refresher training annually and other training during the year in basic interviewing procedures and in the concepts and procedures unique to the NHIS.

Each household address selected for participation in the NHIS is mailed a letter prior to the interviewer's visit. This "advance letter" contains information about the purpose of the NHIS and the amount of time the interview will require, and it assures potential respondents that participation in the NHIS is voluntary. It also informs respondents that the information they provide is protected by law and details how the information will be used. When the interviewer arrives at the household address, he/she provides another copy of the advance letter to each respondent and obtains verbal consent for survey participation. A copy of this letter is included in Appendix I.

For the Household Composition section of the questionnaire, one household member who is at least the age of legal majority for the state of residence is identified as the "household respondent." In most states this age is 18 years, but in Alabama and Nebraska it is 19, and in Mississippi it is 21. 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.

The Family Core questionnaire is administered separately to each family in the household. For each family, 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.

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 increased chances of selection for any black, Hispanic, or Asian persons aged 65 years or older. The sample adult responds for him/herself to the questions in that section unless he/she is physically or mentally unable to do so, in which case a knowledgeable proxy is allowed to answer for the sample adult.

An emancipated minor is any person aged 14 years or older who has not attained 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 to be selected as the sample child or sample adult and are not generally 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 CAPI Reference Questionnaire (CRQ) 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.

All information collected by the NHIS that would permit identification of the individual is held strictly confidential, seen only by persons who work on the NHIS (including related studies carried out by the Public Health Service) with a need to know, and such information is not disclosed or released to anyone for any other purpose without the consent of the respondent. NCHS must adhere to Section 308(d) of the Public Health Service Act (42 U.S.C. 242m), which forbids the disclosure of any information that may compromise the confidentiality promised to survey respondents. In addition, confidentiality protections are also mandated by the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) (44 U.S.C. 3501 note).

Further information about data collection procedures is available in the 2015 NHIS Field Representative's (FR) Manual available on the NHIS website, <http://www.cdc.gov/nchs/nhis.htm>.

## Sample Sizes and Response Rates

The NHIS sample size can vary from year to year. The normal annual sample size (i.e., the number of households and/or persons for whom data are collected and publicly released) for the 2006–2015 sample design is about 35,000 households containing about 87,500 persons. The sample size can be reduced for budgetary reasons or increased because supplementary funding is available. In 2011–2015, 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%; in 2012, by approximately 21%; in 2013, by approximately 18%; in 2014, by approximately 28%; and in 2015, by approximately 19%.

The publicly released data files (also called “public use data files”) for the 2015 NHIS contain data for 41,493 households containing 103,789 persons in 42,288 families. The number of sample children is 12,291 and the number of sample adults is 33,672. In 476 cases, a knowledgeable proxy answered for the sample adult.

The total household response rate was 70.1%: 20.4 percentage points of the 29.9% noninterview rate were the result of respondent refusal and unacceptable “insufficient” partial interviews. The remaining 9.5 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 98.9%, which was calculated by dividing the number of completed family interviews (42,288) by the total number of eligible families (42,762). The unconditional or final response rate of 69.3% for the family component was calculated by multiplying the conditional response rate of 98.9% by the household response rate of 70.1%.

The conditional response rate for the Sample Child component was 91.4%, which was calculated by dividing the number of completed Sample Child interviews (12,291) by the total number of eligible sample children (13,444). The unconditional or final response rate of 63.4% for the Sample Child component was calculated by multiplying the conditional rate of 91.4% by the final family response rate of 69.3%.

The conditional response rate for the Sample Adult component was 79.7%, which was calculated by dividing the number of completed Sample Adult interviews (33,672) by the total number of eligible sample adults (42,270). The unconditional or final response rate of 55.2% for the Sample Adult component was calculated by multiplying the conditional rate of 79.7% by the final family response rate 69.3%.

No compensation or other incentives were provided for participation in the NHIS. However, in May, June and July of 2015 an NHIS incentives experiment was implemented in the states served by three of the six Census regional offices. The three Census regional offices selected were New York, Philadelphia, and Denver. Preliminary results show no impact of incentives on response rates, except when the NY regional office was examined separately. More extensive results from this experiment were presented at the 71<sup>st</sup> Annual Conference of the American Association for Public Opinion Research and will be made available in publications currently in preparation.

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



---

## Overview of the 2015 Data Files and Data File Documentation

In 2015, the following data files are available: Household File, Family File, Person File, Injury and Poisoning Episode File, Sample Child File, Sample Adult File, Sample Adult Cancer File, Adult Functioning and Disability File, Family Disability Questions File, and the Paradata File. (The latter contains data about the NHIS data collection process and is not discussed further in this document.) Data based on most supplementary questions are released in the Core data files in 2015. However, two stand-alone data files have been created and can be linked to other NHIS 2015 data files: the Adult Functioning and Disability File and the Family Disability Questions File. The files and documentation can be found at: <http://www.cdc.gov/nchs/nhis/familydisabilityquestions2015.htm> and <http://www.cdc.gov/nchs/nhis/funcdisb2015.htm>, respectively. A description of each of these files follows this section.

### Editing the Data During and After the Interview

Range and consistency checks may be programmed into the CAPI system to help prevent both interviewer and respondent error. During the interview, if an interviewer enters an out-of-range value (such as 180 years instead of 18 for age), an error message might appear on the interviewer's computer screen, instructing the interviewer that the data cannot be accepted and that a new value must be entered. Such an interruption of the interview is called a "hard edit" if the interview cannot continue without an acceptable response being entered, and a "soft edit" if the interview may continue whether or not a meaningful new response is entered.

Even with such checks built into the CAPI system, data cleaning (data "editing") is still necessary. The first step in the data cleaning process is verification of the valid number of cases in the data file. After verifying the number of cases, initial data frequencies are produced and reviewed for reasonableness. Each variable's range of permissible values is examined for any additional invalid values or unusual distributions. Invalid values, where they occurred, are deleted. If blank values already exist for a variable, they are checked to see whether they are allowable (e.g., due to legitimate skip patterns in the questionnaire) or could be easily corrected based on related questions. Records that are missing responses for unknown reasons are left missing.

### Edits to Protect Confidentiality

As noted earlier, NCHS (including its contractors and agents) collects personally identifiable NHIS and other survey data under a pledge of confidentiality and a promise that the data will be used only for statistical purposes. Section 308d of the Public Health Service Act and Section 512b of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) require that confidentiality be maintained without exception. Violations of CIPSEA are a class E felony, punishable by imprisonment for not more than 5 years, a fine of not more than \$250,000, or both. Strict procedures in survey operations and data dissemination are used by NCHS, its data collection contractors, and other agents to prevent disclosure of survey subjects' identities.

The risk of inadvertent disclosure of confidential information regarding individual respondents is higher when there exists a publicly released data set having detailed geography variables and a detailed and extensive set of survey observations. For this reason, the NHIS does not publicly release state identifiers and some other geographic variables, and the original design strata and PSUs are masked when the data are publicly released (see Appendix III for more detail about the masked design variables). NHIS data sets may also be coarsened by suppressing survey variables, collapsing multiple variables into one, and collapsing response categories. For example, very extreme values of height and weight are masked by being recoded into broad interval categories.

In addition, statistical noise at both the variable level and record level may occasionally be added to protect confidentiality.

The descriptions of the data files that follow include details about edits and data suppression that were done to protect the confidentiality of NHIS participants. However, one important edit is worth noting here because it applies to multiple variables across the survey. To protect confidentiality among the oldest sample adults, all age variables were top-coded to 85+ years and all year variables (such as year of birth or year of diagnosis) were bottom-coded to  $\leq 1931$ . For example, survey questions related to age at diagnosis for cancer (CANAGE1–30) and diabetes (DIBAGE) (“How old were you when you were diagnosed [with this condition]?”) are top-coded to 85+ years. The recode DIFAGE2 is also top-coded to 85+ years to ensure confidentiality.

Analysts interested in working with data that were suppressed or edited to protect confidentiality may apply to access selected unmodified data files through the NCHS Research Data Center (RDC). The RDC is a data enclave established to provide a mechanism whereby researchers can access detailed data files in a secure environment without jeopardizing the confidentiality of survey participants. Information about RDC access options and application procedures is available at: <http://www.cdc.gov/rdc/>.

## 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>. For each survey year, the website provides the *Survey Description*; the Questionnaire Reference Guide, survey questionnaires, the Field Representative Manual, and survey flowchart; information on supplements and their sponsors; summary health statistics (once they become 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 Child, Sample Adult, and Injury/Poisoning Episode Files; Imputed Income Files; the Paradata File; and any supplemental files released that year.

Each of the 2015 data release files includes the following documents. A description of the first three documents follows below the list.

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

### ***Variable Summary Report:***

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.

### ***Variable Layout Report:***

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. Every effort was made to make the variable names in the data 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.

Additional specific information is provided under “Recodes,” “Sources,” “Keywords,” “Notes,” and “Universe.” These terms are defined below.

**Recodes** - A *recode* is a variable derived from the reordering, collapsing, or verbatim coding of another variable. Alternatively, a recode may be constructed from two or more variables. Any variables used to construct a recode are listed as a cross reference. Users will note that a number of standardized variables appear in the NHIS 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 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 number of time units and 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 (i.e., the maximum reported value may be capped) for confidentiality reasons.

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

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

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

- 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.
- 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.

**Universe** - The *universe* refers to the group of persons to whom a specific question applies. For example, the universes for most sample adult variables are persons who were selected as the sample adult who were age 18 or over. This universe is specified on the Variable Layout Report as ASTATFLG = 1 and (AGE GE '018' and AGE not IN ('997', '999')). Sample adults who are not eligible to answer a given question are considered to be not-in-universe. For example, sample adults who reported that they 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 they had surgery in the past 12 months.

Universes for each question are often defined by age. 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, it may appear that a person in-universe was not asked to answer a given question for which he should have been eligible. Though neither the specified universes nor the variables affected are changed when ages are corrected, users can refer to a new variable, AGE\_CHG, which indicates a correction has been made on the record. Occasionally, universe inconsistencies may also exist due to collection or processing errors.

***Variable Frequency Report:***

The Variable Frequency Report provides the frequencies and percents for each NHIS variable. Since 2005, and ongoing in 2015, all response categories are shown in the Variable Frequency Report, including those response categories with a zero count in the data files. For continuous variables, a range of values is provided. 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.

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 coded as 9 (with leading 9’s to the length of the field, as in 9, 99, 999, etc.).

For partially completed interviews (e.g., Sample Adult interviews where the respondent discontinued the interviews any time after completing the first three sections), the responses will appear as 8’s (not ascertained) for the remaining variables in the file (see Appendix II for more information). A code of 8 is also used to indicate “not ascertained” responses when the field was blank or 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).

Analysts are advised to read the notes in the 2015 data release documentation for further information pertaining to any changes that may have occurred, and to compare the 2015 data release documentation to documentation from the 2014 (and earlier) NHIS for any other changes that may have occurred over time to the variables in this section. Changes include but are not limited to: variable name changes due to modifications of question wording or response codes; introduction of new variables during the year; deletion of some of the previous year’s variables; and variables moving from one section to another. The Variable Layout Reports for each year can be found on the NHIS website:

[http://www.cdc.gov/nchs/nhis/quest\\_data\\_related\\_1997\\_forward.htm](http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm).

---

## Household File

The Household File is derived largely from the Household Composition section of the Core and describes characteristics of each household. This file is considered as the base file from which all other files are built. 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. Most “Type A” non-responding households were eligible for the NHIS interview but were 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 II.) 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 59,170 households: 41,493 households were interviewed, while 17,677 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.

## Family File

The Family File contains variables that describe characteristics of the 42,288 families living in households that participated in the 2015 NHIS. The variables contained in the Family file are reconstructions of the person-level data from the Core sections at the family level. 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 2015 survey; this is specified as FM = ALL in the Family File Variable Layout Report. 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 3 indicates, 98.5% 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 3. Number of families per household, 2015 National Health Interview Survey (unweighted counts)**

Families per household	Frequency	Percent
1	40,854	98.5
2	527	1.3
3	84	0.2
4	15	0.0*
5	10	0.0*
6	3	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 4 shows a breakdown of the 42,288 families by number of family members. 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. This definition was first used in 2011. Previously, data were not collected on any infant who was born during the assignment week of the interview.

**Table 4. Size of families, 2015 National Health Interview Survey (unweighted counts)**

Number of members	Frequency	Percent
1	13,109	31.0
2	13,403	31.7
3	6,192	14.6
4	5,360	12.7
5	2,554	6.0
6	1,043	2.5
7	381	0.9
8	150	0.4
9	43	0.1
10	27	0.1
11	14	0.0*
12	9	0.0*
14	1	0.0*
15	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 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 several Family Core sections of the NHIS that collapse the individual-level observations into information about their respective families.

Generally, the Family File consists of two types of recodes: a simple “yes/no” measure that indicates whether any family member falls into a particular category or exhibits a particular characteristic, and a corresponding counter that indicates the number of family members in that category or with that characteristic. Note that counters always consist of values from zero to 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: 16,203 families have one member receiving income from wages/salary, 11,764 families have two members (or  $2(11,764)=23,528$  persons) with wage/salary income, 1,871 families have three members (or  $3(1,871)=5,613$  persons), 505 families have four members (or  $4(505)=2,020$  persons), 84 families have five members (or  $5(84)=420$  persons), 14 families have six members (or  $6(14)=84$  persons), 4 families have seven members (or  $4(7)=28$  members), and 2 families have nine members (or  $2(9)=18$  persons with wage/salary income in 2015). Thus, the sum of persons across the 30,447 families answering “yes” to FSALYN, the associated yes-no indicator,

is 47,914 (16,203 + 23,528 + 5,613 + 2,020 + 420 + 84 + 28 + 18), which is equal to the 47,914 “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 2015 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 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 FM\_STRP and FM\_STRCP, 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.



---

## Person File

The Person File variables are derived from the 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 file for more information.) The sections comprising the Family Core are discussed in greater detail below.

### I. Coverage Section: Telephone (COV)

#### Core Content

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.

Information was collected on whether a telephone number was supplied; presence of working landline inside the home; presence of working cell phone in family; number of working cell phones; and relative frequency of receipt of calls on cell and landline.

### II. Household Composition Section (HHC)

#### Core Content

The Household Composition (HHC) section of the 2015 data file contains information collected in the household roster portion of the NHIS interview, which identifies all members of the household and collects essential demographic information on them.

The HHC section of the 2015 NHIS person file includes the variables sex, race, ethnicity, armed forces status (whether or not a household member is currently serving in the military), month and year of birth, and age, as well as variables describing the relationship of the person in the roster to both the household reference person and the family reference person.

#### Suppression of Data

To protect confidentiality, detailed information on the ages of persons aged over 85 years is not available on the public use data file. These persons have a code of 85 on the AGE\_P variable. Detailed information on these persons' ages is available through the NCHS RDC at <http://www.cdc.gov/rdc>. In addition, to protect confidentiality, month of birth has been suppressed and is not available on the public use files.

## Technical Notes

In accordance with the Office of Management and Budget's 1997 standards for the collection of race and ethnicity data in federal data systems, separate questions are asked about Hispanic origin and race. Persons of Hispanic origin may be of any race or combination of races. Hispanic origin includes persons of Mexican, Puerto Rican, Cuban, Central and South American, or Spanish origin. Race is based on the family respondent's description of his or her own race background, as well as the race background of other family members. More than one race may be reported for a person.

Table 5 summarizes the Hispanic origin and race variables in the 2015 data file. Details on the specific response categories for the race questions and additional details on these variables can be found in the 2015 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.

**Table 5. Race/ethnicity variable names and description, 2015 National Health Interview Survey public use data files**

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	"Other race" category included for bridging purposes. Values were imputed for some records.
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.

For additional important information about the historical context of race and ethnicity data collection in the NHIS, editing, references for NHIS race and Hispanic origin, and important background documents related to race and ethnicity data collection in federal data systems, we recommend that you read Appendix II of the 2012 NHIS *Survey Description*:

[ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2012/srvydesc.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2012/srvydesc.pdf).

---

### III. Family Identification Section (FID)

#### Core Content

The Family Identification (FID) section of the Family Core in the 2015 NHIS collects information on marital status (for all family members age 14 years or older); legal marital status for family members who are living with a partner; relationship data; and whether children under age 18 who are not emancipated and not living with their parents have legal guardians.

Information in the FID section is dependent on the creation of the household roster (RRP) in the Household Composition (HHC) section that precedes the FID section in the Family Core. All relationships specified in RRP are relative to the household reference person, who was identified in the HHC section as the person (or one of the persons) age 18 or older who owns or rents the household being interviewed and who is the first person listed on the roster. Subsequent variables in the FID section describe the type of the relationship between family members and the reference person. For example, if one family member was identified as a brother or sister of the reference person (RRP = 8), SIB\_DEGP indicates the type of sibling relationship (full or adopted, half, step, or in-law) that exists between them. Information is also collected from each family member as to whether his/her mother and father are household members; if so, MOM\_DEGP and DAD\_DEGP describe the type of relationship (biological or adoptive, step, or in-law) between that person and his/her parent(s).

#### Major Recodes

The FID section of the NHIS Person File contains several useful recodes, including MOM\_ED and DAD\_ED, which provide information on the education of the mother and father, respectively, for all family members under age 18. The section also includes a recode called PARENTS, which indicates, for each family member, whether their household contains their mother but no father; their father but no mother; both their mother and father; or neither their mother nor father. Note that this recode does not distinguish between biological, adoptive, step, and foster parents.

In addition, person-level information in the FID section is used to create a number of useful family-level recodes in the NHIS Family File. These include FM\_SIZE (size of family), FM\_KIDS (number of family members under 18 years of age), FM\_ELDR (number of family members age 65 or older), and FM\_TYPE, FM\_STRCP, and FM\_STRP, which provide detailed information about family structure. Unlike the PARENTS recode in the Person File, FM\_STRCP and FM\_STRP take into account the type of relationship between children and parents as well as the marital status of the parents. More information on these recodes can be found in the description of the Family File (under the sub-head “Family Structure Variables”).

#### Suppression of Data

To protect confidentiality, detailed information on family structure, type of sibling and/or parent-child relationship, and the presence and identity of legal guardians for children under age 18 is not included on the public use NHIS data files. Analysts wishing to use the unmodified restricted data can apply for access to the NCHS RDC at: <http://www.cdc.gov/rdc/>.

---

## Technical Notes

The FID section contains the variables CSTATFLG and ASTATFLG, which allow analysts to easily identify the sample children and sample adults in the Person File. These flags can also be used to subset the Person File, which is not generally recommended (see Appendix III) but can make it easier to merge the Person File with the Sample Child or Sample Adult Files.

Information regarding family structure and relationships can be found on both the NHIS Person and Family Files. The Person File contains variables and recodes that describe the relationship of individual family members to the household reference person and to parents residing in the home. The Family File contains counters that indicate the number of family members, as well as the number of children and number of elderly members per family. Detailed recodes also describe each family's structure. Analysts should consult the documentation for both the NHIS Person and Family Files to determine how to effectively use NHIS data on family relationships; Appendix IV provides guidance on how to merge these files.

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

### Core Content

The 2015 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). Since the questions in the FHS section only allow for "yes" or "no" responses, the degree to which a person is limited is not determined.

Since cognitive impairment is increasingly recognized as a source of activity limitations among older adults, the FHS section includes an indicator that identifies persons who are limited because of difficulty remembering or periods of confusion. Other indicators in this section identify persons 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.

### *Conditions Leading to Limitations of Activity*

For each person with a 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 children 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.”

If an adult family member was limited by a condition not listed in one of the 18 fixed categories mentioned above, 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.

## Supplemental Questions

A supplemental question about whether the receipt of Special Education or Early Intervention Services was due to an emotional or behavioral problem was included.

## Major Recodes

The recode LA1AR is a summary measure that identifies persons with 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 PLAPLYM, PSPANDEIS, 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. 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).

## Technical Notes

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 children under age 18, or to one of the 18 fixed categories for adults. 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\_. 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 6 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 initially based on the International Classification of Diseases, Ninth Revision, Clinical Modification. ICD-9-CM codes were converted to the International Classification of Diseases, Tenth Revision, Clinical Modification for 2015. Table 7 shows the final FHS categories with approximate ICD-10-CM ranges. ICD-10-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 adults) remained in the “other impairment problem” categories, and were renumbered “90” and, if necessary, “91” for both children and adults. Note that the verbatim responses associated with the “other impairment problem” categories are not included as a separate field on the public use data file.

**Table 6. Reassignment of “second screen” adult condition categories**

Screen item		Assigned variable	
Question name	Description	Variable name	Description
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/quadruplegia	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

The specific condition categories as well as the “other impairment problem” categories were subsequently transformed into variables indicating whether or not the condition was thought by the respondent to be responsible for the person’s difficulty with any activity (a mention/not-mention format).

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.”

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.”

**Table 7. FHS categories with approximate ICD-10-CM ranges**

Age group and category	ICD-10-CM codes
Adults (age 18 and over)	
1 - Vision or seeing problem	E08.36, E09.36, E10.36, E11.3, E11.36, E13.36, H00–H02, H04–H05, H10–H11, H15–H18, H20–H21, H25–H28, H30–H31, H33–H36, H40–H44, H46–H47, H49–H55, H57, Q15.0, R48.3
2 - Hearing problem	H80, H83, H90–H93
3 - Arthritis / rheumatism	M00, M01X, M02.1, M02.3, M05.0, M05.1, M05.3, M05.6, M06.1, M06.4, M06.9, M08.0, M08.3, M08.4, M11.8, M11.2, M12, M13, M15, M16.1, M16.7, M16.9, M17.1, M17.5, M17.9, M18.9, M19, M35.2, M45.9, M46.0, M46.1, M46.8, M46.9, M47.1, M47.8, M48.1–M48.3, M48.9, M49.8, M70–M79
4 - Back or neck problem	M40, M42.0, M43.2, M43.6, M43.8, M46.4, M48.0, M50.0, M50.2, M50.3, M50.8, M50.9, M51.2–M51.4, M51.8, M51.9, M53, M54, M67.88, M96.1–M96.5
5 - Fractures, bone or joint injury <i>Injury with specific mention of bone or joints</i>	with specific mention of bone or joint : S01–S03, S06, S11–S14, S21–S24, S31–S34, S41–S43, S46, S51–S53, S56, S61–S63, S66, S71–S73, S76, S81–S83, S86, S91–S93, S96, T07, T14.8
6 - Other injury <i>Injury without specific mention of bone or joints</i>	without mention of bone or joint: D78.0–D78.2, E36.0, E36.1, E89.89, G97.0, G97.3–G97.5, G97.8, H59.2, H59.8, H95.3, I97.3–I97.8, J95.7, J95.8, K91.3, K91.7, K91.8, H59.1–H95.4, J95.6, K91.6, L76.0–L76.2, M96.81, M96.83, N99.89, S00, S01, S04, S05, S06.0, S06.1, S06.3–S06.6, S06.8, S06.9, S07.0, S07.8, S08–S11, S14, S15.0, S15.2, S15.3, S15.8, S15.9, S17.9, S19.9, S20, S21, S24, S25, S26.0, S26.1, S26.9, S27.0–S27.4, S27.8–S28.0, S29.8, S30, S31, S34–S37, S38.0, S38.1, S39.8, S40, S41, S44, S45.0–S45.2, S45.8, S45.9, S46, S47.9, S48, S49.8–S51, S54, S55.0, S55.1, S55.8–S57.0, S57.8, S58, S59.8–S61, S64–S67.3, S68, S69.8–S71, S74–S77.2, S78, S79.8–S81, S84–S87.0, S87.8, S88, S89.8–S91, S94–S97.1, S97.8, S98, S99.8, S99.9, T07, T14.8, T15–T34, T36–T55, T57–T65.4, T66–T80, T81.1–T83.6, T83.8–T87.0–87.4, T87.8–T88, T88.8
7 - Heart problem	I20.0, I20.1, I20.8, I20.9, I21, I23–I28, I30–I34.0, I34.8, I35–I39, I42–I45, I46.9, I47, I48.9–I51, I97, Q20–Q24, R00, R01
8 - Stroke problem	G45, I60–I63.5, I65–I67
9 - Hypertension or high blood pressure	I10–I15
10 - Diabetes	E10, E11, E13.1

**Table 7. FHS categories with approximate ICD-10-CM ranges**

Age group and category	ICD-10-CM codes
11 - Lung or breathing problem	A22.1, A37.91, A48.1, B25.0, B44.0, B44.81, J00, J01, J06, J10–J18, J20, J21, J30–J33, J34.2, J35–J37, J39.8, J40–J47, J60–J64, J66–J70, J80–J82, J84–J86, J90, J91, J93–J96, J98, J99, M34.81, R09.1
12 - Cancer	C00–C26, C30–C34, C37–C41, C43, C44, C46, C47, C50–C57, C60–C96, D00–D04, D06, D07, D09
13 - Birth defect <i>Excludes Down syndrome and microcephalus</i>	E78.7, G90.1, P29.3, Q00.0–Q00.2, Q01.9, Q02–Q07.0, Q07.8, Q07.9, Q10–Q18, Q20–Q28, Q30–Q45, Q50–Q52, Q53.1, Q53.2, Q53.9, Q54–Q56, Q60–Q67, Q68.2, Q68.8, Q69–Q85, Q87, Q89–Q99
14 - Intellectual disability, also known as mental retardation <i>Includes Down syndrome and microcephalus</i>	F70–F73, F79, Q02, Q91.4–Q91.6
15 - Other developmental problem <i>Includes learning disabilities</i>	F80–F82, F88, F89, G80, H93.25, R48.0, R62.0, R62.5
16 - Senility (and other cognitive problems)	F01, F03, F05
17 - Depression, anxiety or emotional problem <i>Includes neurotic disorders, personality disorders, and other nonpsychotic mental disorders, excluding alcohol and drug related problems and developmental problems</i>	F07, F09, F21, F32.9, F34, F40–F45, F48, F50.0, F50.2, F50.8, F50.9, F51, F52, F60, F63–F66, F68.1, F68.8, F91.1–F91.3, F91.8, F91.9, F93.0, F93.8, F94.0, F94.1, F94.8, F98, G44.209
18 - Weight problem <i>Indicates a problem with being overweight or obese as defined by the respondent</i>	...
19 - Missing limbs (any part) / amputee <i>Indicates loss of a limb or digit</i>	...
20 - Other musculoskeletal system conditions <i>Diseases of the musculoskeletal system and connective tissue not coded to 3, 4, 5</i>	D48.1, M00–M01, M05, M06, M08, M11–M21, M22.4, M23, M24, M25, M32.1, M33.2, M33.9, M34, M35, M36.2–M36.4, M40, M42.0, M42.1, M43, M45–M48, M49.8, M50, M51, M53, M54, M61, M62, M65, M67.8, M70.0, M70.3, M70.4, M71, M75–M77, M79, M60.0, M60.2, M66, M67.0, M67.4, M67.5, M67.8, M67.9, M75.12, M80.0, M81, M84.3, M84.4, M84.6, M85–M87, M88.9, M89, M90, M91.8, M92–M95, M96.1–M96.5, M99, R25.2, R26.2, R29.4, R29.898, S42.009P, S42.209P, S42.90XP, S52.90XP, S52.90XQ, S52.90XR, S62.90XP, S72.90XP, S72.90XQ, S72.90XR, S82.009P, S82.009Q, S82.009R, S82.90XP, S82.90XQ, S82.90XR, S92.909P, S92.919P, S02.91XK, S02.92XK, S12.000K, S12.001K, S12.100K, S12.101K, S12.200K, S12.201K, S12.300K, S12.301K, S12.400K, S12.401K, S12.500K, S12.501K, S12.600K, S12.601K, S22.9XXK, S32.9XXK, S42.009K, S42.209K, S42.90XK, S52.90XK, S52.90XN, S52.90XM, S62.90XK, S72.90XK, S72.90XM, S72.90XN, S82.90XK, S82.90XM, S82.90XN, S92.909K, S92.919K
21 - Other circulatory system conditions <i>Any diseases of the circulatory system not coded to 7, 8, 9</i>	G45, I00–I02, I05–I13, I15, I20, I21, I24, I28, I30–I51, I60–I63, I65, I66, I67.2, I67.4–I78, I79.8, I80–I83, I85–I87, I89, I95, I97, I99.8, I99.9, K64, M30, M31, R00, R01, R58
22 - Other endocrine system, etc. conditions <i>Any Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders not coded to 10 or 18</i>	C88.0, D47.2, D80–D84, D89, E00, E01, E03–E13, E15, E16, E20.9, E21–E32, E34, E40–E46, E50–E61, E63, E64.3, E65–E67, E70–E72, E74–E80, E83–E89, M10, M1A, M35.9, M83.9, N20.0, N98.1
23 - Other Nervous system conditions <i>Diseases of the nervous system and sense organs not coded to 1, 2, 15, 16</i>	C88.0, D47.2, D80–D84, D89, E00, E01, E03–E13, E15, E16, E20.9, E24–E32, E34, E40, E41, E43–E46, E50–E59, E60, E61, E63, E64.3, E65–E67, E71–E72, E74–E80, E83–E89, M10, M1A, M35.9, M83.9, N200, N98.1



**Table 7. FHS categories with approximate ICD-10-CM ranges**

Age group and category	ICD-10-CM codes
24 - Digestive system conditions	A18.59, B45.1, E08.4, E09.3, E09.4, E10.3, E10.4, E11.3, E11.4, E13.3, E13.4, E75, F51.05, F51.13, F84.2, G00–G12, G20, G21, G23–G25, G30–G32, G35–G37, G40, G43, G44, G47, G50–G52, G54, G56, G57, G58.7, G58.9, G60–G63, G70–G73, G80–G83, G89–G95, G99, H01, H00, H02, H04, H05, H10, H11, H15–H18, H20, H21, H25–H28, H30, H31, H33–H36, H40, H42–H44, H46, H47, H49–H54, H57, H60–H62, H65–H75, H80–H83, H90–H95, G96.0, Q15.0, R48.3, R52
25 - Genitourinary system conditions	N00–N08, N10–N21, N25–N28, N30–N53, N60–N65, N70–N77, N80–N98, R10.2, R31, R36.1, R80.2
26 - Skin & subcutaneous system conditions	K122, L00–L08, L10–L14, L20–L30, L40–L44, L49–L53, L55–L58, L60, L63, L65–L67, L70, L71–L75, L80–L85, L87, L89, L90–L95, L97, L98
27 - Blood & blood-forming organ conditions	D50–D53, D55–D76, D89.2, I88
28 - Tumors & cysts, benign & unspecified <i>Any mention of "tumor" without cancer, malignancy, etc.</i>	C94.4, D10, D12–D19.0, D20, D21, D23–D32, D34–D49, 85.0
29 - Alcohol & drug related problems <i>Any mention of "alcohol," "drugs" (or specific drug types), or substance abuse</i>	F10–F16, F17.2, F18, F19
30 - Other mental conditions <i>Any mental disorders not coded to 14 or 15 or 17</i>	F01, F02.80, F02.81, F03–F06, F20, F22–F24, F25.9, F28–F32, F33.3, F34.8, F39, F44.89, F53, F84, F90
31 - After effects of surgery or other medical treatment <i>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.</i>	...
32 - Old age <i>Any mention of age as the only specified cause</i>	...
33 - Fatigue/Tiredness <i>Any mention of tiredness, stiffness, or weakness without referring to any specific part of the body</i>	...
34 - Pregnancy related conditions <i>Any mention of "pregnancy" or "childbirth"</i>	...
90 - Others Not Elsewhere Classified <i>1st other-specify verbatim, not elsewhere classified</i>	...
91 - Others Not Elsewhere Classified <i>2nd other-specify verbatim, not elsewhere classified</i>	...
Children (age 17 and under)	
1 - Vision or seeing problem	E08.36, E09.36, E10.36, E11.3, E13.36, H00–H02, H04–H05, H10–H11, H15–H18, H20–H21, H25–H28, H30–H31, H33–H36, H40–H44, H46–H47, H49–H55, H57, Q15.0, R48.3
2 - Hearing problem	H80, H83, H90–H93
3 - Speech problem	F63.3, F80, F98.5, H93.25, R45.1, R47.01, R47.02, R47.1, R47.81, R47.82, R47.89
4 - Asthma or breathing problem	A22.1, A37.91, A48.1, B25.0, B44.0, B44.81, J00, J01, J06, J10–J18, J20, J21, J30, J32, J33, J34.2, J39.8, J40–J47, J60–J64, J66–J70, J80–J82, J84–J86, J90, J91, J93–J96, J98, J99, M34.81, R09.1
5 - Birth defect <i>Excludes Down syndrome and microcephalus</i>	E78.71, E78.72, P29.3, Q00.0–Q00.2, Q01, Q03, Q04.1–Q04.3, Q04.5–Q04.9, Q05, Q06, Q07.01, Q07.03, Q07.8, Q07.9, Q10–Q56, Q60–Q84, Q85.1, Q85.8, Q87, Q89, Q91.7, Q91.3, Q92.8, Q93.3, Q93.4, Q93.7, Q93.8, Q95.0, Q96.9–Q98, Q99.9

**Table 7. FHS categories with approximate ICD-10-CM ranges**

Age group and category	ICD-10-CM codes
6 – Injury	D78.0–D78.2, E36.1, G97.0, G97.3–G97.5, G97.8, H59.2, H59.8, H95.3, I97.3, I97.5, I97.7, I97.8, J95.7, J95.8, K91.3, K91.7, K91.8, N99.89, E89.89, E36.0, G97.3, H59.1–H95.4, I97.4, I97.6, J95.6, K91.6, L76.0–L76.2, M96.81, M96.83, S00–S07.0, S07.8, S08–S15.0, S15.2, S15.3, S15.8, S15.9, S17.9, S19.9–S26.1, S26.9, S27.0–S27.4, S27.8–S28.0, S29.8, S30–S38.1, S39.8, S40–S45.2, S45.8–S46, S47.9, S48, S49.8–S55.1, S55.8–S57.0, S57.8, S58, S59.8–S67.3, S68, S69.8–S73–S77.2, S78, S79.8–S88, S89.8–S97.1, S97.8, S98, S99.8, S99.9, T07, T14.8, T15–T34, T36–T55, T57–T65.4, T66–T80, T81.1–T83.6, T83.8–87.4, T87.8–T88
7 - Intellectual disability, also known as mental retardation <i>Includes Down syndrome and microcephalus</i>	F70–F73, F79, Q02, Q91.4–Q91.7
8 - Other developmental problem	G80, R62.0, R62.5
9 - Other mental, emotional, or behavioral problem	F01.5, F03.90, F05, F07, F09–F48, F50–F52, F59, F60, F63–F66, F68.12, F84.0, F84.3–F84.9, F90–F98, R37, R45, Z91.4
10 - Bone, joint or muscle problem	D48.1, M00–M02, M05, M06, M08, M11–M21, M22.4, M23–M25, M32.1, M33.2, M33.9, M34, M35, M36.2–M36.4, M40, M42.0, M42.1, M43, M45, M46.0, M46.1, M46.3, M46.4, M46.8, M46.9, M47, M48, M49.8, M50, M51, M53, M54, M60.0, M60.2, M61, M62, M65–M67.0, M67.4, M67.5, M67.8, M67.9, M70.0, M70.3, M70.4, M71, M75–M77, M79, M80.0, M81, M84.3, M84.4, M84.6, M85–M87, M89, M90, M91.8, M92–M95, M96.1–M96.5, M99, R25.2, R26.2, R29.4, R29.898, S42.009P, S42.209P, S42.90XP, S52.90XP, S52.90XQ, S52.90XR, S62.90XP, S72.90XP, S72.90XQ, S72.90XR, S82.009P, S82.009Q, S82.009R, S82.90XP, S82.90XQ, S82.90XR, S92.909P, S92.919P, S02.91XK, S02.92XK, S12.000K, S12.001K, S12.100K, S12.101K, S12.200K, S12.201K, S12.300K, S12.301K, S12.400K, S12.401K, S12.500K, S12.501K, S12.600K, S12.601K, S22.9XXK, S32.9XXK, S42.009K, S42.209K, S42.90XK, S52.90XK, S52.90XN, S52.90XM, S62.90XK, S72.90XK, S72.90XM, S72.90XN, S82.90XK, S82.90XM, S82.90XN, S92.909K, S92.919K
11 - Epilepsy and seizures	G40, P90, R56
12 - Learning disability	F80–F82, F88, F89, H93.25, R48.0,
13 - Attention Deficit/Hyperactive Disorder (ADD/ADHD)	F90
90 - Others Not Elsewhere Classified <i>1st other-specify verbatim that does not fit in any other category</i>	...
91 - Others Not Elsewhere Classified <i>2nd other-specify verbatim that does not fit in any other category</i>	...
... Category not applicable	

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 to evaluate variables that may be more appropriate.

---

## V. Family Food Security Section (FFS)

### Supplemental Questions

Family food security refers to access by all persons in a family at all times to enough food for active, healthy lives. The Family Food Security (FFS) section, sponsored by the United States Department of Agriculture (USDA), consists of 10 questions addressing adult 30-day food security. The purpose of the questions is to examine the relationship between health and food insecurity. The FFS section has been included in the NHIS since 2011.

The content of the FFS section includes being worried that food would not last, food not lasting until there was money to buy more, not eating balanced meals, eating less than should, being hungry but not eating, losing weight because there was not enough food, cutting or skipping a meal, not eating for a whole day, the number of days that a meal was skipped, and the number of days that did not eat for a whole day. All food security information is provided by the family respondent in reference to the past 30 days.

The procedures for calculating raw food security scores and food security classification variables can be found in Appendix V.

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

### Core Content

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

Questions are included that ask about delay or non-receipt of health care because of worry about the cost, overnight hospital stays, home care, calls to health professionals, and office visits. Questions on home care and health care contacts include 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. Please note that respondents were instructed to exclude dental contacts when considering their responses to the questions.

### Technical Notes

A few large values were found for the variables pertaining to hospitalizations (HOSPNO) and hospital nights (HPNITE). In addition, large numbers may exist for variables pertaining to 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.

## VII. Health Insurance Section (FHI)

### Core Content

The Health Insurance section of the 2015 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 1997–2014 NHIS Family Cores.

The FHI core section covers several different topic areas:

- Type of health care coverage (Medicare, Medicaid, Children’s Health Insurance Program (CHIP), military (TRICARE, VA, CHAMP-VA), other state-sponsored health plans, Indian Health Service, other government programs, private insurance and single service plans);
- Managed care arrangements for those covered by Medicaid, Children’s Health Insurance Program, other state-sponsored health plans and other government programs;
- Medicare managed care (Medicare Advantage plans) and the need for a referral;
- Enrollment in the Medicare Part D program;
- Private insurance characteristics reported by the family respondent, including whether it came through the Federal Health Insurance Marketplace or state-based exchanges, covered individuals’ relationships to the policyholder, coverage of individuals outside the household, HMO, PPO, and POS status, high deductible health plan (annual deductible of \$1,300 for self-only coverage and \$2,600 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 individual/family, managed care detail information, prescription drug benefit, dental coverage;
- Types of single service plans;
- Type of TRICARE coverage;
- Periods of time without health insurance and reasons for no health insurance;
- Out-of-pocket costs in the past year for medical expenses (excluding health insurance premiums); and
- Enrollment in a flexible spending account (FSA) for medical expenses.

New questions were added to the FHI core section in 2014 to obtain details about whether plans (both private and public) were obtained through Healthcare.gov or the Health Insurance Marketplace, whether there is an enrollment fee or premium (public only), and whether the premium is based on income (expanded to include public plans). This follow-up information is now collected for all types of plans—Medicaid, CHIP, other state programs, and other government programs, as well as private plans—and is intended to help correctly classify plans that may have come through the Federal Health Insurance Marketplace or state-based exchanges (MXCHNG, MEDPREM, MDPRINC, PLNEXCH1, PLNEXCH2, PLNPRE1, PLNPRE2, CHXCHNG, STRFPRM1, CHPRINC, OPXCHNG, STRFPRM2, SSPRINC, OGXCHNG, STRFPRM3, OGPRINC).

---

## Supplemental Questions

In addition to the core health insurance questions, additional supplemental questions sponsored by the Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS) were included starting in 2014. The purpose of the FHI supplemental questions is to provide additional information about the following topic areas:

- How private plans were obtained;
- Relationship of the covered individual to the policyholder;
- If premiums were based on income;
- Primary care provider requirement;
- Confidence in purchasing affordable private health insurance on one's own;
- Previous health insurance coverage for persons who are either uninsured or who have had a change in insurance coverage within the past year;
- Difficulty paying medical bills, paying medical bills over time, and having medical bills that cannot be paid at all.

Beginning in 2011, the FHI section has contained numerous 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 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. These variables address a major provision of the ACA that became effective September 23, 2010 that allows parents to keep their children on their health policies until they turn age 26.

FHI variables are also created for the Family File. The inclusion of the FHI section on the Family File is intended to make it easier to analyze insurance data on a family basis.

## Major Recodes

Recodes are created for types of health care coverage and noncoverage because of the complicated editing process that takes place in the FHI section. The recodes are: MEDICARE, MEDICAID, PRIVATE, SCHIP, IHS, MILCARE, OTHPUB, OTHGOV, SINGLE, and NOTCOV. NOTCOV reflects the definition of noncoverage as used in *Health, United States* (in which persons with *only* Indian Health Service coverage are considered uninsured). The EXCHANGE recode indicates whether NCHS has determined that a person has exchange-based coverage.

New for 2015 is the inclusion of three additional recodes (COVER, COVER65, and COVER65O) to make it easier for the NHIS user to match to estimates of health insurance coverage produced by the Division of Health Interview Statistics in annually released products. A health insurance hierarchy was developed in the mid-1990s for ease of analysis and tabulation of estimates of coverage using the National Health Interview Survey for the Summary Health Statistics Reports (now tables) and other analytic projects. For persons under age 65, the

hierarchy has four mutually exclusive categories. Persons under age 65 with more than one type of health insurance coverage were assigned to the first appropriate category in the following hierarchy:

- *Private coverage*—Includes persons who had any comprehensive private insurance plan (including health maintenance organizations, preferred provider organizations, and exchange-based coverage).
- *Medicaid*—Includes persons who do not have private coverage, but who have Medicaid or other state-sponsored health plans including CHIP.
- *Other coverage*—Includes persons who do not have private insurance, Medicaid, or other public coverage, but who have any type of military coverage or Medicare. This category also includes persons who are covered by other government programs.
- *Uninsured*—Includes persons who have not indicated that they are covered at the time of the interview under private health insurance, Medicare, Medicaid, CHIP, a state-sponsored health plan, other government programs, or military coverage. This category also includes person who are covered by Indian Health Service only or who only have a plan that pays for one type of service such as accidents or dental care.

More recently, the health insurance hierarchy for adults aged 65 and over was revised due to an increase in older adults covered by Medicare Advantage plans. This revised hierarchy de-duplicates the report of private health insurance and Medicare Advantage. Those with a report of Medicare Advantage and private coverage are placed in the Medicare Advantage category. For persons aged 65 and over, the hierarchy has six mutually exclusive categories. Those with more than one type of health insurance coverage were assigned to the first appropriate category in the following hierarchy:

- *Private coverage*—Includes older persons who have both Medicare and any comprehensive private insurance plan (including health maintenance organizations, preferred provider organizations, and Medi-gap plans). This category also includes older adults with private insurance only.
- *Medicare and Medicaid*—Includes older persons who do not have any private coverage but have Medicare and Medicaid or other state-sponsored health plans including CHIP.
- *Medicare Advantage*
- *Medicare only excluding Medicare Advantage*
- *Other coverage*—Includes older persons who have not been previously classified as having private, Medicare and Medicaid, Medicare Advantage or Medicare only excluding Medicare Advantage. This category also includes older persons who have only Medicaid, other state-sponsored health plans, or CHIP, as well as persons who have any type of military coverage with or without Medicare.
- *Uninsured*—Includes older persons who have not indicated that they are covered at the time of the interview under private health insurance, Medicare, Medicaid, CHIP, a state-sponsored health plan, other government programs, or military coverage. This category also includes person who are covered by Indian Health Service only or who only have a plan that pays for one type of service such as accidents or dental care.

Beginning in 2015, COVER reflects the hierarchy used for persons under age 65, and COVER65 reflects the hierarchy for persons age 65 and over. The COVER65O recode was included for those who wish to replicate the hierarchy for persons 65 years of age and over that was used prior to 2015. A description of the hierarchy used prior to 2015 is available from: [http://www.cdc.gov/nchs/data/series/sr\\_10/sr10\\_198.pdf](http://www.cdc.gov/nchs/data/series/sr_10/sr10_198.pdf).

Other recodes are WHONAM1 and WHONAM2 (whose name the plan is in). HICOSTR1 and HICOSTR2 are recodes for out-of-pocket premium cost. Other recodes include PRFLG, MAFLG, CHFLG, OPFLG, and OGFLG (flags indicating insurance classification was reassigned); PXCHNG, PSTRFPRM, PSSPRINC, PSTDOC (detailed follow-up information about public plans reassigned to private); EXCHPR1, EXCHPR2, PLEXCHPR, PLEXCHOP, PLEXCHOG (internal coding of plan names as exchange companies, plans, or portals); and MILMANR (type of TRICARE coverage).

In 2015 the following recodes have been removed from the file as the questions they were based on have been removed from the instrument: COHU191, COHU192, COH19251, COH19252, COHO251, COHO252, ECOSTR1, ECOSTR2, PSTCMD, and PSTREF.

## Suppression of Data

To protect confidentiality, detailed information concerning persons' third and fourth plans is not available on the public use data file. Persons with three or more plans have a "yes" response to the PRPLPLUS variable. Detailed information on persons' third and fourth plans is available through the NCHS RDC at <http://www.cdc.gov/rdc/>.

## Technical Notes

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 general, if a family member was reported to have coverage through the Health Insurance Marketplace or a state-based exchange, then that reported coverage is considered accurate unless there is some other information that clearly contradicts that report. Similarly, if a family member was not reported to have coverage through the Marketplace or state-based exchange, then that is considered accurate unless there is some other information that clearly contradicts that report. Specific applications of these general rules are detailed here and were implemented beginning with the release of preliminary 2014 data in September 2014. Additional information can be found at: <http://www.cdc.gov/nchs/nhis/releases.htm>.

The NHIS considers a person reporting private coverage as having exchange-based coverage if they are reported to have a private, non-employment-based, directly purchased plan and the plan name provided is a) an exchange plan name, or b) an exchange portal name (e.g. Healthcare.gov, MNSure in Minnesota), or c) they have provided an exchange company name and the respondent indicated that the plan is through the Health Insurance Marketplace or state-based exchange, or d) the plan name was unknown or refused and the

respondent indicated that the plan was obtained through the Health Insurance Marketplace or state-based exchange. Providing an exchange plan name or an exchange portal name is weighed heavily in the decision to classify a person as having exchange-based coverage. Persons with employment-based coverage were not considered to have exchange coverage unless a very specific exchange plan name was provided.

When a state-sponsored health plan or another government program is reported, the person is classified as having exchange-based coverage if the plan name provided is a) an exchange plan name, or b) an exchange portal name, or c) an exchange company name and the respondent has indicated that the plan was through the Health Insurance Marketplace or state-based exchange, or d) the plan name was unknown or refused and the respondent indicated that the plan was obtained through the Health Insurance Marketplace or state-based exchange and had a premium associated with the plan. The source of coverage is changed from public to private. All individuals classified as having exchange-based coverage are considered to have private coverage, regardless of whether they were reported to have obtained the coverage from a private or public source. Further details on assigning exchange-based coverage can be found at:

[http://www.cdc.gov/nchs/data/nhis/health\\_insurance/2014%20exchange\\_coding\\_%20rules.pdf](http://www.cdc.gov/nchs/data/nhis/health_insurance/2014%20exchange_coding_%20rules.pdf)

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.

Using the most conservative estimate of the uninsured (which would exclude persons with IHS coverage only), a total of 706 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 1,115 persons, the family respondent was not asked either the HILAST or the HINOTYR questions.

Again, analysts are strongly advised to use the recodes MEDICARE, MEDICAID, PRIVATE, SCHIP, IHS, MILCARE, OTHPUB, OTHGOV, and SINGLE for estimates of types of health care coverage and NOTCOV to derive estimates of uninsurance. The variables HILAST and HINOTYR, which reflect periods of noncoverage, cannot be used to estimate the rate of uninsurance. 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.

More details on the editing of health insurance information can be found at:

[http://www.cdc.gov/nchs/nhis/health\\_insurance/hi\\_eval.htm](http://www.cdc.gov/nchs/nhis/health_insurance/hi_eval.htm)

## VIII. Socio-demographic Section (FSD)

### Core Content

The Socio-demographic (FSD) section of the Family Core in the 2015 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.

As with the 2013 and 2014 FSD section, the 2015 section contains a series of variables on specific periods of active duty service in the United States Armed Forces. The questions are similar in format to the 2011–2012 questions but the time periods asked about have been modified. Users should consult the notes in the Variable Layout Report for more information on these changes and the extent to which the 2013–2015 variables are comparable to the 2011–2012 variables. Also, please note that national (i.e., weighted) estimates of active duty service during any given time period may be underestimates because they will not include persons currently serving in the U.S. Armed Forces. The value of the final annual person weight (WTFA) for persons who are currently serving in the U.S. Armed Forces is zero.

## Major Recodes

The FSD section includes two recodes that are based on birthplace information. 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. REGIONBR categorizes all persons into one of 12 geographic categories depending on their country of origin. The CIA online World Factbook was used to place countries into the regional categories shown in Table 8. (For more information about the Factbook, users should refer to <https://www.cia.gov/library/publications/the-world-factbook/index.html>.) Note that persons 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.

In addition, the FSD section contains recodes indicating how long family members born in a country or place other than the United States have been living in the U.S. (YRSINUS); whether persons are U.S. citizens and non-citizens (CITIZENP); and employed adults’ earnings for the previous year using a finite number of categories (ERNYR\_P).

## Suppression of Data

To protect confidentiality, detailed information on state or country of birthplace, year of immigration (for persons born outside the U.S.), citizenship status, Armed Forces service prior to June 1950, and earnings (in dollar amounts) for the previous year are not available on the NHIS public use data files. In addition, data from the Armed Forces variables were modified to protect confidentiality; the public use Person File contains these modified recodes (ARMFTM1P–ARMFTM7P and ARMFDS2P). Analysts wishing to use the unmodified restricted data can apply for access to NCHS’ RDC at: <http://www.cdc.gov/rdc/>.

**Table 8. World regional categories**

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

## Technical Notes

DOINGLWP and WHYNOWKP are the FSD equivalents of DOINGLWA and WHYNOWKA in the ASD section of the Sample Adult File. For the majority of adults, 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 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).

## IX. Income and Assets Section (FIN)

### Imputed Income Files

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. Analysts are strongly encouraged to refer to this technical documentation when using these files.

## Core Content

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. 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 (2014). They are then asked whether anyone in the family received income from a variety of sources, and if so, which family member(s) received income from that source. 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 asked about income received from all family members regardless of age, and inquire 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). In addition, respondents are asked whether any family member(s) had ever applied for SSI or Social Security disability benefits (even if the claim(s) had been denied).

The section also includes questions about the family's total income from all sources in 2014; home tenure status; whether the family was paying lower rent due to governmental rental assistance; and program participation (e.g., participation in the Supplemental Nutrition Assistance Program (SNAP), or the Women, Infants, and Children (WIC) program). Separate questions ask how many months SNAP assistance and/or cash assistance from state or county assistance programs was received during the last calendar year.

The NHIS asks respondents to report their "best estimate" of their family's total income (in dollars) from all sources and for all family members before taxes in the last calendar year (FINCTOT). Because nonresponse to this question tends to be relatively high, the NHIS includes a series of follow-up questions utilizing an unfolding bracket methodology that obtain additional income information. The unfolding bracket method asked a series of closed-ended income range questions (e.g., "is it less than \$75,000, or \$75,000 or more?") if the respondent did not provide an answer to FINCTOT. 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.

In addition to the closed-ended income range questions, other follow-up questions determine broad gradations in federal poverty level (100%, 200%, and 400% of the federal poverty level, or FPL), and additional questions are included in the 2015 questionnaire to determine where a family falls relative to 138% and 250% of the FPL thresholds. These thresholds were first added to the 2014 questionnaire to capture health insurance eligibility and health insurance subsidy eligibility. In some states, individuals in families earning less than 138% FPL may be eligible for Medicaid or other state-run insurance products. Additionally, starting in 2014, adults in families earning between 100% and 250% of the FPL may be eligible for cost-sharing subsidies to pay for private health insurance purchased through health insurance marketplaces. The addition of these questions to the income section allows the eligibility of more persons to be classified for analyses of health insurance eligibility and uptake prior to the release of the imputed income files.

In addition to asking respondents about the family's income relative to specific dollar values (i.e., \$75,000, \$100,000, and \$150,000), some questions ask about the family's income relative to a poverty threshold and take into account each family's size (collected earlier in the interview). These poverty threshold questions use poverty threshold estimates for 2014, given family size. Because weighted Census poverty thresholds for 2014 were not available when the 2015 NHIS instrument was created, the poverty thresholds used in the 2015 NHIS were estimated from several sources: weighted average Census poverty thresholds from 2013; the average Consumer Price Index from 2013; actual Consumer Price Index values for January-July 2014; and projected Consumer Price Index values for August-December 2014. The estimated poverty thresholds used in the 2015 NHIS instrument are shown in Table 9.

**Table 9. Estimated poverty thresholds, 2015 NHIS Instrument**

Family Size	FPOV100	FPOV138	FPOV200	FPOV250	FPOV400
1 person < 66 years	\$12,000	\$17,000	\$25,000	\$31,000	\$49,000
1 person ≥ 66 years	\$11,000	\$16,000	\$23,000	\$28,000	\$46,000
2 persons, both < 66	\$16,000	\$22,000	\$32,000	\$40,000	\$64,000
2 persons, 1 is ≥ 66	\$14,000	\$20,000	\$29,000	\$36,000	\$57,000
3	\$19,000	\$26,000	\$38,000	\$47,000	\$76,000
4	\$24,000	\$34,000	\$49,000	\$61,000	\$97,000
5	\$29,000	\$40,000	\$58,000	\$72,000	\$115,000
6	\$33,000	\$45,000	\$65,000	\$81,000	\$130,000
7	\$37,000	\$51,000	\$74,000	\$92,000	\$148,000
8	\$41,000	\$57,000	\$82,000	\$103,000	\$164,000
9 or more	\$49,000	\$67,000	\$98,000	\$122,000	\$196,000

When the questions about income relative to poverty threshold are 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 in 2014 was less than the applicable NHIS poverty threshold, or if the family's income was greater than or equal to that same poverty threshold.

## Major Recodes

The information from the income follow-up questions is incorporated into several public use income recodes. INCRP4 and INCRP5, which were first introduced in the 2014 NHIS and are retained in the 2015 NHIS, provide information about family income in grouped categories. Analysts should consult the notes in the Family File Variable Layout Report for more information on these recodes and how they compare to earlier versions (e.g., INCRP2 and INCRP3 in the 2007-2013 NHIS).

In addition, RAT\_CAT4 and RAT\_CAT5, which replace RAT\_CAT2 and RAT\_CAT3, contain categories indicating family income relative to the poverty threshold. These recodes collapse the poverty ratio into gradients consisting of a finite number of categories. For both recodes, 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 (go to <http://www.census.gov/hhes/www/poverty/data/threshld/index.html> and select 2014.xls to see federal poverty thresholds by size of family and number of children). Categories "15" through "17" of RAT\_CAT4 and categories "15" through "18" of RAT\_CAT5 are based on the poverty thresholds used in the NHIS survey instrument (shown in Table 9), and reflect the additional poverty ratio information obtained from the income follow-up questions. Because RAT\_CAT5 has one more category than RAT\_CAT4, some families will not be coded

equivalently on the two variables. In the 2015 NHIS, 3.6% of families were coded “17” on RAT\_CAT4 and “18” on RAT\_CAT5, while 1.9% of families were coded “17” on RAT\_CAT4 and “96” on RAT\_CAT5. Analysts should be aware of these discrepancies when analyzing data from RAT\_CAT4 and RAT\_CAT5. In addition, the Family File Variable Layout Report will provide information on how RAT\_CAT4 and RAT\_CAT5 compare to earlier recodes (e.g., RAT\_CAT2 and RAT\_CAT3).

Finally, the NHIS Person File contains three person-level variables relating to the Women, Infants, and Children (WIC) program. The first of these variables, ELIGPWIC, indicates if the person was in a family with at least one WIC age-eligible person (children 0–5 years of age or females 12–55 years of age). If there is at least one WIC age-eligible person in the family, the family respondent is asked if anyone in the family received WIC benefits in the previous calendar year (PWIC). An additional recode, WIC\_FLAG, indicates if persons who received WIC benefits were age-eligible for the WIC program.

## Suppression of Data

To protect confidentiality, family income reported in dollar amounts as well as the variables obtained from the income bracketing questions are not available on the NHIS public use data files. Analysts wishing to gain access to these restricted data can apply for access through NCHS’ RDC at: <http://www.cdc.gov/rdc/>.

## Technical Notes

Income data can be found on both the NHIS Person and Family Files. As previously mentioned, the majority of the questions in the FIN section of the NHIS instrument 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. The Person File contains variables that identify the family member receiving income from a particular source, while the Family File contains recodes that indicate whether the family received income from a source, and if so, how many family members received income from this source. In addition, the recodes INCGRP4, INCGRP5, RAT\_CAT4, and RAT\_CAT5 are also located on the Family File. Analysts should consult the documentation for both the NHIS Person and Family Files to determine how to use NHIS income data most effectively; Appendix IV provides guidance on how to merge these files.

## X. English Language Proficiency Section (FLG)

### Core Content

The English Language Proficiency section (FLG) consists of a single variable, ENGLANG, that provides information on how well persons aged 5 years and over speak English.

---

## Injury and Poisoning Episode File

### Core Content

The Family Core portion of the 2015 survey included questions about medically consulted injuries and poisonings that occurred for any member of the family within a three-month reference period. A medically consulted injury/poisoning episode is defined as an episode that results in a call to a poison control center; visit to an emergency room, doctor's office or other health clinic; or phone call to a doctor or other health care professional. 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.

All injury and poisoning information was provided by the family respondent. The Injury/Poisoning Episode File was created from these data. The Injury/Poisoning Episode File is an episode-based file. Each unique injury and/or poisoning episode is represented in this file.

The Injury/Poisoning Episode File contains 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 initial responses to family level questions FIJ.010\_01.000 to FIJ.028\_00.000. The date of the injury or poisoning episode that was subsequently reported by the family respondent may have been 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).

Up to a total of ten injury and/or poisoning episodes may be recorded. Each episode must have at least one injury condition or poisoning classified according to the nature of injury codes in the Tenth Revision of the International Classification of Diseases (ICD-10-CM) and one external cause of injury code. Valid codes considered for inclusion in the 2015 NHIS are listed below in Table 10. Note that in some instances certain codes function as both an injury and external cause code. At times episodes are removed during the data editing process because responses indicate an injury did not take place (e.g., they consisted solely of health conditions that would not be classifiable according to these nature of injury codes and external cause of injury codes), or because they are duplicate episodes. Codes for other health conditions that were reported as occurring along with a classifiable injury or poisoning are also included in the Injury/Poisoning Episode File, even if they are not injury codes themselves.

**Table 10. Valid ICD-10-CM injury and external cause codes for inclusion in 2015 National Health Interview Survey**

ICD-10-CM codes	Description
Injuries	
All S codes	Anatomic injuries
T07-T34	Foreign bodies, burns, corrosions, frostbite
T36-T50 with any 6 <sup>th</sup> character except 5 or 6	Drug poisoning, except adverse effects and underdosing
T51-T65	Toxic effects of substances nonmedicinal as to source
T66-T77	Other and unspecified effects of external causes (radiation, heat, light, hypo/hyperthermia, asphyxiation, child/adult abuse, lightning, drowning, motion sickness, etc.
T79	Certain early complications of trauma, not elsewhere classified
External Causes*	
V00-V99	Transport accidents
W00-X58	Other external causes of accidental injury
X71-X83	Intentional self-harm
X92-Y08	Assault
Y21-Y33	Event of undetermined intent
Y35-Y38	Legal intervention, operations of war, military operations and terrorism
T14.91	Suicide attempt
T15-T19	Effects of foreign body entering through natural orifice
T36-T50 with any 6 <sup>th</sup> character except 5 or 6	Poisonings by drugs, medicaments, and biological substances
T51-T65	Toxic effects of substances chiefly non-medicinal as to source
T71	Asphyxiation
T73	Effects of deprivation
T74, T76	Adult and child abuse, neglect, and other maltreatment, confirmed or suspected
T75.0, T75.2, T75.3	Effects of lightning, Effects of vibration, Motion sickness

\* 7<sup>th</sup> character of "A," "D" or 7<sup>th</sup> character missing

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, and whether the person missed any days from work or school due to the injury or poisoning. ICD-10-CM diagnostic codes and ICD-10-CM external cause codes will also be provided at a later date; these codes are not immediately available to data users. ICD-10-CM diagnostic and external cause codes are 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.

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-10-CM. Episodes that are not considered poisonings based on ICD-10-CM are no longer included in the Injury/Poisoning Episode 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 2015 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 IV for additional information about merging data files.

## Suppression of Data

To protect confidentiality, selected ICD-10-CM injury and external cause codes may be suppressed or truncated to a specific number of characters. Due to the transition to ICD-10-CM and the delayed availability of these codes, these suppression criteria have not yet been determined. Once released, original ICD-10-CM codes will be available through the RDC at: <http://www.cdc.gov/rdc/>.

## 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-10-CM external cause codes. The category into which an episode was placed was based entirely on the first ICD-10-CM external cause code listed for that episode. Appendix I in the Injury/Poisoning Episode data release documentation contains a list of the ICD-10-CM external cause codes found in each category. Because ECAUS is based on the ICD-10-CM external cause codes which are not yet available, this recode is also not immediately available and will be provided at a later date.

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. 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.

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 2015 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  $\leq 91$ ; they were only constrained to be consistent with the date information reported by the respondent. Variable ETFLG indicates whether the elapsed time (RPD) is  $\leq 91$  days. Variable BEIFLG indicates whether the boundaries (BIETD and EIETD) of the elapsed time interval are  $\leq 91$  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 NCHS has a website devoted exclusively to the NHIS injury and poisoning data: [http://www.cdc.gov/nchs/nhis/injury\\_poisoning.htm](http://www.cdc.gov/nchs/nhis/injury_poisoning.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.

## Recall Period and Weights

Questions in the Injury/Poisoning section of the 2015 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 data release documentation and the specific item in the questionnaire to ensure that annual estimates for these kinds of injury or poisoning episodes have intrinsic meaning.

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.

For additional information about the historical context, editing, frequently asked questions, and references for NHIS injury and poisoning episodes, see: [http://www.cdc.gov/nchs/nhis/special\\_topics.htm](http://www.cdc.gov/nchs/nhis/special_topics.htm).

---

## Family Disability Questions File (FDB)

### Core Content

The Family Disability Questions file contains a set of six core disability questions first developed and used on the American Community Survey (ACS) and subsequently adopted by several federal data collection systems. Originally included on the NHIS as part of a larger effort to develop and test a standard set of disability questions, HHS adopted the question set as a minimum standard for survey questions on disability in 2011.

The six questions in the 2015 Family Disability Questions section (FDB) asked about difficulties hearing, seeing, concentrating, walking, dressing/bathing, and doing errands. These disability questions were initially asked from October 2008 through December 2009 on the Family Disability questionnaire at both the person level (FDB) and the family level (FDA), but the FDA was dropped in 2010. Disability questions were included in the 2011 and 2012 Sample Adult (ADB) and Sample Child (CDB) questionnaires, in addition to the Family Disability questionnaire (FDB), but the ADB and CDB were not included in 2015.

### Background

The U. S. Census has a history of including questions about disability to satisfy a variety of stakeholder needs. The 2000 Census of Population and the 2000–2007 questionnaires of the Census Bureau’s ACS included six disability questions: a combined sensory (vision and hearing) question; separate mobility, self-care, and cognition questions; and two independent living questions (based on daily activities and working). Before each decennial census, other agencies and researchers join with the Census Bureau to develop survey questions most appropriate for their needs and to determine if existing questions need modification. The user community had expressed some dissatisfaction with the disability measures on the 2000 Census and 2000–2007 ACS.

In response, the Census Bureau and other stakeholder agencies refined the disability questions to bring them into line with recent changes in the definition of disability and the conceptualization of the components of that definition. A work group was formed under the auspices of the Office of Management and Budget and led by NCHS. That work group researched the theoretical approach to the definitions based on the latest national and international ideas about disability, analyzed available data to test their conceptualizations, identified and examined agency mandates for collecting disability data, discussed question content and wording, and sponsored cognitive testing of a new question set. The modified disability questions were subsequently tested in the ACS Content Test of 2006, along with other question additions and modifications (Brault et al., 2007).

The underlying concept behind the choice of questions was to identify the subpopulation that is at a greater risk than the general population of experiencing restrictions in social participation, for example, restrictions in employment, education, or civic life. The objective was thus to measure equalization of opportunities. Four basic domains of functioning (vision, hearing, mobility, and cognitive functioning) were identified that would define the largest component of the population of people with disabilities. These four domains could be used individually or combined in order to assess equalization of opportunities for people with disabilities. Also, two more domains were identified that could be used for monitoring independent living and the need for services: the ability to take care of oneself (self-care, in particular, the ability to bathe and dress oneself), and the ability to move around the community (independent living, in particular, the ability to visit a doctor’s office or go shopping). The new set included separate questions for vision and hearing, refined the mobility question, expanded the cognitive functioning question, continued the inclusion of a self-care question, and improved the question on independent living. The question about work, which had been included in earlier censuses, was not retained.

## Technical Notes

A separate, stand-alone public use data file was created for the FDB variables, rather than appending these variables to the 2015 Person File. This file is called the Family Disability Questions File. A random sample of approximately one-half of the respondents from the 2015 Person File was selected for the FDB questions. Therefore, a different weight (WTFA\_FDB) was generated for persons in the FDB File. This weight was designed to produce annual-level estimates calculated based on data included in the files.

Analysts wishing to perform analyses of the NHIS disability questions data (e.g., cross-survey comparisons of disability prevalence rates to gain insight into possible survey context effects on estimates of disability; or produce estimates and perform comparisons within key subgroups such as age, sex, and race/ethnicity) will need to merge the FDB file with one or more health data files. Information on merging data files and combining multiple years of data can be found in Appendix IV.

---

## Sample Child File

The Sample Child section of the 2015 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.

The 2015 Sample Child File includes 12,291 records with 1 record for each sample child. The Sample Child section was fielded with 70 additional sample children, but their records were removed from the file due to poor data quality. A flag, QCCHILD, is included in the Person File to denote these records.

More details about the sections comprising the Sample Child File are provided below.

### I. Child Identification Section (CID)

#### Core Content

The Child Identification section (CID) contains information about the availability of a respondent knowledgeable about the sample child's health, including the relationship of that person to the sample child. This section also includes a variable indicating whether or not the sample child questionnaire was started two or more weeks later than the initial interview. The CID section also contains questions that verify the sample child's sex, age, and date of birth that are initially provided by the family respondent earlier in the interview.

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

#### Core Content

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); Down 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.

The CHS section also includes questions about head colds, chest colds, and intestinal illnesses that started during the 2 weeks prior to the interview. It should be noted that these questions intentionally are measuring fairly broad symptoms and illnesses that may be a result of either acute or chronic conditions (e.g., irritable bowel syndrome or respiratory allergies).

This section also includes questions about the sample child's height and weight. It is important to note that 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. **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.

## Supplemental Questions

A large hearing supplement sponsored by the National Institute on Deafness and Other Communication Disease (NIDCD, NIH) in 2014 was dropped from the survey in 2015.

## Major Recodes

Birth weight was collected in pounds and ounces or in grams. Recodes were created to provide information about total birth weight in ounces (TOTOZ\_P) and total birth weight in grams (BWTGRM\_P), regardless of the units of measurement originally provided. Current height for children aged 12–17 years was collected in feet and inches or in meters and centimeters. A recode was created to provide total current height in inches (CHGHT\_TC), regardless of the units of measurement originally provided. Additionally, these recoded variables, as well as a variable about current weight (CWGHT\_TC), were top and bottom-coded in order to protect the confidentiality of those who had extreme values in height and/or weight.

The CHS section includes several questions pertaining to child behavior for children aged 2–3 years that are derived from the Child Behavior Checklist: 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. The data based on these questions are used to create recodes regarding the child's mental health (MHIBOY2, MHIGRL2); only the recodes are included in the Sample Child File. The items were chosen from the Checklist for their ability to discriminate between young 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. More information on the scale derived from the Child Behavior Checklist is included in Appendix VI of this document.

## Technical Notes

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. See the 2012 NHIS *Survey Description*, CHS section, for more information about children's height, weight, and BMI on the NHIS.

---

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

#### Core Content

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.

#### Supplemental Questions

Supplementary expanded content on access and utilization questions were first added to the survey in 2011 and have been part of this section since then. These questions were sponsored by the Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS) in order to provide more information about the affordability of health care. The questions are embedded throughout the CAU section, and variables based on the responses to these questions include data about difficulty finding providers (CPRVTRYR, CPRVTRFD, CDRNANP, CDRNAI), unmet medical need due to cost (CHCAFYRN, CHCAFYRF, CHCAFYR5, CHCAFYR6), and the timing of and reasons for emergency room visits (CERVISND, CERHOS, CERREA1R–CERREA8R). In 2015, questions about the reasons for not having a usual source of care (CNOUSPL1–CNOUSPL9) were dropped from the survey.

### IV. Child Mental Health Brief Supplement (CMB)

#### Supplemental Questions

The Child Mental Health Brief Supplement (CMB) was sponsored by the Center for Mental Health Services (CMHS, SAMHSA). The purpose of the CMB is to provide information on children's behavior as measured by Dr. Robert Goodman's Strengths and Difficulties Questionnaire (SDQ). 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 VIII of the 2004 NHIS *Survey Description* and/or the SDQ website at <http://www.sdqinfo.org>.

The 2015 CMB is the short version of the SDQ that was originally fielded in 2002 and 2010–2014. For further information about the short version of the SDQ see Appendix VII.

## Major Recodes

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 VII, 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.

## V. Child Influenza Immunization Supplement (CFI)

### Supplemental Questions

The Child Influenza Immunization Supplement (CFI) was sponsored by the National Center for Immunization and Respiratory Diseases (NCIRD, CDC). The purpose of the CFI is to provide information about the timing and type of flu vaccinations for US children. Questions included: 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 later in this report).

There were no changes to the CFI section in 2015. The questions were the same as those used in the latter part of 2010 and 2011-2014. 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 online in the CFI Appendix of the 2010 NHIS Variable Layout Report.



---

## 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.

The 2015 Sample Adult File includes 36,672 records with 1 record for each sample adult. The Sample Adult section was fielded with 163 additional sample adults, but their records were removed from the file due to poor data quality. A flag, QCADULT, is included in the Person File to denote these records.

The sections comprising the Sample Adult File are discussed below.

### I. Adult Identification Section (AID)

#### Core Content

The Adult Identification section (AID) contains information about whether the sample adult was able to respond or whether a knowledgeable proxy was available and completed the Sample Adult section of the interview. In the case of a proxy, the relationship to the sample adult is obtained. This section also includes a variable indicating whether or not the sample adult questionnaire was started two or more weeks later than the initial interview. The AID section also contains questions that verify the sample adult's sex, age, and date of birth that are initially provided by the family respondent earlier in the interview.

### II. Adult Socio-Demographics Section (ASD)

#### Core Content

The 2015 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), 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. Sample adults 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.

## Supplemental Questions

An Occupational Health Supplement sponsored by the National Institute of Occupational Safety and Health (NIOSH, CDC) is included in the 2015 ASD section (Another portion of this supplement is included in the ACN section.) The ASD supplemental questions asked if currently or ever employed adults supervised employees as part of their jobs. In addition, sample adults who said their current job was not the job they had held the longest were asked a series of questions about their longest held jobs (i.e., industry, occupation, work status, how long they worked in this job).

The bulk of the questions in the 2015 supplement were directed at currently employed adults. They were asked about their work arrangements and usual work schedules; whether they had worked between 1:00 AM and 5:00 AM at any time during the past 30 days; and how frequently this happened. In addition, they were asked their opinions regarding various aspects of their current jobs (do the demands of their job interfere with their personal or family life; do they have enough time to get their job done; does their job allow them to make decisions on their own; can they count on their supervisor for support if they need it) and whether they worried about losing their current job. They were also asked several questions about their workplace (how safe they felt it was; whether the health and safety of workers was important to management where they currently worked; whether their employers had made health promotion programs available in the past year; and how often they had participated in these programs in the past year) and their current job (whether they had felt threatened, bullied, or harassed during the past 12 months, and how frequently this happened; how often their job involves repeated lifting, pushing, pulling or bending; how often their job involves standing or walking around; and how often they were exposed to second-hand tobacco smoke while at work during the past 12 months).

## Major Recodes

During the course of the interview, verbatim responses were obtained from each eligible sample adult 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 non-economic Federal surveys, are 4-digit Census codes for industry and occupation consistent with the 2012 North American Industry Classification System (NAICS) and 2010 Standard Occupational Classification (SOC). However, these are not actual NAICS and SOC codes.

While the 2015 NHIS Sample Adult File available for public use 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 2012 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 the Census Bureau.

Importantly, the 2-digit industry and occupation recodes that result from the 2015 Occupational Health Supplement are created in the same manner and with the same categorical structure as the Core recodes. Thus, OCCLH1 is a detailed occupation recode (similar to OCCUPN1) with 94 distinct categories, while OCCLH2 is the associated simple recode (similar to OCCUPN2) with 23 categories. INDSTLH1 is the detailed industry recode and has 79 distinct categories (like INDSTRN1), while the associated simple recode, INDSTLH2, has 21 categories (like INDSTRN2).

Users should consult the Occupation and Industry Appendices at the end of the Sample Adult Variable Layout Report for the response categories and labels of OCCUPN1, OCCUPN2, INDSTRN1, INDSTRN2, and the recodes from the 2015 Supplement (OCCLH1, OCCLH2, INDSTLH1, and INDSTLH2). 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 recodes on the public use data file. These lists should not be used in place of the Occupation and Industry Appendices. For more information about the 2012 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>.

In addition, the 2015 ASD section contains a recode indicating how long currently or ever employed sample adults have worked at their job or business (YRSWRKPA).

### Suppression of Data

To protect confidentiality, the 4-digit Census codes for industry and occupation as well as detailed information regarding the sample adult's years at the (current, longest held, or most recent) job or business are not available on the NHIS public use data files. In addition, the variables WRKARRNG (work arrangement) and NIGHTFRQ (times worked between 1-5 a.m.) were coarsened by collapsing response categories and top-coding. Analysts wishing to gain access to these restricted data can apply for access through NCHS' RDC at: <http://www.cdc.gov/rdc/>.

### Technical Notes

DOINGLWA and WHYNOWKA are the ASD equivalents of DOINGLWP and WHYNOWKP in the FSD section. For the majority of sample adults, 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 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).

## III. Adult Conditions Section (ACN)

### Core Content

The Adult Conditions (ACN) section contains information about whether the sample adult 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). Sample adults 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 recent and current pregnancy status.

For specific details on the health-related conditions included in the 2015 ACN section core content, and the various reference periods covered by these survey questions, please see Table 11.

## Supplemental Questions

In the 2015 NHIS, the ACN section included a series of supplemental questions pertaining to the Million Hearts<sup>®</sup> Initiative sponsored by the Department of Health and Human Services (HHS). This series of questions pertained to aspirin use, hypertension (high blood pressure), and hyperlipidemia (high cholesterol). These questions asked the sample adult if he/she had ever been told by a doctor or other health professional that he/she had hypertension in the past 12 months (HYPYR1) and had high cholesterol in the past 12 months (CHLYR). Sample adults were also asked how long it had been since they had their blood pressure checked by a doctor, nurse, or other health professional (HYBPCKNO and HYBPCKTP), whether they were told their blood pressure was high, normal, or low (HYBPLEV), whether they had ever been prescribed medicine for high blood pressure by a doctor (HYPMDEV2), and whether they are now taking this medicine (HYPMED2). Similar questions were asked for hyperlipidemia (high cholesterol), where sample adults were asked how long it had been since they had their blood cholesterol checked by a doctor, nurse, or other health professional (CLCKNO and CLCKTP), whether they had ever been prescribed medicine by a doctor to lower their cholesterol (CHLMDEV2), and whether they are now taking this medicine (CHLMDNW2). In addition to these questions, four questions asked about low-dose aspirin use among sample adults aged 40 years and over. These four questions included: whether a doctor or other health professional ever told the sample adult to take low-dose aspirin each day (ASPMEDDEV), whether the sample adult is following this advice (ASPMEDAD), whether a doctor or other health professional advised the sample adult to stop taking low-dose aspirin (ASPMEDMED), and whether the sample adult is taking low-dose aspirin on his/her own (ASPONOWN). A subset of these supplemental questions was previously included in the ACN section in the 2012-2014 NHIS.

In the 2015 ACN, the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP, CDC) sponsored a single supplemental question related to inflammatory bowel disease that asked a sample adult if he/she had ever been told by a doctor or other health professional that he/she had Crohn's disease or ulcerative colitis.

The National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP, CDC) also sponsored five supplemental questions related to epilepsy that were embedded in the 2015 ACN section. The first asked a sample adult if he/she had ever been told by a doctor or other health professional that he/she had a seizure disorder or epilepsy (EPILEP1). The remaining four questions were asked of sample adults who reported ever being diagnosed with a seizure disorder or epilepsy. The sample adult was asked if he/she is currently taking medication to control a seizure disorder/epilepsy (EPILEP2), the number of seizures he/she had in the past year (EPILEP3), whether he/she had seen a neurologist or epilepsy specialist in the past year (EPILEP4), and the extent to which epilepsy had interfered with his/her normal activities in the past 30 days (EPILEP5). These questions were last asked in 2013.

**Table 11. Conditions and reference periods in the core 2015 National Health Interview Survey, ACN section**

Question number	Condition	Ever	12 months	3 months	30 days	2 weeks	Now
ACN.010	Hypertension	X					
ACN.020	Hypertension 2+ visits	X					
ACN.031	Coronary heart disease	X					
ACN.031	Angina pectoris	X					
ACN.031	Heart attack (MI)	X					
ACN.031	Other heart condition or heart disease	X					
ACN.031	Stroke	X					
ACN.031	Emphysema	X					
ACN.035	Chronic obstructive pulmonary disease (COPD)	X					
ACN.080	Asthma	X					
ACN.085	Asthma still have						X
ACN.090	Asthma episode / attack		X				
ACN.100	Asthma ER visit		X				
ACN.110	Ulcer ever told	X					
ACN.120	Ulcer recent		X				
ACN.130	Cancer any	X					
ACN.140	Cancer kind	X					
ACN.160	Diabetes	X					
ACN.165	Prediabetes	X					
ACN.180	Insulin						X
ACN.190	Oral agents/pills						X
ACN.201	Hay fever		X				
ACN.201	Sinusitis		X				
ACN.201	Chronic bronchitis		X				
ACN.201	Weak kidneys		X				
ACN.201	Liver condition		X				
ACN.250	Joint symptoms				X		
ACN.260	Joints affected				X		
ACN.270	Joint symptoms chronic			X			
ACN.280	Joints doctor consult	X					
ACN.290	Arthritis (arthritis, gout, fibromyalgia, rheumatoid arthritis, lupus) diagnosis	X					
ACN.295	Limited in activities due to arthritis/joint symptoms						X
ACN.300	Neck pain			X			
ACN.310	Back pain			X			
ACN.320	Leg pain			X			
ACN.331	Jaw, face pain			X			
ACN.331	Migraine			X			
ACN.350	Head/chest cold					X	
ACN.360	Intestinal illness					X	
ACN.370	Pregnant						X
ACN.400	Use hearing aid						X
ACN.410	Use hearing aid	X					
ACN.420	Hearing						X
ACN.430	Vision impairment						X
ACN.440	Blind						X
ACN.451	Lost all teeth						X

A portion of the large supplement focused on occupational and workplace health, sponsored by the National Institute of Occupational Safety and Health (NIOSH, CDC), was also included in the ACN section of 2015 NHIS (the other portion of this supplement was included in the ASD section). A series of questions asked a sample adult if he/she had ever been diagnosed with carpal tunnel syndrome (CTS) (CTSEVER). For a sample adult who had been diagnosed, questions were also asked as to whether he/she had CTS in the past 12 months (CTSYR), and were ever told by (CTSWKREL) or discussed with (CTSWKRL2) a doctor or other health professional that their CTS was work related. Also in the ACN section as part of this supplement was a series of questions that were asked of any sample adult who reported low back pain in the past three months: frequency of low back pain (LBPREQ), severity of low back pain (LBPSEV), if he/she had ever been told by (LBPWKREL) or discussed with (LBPWKRL2) a doctor or other health professional that his or her lower back pain was work-related, is he/she had ever filed a workers' compensation claim for low back pain (LBPWCCLM), if he/she had ever received workers' compensation for low back pain (LBPWCBEN), number of work days in the past three months missed due to low back pain (LBPWKDAY), or if he/she had changed work activities in the past three months due to low back pain (LBPCHJOB).

## Major Recodes

In the ACN section, the survey questions on cancer were asked in a format that allowed a sample adult 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").

One recoded variable was created that captured how long a sample adult has had diabetes (DIFAGE2). This recoded variable DIFAGE2 is calculated as [AGE minus DIBAGE], prior to top-coding either AGE or DIBAGE. 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.

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

### Core Content

The 2015 Adult Health Status and Limitation of Activity section contains information from sample adults on illness behavior, health status, use of special equipment, limitations in functional activities, and the conditions underlying such limitations.

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

While the functional limitation questions in the AHS section may seem similar to questions in the FHS section in the Family Core, 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) only captured whether a person has difficulty walking without using special equipment. In contrast, the Sample Adult question on walking (FLWALK) asked about the degree of difficulty the sample adult has walking a specified distance (a quarter mile, or about three city blocks) by him/herself and without using any special equipment.

The functional limitation questions in the AHS section asked the sample adult to indicate the degree of difficulty he/she would have in performing specific physical tasks (e.g., 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.

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. Respondents were handed a flash card listing various condition 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.”

## Major Recodes

The recode FLA1AR is a summary measure that identifies 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.

## Technical Notes

If the sample adult was limited by a condition not listed in one of the 18 fixed categories mentioned above, 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. Respondents could list any number of applicable conditions.

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\_. 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 6 in the FHS section). The resulting 34 output categories were generally based on the International Classification of Diseases, Tenth Revision, Clinical Modification (see Table 7 in the FHS section).

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.” 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 sample adult’s difficulty with any functional 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 data file.

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.”

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.

In addition, 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 Adult Conditions (ACN) section.



---

## V. Adult Health Behaviors Section (AHB)

### Core Content

The Adult Health Behaviors Section (AHB) contains information on questions related to adult's cigarette smoking, leisure-time physical activity, alcohol use, height, and weight. The same questions have been included in the Sample Adult core every year since 1997 with only minor changes.

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. Current smokers are defined as persons who have ever smoked 100 cigarettes and who currently smoke every day or some days. Those who no longer smoked were asked about the length of time since they had quit.

Sample adults were asked the frequency and duration of doing vigorous leisure-time physical activities; frequency and duration of doing light or moderate leisure-time physical activities; and frequency of doing leisure-time strengthening activities. The questions about leisure-time physical activity are introduced with the following statement: "The next questions are about physical activities (exercise, sports, physically active hobbies...) that you may do in your leisure time." In this section, since sample adults 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).

The alcohol questions include whether the sample adult had at least 12 drinks of any type of alcoholic beverage in any one year, and then in entire life; and the frequency of drinking any type of alcoholic beverage in the past year. There were two additional questions about drinking in the past year: the average number of drinks had on days when drank alcoholic beverages and the number of heavy drinking days in the past year. Since 2014, women have been asked about days in which they had 4 or more drinks of any alcoholic beverage whereas men were still asked about days in which they had 5 or more drinks of any alcoholic beverage in the past year.

In the fourth quarter of 2014, a new question on binge drinking was added to AHB. This question, BINGE1, asked the number of times during the past 30 days that the sample adult had {5+ for men/4+ for women} drinks on an occasion. This was different from the binge drinking question that was asked earlier in 2014; that older question asked the number of times during the past 30 days that a respondent had {5+ for men/4+ for women} drinks in about two hours. This change was made because NHIS researchers believed the data based on BINGE to be an under-report of binge drinking occurrences due to the two hour timeframe. BINGE1, with its "on an occasion" timeframe, continued to be included in 2015.

Sample adults were asked their current height and current weight, without shoes. **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. See the Major Recodes section for additional details.

## Supplemental Questions

In 2015, the AHB section included six supplemental test questions (SMKANY, SMKAGEX, SMKNOWX, SMKNONOX, SMKNOTPX, CIGDAMOX) asked of sample adults who responded "No" or "Don't know" to SMKEV ("Have you smoked at least 100 cigarettes in your entire life?"). The six supplemental questions were designed to identify lifetime and current cigarette smoking among adults who did not meet the 100 cigarette threshold—that is, whether they were new smokers or were adults who had experimented with cigarettes at some point but never smoked as many as 100 cigarettes. The test questions were generally the same as those asked of sample adults who had smoked 100 cigarettes but with some minor wording changes: (1) "Have you ever smoked a cigarette even one time?" (SMKANY); "How old were you the first time you smoked a cigarette?" (SMKAGEX); "Do you now smoke every day, some days, or not at all?" (SMKNOWX). Those who responded "every day" or "some days" were then asked "On how many of the past 30 days did you smoke a cigarette?" (CIGDAMOX). Those who responded "not at all" were asked "How long has it been since you smoked a cigarette?" (SMKNONOX, SMKNOTPX).

## Major Recodes

SMKSTAT2 (Cigarette smoking status) classifies sample adults in terms of their lifetime and current cigarette smoking status. Categories include: current every day; current some days; former; never; {ever} smoker, current status unknown; and unknown if ever smoked {unknown lifetime status}. As noted earlier, current smokers are defined as persons who have ever smoked 100 cigarettes and who currently smoke every day or some days. Never smokers are defined as persons who never smoked any cigarettes or who have ever smoked less than 100 cigarettes.

ALCSTAT (Alcohol drinking status) classifies sample adults in terms of their lifetime and current alcohol drinking status. Categories include:

- Lifetime abstainer (<12 drinks in lifetime);
- Former infrequent (12+ drinks in lifetime but never as many as 12 in 1 year and none in past year);
- Former regular (12+ drinks in lifetime, 12+ drinks in 1 year, but none in past year);
- Former, unknown frequency (12+ drinks in lifetime, none in past year, don't know if 12+ in any 1 year);
- Current infrequent (12+ drinks in lifetime, and 1–11 drinks in past year);
- Current light (12+ drinks in lifetime, and ≤3 drinks per week in past year);
- Current moderate (12+ drinks in lifetime, and (male) >3 drinks per week up to 14 drinks per week or (female) >3 drinks per week up to 7 drinks per week); and
- Current heavier drinkers (12+ drinks in lifetime, and (male) >14 drinks per week in past year or (female) >7 drinks per week in past year).

Categories for current drinking status (frequency/level) unknown, and {lifetime} drinking status unknown are also included. Based on this classification, a current drinker is defined as someone who had had 12+ drinks in their lifetime and at least one drink in the past year.

Body Mass Index (BMI), a measure of body weight relative to height, was calculated using the formula:  $BMI = \text{kilograms} / \text{meters}^2$ . 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 variable values are released with 4 digits and two implied decimals. 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 ( $18.5 \leq \text{BMI} < 25$ ); overweight, but not obese ( $25 \leq \text{BMI} < 30$ ); and obese (BMI  $\geq 30$ ).

SAS and Stata code for creating physical activity recodes reflecting the 2008 Physical Activity Guidelines for Adults are available from the NHIS Adult Physical Activity Information website:

[http://www.cdc.gov/nchs/nhis/physical\\_activity/pa\\_recodes.htm](http://www.cdc.gov/nchs/nhis/physical_activity/pa_recodes.htm).

## Suppression of Data

To protect the confidentiality of NHIS sample adults who might be identifiable by their unusual physical characteristics, values of height for men are limited to 63–76 inches in the public use data file. Values of height for women are limited to 59–70 inches, values of body weight for men are limited to 126–299 pounds, and values of body weight for women are limited to 100–274 pounds. In cases where reported values were outside these limits for either height or weight, the data for both variables have been changed to 96 or 996 (not available) on the public use data file. For additional details, please review the section entitled “Body Weight and Height” within the AHB section and Appendix VIII of the 2006 NHIS *Survey Description*.

The body mass index (BMI) recode 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 data file for reasons of confidentiality.

Although extreme values of height and weight are not publicly released due to confidentiality concerns, analysts can apply for access to that information through the NCHS RDC at: <http://www.cdc.gov/rdc/>.

## Technical Notes

A question asking how often the sample adult 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 sample adult’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.

Respondents have the option of reporting 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. See the 2006 NHIS *Survey Description* and the 2008 NHIS *Survey Description* for historical information.

Please refer to the 2012 NHIS *Survey Description*, AHB section, for greater detail about adult’s smoking, alcohol questions, previous recodes and documentation corrections, height, weight, and BMI on the NHIS.

For additional information about the historical context, editing, frequently asked questions, and references for NHIS tobacco use information for adults and for NHIS physical activity information for adults, see:

[http://www.cdc.gov/nchs/nhis/special\\_topics.htm](http://www.cdc.gov/nchs/nhis/special_topics.htm).

---

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

### Core Content

The Adult Health Care Access and Utilization section (AAU) contains information on access to health care, dental care, health care provider contacts, and immunizations. 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 the place of care; reasons for a delay in getting medical care; and the inability to be able 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.

Questions regarding health care provider contacts include visits and other contacts with doctors and other health care professionals during the past 12 months. Questions about doctor visits allow for visits not only from medical doctors but from a variety of other health care professionals including: 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. Questions asking about surgery or surgical procedures and the number of surgeries are included. Questions about home care also are included and are asked independently of other types of health care visits. 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, and there is a question that asks how long it has been since the sample adult has seen or talked to a doctor.

### Supplemental Questions

There are several supplementary questions related to adult immunizations sponsored by the National Center for Immunization and Respiratory Diseases (NCIRD, CDC) including: 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. The questions pertaining to flu shot and nasal spray flu vaccine were also administered to all sample children (see the CFI section).

There were no changes to the flu vaccination questions in the AAU section in 2015. The questions were the same as those used in the latter part of 2010 and 2011-2014. 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 online in the 2010 NHIS Sample Adult Variance Layout Report, AAU Appendix.

Additional supplementary questions inquire whether sample adults ever had chickenpox and if it had been in the past 12 months; ever had hepatitis; ever lived with someone with hepatitis; ever 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.

Additional questions ask sample adults 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. Also included

are questions that ask sample adults if they have ever received the HPV vaccine and number of HPV shots received.

In 2015, the NHIS continues to include five supplementary questions sponsored by the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP, CDC) which were added to the survey in 2013. The variables based on responses to the questions include data about having been prescribed medication (ARX12MO); having a blood test for hepatitis B and C and reason tested for hepatitis C (AHEPBTST, AHEPCTST, and AHEPCRES); and age of receipt of first HPV shot (AHPVAGE).

Expanded content on health care access and utilization sponsored by the Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS) were first added to the survey in 2011 and have been part of this section since then. The questions are embedded throughout the AAU section, and variables based on the responses to these questions include data about difficulty finding providers (APRVTRYR, APRVTRFD, ADRNANP, ADRNAI); unmet medical need due to cost (AHCAFYR5, AHCAFYR6); worry about paying medical bills (AWORPAY); health insurance compared to a year ago (AHICOMP); strategies used to reduce prescription drug cost (ARXPR\_1-ARXPR\_6); timing of and reasons for emergency room visits (AERVISND, AERHOS, AERREA1R-AERREA8R); health information technology use (HIT1A-HIT5A); preventive care screening which include checking for high blood pressure, high cholesterol, colon cancer, testing for diabetes, a pap test, and mammogram (APSBPCHK, APSCHCHK, APSBSCHK, APSPAP, APSMAM, APSCOL); whether doctor mentioned diet or smoking (APSDIET, APSSMKC); and purchase of health insurance in past 3 years and difficulty purchasing insurance (AINDINS, AINDPRCH, AINDWHO, AINDDIF1-AINDDIF2).

In addition, beginning in Quarter 4 of 2013, a supplementary question that asked if sample adults had looked into purchasing health insurance through the Health Insurance Marketplace was added. This question (AEXCHNG) was sponsored by the Office of the Assistant Secretary for Planning and Evaluation (ASPE, HHS). Starting in 2014, this question has been asked in all four quarters.

Lastly, the supplemental questions ANOUSPL1-ANOUSPL9, AVISLAST, ALASTTYP1-ALASTTYP4, AVISAPT2, AVISAPN2, AWAITRMN, AWAITRMT, LTCFAM, LTCHELP, LTCWHO1-LTCWHO5, AINDENY1-AINDENY3, and AINDNOT1-AINDNOT5 that were added in 2011 were deleted from the AAU section in 2015.

## Technical Notes

In 2014 (but not in 2015), additional blood pressure and cholesterol screening questions were included in the Adult Conditions (ACN) section as part of the supplemental questions pertaining to the Million Hearts<sup>®</sup> Initiative. Valid responses to the blood pressure screening questions (HYBPCKNO and HYBPCKTP) in the ACN section were used to fill the response to the blood pressure screening question in the AAU section (APSBPCH1), and this question was not asked directly of most sample adults. Similarly, valid responses to the cholesterol screening questions (CLCKNO and CLCKTP) in the ACN section were used to fill the response to the cholesterol screening question in the AAU section (APSCHCH1), and this question was not asked directly of most sample adults. Sample adults with “don’t know” or “refused” responses to the screening question in ACN were asked the corresponding AAU question, unless they broke off participation prior to this point in the survey. Values for persons who broke off participation prior to this point are coded as “not ascertained,” even if there were valid responses to the ACN questions. The variable names were changed from APSBPCHK to APSBPCH1 and from APSCHCHK to APSCHCH1 to reflect these modifications. Estimates based on APSBPCH1 and APSCHCH1 were not comparable with earlier years.

In addition, because APSBPCHK and APSCHCHK were not asked directly of most sample adults, the diabetes screening question (APSBCHK) generally followed WRKHLTH or WRKDIR rather than the two blood pressure

and cholesterol screening questions. The variable name was changed from APSBSCHK to APSBSCH1 to reflect this change. Estimates based on APSBSCH1 were not comparable with earlier years.

In 2015, the supplemental questions which led to the above-mentioned auto-fills and question placement changes have been removed, and the variable names for blood pressure, cholesterol, and diabetes screenings have returned to those from 2013 and earlier: APSBPCHK, APSCHCHK, and APSBSCHK, respectively. Preliminary analyses suggest that estimates based on these three variables are not comparable with 2014; however, they are comparable with years earlier than 2014.

## VII. Adult Selected Items Section (ASI)

### Core Content

The Adult Selected Items (ASI) section collected information on the frequency of computer usage (ASICPUSE), satisfaction with healthcare received (ASISATHC), neighborhood tenure and perceived neighborhood social cohesion (ASITENUR, ASINHELP, ASINCNT0, ASINTRU, ASINKNT), sexual orientation (ASISIM, ASISIF), financial worries (ASIRETR, ASIMEDC, ASISTLV, ASICNHC, ASICCOLL, ASINBILL, ASIH CST, ASICCMP), hours slept and trouble falling/staying asleep (ASISLEEP, ASISLPFL, ASISLPST, ASISLPMD, ASIREST) and HIV testing (ASIHIVT, ASIHIVWN).

The ASI section continued with almost all the core questions asked since 2013 when this section was first introduced as part of the Sample Adult component. Some of the continuing questions in the ASI section were included in other sections of the NHIS prior to 2013. Questions in the NHIS prior to the launch of the ASI section included hours slept (ASISLEEP, previously in the AHB section) and HIV testing (ASIHIVT, ASIHIVWN, previously in the AAU and AIDS Knowledge and Attitudes section (ADS) sections, respectively). Data users should note that with the movement of these questions to the ASI section, the corresponding variable names were changed.

Six core questions in the ASI section were removed in 2015 (ASISMEELS, ASISIMDK, ASIMSESP, ASISFELS, ASISIFDK, and ASIFSESP). These questions were follow-up questions to the initial question asking men about their sexual orientation (ASISIM) and the initial question asking women about their sexual orientation (ASISIF). It was found that the information collected does not alter the interpretation of the initial sexual orientation questions for either men or women. Prior to 2015, these data were suppressed for confidentiality reasons.

### Supplemental Questions

Since 2013, the ASI section has included seven supplemental questions about the sample adult's current mental or emotional health and the extent to which these feelings interfere with his or her life or daily activities (ASISAD, ASINERV, ASIRSTLS, ASIHOPLS, ASIEFFRT, ASIWTHLS, and ASIMUCH). These questions were fielded in the ACN section previously.

### Cognitive Testing

Some of the topics in this section were the subject of cognitive testing conducted by the National Center for Health Statistics' Questionnaire Design Research Laboratory (QDRL). A report (Miller and Ryan, 2011) describing the results of this testing for the survey questions pertaining to sexual orientation is available at [http://wwwn.cdc.gov/qbank/report/Miller\\_NCHS\\_2011\\_NHIS%20Sexual%20Identity.pdf](http://wwwn.cdc.gov/qbank/report/Miller_NCHS_2011_NHIS%20Sexual%20Identity.pdf).

## Technical Notes

A quality assessment of the sexual identity data collected in the ASI section as part of the 2013 National Health Interview Survey is available at [http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_169.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_169.pdf).

Six survey questions related to non-specific psychological distress were asked to adults in the ASI section, related to whether he or she experienced feelings of sadness, nervousness, restlessness, hopelessness, worthlessness, or that everything was an effort in the past 30 days. These six survey questions can be used together to create Kessler's "K6" screen for non-specific psychological distress. For more information about Kessler's K6 screen, please refer to [http://www.hcp.med.harvard.edu/ncs/k6\\_scales.php](http://www.hcp.med.harvard.edu/ncs/k6_scales.php).

## VIII. Adult Internet and Email Usage Section (AWB)

### Supplemental Questions

The Adult Internet and Email Usage section (AWB) is a supplement that began in 2012. It was sponsored by the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services (HHS) in order to collect information on internet usage. This section contains information on internet and email usage for sample adults aged 18 years and over, including whether or not the adult uses the internet or email and the frequency at which they are used. This section also includes a question (AWEBORP) that asks sample adults if they are currently a member of an online research panel.

---

## Adult Functioning and Disability Supplement File (AFD)

### Supplemental Questions

The Adult Functioning and Disability supplement (AFD) fielded with the 2015 NHIS Sample Adult module is part of an international project to develop and test improved measures of functioning. Questions are asked about sample adults' 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). Follow-up questions on the degree of difficulty, use of assistive devices, and functioning with assistance were included for most domains. More information on the international project and the questions used in this supplement can be found on the NCHS website at: [http://www.cdc.gov/nchs/washington\\_group.htm](http://www.cdc.gov/nchs/washington_group.htm).

Approximately one half of all sample adults were randomly selected to receive the 2015 AFD supplement. However, 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 persons 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 92% of AFD sample adults are coded "4" on RCS\_AFD.

### Technical Notes

Because the AFD supplement was only administered to one half of all sample adults, different weights (WTFA\_AFD) were generated for sample adults in the AFD File. In addition, a separate, stand-alone public use file was created for the AFD variables, rather than appending these variables to the 2015 Sample Adult File. 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 AFD File with one or more NHIS data files (e.g., Sample Adult, Person, etc.). Information on merging data files can be found in Appendix IV of this document. The weight WTFA\_AFD provided with the AFD File is designed to produce annual-level estimates calculated based on data included in this file.

The 2015 AFD supplement is very comparable to the 2012-2014 AFD supplements in terms of the questions asked on the NHIS instruments across these survey years. However, the 2012-2015 AFD instruments contained fewer detailed follow-up questions than the 2011 AFD instrument, which in turn contained fewer detailed follow-up questions than the first supplement in 2010, the Quality of Life (QOL) supplement. Thus, analysts interested in combining years of data should read the notes in the Variable Layout Report, which provide information regarding the comparability of 2015 variables relative to variables in the 2011–2014 AFD files and in the 2010 QOL file. Users should also be aware that only one-quarter of sample adults were selected to receive the 2010 QOL and 2012 AFD supplements, whereas one-half of sample adults were selected to receive the 2011 and 2013-2015 AFD supplements.



---

## Cancer Control Supplement File

In 2015, a Cancer Control Supplement was sponsored by NIH's National Cancer Institute, CDC's National Center for Chronic Disease Prevention and Health Promotion, and FDA's Center for Tobacco Products. The supplement consisted of six sections covering diet and nutrition, physical activity, tobacco, cancer screening, genetic testing, and family history. These sections are described in greater detail below. The sample adults for each family also participated in the Cancer Control Supplement. As a result of the large number of variables derived from the Cancer Control Supplement, NHIS staff decided to create a separate, stand-alone file for the cancer variables, rather than append the variables to the 2015 Sample Adult file.

### I. Diet and Nutrition Section (NAC)

The Diet and Nutrition section (NAC) of the 2015 Cancer Control Supplement collected information about selected foods and vitamins consumed during the past month. Sample adults were asked about the quantity and frequency of consumption of hot or cold cereal, milk, soda, 100% fruit juice, coffee and tea, sports and energy drinks, fruit drinks, fruit, salad, fried potatoes, other kinds of potatoes, beans, brown rice, other vegetables, salsa, pizza, tomato sauces, cheese, red meat, processed meat, whole grain bread, candy, donuts, cookies, ice cream, and popcorn. Responses for the number of times a particular item was consumed were reported in terms of daily, weekly, or monthly consumption. Respondents who consumed hot or cold cereal were also asked for the name of up to two cereals usually consumed, and those who consumed milk were asked about milk fat type. Sample adults were asked whether they use any kind of vitamins, mineral pills or supplements. Those who reported any vitamin consumption were asked whether they consumed multivitamins, calcium supplements and vitamin D, on how many days in the past month each vitamin type was taken, and the main reason for taking vitamin D.

Most of the questions administered in the 2015 NAC Cancer Control Supplement were the same as those administered in the 2010 NAC Cancer Control Supplement. Changes in the 2015 NAC Supplement include a revised wording for the question about coffee consumption, (replacing COFFEENO and COFFEETP with COFFEEN1 and COFFEET1) and the removal of a question about how often light or low fat cheese was consumed among those who ate cheese in the past month.

#### Suppression of Data

To protect the confidentiality of NHIS sample adults who might be identifiable by their infrequent responses, verbatim responses for languages other than English or Spanish (OTHLANG), cereals names other than the top 10 most common cereal names consumed (CERTYP, CERTPSP, CERTYP3 and CERTP3SP) and other reasons for taking vitamin D (VITDSPEC) were suppressed. These variables are only available via the NCHS - Research Data Center (<http://www.cdc.gov/rdc/>).

### II. Physical Activity Section (NAD)

The Physical Activity section (NAD) of the 2015 Cancer Control Supplement collected information about daily physical activity that supplemented the basic physical activity data collected in the Sample Adult core. All sample adults were asked about walking for transportation, and those who were able to walk were also asked about walking for fun, relaxation, or exercise. In 2015, a new set of questions was introduced to assess transportation access and the walkability of respondents' communities. All sample adults were then asked the average number of hours per day that they spend in a sitting position (separate questions distinguish weekdays from weekends) and whether a doctor or other health professional had recommended in the past 12 months that they begin or continue exercise or physical activity.

### III. Tobacco Section (NAE)

The Tobacco section (NAE) of the 2015 Cancer Control Supplement collected tobacco use information that supplemented the basic cigarette smoking data collected in the Sample Adult core. Questions asked of current and former cigarette smokers (identified in the core) included their use of menthol cigarettes, number of cigarettes smoked (by former smokers) and methods used to quit or attempt to quit smoking. Current smokers who had not tried to quit in the past year were asked if they had ever tried to quit in their lifetime and all current cigarette smokers were asked whether they wanted to quit.

Electronic cigarette (e-cigarette) use questions, previously asked in the Sample Adult core in 2014, were asked in the Cancer NAE section. These included whether e-cigarettes were ever used (even once), current use (every day, some days, not at all), and number of days used in past 30 days.

Questions about non-cigarette tobacco included lifetime use of cigars, pipes, and smokeless tobacco. For cigars and smokeless tobacco, sample adults were asked (1) if they ever used/smoked the product even one time, (2) if they smoked a least a minimum threshold (50 cigars; used smokeless at least 20 times); (3) if they currently used the product every day some days or not at all; and (4) the number of days in the past 30 days that they smoked or used the tobacco product. Smokeless tobacco users were asked the brand used. Questions for pipes were limited to lifetime (ever used even one time) and current use (every day, some days, or not at all.)

Current tobacco users and recent former cigarette smokers who had seen or talked to a doctor or other health care professional in the past 12 months were asked whether that health care professional had advised them to quit smoking (or quit using other tobacco products) and the type of health professional who gave the advice.

The final five questions in the section asked all female sample adults 18-49 years of age who had given birth to a live born infant within the past five years about their smoking habits during their pregnancy. Questions included smoking before and during pregnancy, quit attempts and success of any attempts.

#### Suppression of Data

The verbatim fields in the NAE section include smokeless tobacco brand names (SMKLBRSP) and types of other health professionals who advised the respondent to quit using tobacco (HPTOTH1). These variables are only available via the NCHS - Research Data Center (<http://www.cdc.gov/rdc/>) to protect respondent's confidentiality.

### IV. Cancer Screening Section (NAF)

The Cancer Screening section (NAF) of the 2015 sample adult questionnaire collected information about selected cancer screening examinations received by the sample adult and behaviors linked to some types of cancer. Since 2000, information on cancer screening has been collected as a part of a recurring Cancer Control Supplement every five years. Additionally, the NAF is administered as a brief supplement three years after each previous full supplement. Prior to 2015, NAF was last administered in 2013 (a partial supplement year).

The 2015 NAF section includes questions regarding Pap smear tests and hysterectomy (for all sample adult women); mammograms (for sample adult women aged 30 years and over); prostate specific antigen (PSA) tests (for sample adult men aged 40 years and over); and colorectal screening exams and fecal occult blood (FOB) tests (for sample adults aged 40 years and over). Beginning in 2015, keeping in line with the new recommendations of the U.S. Preventive Services Task Force (USPSTF), women were asked if they received an

HPV test with their most recent Pap test. For each type of cancer screening exam, sample adults who indicated that they had the exam were asked when the most recent screening exam occurred (month/year, number of days/weeks/months/years ago, or time interval grouping). Sample adults were also asked their reasons for having the particular exam and questions regarding recommendations for the exam by a doctor or other healthcare professional. The 2015 NAF section also includes questions on the use of tanning devices in the past 12 months.

In 2015, new questions on the use of over the counter and prescription pain medications that contain acetaminophen were added to the survey, as well as questions on how much was spent out of pocket on screening examinations.

## Major Recodes

Two sets of time-since-screening recodes were created for the NAF. The set of “A” recoded variables (RMAM3A, etc.) was created for 2005, 2008, 2010, 2013 and 2015 from respondents’ answers to previous questions; no data were imputed. Information was collected in three formats: month and year, number and time unit (e.g. days, weeks, months, years), or time category. In addition, any respondent giving an incomplete or partial date of a screening test was also asked to answer the time category question (e.g., RMAM2). Available data were combined to create time-since-screening recodes. Because these data were provided by the respondent, these time-since-screening recodes are recommended by NCHS and the National Cancer Institute for the 2005, 2008, 2010, 2013 and 2015 NAF data.

A set of “B” recoded variables (RMAM3B, etc.) was created for analysts who wish to study trends in time-since-screening for 2000–2015. Although the time-since-screening questions and answer categories for the various cancer screening examinations have been fairly consistent since the 2000 NHIS, there were differences in the skip patterns of the questions between earlier (2000 and 2003) and later years (2005, 2008, 2010, 2013 and 2015). For the earlier years, missing time-since-screening data were imputed; for the later years, it was not. Therefore, a common method was used to estimate the missing time-since-screening data for analysts who wish to study trends for 2000–2015. In the set of “B” recoded variables, all variables from 2000–2015 were created using the imputation method used for the earlier years.

Starting with the 2010 NHIS and continuing in 2015, time-since-screening recodes for the various cancer screening examinations are included in the released data files. In 2005 and 2008, time-since-screening recodes were not included in the released data files, but SAS programming code to produce these recodes was included.

## Suppression of Data

To protect the confidentiality of NHIS sample adults who might be identifiable by their unusual responses, BIRTHNUM (number of live births) and BIRTHAGE (age when first child was born) were top-coded. TAMOXSP (other reason for taking tamoxifen) and RALOXSP (other reason for taking raloxifen) were suppressed, but analysts wishing to gain access to these restricted data can apply for access through NCHS’ RDC at: <http://www.cdc.gov/rdc/>.

## V. Genetic Testing Section (NAG)

The Genetic Testing Section (NAG) is part of the Cancer Control Supplement File in the 2015 NHIS. It contains a series of questions asking sample adults if they had ever had genetic counseling or testing for cancer risk, the

reason for such counseling and/or testing, and the types of cancer for which the counseling and/or testing was performed.

The specific questions asked of a sample adult in the NAG section regarding genetic counseling include whether he/she ever received genetic counseling for cancer risk (GCEVER), the reason for this counseling (GCMREAS), and the type of cancer for which it was received – breast (GCBREAST), ovarian (GCOVRN), colon or rectal (GCCOLON), or another type (GCANOTH). For genetic testing, the specific questions asked whether a sample adult had discussed with a doctor or other health professional getting genetic testing (GTPOSS1), been advised to have testing performed (GTADVIS1), and ever had genetic testing completed to determine if he/she is at a greater risk for developing cancer (GTGRISK). Additional questions on genetic testing were also asked and included what type of cancer the testing was performed – breast (GTBRE), ovarian (GTOVA), colon or rectal (GTCOL) or another type (GTOTH) – when the testing was performed (GTRSK\_MT, GTRSK\_YR, GTRSKN, GTRSKT, and GTRSK2), and the likelihood of having either colon or rectal cancer (GTCCLOM) or (for women only) breast cancer (GTCBCM).

### Suppression of Data

To protect the confidentiality of NHIS sample adults, GCSPEC (type of cancer for which genetic counseling received) and GTRKOTH1 (other genetic tests for cancer) were suppressed, but may be accessed through the RDC (<http://www.cdc.gov/rdc/>).

## VI. Family History Section (NAH)

The questions in the Family History section (NAH) of the 2015 Cancer Control Supplement were asked of all sample adults. Respondents were asked whether their biological father or mother had ever had cancer of any kind. If a "yes" response was obtained, the respondent was then asked to specify the type of cancer (up to three specific types) and if there were more than three types, and whether the parent in question was less than 50 years of age when the cancer was first diagnosed. The NHIS also collected information on the number of "full" brothers and sisters (same biological mother and father), the number of siblings diagnosed with a particular type of cancer (up to three specific types) and if there were more than three types, and the number of siblings who were less than 50 years of age when the cancer was first diagnosed. The NHIS further collected information on the number of "full" (biological) sons and daughters, the number of children diagnosed with a particular type of cancer (up to two specific types) and if there were more than two types, and the number of children who were less than 50 years of age when the cancer was first diagnosed.

### Suppression of Data

To protect confidentiality, information on several rare cancer types were suppressed. In addition, responses indicating colon or rectal cancer were combined into one variable, as were responses indicating head or neck cancer. FHBNUM (number of full brothers) and FHSNUM (number of full sisters) were top-coded; FHDNUM (number of biological daughters) and FHNNUM (number of biological sons) were suppressed; and variables concerning types of cancer not applicable to men (for example, cervical cancer, uterine cancer, and ovarian cancer) were suppressed. Suppressed variables are available through the RDC (<http://www.cdc.gov/rdc/>).

---

## Weighting

The NHIS 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 are reflected in the sample weights that are provided in the accompanying data files.

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.

In 2015, NCHS weights were derived from 2010-census-based population estimates.

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. If data are not weighted, severely biased estimates may result.

Each file has weights based on the unit of analysis, as described below. For more detail about the calculation of the sampling weights, please see "*Design and estimation for the National Health Interview Survey, 2006–2015*," available from: [http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_165.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_165.pdf).

### Person-Level Analyses

Two sets of weights are provided on the Person 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.

Analysts should be aware that, in 2015, data about 269 persons who were active duty members of the Armed Forces at time of interview are on the Person 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.

---

## Sample Child or Sample Adult Analyses

Two sets of weights are included on the Sample Child 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 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.

A non-response adjustment is included in 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.

## Household-Level Analyses

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.

## Family-Level Analyses

Lastly, the Family File contains one set of weights. 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 2015 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 2015 NHIS Family File, WTFA\_FAM, corresponds to the 2015 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.

### 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, multistage sample design that involves stratification, clustering, and oversampling of specific population subgroups. 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.

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® with Taylor series linearization methods to produce standard error estimates and hypothesis test results (such as p values). SUDAAN® is available from: <https://www.rti.org/impact/sudaan-statistical-software-analyzing-correlated-data> .

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. 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.

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.



## Merging Data Files and Combining Years of Data in the NHIS

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 persons 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 increase the number of persons and the precision of estimates.

Each data file contains household, family, and person numbers that make merging the files possible, if needed. Appendix IV provides more detailed information, as well as sample code for merging the files and combining years of data. It is also important to note, however, that some frequently used variables are repeated on various data files; therefore, merging of files may not be required.

---

## NCHS Record Linkage Program

NCHS has developed a record linkage program designed to maximize the scientific value of the Center's population-based surveys. Linked data files enable researchers to examine the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality. NCHS is currently linking NHIS survey data with death certificate records from the National Death Index (NDI); enrollment and claims data from the Centers for Medicare and Medicaid Services (CMS); Retirement, Survivor, and Disability Insurance (RSDI) and Supplemental Security Income (SSI) benefit data from the Social Security Administration (SSA); End Stage Renal Disease (ESRD) data obtained from the United States Renal Data System (USRDS); and administrative data for the Department of Housing and Urban Development's (HUD) largest rental housing assistance programs. For more information, see [http://www.cdc.gov/nchs/data\\_access/data\\_linkage\\_activities.htm](http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm).

NHIS interviewed households also serve as a sampling frame for the Medical Expenditure Panel Survey (MEPS). MEPS, conducted by the Agency for Healthcare Research and Quality (AHRQ), collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of health insurance held by and available to U.S. workers. The MEPS Household Component collects data from a nationally representative subsample of households that participated in the prior year's NHIS. Crosswalks that will allow data users to merge the MEPS full-year population characteristics public use data files with the NHIS person-level public use data files are available from AHRQ: [http://meps.ahrq.gov/mepsweb/data\\_stats/more\\_info\\_download\\_data\\_files.jsp#hc-nhis](http://meps.ahrq.gov/mepsweb/data_stats/more_info_download_data_files.jsp#hc-nhis).

---

## Guidelines for Data Use

With the goal of mutual benefit, NCHS requests that recipients of data files cooperate in certain actions related to their use.

Any published material derived from the data should acknowledge NCHS as the original source. The suggested citation, "SOURCE: NCHS, National Health Interview Survey, 2015," should appear at the bottom of all tables and graphs. Published material derived from the data should also include a disclaimer that credits the author's analyses, interpretations, and conclusions 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 ensure that the description is not inconsistent with that published by NCHS.

CIPSEA and the Public Health Service Act (Section 308d) provide that NHIS data collected by NCHS may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by these laws. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users must:

- Use the data in these data files for statistical reporting and analysis only.
- Make no use of the identity of any person discovered, inadvertently or otherwise, and advise the Director, NCHS, of any such discovery (301-458-4500).
- Not link these data files with individually identifiable data from any other NCHS or non-NCHS data sets.

Use of the NHIS data files signifies users' agreement to comply with the above-stated statutory-based requirements.

---

## References

Achenbach TM, Edelbrock C. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont, Department of Psychiatry. 1983.

Angold A, Costello EJ. The Child and Adolescent Psychiatric Assessment (CAPA). Journal of the American Academy of Child and Adolescent Psychiatry 39(1):39–48. 2000.

Brault M, Stern S, Raglin D. Evaluation Report Covering Disability. 2006 American Community Survey Content Test Report P.4. U.S. Census Bureau. Available from:  
[http://www.census.gov/content/dam/Census/library/working-papers/2007/acs/2007\\_Brault\\_01.pdf](http://www.census.gov/content/dam/Census/library/working-papers/2007/acs/2007_Brault_01.pdf)

Egger H, Ascher BH, Angold A. 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. 2004.

Goodman R. The Strengths and Difficulties Questionnaire: A research note. Journal of Child Psychology and Psychiatry 38(5):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.

Korn EL, Graubard BI. Analysis of health surveys. New York: John Wiley & Sons. 1999.

Miller, K, Ryan, JM. Design, Development and Testing of the NHIS Sexual Identity Question. National Center for Health Statistics. Hyattsville, MD. 2011. Q:Bank (on-line access to question evaluation reports). Available from:  
[http://wwwn.cdc.gov/qbank/report/Miller\\_NCHS\\_2011\\_NHIS%20Sexual%20Identity.pdf](http://wwwn.cdc.gov/qbank/report/Miller_NCHS_2011_NHIS%20Sexual%20Identity.pdf)

Moriarity C, Parsons, V. Expanding the number of primary sampling units for the National Health Interview Survey. Proceedings of the Joint Statistical Meetings. Alexandria, VA: American Statistical Association. 2013. Available from: [http://www.amstat.org/sections/srms/proceedings/y2013/Files/308525\\_81035.pdf](http://www.amstat.org/sections/srms/proceedings/y2013/Files/308525_81035.pdf).

National Center for Health Statistics. Health, United States, 2014: With Special Feature on Adults Aged 55–64. Hyattsville, Maryland. 2015. Available from: <http://www.cdc.gov/nchs/hus.htm> .

Parsons VL, Moriarity C, Jonas K, et al. Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. Vital Health Stat 2(165). 2014. Available from:  
[http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_165.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_165.pdf).

Ringeisen H, Aldworth J, Colpe LJ, Pringle B, Simile C. Estimating the prevalence of any impairing childhood mental disorder in the National Health Interview Survey. International Journal of Methods in Psychiatric Research 24(4): 266–274. 2015.

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. Advance Letter

### Department of Health and Human Services, Centers for Disease Control and Prevention letterhead

From the Director of the United States National Center for Health Statistics

I'm Charles Rothwell and I head the National Center for Health Statistics (NCHS), part of CDC (the United States Centers for Disease Control and Prevention). In partnership with the U.S. Census Bureau, my agency is conducting a major survey about the nation's health and we need your help.

In the next few days, a Census Bureau interviewer will ask you some questions to see if you are eligible for the National Health Interview Survey. For your protection, the interviewer will show you an official identification card.

Please know that everything you tell us will be kept strictly private. Your answers are used only for health research, and to help understand and solve today's health problems and anticipate future health issues. Quality health information is necessary to make good decisions and sound policies. In this way, taking part in the survey indirectly benefits all Americans.

Strict federal laws protect your information. Question 6 on the back of this letter describes these laws and who may see your personal information.

I hope you will want to take part in the survey—it is your choice. No penalties or loss of benefits will come from refusing.

Some interviews take about five minutes. Most interviews will take about an hour to do all parts, depending on the size and health of your family. You may choose not to answer any question and, of course, you can stop at any time. Health and health care information from other records may be combined with your survey answers. These data also will be kept strictly private. You also may be given the choice to take part in other surveys sponsored by the National Center for Health Statistics.

Please contact the Census Bureau, toll-free, at 1-800-991-2520, Ext. 43480, if you have questions about the survey or to schedule an interview. About a week after the interview, some households will be asked a few extra questions for quality purposes.

You can learn more about the survey at our website: <http://www.cdc.gov/nchs/nhis.htm>.

I know your time and privacy are valuable so I am very grateful for your help. Thank you for your cooperation.

Sincerely,

/Charles J. Rothwell/

Charles J. Rothwell  
Director, National Center for Health Statistics  
Centers for Disease Control and Prevention

---

## FREQUENTLY ASKED QUESTIONS ABOUT THE NATIONAL HEALTH INTERVIEW SURVEY (NHIS)

### 1. HOW WAS I CHOSEN FOR THE SURVEY?

Every month we pick between 5,300 and 5,500 home addresses across the entire United States. We pick addresses using scientific methods so they represent all communities in the U.S.

### 2. WHY NOT INTERVIEW AT THE HOUSE ACROSS THE STREET? WHY IS MY PARTICIPATION IMPORTANT?

It is important that the people living at the address selected be in the survey. Due to the scientific methods used to pick addresses, we cannot exchange one address for another. If we did that, the survey results would not describe the entire country.

### 3. I AM NOT SICK – WHY SHOULD I TAKE PART IN A HEALTH SURVEY?

This is a survey of the Nation's health. We want to know how many people are sick and why they are sick, but it is also important to know how many people are healthy and why they are healthy. Everyone's answers are important.

### 4. WHAT ARE YOU GOING TO ASK ME?

The NHIS covers a wide range of topics like doctor visits, medical conditions, health insurance, physical activity, and injuries. We also ask questions that help us better understand the health information you give us. For example, we ask about race, income, and permission to combine your answers with information from other places, like medical records. Most people have no difficulty with any of the questions in the NHIS. However, others find some questions to be sensitive. You do not have to answer any questions you don't want to.

### 5. WHY DO YOU ASK ABOUT IMMUNIZATIONS?

Immunizations help prevent infectious disease, disability, and death. To get the most accurate picture of children's immunization levels, we may ask you for permission to contact your child's immunization providers. This is only for young children and teenagers, and we ask their providers only for immunization dates and doses. Like all the data we collect, this information is treated as confidential.

### 6. WHO WILL SEE MY ANSWERS?

We take your privacy very seriously. Only those NCHS employees, our specially designated agents including the U.S. Census Bureau, and our full research partners who must use your personal information for a specific reason can see your answers. Everyone else who uses your data can do so only after all information that could identify you and your family is removed. The answers you give us are used for statistical research only. This means that your answers will be combined with those given by other people in a way that protects everyone's identity.

Strict laws prevent us from releasing information that could identify you or your family to anyone else without your consent. Congress authorized the NHIS data collection in Section 306 of the Public Health Service Act (42 United States Code 242k). The federal laws that require all information we collect to be held in strict confidence are Section 308(d) of the Public Health Service Act [42 United States Code 242m (d)] and the Confidential Information Protection and Statistical Efficiency Act (PL 107-347). If any federal employee, contractor, or agent gives out confidential information not authorized by law, he or she can be fired, fined and/or imprisoned.

### 7. WHO LOOKS OUT FOR THE INTERESTS OF SURVEY PARTICIPANTS?

Every year, the Research Ethics Review Board (ERB) of the National Center for Health Statistics reviews survey content and methods to protect study participants. You may call the ERB if you want to ask about your rights as a participant in this research study. The toll-free number is 1-800-223-8118. Please leave a brief message with your name and phone number. Say you are calling about Protocol # 2009-16. Your call will be returned promptly.

HIS-600(L) (2-2014)

---

## Appendix II. Calculation of Response Rates

The response rates calculated here pertain to the Core questions in the 2015 NHIS. Note that the categories “interviewed households,” “interviewed families,” “interviewed sample children,” and “interviewed sample adults” include those with completed interviews or acceptable “sufficient” partial interviews.

### Household Response Rate

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).

### Conditional Family Response Rate

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

The Final Family Response Rate is the rate for those families identified as eligible that **takes into account household non-response**. The Final Family Response Rate is calculated by dividing the number of 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

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

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

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

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.

**Table I. Conditional response rates, National Health Interview Survey 1997–2015**

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
2013	75.7	99.0	92.1	81.7
2014	73.8	99.0	91.2	80.5
2015	70.1	98.9	91.4	79.7



**Table II. Unconditional response rates, National Health Interview Survey 1997–2015**

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
2013	75.7	74.9	69.0	61.2
2014	73.8	73.1	66.6	58.9
2015	70.1	69.3	63.4	55.2

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

Some users may wish to calculate a single response rate across multiple years. This is reasonable to do because the NHIS is continuously in the field. The response rates for combined NHIS data years are calculated in the same basic way as for a single year. For example, the Household Response Rate for Combined Data Years can be 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.

The Conditional Family Response Rate for Combined Data Years can be 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.

Then, the Final Family Response Rate for Combined Data Years can be 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.

Similar methods apply for more than two years of data and for calculating sample child and sample adult response rates. The 2015 counts for eligible and interviewed sample units are given in Table III. Counts for 2013 and 2014 are available in the 2013 and 2014 *NHIS Survey Descriptions*, and counts for data years 1997–2012 are available in the 2012 *NHIS Survey Description*.

**Table III. Number of households, families, and persons eligible and interviewed by module, National Health Interview Survey, 2015**

File / Type of Records	Eligible	Interviewed
Household / households*	59,170	41,493
Family / families	42,762	42,288
Sample Child / persons	13,444	12,291
Sample Adult / persons	42,270	33,672

\*Number of eligible households includes a small number of Type A non-response households that are ineligible for the survey. See description of the household response rate in Appendix II.

### Partial and Break-off Interviews

In the NHIS, an interview is a “partial” when all sections are not completed. Partials can be due to many reasons, such as the unavailability of a needed respondent or the interviewer failing to complete the interview during the allotted assignment period. Another reason for a partial is a “break-off,” which occurs when a respondent stops the interview in-progress before completion. The partial rate is the percent of all sufficiently complete interviews that are not entirely complete; these “sufficient partials” count towards the completion rate. Partials that are not far enough along in the interview, known as “insufficient partials,” are included in the refusal rate. Additional information about partial and break-off cases is available from: [http://fcsmsites.usa.gov/files/2014/05/2003FCSM\\_Stussman.pdf](http://fcsmsites.usa.gov/files/2014/05/2003FCSM_Stussman.pdf).

In the 2015 data file, for example, almost 8% of completed sample adult cases were partially completed interviews that were deemed sufficient for analysis (that is, the sample adult discontinued the interview but did complete at least the first three sections). This represents an increase from percentages seen over the last several years, when slightly more than 2% of completed sample adult cases were partially completed. The unusually large percentage of partially completed interviews in 2015 has been observed in previous years when large numbers of supplementary questions have also been fielded.

## Appendix III. Variance Estimation Method for Public Use Data

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. The method described below is applicable to the 2015 NHIS Household, Person, Sample Child, Sample Adult, and supplement public use data files.

Data 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.

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® to produce standard error estimates. (SUDAAN® is available from: <https://www.rti.org/impact/sudaan-statistical-software-analyzing-correlated-data>.) 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. 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.

### Variables Available for Variance Estimation

The 2015 Household, Person, Sample Child, Sample Adult, and supplement public use files contain the design variables necessary for variance estimation; Table IV 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.

**Table IV. Variables used for variance estimation, 2015 National Health Interview Survey Person File**

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

It should be noted that the STRAT\_P and PSU\_P levels are pseudo-levels or simplified versions of the true NHIS sample design variables. NCHS must adhere to laws that forbid the disclosure of any information (such as county of residence) that may compromise the confidentiality promised to its survey respondents. Consequently, much of the NHIS design information cannot be publicly released, and other data are either suppressed or recoded to ensure confidentiality. In order to satisfy this disclosure constraint, many of the original design strata, substrata, PSUs, and SSUs 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 variance estimation variables (i.e., STRAT\_P and PSU\_P).

The use of these publicly available variance estimation variables tends to provide slightly more conservative (larger) standard errors than the use of the confidential variance estimation variables used by analysts at NCHS. Moreover, analysts are cautioned that these publicly available variance estimation variables are not designed to support geographic analyses below the Census region level. Data users who want access to the confidential variance estimation variables used by analysts at NCHS may apply to the NCHS RDC: <http://www.cdc.gov/rdc/>.

The STRAT\_P and PSU\_P values for 2006–2015 have no connection with the STRATUM and PSU values for 2005 and earlier years. Analysts should refer to Appendix IV for variance estimation guidance for pooled analyses of adjacent years of the NHIS, including pooling 2006–2015 data with data for 2005 and earlier years.

## Variance Estimation for Analyses of Single Years of the NHIS

The limited public release design information requires a mathematical simplification that the PSUs be treated as if they were sampled with replacement (WR). 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.:

```
CSDSCRIPTIVES
/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 that 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 that 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>.

## Degrees of Freedom

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. This rule of thumb is not directly applicable to the NHIS design because many NHIS PSUs are self-representing; they are selected with certainty. The applicability of this rule of thumb 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 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, analyses 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 available to the software is incomplete; subsetting data deletes important design information needed for variance estimation. Note that SUDAAN has a SUBPOPN statement 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 statement 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 SUDAAN 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.

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 cdescriptives 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 cdescriptives or cstabulate, e.g.:

```
CSDSCRIPTIVES (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=subgrp)
```

### *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”).

### **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.

---

## Appendix IV. Merging Data Files and Combining Years of Data in the NHIS

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 persons 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 increase the number of persons and the precision of estimates.

### Merging Data Files

Each data file contains one or more of the household, family, and person record identifiers that make merging the data files possible, if needed. The Household File contains the household record identifier. The Family File contains the household and family record identifiers. Person-level files contain all three identifiers. 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 2015 Household, Family, Person, and Sample Child Files.

```
/* Merge the 2015 Household File and the 2015 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 NHIS2015.HOUSEHLD;
/*The SET statement reads data from the 2015 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 NHIS2015.FAMILYXX;
/*The SET statement reads data from the 2015 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).

```
/* Merge the 2015 Person File and the combined 2015 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 NHIS2015.PERSONSX;  
/*The SET statement reads data from the 2015 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 2015 Sample Child File and the combined 2015 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 NHIS2015.SAMCHILD;  
/*The SET statement reads data from the 2015 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

The purpose of combining, pooling, or concatenating years of data (in SAS terminology) is to increase the number of observations or persons for the same number of variables and thus to 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. 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.

**It is important to note that the name of a variable may change from one year to another. This usually occurs when some kind of modification, even a small one, has been made to the variable. 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.**

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).

The sampling weights for pooled data should be adjusted; otherwise, weighted 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. To achieve annualized results, a simple and 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 reasonably similar in the data years being combined, the simple procedure and the sophisticated alternative would yield a similar adjustment.

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 2014 NHIS and the 2015 NHIS. This same program can be used to combine the 2004 and 2005 NHIS Person Files after minor modifications (e.g., change “2014” and

“14” to “2004” and “04”, change “2015” and “15” to “2005” and “05”, and change STRAT\_P PSU\_P to STRATUM PSU).

**Please note that the person identifier was called PX in the 2003 (and earlier) NHIS and FPX in the 2004 (and later) NHIS; analysts 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_14; /* Create SAS dataset PER_14.*/
```

```
SET NHIS2014.PERSONSX (KEEP=HHX FMX FPX AGE_P SEX WTFA STRAT_P PSU_P);
```

```
/* The SET statement reads data from an existing SAS dataset, e.g., the 2014 Person File. The KEEP statement retains only the listed variables for processing. */
```

```
RUN;
```

```
PROC SORT DATA=PER_14; /* Sort SAS dataset PER_14. */
```

```
BY HHX FMX FPX;
```

```
RUN;
```

```
DATA PER_15; /* Create SAS dataset PER_15.*/
```

```
SET NHIS2015.PERSONSX (KEEP=HHX FMX FPX AGE_P SEX WTFA STRAT_P PSU_P);
```

```
/* The SET statement reads data from an existing SAS dataset, e.g., the 2015 Person File. The KEEP statement retains only the listed variables for processing. */
```

```
RUN;
```

```
PROC SORT DATA=PER_15; /* Sort SAS dataset PER_15. */
```

```
BY HHX FMX FPX;
```

```
RUN;
```

```
DATA COMB1415; /* New, combined SAS dataset */
```

```
SET PER_14 PER_15;
```

```
/* Concatenate selected variables from 2014 and 2015 datasets. */
```

```
WTFA_2YR=WTFA/2;
```

```
/*Create a new weight by dividing the existing Person File weight (WTFA) by 2, the number of Person 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 “COMB1415” with the combined 2014 and 2015 Person Files. As part of creating a dataset named “COMB0414” containing the combined 2004, 2005, 2014 and 2015 Person Files, two issues need to be addressed:

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

The weights in COMB0405 and COMB1415 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 COMB1415 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 2014–2015. 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 2014–2015 data file */
```

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

```
DATA COMB1415;  
SET COMB1415;  
DROP STRAT_P PSU_P;  
NSTRATUM=STRAT_P+2000;  
NPSU=PSU_P;  
RUN;
```

```
DATA COMB0414;  
SET COMB0405 COMB1415;  
DROP WTFA_2YR;  
WTFA_4YR=WTFA_2YR/2;  
RUN;
```

## Variance Estimation for Pooled Analyses of Adjacent Years

This section of this appendix provides information about variance estimation methods when combining datasets from more than one year. In short, if data from 1997–2005 are combined or data from 2006–2015 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 if data from 2005 or before are combined with data from 2006 and beyond, variance estimation is more complicated.

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, 2006–2015).
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., usually the same PSUs and nearby segments). 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.

## Appendix V. Analysis of the Family Food Security Variables

The ten questions on the Family Food Security Supplement (FFS) section can be used to determine the food security status of families as recommended by the US Department of Agriculture (USDA) Economic Research Service. See <http://www.fns.usda.gov/sites/default/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.

### Creating Food Security Scores

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”, “sometimes true”, and “yes” are considered affirmative. Responses to questions that ask about the frequency of occurrence in the past 30 days 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. Less than 1% of families did not receive additional questions after answering “don't know” or “refused” to the first three questions. 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. For additional information, see: <http://www.fns.usda.gov/sites/default/files/FSGuide.pdf>. Analyses with previous 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. Scale scores are not included in the 2015 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*

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

#### *Option 2*

- High food security (raw score 0)
- Marginal food security (raw score 1–2)
- Low food security (raw score 3–5)
- 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/measurement.aspx>.

## Appendix VI. Mental Health Indicator (MHI) for Children Aged 2–3 Years

This preliminary evaluation and these recommendations for use are based on a report by Thomas M. Achenbach, Ph.D., which was submitted to NCHS's 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 VII. 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 (Goodman, 1997, 1999). 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 2004 NHIS *Survey Description* and on the SDQ website at: <http://www.sdqinfo.org>.

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

- CMHMF\_1 (Generally obedient) correlates 0.69 with the long SDQ conduct score.
- CMHMF\_2 (Many worries) correlates 0.71 with the long SDQ emotion score.
- CMHMF\_3 (Often unhappy) correlates 0.64 with the long SDQ emotion score.
- CMHMF\_4 (Gets along better) correlates 0.69 with the long SDQ peer problems score.
- CMHMF\_5 (Sees tasks through) correlates 0.72 with the long SDQ hyperactivity-inattention score.

### Creating SDQ Scores

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, and then all values are summed to produce a total score. A total score correlates 0.84 with the long SDQ total difficulties score.

**Table V. Score assigned for each response, by item: Short Strengths and Difficulties Questionnaire**

Variable Name	Not true	Somewhat true	Definitely true
CMHMF_2, CMHMF_3, CMHMF_4	0	1	2
CMHMF_1, CMHMF_5	2	1	0

CMHDIFF 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. CMHDIFF 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 and 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, a subsample of 217 NHIS respondents completed a follow-up CAPA or PAPA interview. The CAPA or PAPA were conducted over the telephone between February and August 2012 with a parent of children aged 4–11 years and with adolescents aged 12–17 years. Logistic regression models were developed to model presence of any impairing child mental disorder in terms of responses only to the SDQ items. Findings suggest that SDQ is a useful instrument for detecting mental disorders among children (Ringeisen et al., 2015).

## Appendix VIII. Core Changes/Additions/Deletions in 2015

**Table VI. Core changes, additions, and deletions, National Health Interview Survey, 2015**

Questionnaire, section, and variable name	Brief variable description	Brief description of change
Family core, family health insurance section (FHI)		
MAPCMD	Primary care physician for routine care (Medicaid)	Deleted variable
MAREF	Need a referral for special care (Medicaid)	Deleted variable
PLNWRKR1	How plan was originally obtained (Plan 1)	Deleted variable; replaced by PLNWRKS1
PLNWRKS1	How plan was originally obtained (Plan 1)	New variable; replaced PLNWRKR1
PLNWRKR2	How plan was originally obtained (Plan 2)	Deleted variable; replaced by PLNWRKS2
PLNWRKS2	How plan was originally obtained (Plan 2)	New variable; replaced PLNWRKR2
MGPREF1	Private referral (Plan 1)	Deleted variable
MGPREF2	Private referral (Plan 2)	Deleted variable
STPCMD1	Primary care physician for routine care (SCHIP)	Deleted variable
STREF1	Need a referral for special care (SCHIP)	Deleted variable
STPCMD2	Primary care physician for routine care (OTHPUB)	Deleted variable
STREF2	Need a referral for special care (OTHPUB)	Deleted variable
STPCMD3	Primary care physician for routine care (OTHGOV)	Deleted variable
STREF3	Need a referral for special care (OTHGOV)	Deleted variable
MILMAN	Type of TRICARE coverage	Deleted variable; replaced by MILMANR
MILMANR	Type of TRICARE coverage	New variable; replaced MILMAN
Sample child core, child conditions section (CHS)		
CHEARST1	Hearing is excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, or deaf	New variable; replaced CHEARST2
CHEARST2	Hearing is excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, or deaf	Deleted variable; replaced by CHEARST1
Sample adult core, adult conditions section (ACN)		
AHEARST1	Hearing is excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, or deaf	New variable; replaced AHEARST2
AHEARST2	Hearing is excellent, good, a little trouble hearing, moderate trouble, a lot of trouble, or deaf	Deleted variable; replaced by AHEARST1
Sample adult core, Adult health behaviors (AHB)		
BINGE1	Number of times men had 5+ / women had 4+ drinks on an occasion	New variable
Sample adult core, adult selected items section (ASI)		
ASISMELS	What is meant by something else (sexual orientation; male)	Deleted variable
ASISIMDK	What is meant by don't know (sexual orientation; male)	Deleted variable

**Table VI. Core changes, additions, and deletions, National Health Interview Survey, 2015**

Questionnaire, section, and variable name	Brief variable description	Brief description of change
ASIMSESP	What is meant by something else, verbatim (sexual orientation; male)	Deleted variable
ASISFELS	What is meant by something else (sexual orientation; female)	Deleted variable
ASISIFDK	What is meant by don't know (sexual orientation; female)	Deleted variable
ASIFSESP	What is meant by something else, verbatim (sexual orientation; female)	Deleted variable