

2022  
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 2023

## Table of Contents

NCHS Website.....	6
NHIS Electronic Mail List.....	6
Questions about NHIS?.....	6
Guidelines for Citation of Data Source .....	7
Data User Agreement .....	8
What’s New in 2022?.....	9
About NHIS .....	11
Overview of 2022 Survey Methods .....	12
NHIS Methods.....	13
I. Sample Design .....	13
II. Interviewing Procedures.....	14
Data Collection .....	14
Field Results .....	15
Rostering and Respondents .....	16
Confidentiality .....	16
III. NHIS Design .....	17
Structure of NHIS .....	17
Content of the Household Roster .....	17
Content of the Sample Child Interview.....	18
Content of the Sample Adult Interview.....	20
Types of Questions .....	20
Questionnaire Sections.....	20
Description of the 2022 Questionnaire.....	20
IV. Sponsors .....	23
NHIS Sustaining Sponsors .....	23
Other NHIS Sponsors .....	24
V. Sample Sizes and Response Rates.....	25
Response Rate Method .....	25
2022 Sample.....	26
Reporting Household, Sample Adult, and Sample Child Response Rates.....	27
VI. Weighting .....	28
Weights.....	29
Variance Estimation.....	29

VII.	Editing the Data During and After the Interview .....	31
	Edits to Protect Confidentiality .....	31
	Family-Level Replicate .....	31
	Hard and Soft Edits .....	32
	Question-Specific Replication .....	32
	Recode into Different Variable .....	33
	Orientation to How to Use NHIS data.....	34
	Documentation.....	34
	Data Files .....	37
	Variable Conventions.....	39
	Definitions .....	39
	Analyzing 2022 NHIS and Software Examples.....	41
	Applying Sample Weight and Variance Estimation Variables in Analysis .....	41
	Examples of Analysis and Weighting Procedures for Sample Adults and Sample Children.....	43
	Scenario 1: Concatenating 2021 and 2022 (pooling) Sample Child data to increase sample size .....	44
	Scenario 2: Comparing semi-annual intervals (trends):.....	46
	Scenario 3: Concatenating 4 years of data: 2019–2022 (pooling) to increase sample size .....	49
	Merging Survey Data and Paradata Files.....	53
	Merging Survey Data and Imputed Income Files .....	55
	Appending Sample Adult and Sample Child Files.....	60
	Variance Estimation for Subsetted Data Analysis .....	63
	Questionnaire and Codebook Section Acronyms .....	66
	Sample Adult’s Health .....	69
I.	Health Status and Conditions .....	69
	Annual Core .....	69
	Rotating Content.....	70
	Emerging Content .....	70
	Sponsored Content.....	70
II.	Functioning and Disability.....	73
	Annual Core .....	73
	Sponsored Content.....	75
III.	Health Care Access and Health Service Utilization .....	76
	Annual Core .....	76
	Rotating Content .....	77
	Emerging Content.....	77
	Sponsored Content.....	77
IV.	Health-Related Behaviors .....	80
	Annual Core .....	80
	Rotating Core.....	80
	Sponsored Content.....	82
V.	Health Promotion.....	85
	Rotating Core.....	85

Major Recodes.....	85
VI. Mental Health.....	86
Rotating Core .....	86
Sponsored Content.....	87
VII. Barriers to care.....	88
Emerging Content.....	88
VIII. Community engagement .....	88
Emerging Content.....	88
Sample Child’s Health.....	89
I.    Health status and conditions .....	89
Annual Core .....	89
Rotating Core .....	89
Emerging Content.....	90
II.   Functioning and Disability.....	92
Annual Core .....	92
III.  Health Care Access and Health Service Utilization .....	94
Annual Core .....	94
Rotating Core .....	95
Emerging Content .....	95
Sponsored Content.....	96
IV.   Behavioral and Mental Health .....	97
Annual Core .....	97
Rotating Core.....	97
Emerging Content .....	98
V.    Stressful Life Events .....	100
Rotating Core .....	100
Emerging Content .....	100
VI.  Health-Related Behaviors .....	101
Rotating Core.....	101
VII.  Injuries .....	101
Emerging Content .....	101
VIII.  Health Promotion.....	101
Rotating Core.....	101
Health Insurance.....	102
Annual Core .....	102
Characteristics about the Sample Adult and Sample Child.....	106
Annual Core .....	106
Emerging Content.....	111

---

Characteristics about the Parents of the Sample Child .....	112
Annual Core .....	112
Characteristics about the Spouse or Partner of the Sample Adult .....	114
Annual Core .....	114
Characteristics about the Family and Household of the Sample Adult and Sample Child.....	115
Annual Core .....	115
Sponsored Content.....	120
NCHS Data Linkage Program.....	123
Medical Expenditure Panel Survey (MEPS) .....	125
References .....	126
Appendix: Availability of Health, Health Insurance and Selected Demographic Questions Not Included in the Sample Adult and Sample Child Public-use Files.....	129

---

## NCHS Website

Data users can obtain the latest information about the National Health Interview Survey (NHIS) by periodically checking our website: <https://www.cdc.gov/nchs/nhis.htm>

The website features downloadable data and documentation for the 2022 NHIS and previous years, as well as important information about any modifications or updates to the data or documentation. Published reports from previous years' surveys are also available, as are updates about future surveys and datasets. Data files and documentation can be found at: <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>

## NHIS Electronic Mail List

Data users are encouraged to join the NHIS Listserv, an electronic mailing list. The Listserv is made up of over 3,150 NHIS data users located around the world who receive news about NHIS surveys (e.g., new releases of data or modifications to existing data), publications, workshops, and conferences. To join, go to [https://www.cdc.gov/nchs/products/nchs\\_listservs.htm](https://www.cdc.gov/nchs/products/nchs_listservs.htm), and select "National Health Interview Survey (NHIS) Researchers" as one of your options, or click on "Contact Us" on the NHIS website and scroll to "Listserv" or "How to Subscribe to the HISUSERS List."

## Questions about NHIS?

The staff of the Division of Health Interview Statistics at the National Center for Health Statistics respond to data users' questions about NHIS. Users may call us at 301-458-4901 and leave a voice message or e-mail us at [nhislist@cdc.gov](mailto:nhislist@cdc.gov). A response may take 1-2 business days.

## 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 2022 NHIS data should acknowledge “National Center for Health Statistics, National Health Interview Survey” as the original source. The full spelling of the source without the use of acronyms is preferred.

The suggested citation to appear at the bottom of all tables and graphs is as follows:

Data Source: National Center for Health Statistics, National Health Interview Survey, 2022.

In a bibliography, the suggested citation for this document is:

National Center for Health Statistics. National Health Interview Survey, 2022 survey description. 2023. Available from:  
[https://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2022/srvydesc-508.pdf](https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2022/srvydesc-508.pdf)

The suggested citation for 2022 NHIS survey data and other documentation is:

National Center for Health Statistics. National Health Interview Survey, 2022. Public-use data file and documentation. <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>. 2023.

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 NCHS publications and electronic media is available at:  
<https://www.cdc.gov/nchs/products/citations.htm>.

---

## Data User Agreement

### Please Read Carefully Before Using the National Health Interview Survey

The National Health Interview Survey (NHIS) is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC).

NCHS, CDC conducts statistical and epidemiological activities under the authority granted by the Public Health Service Act (42 U.S.C. § 242k). NCHS survey data such as NHIS are protected by Federal confidentiality laws including Section 308(d) Public Health Service Act [42 U.S.C. 242m(d)] and the Confidential Information Protection and Statistical Efficiency Act or CIPSEA [44 U.S.C. 3561-3583]. These confidentiality laws state the data collected by NCHS may be used only for statistical reporting and analysis. Any effort to determine the identity of individuals and establishments violates the assurances of confidentiality provided by federal law.

### Terms and Conditions

NCHS does all it can to assure that the identity of individuals and establishments cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the dataset. Any intentional identification or disclosure of an individual or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will:

1. Use the data in this dataset for statistical reporting and analysis only.
2. Make no attempt to learn the identity of any person or establishment included in these data.
3. Not link this dataset with individually identifiable data from other NCHS or non-NCHS datasets.
4. Not engage in any efforts to assess disclosure methodologies applied to protect individuals and establishments or any research on methods of re-identification of individuals and establishments.

By using these data, you signify your agreement to comply with the above-stated statutorily based requirements.

### Sanctions for Violating NCHS Data Use Agreement

Willfully disclosing any information that could identify a person or establishment in any manner to a person or agency not entitled to receive it, shall be guilty of a class E felony and imprisoned for not more than 5 years, or fined not more than \$250,000, or both.



## What's New in 2022?

- Continuing since 2019, the NHIS questionnaire includes four types of content: 1) Annual core is content that is included in the NHIS every year; 2) Rotating core is scheduled to appear on a periodic basis of every other year, one out of every three years, or two out of every three years; 3) Sponsored content includes questions that are funded by other agencies; and 4) Emerging content is content that is sponsored by NCHS on new health topics. (See a visualization of the [Sample Adult](#) and [Sample Child](#) content by periodicity for 2019-2027).
- Beginning in 2022, two new items regarding Medicare Advantage covering vision and dental plans were added as annual core content to the health insurance section. These items are available in the public-use file for Sample Adults and via the Research Data Center for Sample Children.
- In addition to annual content, 2022 rotating, sponsored and emerging content is listed below:

### Sample Adult Rotating Core

Mental health assessment for depression and anxiety — previously fielded in 2019.

Dental care, physical and other therapeutic care— previously fielded in 2019–2020.

Health-related behaviors: sleep, physical activity, walking for transportation and leisure, alcohol use, cigarette history and advice to quit smoking, fatigue, health-related advice from a health care provider— previously fielded in 2020.

### Sample Child Rotating Core

Strength and difficulties questionnaire— previously fielded in 2019.

Health-related behaviors: physical activity, sleep, and screen time; Height and weight; Neighborhood characteristic— previously fielded in 2020.

Dental care, physical and other therapeutic care— previously fielded in 2019–2020.

Mental health care— previously fielded in 2019–2020, with partial content also in 2021.

Stressful life events— previously fielded in 2021.

### Sample Adult Sponsored Content

Age of disability onset; chronic fatigue syndrome; complementary and integrative health; communication disorders related to voice, swallowing, speech, and language; diet and nutrition; epilepsy; food security; food related programs; immunization (COVID-19\*, flu, HPV, shingles, tetanus); immunosuppression; insulin use; life satisfaction; menthol cigarette use, noncigarette tobacco product use, smoking cessation methods and advice to quit smoking from health care professional; and working or volunteering in a health care setting.

\*In October (Quarter 4), there was a change to the question on COVID-19 brand.

### Sample Child Sponsored Content

Complementary and integrative health; communication disorders related to voice, swallowing, speech, and language; food security; food related programs; immunization (COVID-19\*\*, HPV); life satisfaction; and stressful life events (additional measures).

\*\* Asked of children aged 5–17 in January–July and asked of all children in August–December.

### Sample Adult Emerging Content

Access and transportation barriers to health care; gender identity\*\*\*; history of COVID-19; Long COVID; and telehealth. July–December: civic engagement; internet access and health information technology; and volunteer activities.

\*\*\*Experimental questions available via the Research Data Center.

### Sample Child Emerging Content

Bullying; concussions; history of COVID-19; Long COVID; interaction with health care providers stressful life events (additional measures) telehealth and social and emotional support.

---

## About NHIS

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 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.

Since 1957, the content of the survey has been updated about every 10–15 years to incorporate advances in survey methodology and coverage of health topics. In January 2019, NHIS launched a redesigned content and structure that differs from the 1997–2018 NHIS.

---

## Overview of 2022 Survey Methods

NHIS is a nationally representative household survey of the U.S. civilian noninstitutionalized population. It is conducted continuously throughout the year by the National Center for Health Statistics (NCHS). Interviews are typically conducted in respondents' homes, but follow-ups to complete interviews may be conducted over the telephone. In 2022, 55.7% of the Sample Adult interviews and 56.1% of the Sample Child interviews were conducted at least partially by telephone.

Information about the Sample Adult is self-reported, unless physically or mentally unable to do so and a knowledgeable proxy can answer for the Sample Adult. Information about the Sample Child is collected from a parent or adult who is knowledgeable about and responsible for the health care of the Sample Child. In 2022, there were 27,651 Sample Adult interviews and 7,464 Sample Child interviews. The Sample Adult response rate was 47.7% and the Sample Child response rate was 45.8%. The NHIS includes annual content that appears on the survey every year. The survey also includes rotating core content, sponsored content, and emerging content that appears periodically. Visit [https://www.cdc.gov/nchs/nhis/2019\\_quest\\_redesign.htm](https://www.cdc.gov/nchs/nhis/2019_quest_redesign.htm) for a description of content in any given year. For more information about NHIS, visit: <https://www.cdc.gov/nchs/nhis.htm>.

## NHIS Methods

### I. Sample Design

NHIS is a cross-sectional household interview survey. The target population for the NHIS is the civilian noninstitutionalized population residing within the 50 states and the District of Columbia at the time of the interview. The NHIS universe includes residents of households and noninstitutional group quarters (e.g., homeless shelters, rooming houses, and group homes). Persons residing temporarily in student dormitories or temporary housing are sampled within the households that they reside in permanently. Persons excluded from the universe are those with no fixed household address (e.g., homeless and/or transient persons not residing in shelters), active duty military personnel and civilians living on military bases, persons in long-term care institutions (e.g., nursing homes for the elderly, hospitals for the chronically ill or physically or intellectually disabled, and wards for abused or neglected children), persons in correctional facilities (e.g., prisons or jails, juvenile detention centers, and halfway houses), and U.S. nationals living in foreign countries. While active-duty Armed Forces personnel are not sampled for inclusion in the survey, any civilians residing with Armed Forces personnel in non-military housing are eligible to be sampled.

Because the NHIS is typically 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 for cost-effective interviewing. To keep survey operations manageable, cost-effective, and timely, the NHIS uses geographically clustered sampling techniques to select the sample of dwelling units for the NHIS. The sample is designed in such a way that each month's sample is nationally representative. Data collection on the NHIS is continuous, i.e., from January to December each year.

The sampling plan is redesigned after every decennial census. A new sampling plan for the 2016–2025 NHIS was designed with results of the 2010 decennial census. The sampling process starts with partitioning the United States into 1,689 geographic areas. These geographic areas are defined as counties, county equivalents, or groups of counties, are almost always contiguous, and do not cross state boundaries. Next, within some states, the geographic areas are divided into two strata defined by population density (generally, urban counties and rural counties). For the remaining states, all the geographic areas form one stratum. Clusters of addresses were then defined within each stratum. The sizes of the clusters correspond generally to the size of an interviewer's workload over the course of the 10-year sample design period; the approximate size is 2,500 addresses per cluster. Each cluster is located entirely within one of the 1,689 originally defined geographic areas. Within each stratum, a specific number of clusters is systematically selected for the NHIS sample. The number selected is generally proportional to the number of clusters in the strata, e.g., larger strata have more clusters selected within the strata. The exception is in the 10 least populous states and the District of Columbia, where a slightly higher number of clusters are selected in order to ensure that all states have a minimum number of addresses in the sample.

Commercial address lists were used as the main source of addresses, supplemented by field listing. As of the beginning of 2016, the NHIS sampling frame consists of two non-overlapping parts: the unit frame (a list of addresses purchased from a vendor), and the area frame (generated by traditional field enumeration). At that time approximately 11% of the counties in the sample were part of the area frame. These area frame counties are typically counties with relatively few city-style addresses, and counties where the unit frame did not have acceptable coverage, i.e., where the vendor-supplied list did not adequately include all eligible households. For more information about the 2016–2025 NHIS sample design and estimation structures, see [Vital and Health Statistics, Series 2, Number 191 \(cdc.gov\)](#).

## II. Interviewing Procedures

### *Data Collection*

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. In 2022, about 827 interviewers, also called “Field Representatives” or “FRs,” conducted NHIS interviews nationwide. FRs 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. The “Advance letter” is mailed one week prior to the start of the interview period (one week before the 1st of the month) with the goal that it might be fresh on people’s mind when the FR makes contact the first few days of the month. 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 the current “Advance letter” and other NHIS materials available for distribution by FRs in the field are available at the NHIS participants’ page: <https://www.cdc.gov/nchs/nhis/participant.htm>.

The NHIS is conducted using computer-assisted personal interviewing (CAPI). The CAPI data collection method employs Blaise computer software that presents questions on computer screens to each interviewer. The instrument 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.

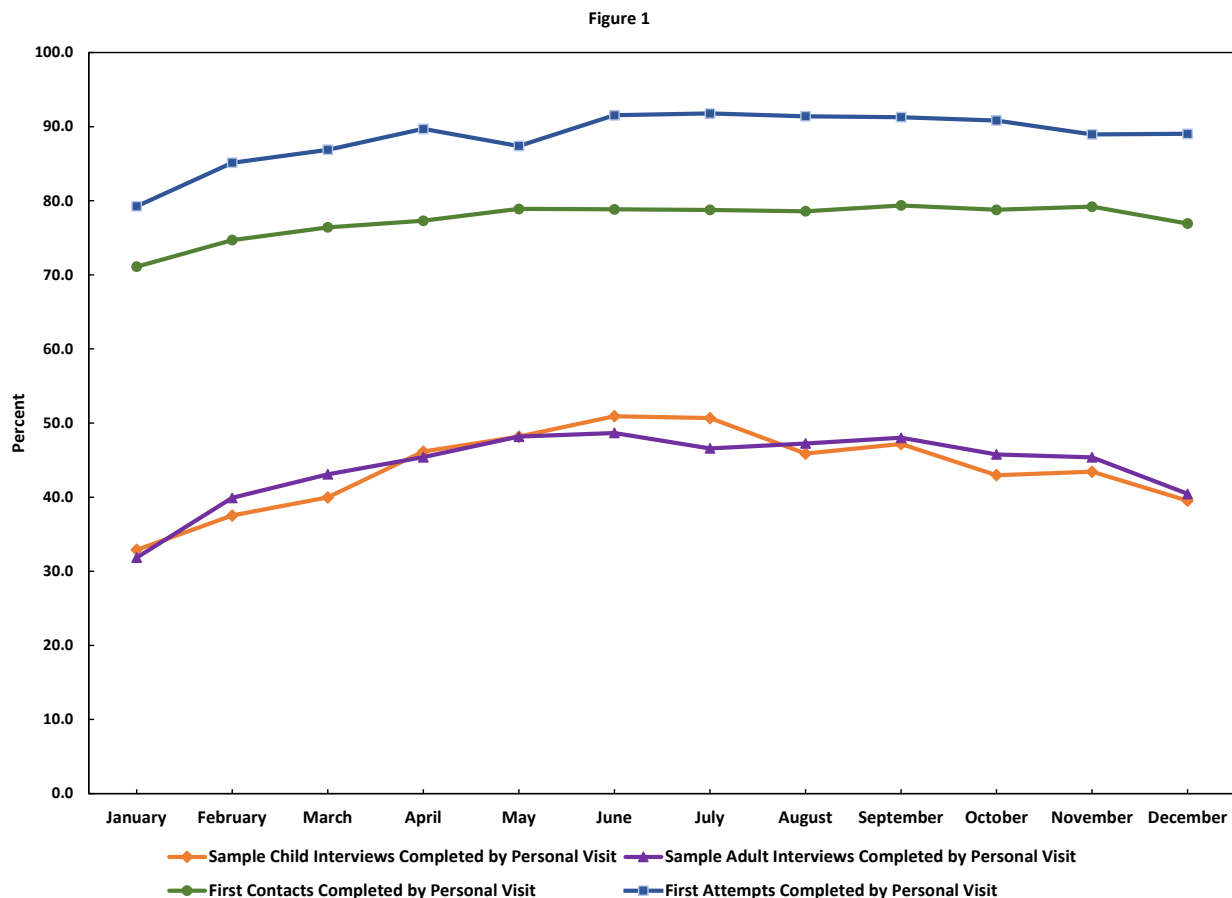
Typically, face-to-face interviews are conducted in respondents’ homes, but follow-ups to complete interviews may be conducted over the telephone. A telephone interview may also be conducted when the respondent requests a telephone interview or when road conditions or travel distances would make it difficult to schedule a personal visit before the required completion date.

## Field Results

In 2022, 55.7% of the Sample Adult interviews and 56.1% of the Sample Child interviews were conducted at least partially by telephone. This is lower than the percentage of telephone interviews that were conducted at least partially by telephone in 2021 (62.8% for Sample Adult interviews and 61.4% for Sample Child interviews), but still considerably higher than interviews in 2019 (pre-pandemic) for both the Sample Adult (34.3%) and Sample Child (31.7%).

Figure 1 shows metrics summarizing the extent to which the NHIS FRs conducted personal visits throughout the calendar year. It includes the percent of Sample Adult and Sample Child interviews completed by personal visit, the percent of first attempts made by personal visit, and the percent of first contacts completed by personal visit. First attempt refers to the first attempt that an FR makes, in-person or by phone, to determine eligibility of a household at a sample address and recruit a household participant in the survey, regardless of whether contact was made. First contact refers to the first interaction between the FR and anyone from the household at a sample address, in-person or by phone.

Figure 1 shows that the percentage of both Sample Adult and Sample Child interviews completed by personal visit was lowest at the beginning of year at around 30% for each interview. The percentage of Sample Adult and Sample Child interviews completed by personal visit peaked at around 50% in June and July before declining to approximately 40% by the end of the year. First attempts by personal visit increased from about 80% in January to about 90% in June before leveling off for the rest of the year. First contact by personal visit increased from 70% in January to approximately 80% in May before leveling off for the rest of the year.



### *Rostering and Respondents*

For the Household Roster section of the questionnaire, any responsible household member aged 18 years or over is identified to act as the “household respondent.” The household respondent provides names, age, sex, race, and ethnicity for all household members. The highest level of education completed, and active military status is asked for all adult household members aged 18 years or over. In addition to collecting this basic demographic information, the household roster interview also identifies whether all persons in the household are members of the same or different family. Note that in a multi-family household, a single “household respondent” provides household information for all families.

NHIS has consistently defined a family as an individual or a group of two or more people residing together who are related by birth, marriage, or adoption. A family additionally includes any unrelated children who are cared for by the family (such as foster children) and any unmarried cohabiting partners and their children. After the household roster is completed, data are collected on one adult and child per household.

A “Sample Adult” is randomly selected by the computer from each household with at least one household member aged 18 years or over and is asked more detailed health related questions. 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 may answer for the Sample Adult. Students aged 18 and over living away at college, trade, or commercial schools in on-campus housing are eligible to be interviewed in the location they consider to be their usual residence, such as their parent’s or other family member’s household. Students living away at school or college in off-campus housing will not be included as members of the household, since they could be sampled at their off-campus location.

A “Sample Child” is randomly selected by the computer from each household with at least one child 17 years of age or younger. An adult respondent who was previously indicated to be knowledgeable and responsible for the Sample Child’s health will be asked questions about that child. In 2022, 93.8% of the Sample Child respondents were the child’s parent, either a biological, adoptive or stepparent, while 4.1% were a grandparent, 0.4% were an aunt or uncle, 0.7% were an adult sibling, 0.9% were another relative or other non-relative, and 0.04% were not ascertained. For each sampled household address, interviewers also maintain electronic documentation about the NHIS interview process, including contact attempts, observed characteristics about the exterior of the sample unit or vicinity, and descriptive information about the interview outcome.

### *Confidentiality*

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(d)), 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 of 2018 (44 U.S.C. 3561-3583).

Further information about data collection procedures is available in the Field Representative Manual available on the NHIS website, <https://www.cdc.gov/nchs/nhis.htm>.



### III. NHIS Design

Since 2019, the structure and content of the NHIS questionnaire was redesigned to better meet the needs of data users, the Centers for Disease Control and Prevention, and the Department of Health and Human Services (DHHS). The goals of the redesign were to reduce respondent burden by shortening the length of the questionnaire, harmonize overlapping content with other federal health surveys, establish a long-term structure of ongoing and periodic topics, and incorporate advances in survey methodology and measurement. Visit the [2019 Questionnaire Redesign](#) webpage for more information. For information about the previous NHIS design conducted during 1997–2018, refer to year-specific NHIS documentation: <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>

#### *Structure of NHIS*

The redesigned NHIS questionnaire consists of three modules: (1) Household Roster; (2) Sample Adult Interview; and (3) Sample Child Interview. Following a brief household rostering interview that collects basic demographics of everyone who usually lives or stays in the household and identifies which members of the household are in the same or a separate family, one “Sample Adult” aged 18 years or older and one “Sample Child” aged 17 years or younger (if any children live in the household) are randomly selected from each household. Information about the household members is collected from a person aged 18 and over residing at the sample address. Information about the Sample Adult is collected from the Sample Adult themselves unless they are physically or mentally unable to do so, in which case a knowledgeable proxy can answer for the Sample Adult. Information about the Sample Child is collected from a parent or adult who is knowledgeable and responsible for the health care of the Sample Child. The respondent for the Household Roster, Sample Adult and Sample Child modules may or may not be the same person.

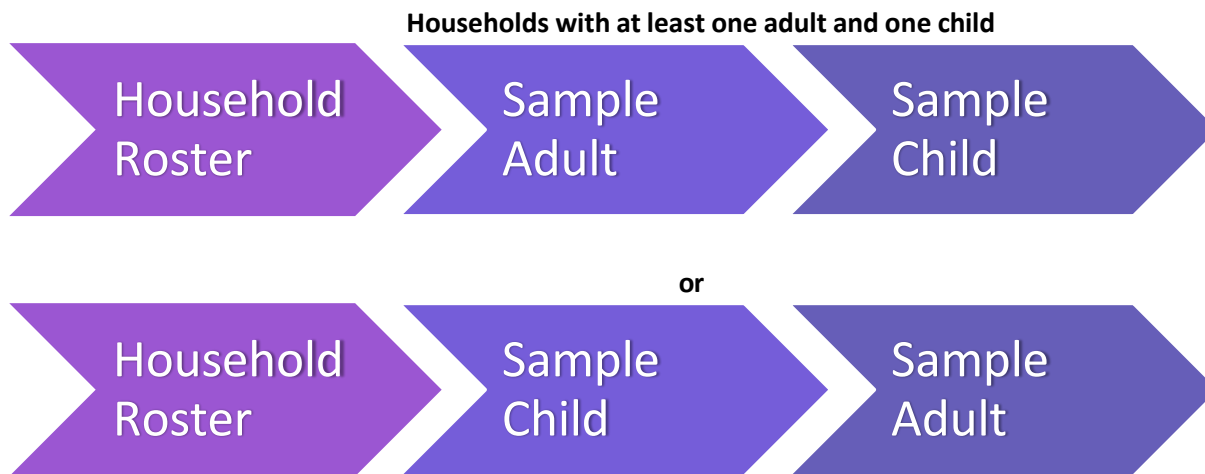
Figure 2 (on the next page) illustrates the interviewing flow of the Household Roster, Sample Adult and Sample Child interviews, or modules, in the NHIS. The Sample Adult and Sample Child may be part of the same family or be part of different families in the household. Family-level content (e.g., family income, food insecurity) is collected in the Sample Adult or Sample Child questionnaire module.

#### *Content of the Household Roster*

In the Household Roster, an adult (aged 18 years and over) living in the household provides basic information about themselves and other people living in the household. The names, age, sex, race, and ethnicity of everyone in the household are collected. Additionally, the parents of all children are identified. The instrument then randomly selects one adult (Sample Adult) and one child (Sample Child), if any children live in the household, to be given follow-up questions. The Sample Adult is selected randomly among persons aged 18 years and over living in the household, and the Sample Child is selected randomly among those aged 17 years or younger. Questions are asked to determine who is in the family of the Sample Adult and Sample Child. The Sample Adult and Sample Child do not need to be in the same family. No health information is collected in this section. When the Household Roster is complete, the interviewers can then proceed with the Sample Adult or Sample Child interview (if a child lives in the household).

Figure 3 (on the second next page) illustrates how topics or content in 2022 NHIS are organized by interviewing modules (i.e., Household Roster, Sample Adult and Sample Child).

Figure 2. NHIS Structure



Note. In households where there is an eligible Sample Adult and Sample Child, either the Sample Adult or Sample Child interview is administered first. Once both the Sample Adult and Sample Child interviews have been completed the interview is complete.

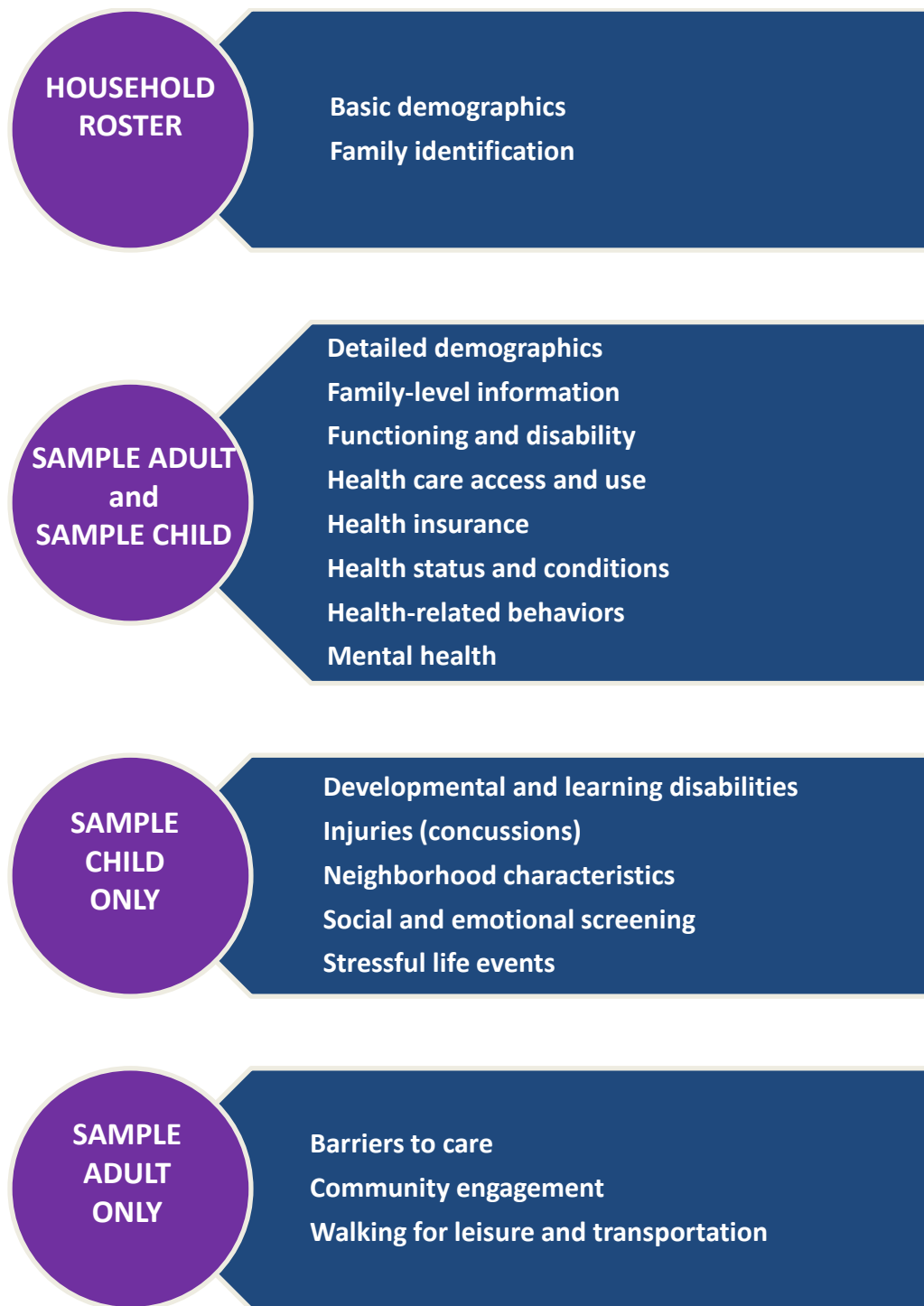


Note: The NHIS is a survey of the civilian noninstitutionalized population, and active-duty military personnel are not included. In the rare case where a child lives in a household consisting of only active-duty military personnel, the Household Roster would be completed followed by a Sample Child interview.

### *Content of the Sample Child Interview*

An adult knowledgeable and responsible for the health of the child is asked a set of questions about the Sample Child. Some questions asked of the Sample Child vary by age, as younger and older children have different health needs. Additional demographic information is also collected about the child and their family.

Figure 3. Topic organization in the Household Roster, Sample Adult and Sample Child modules: NHIS 2022.



## *Content of the Sample Adult Interview*

The Sample Adult is asked a series of health questions about themselves. Some questions vary by age or sex of the Sample Adult, but most are the same for all Sample Adults. Additional demographic information is also collected about the Sample Adult and their family.

### *Types of Questions*

The redesigned NHIS incorporates a long-term structure of fixed and periodic content. The long-term structure for the Sample Adult and Sample Child questionnaires organizes question topics by year and by type of content for the survey years 2019–2027. Additional information about periodicity of question topics for 2019–2027, see, [https://www.cdc.gov/nchs/nhis/2019\\_quest\\_redesign.htm](https://www.cdc.gov/nchs/nhis/2019_quest_redesign.htm). There are four types of content: (1) Annual core; (2) Rotating core; (3) Sponsored content; and (4) Emerging topics.

**Annual core** are consistent questions that are asked every year.

**Rotating core** are questions that are asked some, but not all years. These questions are scheduled to appear on a periodic basis of every other year, one out of every three years, or two out of every three years.

**Sponsored content** are questions funded by other federal agencies or other centers within CDC about topics of interest to the sponsor. Sustaining sponsors fund content every year, whereas other sponsors fund content periodically.

**Emerging topics** are questions about areas of interest to NCHS, CDC, or DHHS. These are newer subject areas that have generally not been researched in the general population.

### *Questionnaire Sections*

The NHIS is divided into many questionnaire sections within each module, each with a different focus. The sections may include any combination of annual core, rotating core, sponsored content, or emerging topics. When the same questions or same types of questions are asked in a Sample Adult and Sample Child interviews, the sections are given the same name for both interviews.

The names of the questions asked of the Sample Adult or pertaining to the Sample Adult's family all end in “\_A” whereas those asked of the Sample Child or about the Sample Child's family end with “\_C.” Section names have a 3-letter abbreviation (e.g., INS for Health Insurance), and questions are grouped by module and section.

### *Description of the 2022 Questionnaire*

A description of the 2022 topics and type of questions are described in this report under Sample Adult's Health, Sample Child's Health, and Health Insurance, while all demographic information has been portioned into four sets of characteristics: 1) those about the Sample Adult and Sample Child; 2) those about the parents or guardian residing in the household with the Sample Child; 3) those about the spouse or partner residing in the

household with the Sample Adult (if married or cohabiting); and 4) those about the family and household of the Sample Adult and Sample Child. In this document, multiple questionnaire sections with annual, rotating, emerging and sponsored content are described in each of the health topics included under Sample Adult's Health and Sample Child's Health.

**Sample Adult health topics for 2022 are:**

**I. Health Status and Conditions**

self-reported health status, height, weight, pregnancy status, and the following health conditions: angina pectoris, anxiety disorder, arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia, asthma, cancer and cancer type, chronic fatigue syndrome, chronic obstructive pulmonary disease, coronary heart disease, dementia including Alzheimer's disease, depression, diabetes, epilepsy or seizure disorder, gestational diabetes, hyperlipidemia, hypertension, immunosuppression, history of COVID and Long COVID, myocardial infarction, pre-diabetes, stroke, and voice, swallowing, speech, and language communication disorders. It also includes medication use for hyperlipidemia, hypertension, and diabetes, and health care services for asthma, epilepsy, and speech and language communication disorders.

**II. Functioning and Disability**

anxiety, cognition, communication, depression, hearing, mobility, self-care and upper body, social functioning (participation), vision, and age of disability onset.

**III. Health Care Access and Health Service Utilization**

difficulty paying for health care; health care use (medical care visit, dental, home, mental health and vision care, physical/speech/rehabilitative/occupational therapy, wellness visit, urgent care, emergency care, hospitalization); immunizations (COVID-19, flu, human papilloma virus (HPV) pertussis/tetanus/diphtheria, Pneumonia, shingles); prescription medication; use of complementary and integrative health, telehealth, usual place for medical care; unmet medical needs (dental, medical, mental health, and prescriptions) due to cost; and work or volunteer activities in health care settings.

**IV. Health-Related Behaviors**

alcohol use, cigarette smoking (initiation, cessation, quitting methods), electronic cigarette use, use of other tobacco products, diet and nutrition, physical activity, sleep, and advice from a health care professional to engage in physical activity and to quit smoking and tobacco.

**V. Health Promotion**

walking for leisure and transportation.

**VI. Mental Health**

General Anxiety Disorder-7 (GAD-7), Patient Health Questionnaire-8 (PHQ-8), and life satisfaction.

**VII. Barriers to care**

accessibility, availability, and transportation barriers, internet access and health information technology.

**VIII. Community engagement**

civic engagement, and volunteer activities.

**Sample Child health topics for 2022 are:****I. Health Status and Conditions**

reported health status of the Sample Child, height, weight, and the following health conditions: asthma, developmental conditions, diabetes, and learning disabilities, history of COVID and Long COVID, and voice, swallowing, speech, and language communication disorders. It also includes health care services for asthma, and speech and language communication disorders.

**II. Functioning and Disability**

anxiety, behavior, cognition, communication, depression, hearing, mobility, self-care and upper body, and vision.

**III. Health Care Access and Health Service Utilization**

difficulty paying for health care; health care use (medical care visit, dental, home, mental health and vision care, physical/speech/rehabilitative/occupational therapy, wellness visit, urgent care, emergency care, hospitalization); immunizations (COVID-19, flu, and human papilloma virus (HPV)); presence of a personal health care professional and receipt of care without parent/guardian; prescription medication; use of complementary and integrative health, telehealth, usual place for medical care; unmet medical needs (dental, medical, mental health, and prescriptions).

**IV. Behavioral and Mental health**

Baby Pediatric Symptom Checklist (BPSC), bullying, life satisfaction, Strengths and Difficulties Questionnaire (SDQ), and social support.

**V. Stressful Life Events**

Experienced verbal abuse, had a parent who was incarcerated after birth, had unmet basic needs, lived with someone with a mental illness, lived with someone with a drug or alcohol problem, witnessed or experienced neighborhood violence, was treated or judged unfairly because of their sexual orientation or gender identity, and was treated or judged unfairly because of their race or ethnic group.

**VI. Health-Related Behaviors**

physical activity, sleep, and screen time.

**VII. Injuries**

concussions.

**VIII. Health Promotion**

neighborhood characteristics.

## IV. Sponsors

Some 2022 NHIS content is sponsored by other federal agencies or other centers within the Centers for Disease Control and Prevention. Sponsored content may be used to collect data on new topics or to go into more depth about subjects already on the NHIS. Sustaining sponsors add content every year. Other sponsors add content for selected years.

### *NHIS Sustaining Sponsors*

#### *Cancer Control and Prevention*



The National Cancer Institute at the National Institutes for Health (NIH/NCI) and the National Center for Chronic Disease Prevention and Health Promotion at the Centers for Disease Control and Prevention (CDC/NCCDPHP) sponsored 20 questions asked of Sample Adults about smoking cessation methods and advice to quit smoking, and 14 questions on diet and nutrition.



#### *Immunization and Employment in Health Care Settings*



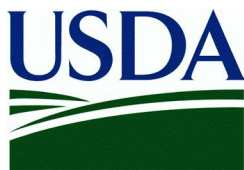
The National Center for Immunization and Respiratory Diseases at the Centers for Disease Control and Prevention (CDC/NCIRD) sponsored 24 Sample Adult questions about flu vaccination during pregnancy, vaccination for COVID-19, shingles, tetanus, and HPV, immunosuppression and working or volunteering in the health care industry. NCIRD also sponsored seven questions on COVID-19 and HPV vaccination for the Sample Child.

#### *Noncigarette Tobacco Product Use/Menthol Cigarette Use*



The Center for Tobacco Products at the Food and Drug Administration (FDA) sponsored seven Sample Adult questions about the use of cigars, pipes, and smokeless tobacco and two questions on the use of menthol cigarettes.

#### *Food Security and Food Stamp Benefits*



The United States Department of Agriculture (USDA) sponsored 10 questions that can be used to determine food security or insecurity and degree of insecurity in the Sample Adult or Sample Child's family. Additionally, the USDA sponsored a question about use of SNAP in the past 30 days. (Use of SNAP in the past year is part of the annual core content.)

*Insulin Use*

**National Institute of  
Diabetes and Digestive  
and Kidney Diseases**

The National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes for Health (NIH/NIDDK) and the National Center for Chronic Disease Prevention and Health Promotion at the Centers

for Disease Control and Prevention (CDC/NCCDPHP) sponsored three Sample Adult questions about insulin initiation among adults with diabetes who take insulin.

*Voice, Swallowing, Speech and Language*

**National Institute on  
Deafness and Other  
Communication Disorders**

National Institute on Deafness and Other Communication Disorders at the National Institutes of Health (NIH/NIDCD) sponsored 17 questions for the Sample Adult and 17 questions for the Sample Child questions on voice, swallowing, speech, and

language.

*Other NHIS Sponsors**Age of Disability Onset*

The Administration for Community Living at the Department of Health and Human Services (ACL/HHS) sponsored for 2022 one Sample Adult question about the age of disability onset.

*Chronic Fatigue Syndrome/ME*

The National Center for Emerging and Zoonotic Infectious Diseases (CDC/NCEZID) sponsored for 2022 two Sample Adult questions on chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME).

*Complementary and Integrative Health*

The National Center for Complementary and Integrative Health (NIH/NCCIH) at the National Institutes of Health sponsored for 2022 25 Sample Adult questions and 25 Sample Child questions on complementary and integrative health practices.

*Epilepsy*

The National Center for Chronic Disease Prevention and Health Promotion at the Centers for Disease Control and Prevention (CDC/NCCDPHP) sponsored for 2022 four Sample Adult questions on epilepsy.



### Satisfaction With Life

The National Center for Chronic Disease Prevention and Health Promotion at the Centers for Disease Control and Prevention (CDC/NCCDPHP) and the Office of Disease Prevention at the National Institutes of Health (NIH/ODP) sponsored one Sample Adult question on satisfaction with life.

### Stressful Life Events

The National Center for Injury Prevention and Control at the Centers for Disease Control and Prevention (CDC/NCIPC) sponsored for 2022 two Sample Child questions on stressful life events.

## V. Sample Sizes and Response Rates

When the NHIS sample was redesigned for 2016–2025, was expected to yield about 27,000 Sample Adult interviews and 9,000 Sample Child interviews in roughly 35,000 households each year. However, NHIS sample size may vary from year to year. Table 1 provides a breakdown of sample sizes for the various components of the 2022 NHIS public-use data release.

Table 1. Final sample sizes for the 2022 NHIS public-use data release	
Interview Unit	2022 Sample
Households	28,854
Sample Adults	27,651
Sample Children	7,464
<b>* For 488 of the 27,651 Sample Adults on the 2022 sample, a knowledgeable proxy answered for the Sample Adult because he/she was mentally or physically incapable of answering for himself/herself.</b>	

### Response Rate Method

Response rates presented below conform to the American Association of Public Opinion Research (AAPOR) Response Rate Definition # 2, or AAPOR RR2 (AAPOR, 2016). “Interviewed households,” “interviewed Sample Adults,” and “interviewed Sample Children” include those with completed interviews or acceptable “sufficient partial” interviews.

In the NHIS, a Sample Adult or Sample Child interview is considered fully complete when respondents complete all sections, and therefore questions, for which they are eligible. Conversely, an interview is considered a “partial” when all sections are not completed. The most common reason for a partial is a “break-off,” which occurs when a respondent stops the interview in-progress before completion and the interviewer fails to complete the interview during the allotted assignment period. The partial interview rate is the percent of all sufficiently complete interviews that are not entirely complete. These “sufficient partials” are counted as interviews in the computation of response rates. Partial interviews that are not far enough along in the interview, known as “insufficient partials,” are considered refusals (Stussman et al., 2003) and therefore included as eligible, non-respondents in the computation of response rates.

## 2022 Sample

### Household Response Rate

For the 2022 Sample, the household response rate was calculated by dividing the number of interviewed households (n=28,854) by the sum of the number of interviewed households (n=28,854) and the number of nonresponding households (n=29,361). Nonresponding households are eligible households that were not interviewed for a variety of reasons, including language barriers, no one home after repeated contact attempts, refusal, household records rejected for insufficient data, or other reasons for no interview.

**The total Household Response Rate for the 2022 Sample was 49.6%.**

It is important to note that the definition of an interviewed household differs from the past design (1997-2018). Previously, an interviewed household was defined as one where at least one family in the household completed a substantial portion of the family interview. With the family interview removed from the redesigned NHIS, an interviewed household is now defined as one where the household roster and a substantial portion of either the Sample Adult interview or the Sample Child interview (if one or more children reside in the household) is completed. A household response rate obtained during the 1997-2018 NHIS and 2019-2022 NHIS should be presented separately.

### Household Roster Completion

The completion of the Household Roster is defined as the enumeration of all persons in an eligible household, with basic demographic information collected about each household member. The Household Roster Completion Rate is calculated by dividing the number of eligible households with a completed household roster (n=31,654) by the number of eligible households (n=58,215). For the 2022 Sample, the Household Roster Completion Rate was 54.4%. Based on demographic information obtained from completed household rosters, there were 8,879 eligible Sample Children and 31,579 eligible Sample Adults in the 2022 Sample.

### Sample Child Response Rates

Sample Child response rates can be computed two ways, resulting in either a conditional or final response rate. The Conditional Sample Child Response Rate is calculated by dividing the number of interviewed Sample Children (n=7,464) by the number of eligible Sample Children from households with completed rosters (n=8,879). For the 2022 Sample, the Conditional Sample Child Response Rate was 84.1%.

The Final Sample Child Response Rate accounts for the Household Roster Completion Rate and is calculated by dividing the number of interviewed Sample Children (n=7,464) by the number of eligible Sample Children (n=8,879) from households with completed rosters, and then multiplying this quotient by the Household Roster Completion Rate (54.4%). In 2022, 2.8% of Sample Child interviews were sufficient partials.

**For the 2022 Sample, the Final Sample Child Response Rate was 45.8%.**

### Sample Adult Response Rates

As with Sample Children, both a conditional and final response rate can be computed for Sample Adults. The Conditional Sample Adult Response Rate is calculated by dividing the number of interviewed Sample Adults (n=27,651) by the number of eligible Sample Adults from households with completed rosters (n=31,579). For the 2022 Sample, the Conditional Sample Adult Response Rate was 87.6%.

The Final Sample Adult Response Rate is calculated by dividing the number of interviewed Sample Adults (n=27,651) by the number of eligible Sample Adults from households with completed rosters (n=31,579), and then multiplying this quotient by the Household Roster Completion Rate (54.4%). In the 2022 Sample, 4.8% of Sample Adult interviews were sufficient partials.

**For the 2022 Sample, the Final Sample Adult Response Rate was 47.7%.**

Note that numbers of households, Sample Children, and Sample Adults eligible and interviewed were used for the calculations of response rates shown and rounding discrepancies may occur when using the percentages.

### Reporting Household, Sample Adult, and Sample Child Response Rates

Which response rate to report depends on the focus of one's analysis. When reporting on analyses performed with the Sample Adult data file, the data user should report the Final Sample Adult Response Rate. In addition, it is good practice to also report the Household and Conditional Sample Adult Response Rates. Similarly, if the focus of one's analysis is the Sample Child, the Final Sample Child Response Rate should be reported. Again, the Household and Conditional Sample Child Response Rates could also be reported.

The Household Response Rate would also be reported if one is performing a household-level analysis using the public-use paradata file.

### Summary of Household, Sample Adult, and Sample Child Response Rates Since 2019

The 2019-2022 counts for eligible and interviewed sample units used in the calculation of response rates for combined data years are shown in Table 2. Tables 3 and 4 present the conditional and unconditional response rates for the household, Sample Adult, and Sample Child modules for 2019-2022.

**Table 2. Number of eligible and interviewed households, Sample Adults and Sample Children, National Health Interview Survey, 2019-2022**

Year	Household			Sample Adult		Sample Child	
	Eligible	Roster Complete	Interviewed	Eligible	Interviewed	Eligible	Interviewed
2019	54,231	35,404	33,138	35,365	31,997	10,155	9,193
2020	43,280	23,694	21,930	23,694	21,153	6,626	5,790
2021	58,068	33,325	30,673	33,264	29,482	9,511	8,261
2022	58,215	31,654	28,854	31,579	27,651	8,879	7,464

**Table 3. Conditional response rates, National Health Interview Survey, 2019-2022**

Year	Sample Adult	Sample Child
2019	90.5	90.5
2020	89.3	87.4
2021	88.6	86.9
2022	87.6	84.1

**Note: The Conditional Response Rate is calculated by dividing the number of interviewed by the number of eligible from households with completed rosters.**

**Table 4. Unconditional response rates, National Health Interview Survey, 2019-2022**

Year	Household	Sample Adult	Sample Child
2019	61.1	59.1	59.1
2020	50.7	48.9	47.8
2021	52.8	50.9	49.9
2022	49.6	47.7	45.8

**Note: This is the Final Response Rate, and it is calculated by dividing the number of interviewed by the number of eligible from households with completed rosters, and then multiplying this quotient by the Household Roster Completion.**

## VI. Weighting

NHIS is a sample survey. That is, only a sample (subset) of the civilian noninstitutionalized population is selected to participate in the survey. Additionally, not everyone selected to participate agrees to participate, which can affect the representativeness of the sample. To account for these two factors, sampling weights are created.

These sampling weights are used to produce representative national estimates. The data must be weighted to obtain population estimates for survey outcomes in the population represented by the NHIS. The value of the weight for a given respondent can be interpreted as the number of persons in the NHIS target population represented by that respondent. The sum of the weights over all respondents is used to estimate the size of the total target population. The weights reflect several steps of adjustments starting with a base weight, which is inverse to the probability of selection. Households and persons that are more likely to be selected are given lower weights so that the final estimates are not biased by their increased likelihood of being selected. For example, in a household of two eligible adults, the Sample Adult has a selection probability of one-half, and therefore their base weight will be increased by two. However, in a household of four eligible adults, the Sample Adult has a selection probability of one-fourth, and therefore their base weight will be increased by four, since roughly speaking they represent more people from the household. The base weights are then adjusted for nonresponse patterns, that is, the different response rates among different household and person-level subgroups.

The 2019 questionnaire redesign provided an opportunity to evaluate the adjustment approach that had been in place since 1997. For 1997-2018, the adjustment approach was based on geography; the weights for households and persons in geographic areas with lower response rates were increased more than for those in areas with higher response rates. That way, final estimates were not biased by the latter group's increased likelihood of participation. More sophisticated methods to decrease potential nonresponse bias are now available (Olson, 2013; Valiant et al., 2018), and based on the evaluation, the weighting process for 2019 was updated. In 2019 the updated approach for nonresponse adjustment used multilevel regression models that include paradata variables that are predictive of both survey response and selected key health outcomes, the key criteria for effective bias reduction. Starting in

2021, Recursive Partitioning for Modeling Survey Data (RPMS) was used within R programming language to build classification trees of household, Sample Adult, and Sample Child response (Toth, 2019). The inverse of the response rate within terminal nodes of each resulting tree serves as the nonresponse adjustments to household, Sample Adult, and Sample Child base weights. With RPMS, tree and random forest models can account for stratification and clustering as well as unequal probabilities of selection. A major advantage of RPMS is the automatic consideration and estimation of complex interactions.

Finally, the nonresponse adjusted weights are typically calibrated to U.S. Census Bureau population projections and American Community Survey (ACS) one-year estimates. Starting in 2019, calibration is performed with iterative proportional raking using the following raking dimensions: age by sex (population projections), age by race and ethnicity (population projections), educational attainment (ACS one-year estimate), and Census division by Metropolitan Statistical Area (MSA) status (ACS one-year estimates). In 2020, housing tenure (ACS one-year estimate) was added to the calibration step, and in 2022, Census division was replaced with region. Note that for the 2021 survey year, the U.S. Census Bureau did not release single-year ACS estimates by housing tenure, education level, and MSA by Division. Therefore, substitute calibration totals for these variables were obtained from the 2021 Current Population Survey (CPS) March Annual Social and Economic (ASEC) Supplement. ACS one-year estimates are once again being used for raking dimensions in 2022. These changes to the nonresponse adjustment approach and the calibration methods have the potential to impact comparisons of the weighted survey estimates over time.

### *Weights*

The Sample Adult and the Sample Child file each have a unique and separate final annual weights and variance estimation variables.

The Final Annual Weight should be used to generate national estimates. This weight includes the design, ratio, nonresponse and calibration adjustments.

**The final annual weight in the Sample Adult file is WTFA\_A.**

**The final annual weight in the Sample Child file is WTFA\_C.**

Further information on how to implement sampling weights is found in the section “Analyzing 2022 NHIS.”

The paradata files do not contain weights. Depending on the analysis of the paradata, if weights are needed, they can be pulled from public-use data files.

Interim weights are those sampling weights that do not include the final standard calibration adjustment for age, sex, race and ethnicity, education, and region by MSA status raking to population control totals. Since 2020, interim weights are not included in the public use files and are available through the NCHS Research Data Center (RDC): <https://www.cdc.gov/rdc/>.

### *Variance Estimation*

In a data collection, estimates based on different samples will vary and can differ from the true population values. The estimated difference between the true target population value and the estimate from a random

sample is the sampling error. Sampling error cannot be directly calculated because the true target population value is unknown. Rather, sampling error is estimated and expressed as a standard error (SE), the average degree to which estimates based on random samples differ from each other and the true target population value due to sampling. This measure is incorporated in common statistical methods such as significance testing and estimating confidence intervals.

Because of the complex nature of the NHIS sampling design (specifically, the use of stratified cluster sampling), key nesting variables were created to capture explicit stratification and to identify clustering for a more accurate estimation of the sampling error.

**For both the Sample Adult and the Sample Child files the stratum and primary sampling unit (PSU) variable names are PSTRAT and PPSU.**

PSTRAT and PPSU are simplified versions of the true NHIS sample design variables created for the public-use files in order to protect the identity of survey respondents. The strata identifier is not directly related to state or density strata. When using the publicly available data files for estimation purposes, strata and PSU identifiers provided by NCHS are required to properly estimate variances. The use of these publicly available variance estimation variables may provide slightly different standard errors than the use of the confidential variance estimation variables used by analysts at NCHS. Data users who want access to the confidential variance estimation variables used by analysts at NCHS may apply to the RDC.

Analysts should be aware that the use of standard statistical procedures that are based on the assumption that data are generated via simple random sampling (SRS), instead of a complex sample design, generally will produce incorrect estimates of variances and standard errors when used to analyze data from the NHIS. 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.

### *Degrees of Freedom*

The number of degrees of freedom is used to determine the t-statistic, its associated percentage points, p-values, standard error, and confidence intervals. 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 at least 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. 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). 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 address clusters, 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.

## VII. Editing the Data During and After the Interview

### *Edits to Protect Confidentiality*

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 302 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 primary sampling units (PSUs) are masked when the data are publicly released. NHIS data sets may also be coarsened by suppressing survey variables, collapsing multiple variables into one, and collapsing response categories. In addition, statistical noise at both the variable level and record level may occasionally be added to protect confidentiality.

Notes fields in the Codebook report may include information 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 adults, all age variables were top-coded to "85 years and older" (85+). For example, survey questions related to age at diagnosis for a type of cancer (e.g., LUNGAGE\_A) and diabetes (DIBAGE\_A) ("How old were you when you were diagnosed with [this condition]?") are top-coded to 85+ years.

To further protect confidentiality, detailed information for some variables is not available on the NHIS public-use data files. For a list of questions not available on the public-use file, see the annual restricted-use codebook available on the data release webpage, and the Appendix in this document.

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 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: <https://www.cdc.gov/rdc/>.

### *Family-Level Replicate*

In the field, the interviewer can conduct either the Sample Adult or Sample Child interview first in households where both eligible adults and children reside. In instances where the Sample Adult and the Sample Child belong to the same family, the instrument is optimized to only ask family level questions in the first interview. This helps to minimize respondent burden by eliminating repetition for family level questions, such as family income and food security. If, however, the respondent of the first interview refuses or doesn't know the answer to a significant number of questions within a family-level section, that section is repeated in the second interview when the respondent of the second interview is not the same individual.

The family level data collected are then replicated (i.e., copied) to the other interview to a replicate variable of the same name (but a different suffix) in a post-processing step. For example, if the adult interview preceded the child interview and they are in the same family, the question about whether anyone in the family had problems paying medical bills is collected in the adult variable PAYBLL12M\_A and replicated to the child variable PAYBLL12M\_C. The Questionnaire report identifies a variable as being replicated in the “Replicate to:” field.

Searching the document for the string “Replicate” will identify the variables that underwent replication.

When the Sample Adult and the Sample Child are in different families within the household, both the Sample Adult and the Sample Child respondent will be asked family level questions about their respective families. In households where there are no children or there are no eligible adults (e.g., all active Armed Forces), there is no replication involved.

Annual core sections of the questionnaire with instrument optimizations and replicated variables include Family Income (INC), Family Employment (FEM), Difficulty Paying for Health Care (PAY), Food-Related Programs (FOO), Housing (HOU), and Telephone Use (TEL). They can also be found in some sponsored sections of the questionnaire such as Food Security (FDS) and Food-Related Programs (FOO).

Replicate measures are used in analyses the same way as any other measure available in the Sample or Sample Child files. Analyses of replicate measures can be interpreted as an estimate of Sample Adults/Sample Children/persons who are in a family meeting a specific outcome or characteristic (e.g., percentage of persons aged 0–64 years who are in a family that is having problems paying medical bills).

### *Hard and Soft Edits*

To help prevent both interviewer data entry error and respondent error, range values and consistency checks may be programmed into the CAPI system. During the interview, if an interviewer enters an out-of-range value (such as 180 years instead of 18 for age), an error message instructs the interviewer to enter a new value. 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 with or without a new response being entered. Soft edits may apply to questions for which the response entered is plausible (such as an extreme height value).

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 and the review of frequencies for reasonableness. Each variable is examined to determine if its values are within its range of permissible values. Values not in that range are verified as missing if they are not in the universe due to legitimate skip patterns in the questionnaire or set to the special value of “not ascertained” if there was a break-off in the interview.

### *Question-Specific Replication*

An optimization edit is an edit that fills-in values for variables that were skipped in the instrument because the information could be inferred from the Sample Adult or Sample Child interview, whichever went first. For example, the marital status of the Sample Adult would be known if the Sample Child interview preceded the Sample Adult interview, if the Sample Adult and Sample Child were in the same family, and if the Sample Adult was also a parent of the Sample Child. Select questions in health insurance relating to detailed characteristics of



shared private plans between the Sample Child and Sample Adult in the same family were also filled-in from responses of the interview that came first.

### *Recode into Different Variable*

Recodes have been created for select questions to make the data more analytically useful. One example of this is a recode that converts a single variable allowing selection of as many answers as are applicable into a series of variables (one for each possible response) with yes/no or mentioned/not mentioned responses. Other recodes have been created to summarize information obtained from multiple questions available to the public (e.g., summary scores of validated scales), or to combine information from which some information may not be publicly available (e.g., multiple race categories).

Data users are recommended to review the description of 'Major Recodes' for the topic of interest in this document, and the codebook documentation for additional recode information.

## Orientation to How to Use NHIS data

All datasets and associated documentation for 2022 are available on the NHIS website:

<https://www.cdc.gov/nchs/nhis/2022nhis.htm>

### Documentation

Documents in Portable Data Format (PDF) describing NHIS methods, survey implementation materials and other background information are included under the tabs “Using the NHIS” and “Survey Implementation Materials.”

The following documents are included in the tab “Using the NHIS:”

**Survey Description document (PDF):** A description of NHIS methods, year-specific response rates and content, and other useful resources for NHIS data users.

**Imputed income technical document (PDF):** A description of the methodology for creating the 2022 NHIS imputed income variables.

**Paradata Survey Description document (PDF):** A year-specific description of the interview process information collected from sampled households.

**File record length and size summary (PDF):** A year-specific list summarizing the number of records, file size, and record length for each of the ASCII data files released.

**Checksum (PDF):** A list of year-specific reference values for each ASCII and CSV data file released to allow data users to verify the integrity of downloaded files.

The following documents are included in the tab “Survey Implementation Materials:”

**Survey Questionnaire - English (PDF):** Year-specific NHIS questions fielded.

**Survey Questionnaire - Spanish (PDF):** Spanish version of the year-specific NHIS questions fielded.

**Field Representative Manual (PDF):** The manual on Computer-Assisted Personal Interview (CAPI) for NHIS interviewers.

**NHIS Instrument Flowchart(s) (PDF):** A graphical view of the questionnaire content during the year.

**NHIS Sponsored Content (PDF):** A year-specific list of sponsoring agencies and associated sponsored questions in NHIS.

**Survey Brochure (PDF):** A booklet explaining the NHIS, planned periodicity of topic areas for the Sample Adult and Sample Child, and a list of agencies sustaining survey context.

**Advance Letter (PDF):** A letter that explains the NHIS mailed to sampled households prior to interviewer contact and provided to survey respondent at the time of the interviewers’ visit.

## *Survey Questionnaire*

The 2022 survey questionnaire (PDF) lists the questions in the survey and descriptive information about them. The information in the document is organized in two panels: a hierarchical bookmarks panel on the left for navigation, and a main panel on the right for displaying detailed content. The bookmarks themselves are organized as:

- a) Link to a contents page that explains the document's overall structure
- b) List of questions added during 2022 that were not included at the start of the year.
- c) Hierarchical section index that lists the sections, their descriptions, the content type (Annual Core, Rotating Core, Sponsored Content or Emerging Content) and the page range in the PDF for each section for ease of printing
- d) Hierarchical questionnaire organized by module, section and variable and appearing in the order that the questions are asked.
- e) A glossary of variables mentioned in the skip and fills instructions of some survey questions that are used for the programming the interview.

Selecting a bookmark for a module or a section navigates to the first variable in the module or section, respectively. When selecting a variable of interest, detailed information for that variable is displayed in the main panel. For each variable, the main panel heading has the year and title of the survey along with the section abbreviation and description. The body of the main panel starts with a header (in blue background) with the Question ID (used for ordering questions in the questionnaire), the variable name, the interview module and the content type. This is followed by the question text as it appears in the instrument. The question text may contain one or more context-sensitive fills, indicated by text with a leading caret (^) symbol. A fill is text that is conditionally generated to modify the question text to make it more suitable to the context of the interview. For example, the fill whose name is ^heshe\_C will expand to "he" if the Sample Child respondent indicated that the Sample Child is male, "she" if female or "they" if sex was not provided. All fills in the question text appear in the fills table in order along with their description and rule-based instructions on how the fill text is generated in the instrument. For some variables, as part of the question text, there may be interviewer instructions in bolded blue text with any optional text appearing in italics and gray font. Below the fills table, there is another table with valid response categories and their descriptions, followed by the universe description, and if present, any skip instructions, hard or soft edits.

Both English and Spanish versions of the questionnaire are available on the NHIS website. The Spanish version of the questionnaire has Spanish translations for the question text, the fills and the response choices.

## *NHIS Sponsored Content*

The NHIS Sponsored Content (PDF) lists the sponsoring organizations and the questions that they have sponsored in the NHIS for the current year. The information in this document is organized in two panels: a bookmarks panel on the left listing the sponsoring organization names alphabetically (short form), and a main panel on the right listing the variables and some descriptive information about them such as question ID, question text and the universe description. When content is sponsored by multiple organizations, the names of all the sponsors are shown on the bookmarks panel. Selecting an organization's bookmark takes you to the content sponsored by that organization. The document's page header lists the sponsoring organizations' full name(s).

### [Codebook \(PDF\)](#)

The Codebook report combines all the detailed information for a variable with the unweighted frequencies (counts and percentages) found in the data. The report has a navigational bookmarks panel on the left with expandable module and section bookmarks in questionnaire order. The main panel on the right contains the variable detail. When a variable bookmark is selected, the detailed display includes its module, section, file, data type, question text (if present), question fill information, universe and universe description, the variable description or label, question ID, keywords, and notes. This is followed by a table that provides the unweighted frequencies and percentages for the variable. All response categories are shown in the table, including those 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 cases that are not in the universe.

In the NHIS, the same codes are used across all files to designate “refused” (RF) and “don’t know” (DK) responses: refusals are coded as 7 (with leading 9’s added 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 added 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 interview before reaching the question), the responses will appear as 8’s for “not ascertained,” again with leading 9’s added to the length of the field, as in 8, 98, 998, etc., for the remaining variables in the file. 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 may be collapsed into a single category called “Unknown,” which is typically designated with a 8 or 9 (with leading 9’s to fill out the field, if necessary). Data users are advised to read the notes in the data release documentation for further information about the variables of interest.

### [Summary \(PDF\)](#)

The summary report is a PDF document with a bookmarks panel on the left organized by module and section in questionnaire order, and a main panel that displays the variable list by section. When the section is selected in the bookmarks panel, the following information for all variables in that section is displayed in the main panel in tabular form. The lead-in header has the module name, the 3-letter section abbreviation, and the section description. The table has rows with the following information:

**Question ID:** if the variable is in the questionnaire, the unique ID for that variable is displayed. For recodes, the word “Recode” is displayed, and for any that are neither (e.g., identifiers), this column is blank. Question ID may change by survey year and should not be used for data management purposes across years.

**Variable name:** the name of the variable in the data. Generally, the variable name in the questionnaire has the same variable name in the dataset.

**Source variables:** for recodes, this column lists the names of variables used to create the recode

**Description:** the variable label

**Type:** the data type for this variable, i.e., character or numeric

**Location:** the column range in the ASCII file (column numbers) where this variable is stored

**Length:** the length of the variable as a character data type

### *Codebook for restricted-use variables (PDF)*

This document lists the restricted-use (or inhouse) variables that are available to analysts in the RDC. It does not include any variables that are in the public data files. The format is similar to the codebook, except that no frequencies are shown.

### *Data Files*

The 2022 NHIS data release includes files for the annual Sample Adult, annual Sample Child, Imputed Income (as a poverty ratio) for the Sample Adult and Sample Child, and Paradata.

The data files are released as both a column-delimited text (ASCII) file and a comma-separated values (CSV) file. Programs that contain input statements in SAS, Stata and SPSS environments are provided to help load the ASCII files into datasets with proper data formats and labels in the respective computing environments. Table 5 lists the names of data files, programs and documentation in the 2022 NHIS data release.

Files corresponding to the 2022 NHIS have a two-digit suffix at the end that represents the survey year, e.g., 22 for 2022, or adult22. In years prior to 2019, separate files were available for household, family, and person level information due to the different survey design. In 2020, additional files were available for the Sample Adult (Partial and Longitudinal) due to the longitudinal component implemented that year but not continued for 2022. Imputed income files for Sample Adult and Sample Child can be merged with their respective Sample Adult and Sample Child files to create a single dataset (see, “Merging Survey Data and Imputed Income Files” in next section). Sample Adult and Sample Child files can also be merged with Paradata.

### *Sample Adult and Sample Child files*

The 2022 Sample Adult and Sample Child files include all publicly available questionnaire variables and associated recodes, and household and family-level variables. The Codebook and Summary reports for each file describe their contents in detail.

### *Imputed Income Files*

The 2022 Imputed Income files for Sample Adult and Sample Child contain 10 imputations of family poverty ratio as both continuous and categorical variables. An example with sample code that demonstrates using the imputed income data file in an analysis is described in this report in the section “Merging Files” under the heading “Using Imputed Income Data Files.”

**Table 5. Data release files names for the Sample Adult, Sample Child, Imputed Income and Paradata files: NHIS 2022.**

Type of file	Sample Adult Annual files	Sample Child Annual files	Imputed Income files	Paradata files
Data in column-delimited ASCII format	adult22.dat	child22.dat	Adultinc22.dat Childinc22.dat	paradata22.dat
Data in comma separated Values (CSV) format	adult22.csv	child22.csv	Adultinc22.csv Childinc22.csv	paradata22.csv
SAS program with input statements	adult.sas	child.sas	adultinc.sas childinc.sas	paradata.sas
STATA program with input statements	adult.do	child.do	adultinc.do childinc.do	paradata.do
SPSS program with input statements	adult.sps	child.sps	adultinc.sps childinc.sps	paradata.sps
Summary	Adult-summary.pdf	Child-summary.pdf	Adultinc-summary.pdf Childinc-summary.pdf	Paradata_summary.pdf
Codebook	Adult-codebook.pdf	Child-codebook.pdf	Variables included in the Adult and Child Codebooks	Paradata_codebook.pdf

Variables based on the first imputation were also added to the Sample Adult and Sample Child files for convenience of users who choose not to use multiply imputed data in their analyses. While each of the 10 imputations has been drawn from a valid distribution based on a regression model, the first imputation included in the Sample Adult and Sample Child files may be slightly different from the other sets of imputations. Single imputation analyses result in estimated standard errors that are too small because the imputed values are treated as if they were observed. This ignores the inherent uncertainty resulting from lack of knowledge about the true (unobserved) value, but it is superior to analyses that use only cases with observed values.

Information about income measurements collected in NHIS and income recodes are described in this document in the sections “Family Income” and “Recodes of Family Income and Imputed Family Income.” Methodology for creating the imputed income variables is described in the “Imputed Income Technical Document” available with the 2022 file releases on the NHIS website, under “Using the NHIS.”

### *Paradata File*

The NHIS Paradata file contains information about the interview process. The data from the Paradata file are collected as part of the NHIS interview, using computer-assisted personal interviewing (CAPI). The NHIS paradata come from a number of sources:

- The Contact History Instrument (CHI), a supplemental piece to the NHIS that collects data from the interviewer about each contact attempt. Data include strategies used for gaining participation and reasons for respondent reluctance.
- The Back section of the NHIS, where a series of questions are asked of the interviewer, including mode of interview (in-person visit vs. phone interview), and reasons for partial interviews/breakoffs.
- Date and time variables from each module of the instrument (Household, Sample Child, Sample Adult). The date and time information are collected each time a module is started and completed.

The Paradata file is on a case (household) level, where one record represents one case. Unlike the NHIS public-use Sample Adult and Sample Child data files, which contain information on fully complete and sufficiently complete interviewed cases only, the Paradata file also contains data on other types of cases, including cases that were ultimately refusals, insufficient partials, and other types of nonresponse.

The Paradata file is intended as both a stand-alone data file and one whose fully complete and sufficiently complete cases can be linked with the Sample Adult and/or Sample Child data files. For more information about the Paradata file, including linking Paradata files with other data files, see the Paradata Survey description document available with the 2022 file releases on the NHIS website, under “Using the NHIS.”

### *Variable Conventions*

Variable labels are restricted to 80 characters due to limits in some programming languages. All variables have a length of 12 characters or less. Variables names in the Sample Adult file have the suffix `_A`, e.g., `DIBEV_A`, to indicate that they refer to the Sample Adult or were asked of the Sample Adult’s family. Variables associated with the Sample Child will analogously have the suffix `_C`, e.g., `DIBEV_C`. Variables that do not have these suffixes are household or family level variables or identifiers, e.g., `HHX`.

### *Definitions*

The following defines some terms used in the different reports:

**F1:** Indicates there is a ‘help screen’ available to the interviewers with definitions about the question topic that can be used for answering respondents’ questions.

**Fills:** Text that modifies the question, based on previously collected information and using conditional logic. Fills are indicated by a caret (^) symbol followed by the name of the fill, e.g., `^SCNAME`.

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

**Notes:** Additional information that analysts need to know about a variable, such as assumptions, limitations, caveats, and differences between instrument versions. Analysts are encouraged to read the notes pertaining to variables of interest. Notes may contain cross-references to other pertinent variables.

**Recode:** A variable derived from the reordering, collapsing, or verbatim coding of another variable. Alternatively, a recode may be constructed from two or more variables. All variables used to construct a recode are listed as a cross reference in Sources. Examples of recodes based on annual core content include MARSTAT\_A (current marital status of the Sample Adult) and SMKCI GST\_A (cigarette smoking status). Additionally, a recode may be created when a continuous variable needs top- and/or bottom-coding for confidentiality reasons (for example, SCHDYMSSTC\_C, school days missed due to illness or injury in the past 12 months). In some years, the NHIS may also contain content on the frequency of a health behavior in two parts: number and type of time unit. Recodes are then created to combine and standardize responses into a single time unit.

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

**Universe:** The group of adults or children to whom a specific question applies. For example, the universes for most Sample Adult variables are adults who were age 18 or over. This universe is specified on the Codebook report as HHSTAT\_A=1. 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 never had high cholesterol, e.g., CHLEV\_A having a value of 2, or RF or DK the response would not be eligible for a follow-up question CHL12M\_A about whether they had high cholesterol in the past 12 months. Universes for many questions are often age specific. In the redesigned NHIS, missingness in the Sample Adult or Sample Child's age is possible, and in those few cases the individuals would be ineligible for the question. Note that during rostering, when a person's age is not known, there are age-related follow-up questions to get at whether they are adults or children so the Sample Adult or Sample Child selections can be made. If the age is still not known, the interview terminates. Similarly, the sex variable (SEX\_A or SEX\_C) also allows for missing values, but the interview can proceed. Sex-specific questions for the Sample Adult and the Sample Child are not in universe when sex is unknown.



## Analyzing 2022 NHIS and Software Examples

**To appropriately analyze NHIS data, it is necessary to utilize weights and variance estimation variables.** This is because the NHIS uses a complex sample design involving stratification and clustering designed to represent the civilian noninstitutionalized population of the United States and not all sampled respondents respond. If data are not weighted, severely biased estimates may result, such as producing estimates that are not representative of the NHIS target population. If the correct variance estimation variables are not used, then estimates of precision, such as standard errors, will likely be smaller than they should be. This will make the data appear to be more precise and will result in more statistically significant differences between estimates and in other analyses that are subject to excessive Type I error (rejection of a true null hypothesis).

### *Applying Sample Weight and Variance Estimation Variables in Analysis*

Several software packages are available for analyzing complex samples. Below are examples of computer code for specifying sample weight and variance estimation variables for standard error calculation code of means, percentages and totals with the NHIS data using SUDAAN, Stata, SPSS, SAS, and R software packages for illustrative purposes. The examples below use the Sample Adult sampling weight (WTFA\_A) or for Sample Child analysis, use the Sample Child sampling weight (WTFA\_C).

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 for the file with the following statements in selected software packages.

### *Example using SUDAAN*

```
PROC <DESCRIPT, CROSSTAB, ...> ... DESIGN = WR;  
NEST PSTRAT PPSU;  
WEIGHT WTFA_A;
```

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

### *Example using STATA*

```
Stata svy  
svyset [pweight=wtfa_a], strata(pstrat) psu(ppsu)  
svy: mean <name of variable to be analyzed for average>  
or  
svy: proportion <name of variable to be analyzed for percentage/proportion>
```

### Example using SPSS

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_A
/DESIGN STRATA=PSTRAT CLUSTER=PPSU
/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.
```

### Example using SAS

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

```
PROC SURVEYMEANS;
STRATA PSTRAT;
CLUSTER PPSU;
WEIGHT
WTFA_A;
VAR <name of variable to be analyzed>;
RUN;
```

```
PROC
SURVEYFREQ;
STRATA PSTRAT;
CLUSTER PPSU;
WEIGHT WTFA_A;
TABLES <name of variable to be
analyzed>; RUN;
```

### Example using R

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=~PPSU, strata=~PSTRAT,
  nest = TRUE,
  weights=~WTFA_A,
  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.

### Examples of Analysis and Weighting Procedures for Sample Adults and Sample Children

Table 6 includes examples of types of research questions that analysts may use 2022 Sample Adult and Sample Child NHIS data for, and guidance on additional adjustments needed to the sampling weights when analyzing in combination with the 2019–2021 NHIS. Some of these analytic examples are interchangeable between the Sample Adult and Sample Child files but analyst must use the respective Sample Adult and Sample Child weight.

<b>Analytic Goal</b>	<b>Example</b>	<b>Weight Variable and Modifications Needed</b>	<b>Variance Estimation</b>
<b>Produce official estimates for 2022</b>	Percentage of adults ever told by doctor or other health professional that they had diabetes	Use WTFA_A	Use standard variance estimation variables PSTRAT and PPSU for 2022
<b>Compare estimates between 2022 and 2021</b>	Percentage of adults ever told by doctor or other health professional that they had diabetes, 2022 compared with 2021	Use WTFA_A for both 2022 and 2021	Use standard variance estimation variables PSTRAT and PPSU for 2022 and 2021
<b>Produce estimates and population totals for topics only available in quarters 3 and 4 of 2022</b>	Percentage of adults who have access to the Internet.	Create a new weight variable where the observations in quarters 1 and 2 are assigned a value of zero, and for observations in quarters 3 and 4, multiply the weight (WTFA_A) by 2	Use standard variance estimation variables PSTRAT and PPSU for 2022
<b>Pooling 2022 and 2021 to increase sample size <sup>1</sup></b>	Percentage of children who have Attention-Deficit/Hyperactivity Disorder or ADD or Attention-Deficit Disorder.	Concatenate the 2021 and 2022 files. Create a new weight using WTFA_C and divide it by 2	Use standard variance estimation variables PSTRAT and PPSU for 2022 and 2021

Table 6 continued. Examples of Analysis and Weighting Procedures for Sample Adults and Sample Children—2022 NHIS— Continued.

Analytic Goal	Example	Weight Variable and Modifications Needed	Variance Estimation
<b>Produce estimates for topics available from July 2020 through December 2022 as 6-month intervals. Compare semi-annual intervals (trends).<sup>2</sup></b>	Percentage of adults diagnosed with COVID-19: July–December 2020, January–June 2021, July–December 2021, January–June 2022, July–December 2022	Create a new weight variable where the observations in quarters 1 and 2 of 2020 are assigned a value of zero; and for each remaining 6-month interval, multiply the weight by 2.	Use standard variance estimation variables PSTRAT and PPSU for 2022, 2021 and 2020
<b>Pooling 2019, 2020, 2021 and 2022 data to increase sample size.<sup>3</sup></b>	Percentage of adults ever told by doctor or other health professional that they had a stroke	Use WTFA_A in 2019, 2021, and 2022, and the partial weight WTSA_P for 2020. Create a new weight using these weight variables and divide it by 4	Use standard variance estimation variables PSTRAT and PPSU for 2019–2022
<sup>1</sup> See scenario 1 for example using SAS, STATA, or R. <sup>2</sup> See scenario 2 for example using SAS, STATA, or R. <sup>3</sup> See scenario 3 for example using SAS, STATA, or R.			

Note that the 2020 NHIS included three different sampling weights for the Sample Adult, and these were designated for the analyses of the longitudinal sample, the partial sample (2020 sample only), and the combination of those two samples. The Sample Child annual files did not include a longitudinal sample as done with Sample Adults in 2020, and any procedure described below referencing the longitudinal sample is not applicable to Sample Child files.

### *Scenario 1: Concatenating 2021 and 2022 (pooling) Sample Child data to increase sample size*

This example illustrates how to create a dataset that pools data from 2021 and 2022 to increase sample size and produce a cross-sectional estimate (e.g., Attention-Deficit/Hyperactivity Disorder or ADHD or Attention-Deficit Disorder or ADD). An analyst can pool 2021 and 2022 data when interested in increasing the precision of an estimate among a population that might have a small sample size with a single year of data. For this example, a new file called CHILD21\_22 will include all Sample Children in 2021 and 2022 data. It will be derived from the data files, child21 and child22. This new file will only contain the variables of interest: a doctor or other health professional diagnosed the Sample Child with ADHD/ADD, variance structure variables, and sampling weights from the child21 and child22 files. The new weight WTFA\_ADJ will be created to account for 2 years of data (i.e., that is dividing each sample weight by the number of years that are being pooled).

### Example using SAS

```
*Create temporary datasets of 2021 and 2022 data by selecting weight, analytic variables and variance structures (PSTRAT and PPSU);
```

```
DATA TEMPCHILD21;  
SET CHILD21; *2021 Sample Child file;  
KEEP WTFA_C PSTRAT PPSU ADHDEV_C;  
RUN;
```

```
DATA TEMPCHILD22;  
SET CHILD22; *2022 Sample Child file;  
KEEP WTFA_C PSTRAT PPSU ADHDEV_C;  
RUN;
```

```
*Concatenate the temporary 2021 and 2022 datasets; DATA CHILD21_22;  
SET TEMPCHILD21 TEMPCHILD22;  
WTFA_ADJ=WTFA_C/2; *Divide the weight by 2;  
RUN;
```

### Example using Stata

```
*Create temporary datasets of 2021 and 2022 data by selecting weight, analytic variables and variance structures (PSTRAT and PPSU)
```

```
*Child21 – 2021 Sample Child file use child21  
keep wtfa_c pstrat ppsu adhdev_c  
save tempchild21
```

```
*Child22 – 2022 Sample Child file use child22  
keep wtfa_c pstrat ppsu adhdev_c  
save tempchild22
```

```
*Concatenate temporary 2021 and 2022 datasets  
append using tempchild21  
gen wtfa_adj=.  
*Divide the weight by 2;  
replace wtfa_adj=wtfa_c/2  
save child21_22
```

### Example using R

```
# Keep only the selected variables  
TEMPCHILD21 <- subset(CHILD21, select = c("WTFA_C", "PSTRAT", "PPSU", "ADHDEV_C"))
```

```
# Keep only the selected variables  
TEMPCHILD22 <- subset(CHILD22, select = c("WTFA_C", "PSTRAT", "PPSU", "ADHDEV_C"))
```

```
# Concatenate the two temporary datasets
CHILD21_22 <- rbind(TEMPCHILD21, TEMPCHILD22)

# Create a new variable by dividing WTFA_C by 2
CHILD21_22$WTFA_ADJ <- CHILD21_22$WTFA_C / 2
```

### *Scenario 2: Comparing semi-annual intervals (trends):*

#### *Produce estimates of COVID-19 diagnosis from July 2020 through December 2022 as 6-month intervals*

This example illustrates how to create one dataset to examine semi-annual intervals (trends) over a 30-month period of data collection. Data from the third and fourth quarters of 2020 and a full year of data from 2021 and from 2022 will be used in this example to produce cross-sectional 6-month interval estimates (e.g., COVID-19 diagnosis).

For this example, one new file will be created called NEWADULT20\_22. This new file will only contain the variables of interest: a doctor or other health professional told the Sample Adult that they had or likely had COVID-19, variance structure variables, modified sampling weights from the adult20, adult21, and adult22 files and an indicator for each 6-month interval (TIME6M). For each 6-month interval, a NEWWT will be created by multiplying the weight (WTFA\_A) by 2 to account for only using 6-months of data. Estimates in this dataset should only be examined by 6-month intervals, which is indicated by TIME6M variable (i.e., value 1 is the first 6-month period, value 2 is the second 6-month period, value 3 is the third 6-month period, value 4 is the fourth 6-month period, and value 5 is the fifth 6-month period). The annual weight has been adjusted to treat 6 months of data so that weighted frequencies for a particular measure (e.g., COVID-19 diagnosis) will reflect the number of adults who were diagnosed in the U.S. civilian noninstitutionalized population. Using WTFA\_A without adjustment for a 6-month analysis will result in population estimates that are too low and will not reflect the number of people in the U.S. civilian noninstitutionalized population with the outcome of interest.

### *Example using SAS*

```
*Create temporary datasets of 2020 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU);
DATA NEWADULT20;
SET ADULT20; *2020 Sample Adult file;
KEEP PSTRAT PPSU CVDDIAG_A NEWWT TIME6M;
IF INTV_QRT IN (1, 2) THEN NEWWT = 0; *For the first and second quarters in 2020, assign weights of
zero;
ELSE NEWWT=WTFA_A * 2; *For 6-months of data, multiply the weight by 2;
IF INTV_QRT IN (3, 4) THEN TIME6M=1; *This value indicates the first 6-month time period that data was
collected;
RUN;

*Create temporary dataset of 2021 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU);
DATA NEWADULT21;
SET ADULT21; *2021 Sample Adult file;
```

```
KEEP PSTRAT PPSU CVDDIAG_A NEWWT TIME6M;
IF INTV_QRT IN (1, 2) THEN TIME6M=2; *For the first and second quarters of 2021, this is the second 6-
month time period that data was collected;
IF INTV_QRT IN (3, 4) THEN TIME6M=3; *For the third and fourth quarters of 2021, this is the third 6-
month time period that data was collected;
NEWWT = WTFA_A *2; *For 6-months of data, multiply the weight by 2;
RUN;
```

```
*Create temporary dataset of 2022 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU);
DATA NEWADULT22;
SET ADULT22; *2022 Sample Adult file;
KEEP PSTRAT PPSU CVDDIAG_A NEWWT TIME6M;
IF INTV_QRT IN (1, 2) THEN TIME6M=4; *For the first and second quarters of 2022, this is the second 6-
month time period that data was collected;
IF INTV_QRT IN (3, 4) THEN TIME6M=5; *For the third and fourth quarters of 2022, this is the third 6-
month time period that data was collected;
NEWWT = WTFA_A *2; *For 6-months of data, multiply the weight by 2;
RUN;
```

```
*Concatenate temporary 2020, 2021 and 2022 datasets;
```

```
DATA NEWADULT20_22;
SET NEWADULT20 NEWADULT21 NEWADULT22;
KEEP PSTRAT PPSU CVDDIAG_A NEWWT
TIME6M; RUN;
```

### *Example using Stata*

```
*Create temporary datasets of 2020 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU)
use adult20
keep pstrat ppsu cvddiag_a intv_qrt wtfa_a
gen newwt=.
*For the first and second quarters in 2020, assign weights of zero
replace newwt=0 if intv_qrt == 1 | intv_qrt == 2
*For 6-months of data, multiply the weight by 2
replace newwt=wtfa_a * 2 if intv_qrt == 3 | intv_qrt == 4
*This value indicates the first 6-month time period that data was collected
gen time6m=1 if intv_qrt == 3 | intv_qrt == 4
drop wtfa_a intv_qrt
save newadult20

*Create temporary dataset of 2021 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU)
use adult21
keep pstrat ppsu cvddiag_a intv_qrt wtfa_a
```

```

gen time6m=.
*For the first and second quarters of 2021, this is the second 6-month time period that data was collected
replace time6m=2 if intv_qrt == 1 | intv_qrt == 2
*For the third and fourth quarters of 2021, this is the third 6-month time period that data was collected
replace time6m=3 if intv_qrt == 3 | intv_qrt == 4
*For 6-months of data, multiply the weight by 2
gen newwt = wtfa_a * 2
drop wtfa_a intv_qrt
save newadult21

*Concatenate temporary 2020 and 2021 datasets
append using newadult20
keep pstrat ppsu cvddiag_a newwt time6m
save newadult20_21

*Create temporary dataset of 2022 data by selecting weight, analytic variables and variance structures
(PSTRAT and PPSU)
use adult22
keep pstrat ppsu cvddiag_a intv_qrt wtfa_a
gen time6m=.
*For the first and second quarters of 2022, this is the fourth 6-month time period that data was collected
replace time6m=4 if intv_qrt == 1 | intv_qrt == 2
*For the third and fourth quarters of 2022, this is the fifth 6-month time period that data was collected
replace time6m=5 if intv_qrt == 3 | intv_qrt == 4
*For 6-months of data, multiply the weight by 2
gen newwt = wtfa_a * 2
drop wtfa_a intv_qrt
save newadult22

*Concatenate temporary 2020, 2021 and 2022 datasets
append using newadult20_21
keep pstrat ppsu cvddiag_a newwt time6m
save newadult20_22

```

### Example using R

```

# Create temporary datasets of 2020 data
NEWADULT20 <- ADULT20[, c("PSTRAT", "PPSU", "CVDDIAG_A", "WTFA_A", "INTV_QRT")]
NEWADULT20 <- transform(NEWADULT20, NEWWT = ifelse(INTV_QRT %in% c(1,2), 0, WTFA_A * 2),
  TIME6M = ifelse(INTV_QRT %in% c(1,2), 1, 0))
NEWADULT20 <- NEWADULT20[, c("PSTRAT", "PPSU", "CVDDIAG_A", "NEWWT", "TIME6M")]

# Create temporary dataset of 2021 data
NEWADULT21 <- ADULT21[, c("PSTRAT", "PPSU", "CVDDIAG_A", "WTFA_A", "INTV_QRT")]
NEWADULT21 <- transform(NEWADULT21, NEWWT = WTFA_A * 2,
  TIME6M = ifelse(INTV_QRT %in% c(1,2), 2, ifelse(INTV_QRT %in% c(3,4), 3, 0)))
NEWADULT21 <- NEWADULT21[, c("PSTRAT", "PPSU", "CVDDIAG_A", "NEWWT", "TIME6M")]

```



```
# Create temporary dataset of 2022 data
NEWADULT22 <- ADULT22[, c("PSTRAT", "PPSU", "CVDDIAG_A", "WTFA_A", "INTV_QRT")]
NEWADULT22 <- transform(NEWADULT22, NEWWT = WTFA_A * 2,
  TIME6M = ifelse(INTV_QRT %in% c(1,2), 4, ifelse(INTV_QRT %in% c(3,4), 5, 0)))
NEWADULT22 <- NEWADULT22[, c("PSTRAT", "PPSU", "CVDDIAG_A", "NEWWT", "TIME6M")]

# Concatenate temporary 2020, 2021 and 2022 datasets
NEWADULT20_22 <- rbind(NEWADULT20, NEWADULT21, NEWADULT22)
NEWADULT20_22 <- NEWADULT20_22[, c("PSTRAT", "PPSU", "CVDDIAG_A", "NEWWT", "TIME6M")]
```

*Scenario 3: Concatenating 4 years of data: 2019–2022 (pooling) to increase sample size (and excluding 2020 observations from Sample Adults who were interviewed both in 2019 and 2020)*

This example illustrates how to create a dataset that pools data from 2019, 2020, 2021 and 2022 to increase sample size and produce a cross-sectional estimate (e.g., ever told by a doctor or other health professional that they had a stroke). An analyst can pool 2019, 2020, 2021, and 2022 data when interested in increasing precision of an estimate among a population that might have a small sample size with a single year of data, especially when examining among subgroups (e.g., adults younger than 65 years of age, living in nonmetropolitan areas).

This analysis excludes the responses from the second interview of the longitudinal sample. For this example, a new file called NEWADULT19\_20\_21\_22 will include all Sample Adults in 2019, Sample Adults only interview in 2020 (and not part of the longitudinal sample), all Sample Adults in 2021 and all Sample Adults in 2022. It will be derived from the data files, adult19, adult20, adultpart20 and adult21. A temporary dataset will be created from the adult19 file that will only contain the 2019 variables for stroke, variance structure and the sampling weight. The adultpart20 file will be merged with the adult20 file to create a temporary 2020 dataset that will only contain the variables of interest: stroke and the variance structure variables from the adult20 file, and the sampling weight from the adultpart20 file. The adultpart20 file contains the weight variable for Sample Adults in 2020 sample who were not part of the longitudinal sample. A third temporary dataset will be created from the adult21 file that will only contain the 2021 variables for stroke, variance structure and the sampling weight. A fourth temporary dataset will be created from the adult22 file that will only contain the 2022 variables for stroke, variance structure and the sampling weight. The 2019, 2020, 2021, and 2022 temporary files will be combined into a new dataset, NEWADULT19\_20\_21\_22. It will have a new weight (NEWWT\_ADJ) where the annual weight (WTFA\_A) will apply to 2019, 2021 and 2022 cases, and the partial weight (WTSA\_P) will apply to 2020 cases. This weight will also be adjusted to account for 4 years of data (i.e., that is dividing each sample weight by the number of years that are being pooled).

*Example using SAS*

```
*Create a temporary subset of 2019 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU);
DATA NEWADULT19;
SET ADULT19; *2019 Sample Adult file;
KEEP WTFA_A PSTRAT PPSU STREV_A NEWWT; NEWWT=WTFA_A;
RUN;
```

```
*Create temporary subsets of 2020 data by selecting household ID (HHX), weight, analytic variables and variance structures (PSTRAT and PPSU);  
DATA TEMPADULT20;  
SET ADULT20; *2020 Sample Adult file;  
KEEP HHX PSTRAT PPSU STREV_A NHHX;  
NHHX=HHX; *Rename HHX to NHHX to match with household ID in 2020 Sample Adult partial file;  
RUN;
```

```
DATA TEMPPART20;  
SET ADULTPART20; *2020 Sample Adult partial file;  
KEEP HHX_2020 WTSA_P NHHX;  
NHHX=HHX_2020; *Rename HHX_2020 to NHHX to match with household ID in 2020 Sample Adult file;  
RUN;
```

```
*Sort each temporary dataset by the merge variable;  
PROC SORT DATA=TEMPADULT20;  
BY NHHX;  
PROC SORT DATA=TEMPPART20; BY NHHX;  
RUN;
```

```
*Merge the two temporary 2020 Sample Adult files;  
DATA NEWADULT20;  
MERGE TEMPADULT20 TEMPPART20; BY NHHX;  
IF WTSA_P > 0; *Keep records with a value in the partial weight;  
NEWWT=WTSA_P; *Rename WTSA_P to NEWWT to match the variable name for sampling weight when combining files;  
RUN;
```

```
*Create a temporary subset of 2021 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU);  
DATA NEWADULT21;  
SET ADULT21; *2021 Sample Adult file;  
KEEP WTFA_A PSTRAT PPSU STREV_A NEWWT;  
NEWWT=WTFA_A;  
RUN;
```

```
*Create a temporary subset of 2022 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU);  
DATA NEWADULT22;  
SET ADULT22; *2022 Sample Adult file;  
KEEP WTFA_A PSTRAT PPSU STREV_A NEWWT;  
NEWWT=WTFA_A;  
RUN;
```

```
*Concatenate temporary 2019, 2020, 2021, and 2022 datasets;  
DATA NEWADULT19_20_21_22;  
SET NEWADULT19 NEWADULT20 NEWADULT21 NEWADULT22;  
KEEP PSTRAT PPSU STREV_A NEWWT_ADJ;  
NEWWT_ADJ=NEWWT/4; *Divide the new weight by 4 for the three years of data being combined.  
Otherwise, weighted estimates of totals will be higher than the estimated total U.S. civilian
```

noninstitutionalized population;  
RUN;

### *Example using Stata*

\*Create a temporary subset of 2019 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU)

```
use adult19
```

```
keep wtfa_a pstrat ppsu strev_a
```

\*Rename wtfa\_a to newwt to match the variable name for sampling weight when combining files

```
gen newwt=.
```

```
replace newwt=wtfa_a
```

```
drop wtfa_a
```

```
save newadult19
```

\*Create temporary subsets of 2020 data by selecting household ID (HHX), weight, analytic variables and variance structures (PSTRAT and PPSU)

\*Adult20 – 2020 Sample Adult file

```
use adult20
```

```
keep hhx pstrat ppsu STREV_A
```

\*Rename hhx to nhhx to match with household ID in 2020 Sample Adult partial file;

```
gen nhhx=""
```

```
replace nhhx=hhx
```

```
save tempadult20
```

\*Adultpart20 – 2020 Sample Adult partial file

```
use adultpart20
```

\*Rename hhx\_2020 to nhhx to match with household ID in 2020 Sample Adult file

```
gen nhhx=""
```

```
replace nhhx=HHX_2020
```

```
keep HHX_2020 WTSA_P nhhx
```

```
save temppart20
```

\*Sort each temporary dataset by the merge variable use tempadult20

```
sort nhhx
```

```
save temppart20, replace
```

```
use tempadult20
```

```
sort nhhx
```

```
save tempadult20, replace
```

\*Merge the two temporary 2020 Sample Adult files use tempadult20

```
merge 1:1 nhhx using temppart20
```

\*Keep records with a value in the partial weight

```
keep if WTSA_P !=.
```

\*Rename WTSA\_P to newwt to match the variable name for sampling weight when combining files

```
gen newwt=.
```

```
replace newwt=W TSA_P
drop nhhx
save newadult20
```

```
*Concatenate 2019 and 2020 datasets
append using newadult19
keep pstrat ppsu strev_a newwt
save newadult19_20
```

```
*Create a temporary subset of 2021 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU)
```

```
use adult21
keep WTFA_A pstrat ppsu STREV_A
*Rename wtfa_a to newt to match the variable name for sampling weight when combining files
gen newwt=.
replace newwt=WTFA_A
drop WTFA_A
save newadult21
```

```
*Concatenate 2019, 2020 and 2021 datasets
append using newadult19_20
keep pstrat ppsu STREV_A newwt
save newadult19_20_21
```

```
*Create a temporary subset of 2022 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU)
```

```
use adult22
keep WTFA_A pstrat ppsu STREV_A
*Rename wtfa_a to newt to match the variable name for sampling weight when combining files
gen newwt=.
replace newwt=WTFA_A
drop WTFA_A
save newadult22
```

```
*Concatenate 2019, 2020, 2021 and 2022 datasets
append using newadult19_20_21
```

```
*Divide the new weight by 4 for the four years of data being combined. Otherwise, weighted estimates of totals will be higher than the estimated total U.S. civilian noninstitutionalized population
```

```
gen newwt_adj=.
replace newwt_adj=newwt/4
keep pstrat ppsu STREV_A newwt_adj
save newadult19_20_21_22
```

### *Example using R*

```
#Create a temporary subset of 2019 data by selecting weight, analytic variables, and variance structures (PSTRAT and PPSU)
#renaming option uses library (dplyr)
```

```

NEWADULT19 <- subset(ADULT19, select = c(WTFA_A, PSTRAT, PPSU, STREV_A))
NEWADULT19$NEWWT <- NEWADULT19$WTFA_A

#Create temporary subsets of 2020 data by selecting household ID (HHX), weight, analytic variables and
variance structures (PSTRAT and PPSU)
TEMPADULT20 <- subset(ADULT20, select = c(HHX, PSTRAT, PPSU, STREV_A))
TEMPADULT20$NHHX <- TEMPADULT20$HHX

TEMPPART20 <- subset(ADULTPART20, select = c(HHX_2020, WTSA_P))
TEMPPART20$NHHX <- TEMPPART20$HHX_2020

#Sort each temporary dataset by the merge variable
TEMPADULT20 <- TEMPADULT20[order(TEMPADULT20$NHHX), ]
TEMPPART20 <- TEMPPART20[order(TEMPPART20$NHHX), ]

#Merge the two temporary 2020 Sample Adult files
NEWADULT20 <- merge(TEMPADULT20, TEMPPART20, by = "NHHX")
NEWADULT20 <- subset(NEWADULT20, WTSA_P > 0)
NEWADULT20$NEWWT <- NEWADULT20$WTSA_P
NEWADULT20 <- NEWADULT20 %>% rename("WTFA_A"="WTSA_P")
NEWADULT20 <- subset(NEWADULT20,select = c(WTFA_A, PSTRAT, PPSU, STREV_A, NEWWT))

#Create a temporary subset of 2021 data by selecting weight, analytic variables, and variance structures
(PSTRAT and PPSU)
NEWADULT21 <- subset(ADULT21, select = c(WTFA_A, PSTRAT, PPSU, STREV_A))
NEWADULT21$NEWWT <- NEWADULT21$WTFA_A

#Create a temporary subset of 2022 data by selecting weight, analytic variables, and variance structures
(PSTRAT and PPSU)
NEWADULT22 <- subset(ADULT22, select = c(WTFA_A, PSTRAT, PPSU, STREV_A))
NEWADULT22$NEWWT <- NEWADULT22$WTFA_A

#Concatenate temporary 2019, 2020, and 2021 datasets
NEWADULT19_20_21_22 <- rbind(NEWADULT19, NEWADULT20, NEWADULT21, NEWADULT22)
NEWADULT19_20_21_22$NEWWT_ADJ <- NEWADULT19_20_21_22$NEWWT / 4
NEWADULT19_20_21_22 <- subset(NEWADULT19_20_21_22, select = c(PSTRAT, PPSU,
STREV_A,NEWWT_ADJ))

```

### *Merging Survey Data and Paradata Files*

Data users can merge the Paradata file with the Sample Adult file (or the Sample Child file) to explore associations between a wide range of methodological measures and survey data. To merge 2022 files, use variable HHX as the unique identifier between the two files. Sample code in SAS, STATA and R is provided below to illustrate merging the Sample adult file with the Paradata file.

Note that data files should be merged within the same year before combining (pooling) data files for multiple years.

*Example using SAS*

```
PROC SORT DATA=ADULT22;
BY HHX;
RUN;
PROC SORT DATA=PARADATA22;
BY HHX;
RUN;

/* creates a new file with Sample Adult and Paradata variables for each household */
DATA ADULT22_PLUS_PARA;
MERGE ADULT22 PARADATA22; BY HHX;
RUN;
```

*Example using STATA*

```
*Change working directory to location of data files cd c:\nhis2022\

use adult22
sort hhx
save adult22, replace

use paradata22
sort hhx
save paradata22, replace

merge 1:1 hhx using adult22
save adultpara22
* Creates a new file with Sample Adult and Paradata variables for each household
```

*Example using R*

```
library(dplyr)

#Reading data from ADULT22 and sorting it by HHX
ADULT22 <- read.csv("ADULT22.csv")
ADULT22 <- arrange(ADULT22, HHX)

#Reading data from PARADATA22 and sorting it by HHX
PARADATA22 <- read.csv("PARADATA22.csv")
PARADATA22 <- arrange(PARADATA22, HHX)

#Merging the sorted datasets by HHX and creating a new dataset ADULT22_PLUS_PARA
ADULT22_PLUS_PARA <- merge(ADULT22, PARADATA22, by = "HHX")
```

*Merging Survey Data and Imputed Income Files*

Data users can merge the Sample Adult file (or the Sample Child file) with their respective Imputed Income file to apply the imputation variable to the analyses for the appropriate calculation of standard error of the imputed variable. Variable HHX is the unique identifier between the two files.

**Table 7. Variables in the imputed income example.**

Survey question	Original variable name	Original values	Recoded variable name	Recoded values
<b>Imputation number</b>	IMPNUM_A	Values 01-10	(not recoded)	(not recoded)
<b>Randomly assigned household number unique to household</b>	HHX	Range of unique of values	(not recoded)	(not recoded)
<b>Would you say your health in general is excellent, very good, good, fair, or poor?</b>	PHSTAT_A	1. Excellent 2. Very good 3. Good 4. Fair 5. Poor 7. Refused 8. Not ascertained 9. Don't know	HEALTH	1. Excellent, very good or good 0. Fair or poor . (missing)
<b>Was the last doctor's visit a wellness visit, physical, or general-purpose check-up?</b>	WELLNESS_A	1. Yes 2. No 7. Refused 8. Not ascertained 9. Don't know	WELLCHK	1. Yes 2. No . (missing)
<b>Ratio of income to poverty threshold</b>	RATCAT_A	1. 0.00–0.49 2. 0.50–0.74 3. 0.75–0.99 4. 1.00–1.24 5. 1.25–1.49 6. 1.50–1.74 7. 1.75–1.99 8. 2.00–2.49 9. 2.50–2.99 10. 3.00–3.49 11. 3.50–3.99 12. 4.00–4.49 13. 4.50–4.99 14. 5.00 or greater	(not recoded)	(not recoded)

The following code (in SAS, SUDAAN, STATA, and R) is for merging the Sample Adult data file and the Sample Adult imputed income file and conducting an analytic procedure (i.e., logistic regression using survey data). The variables used in this example were recoded as illustrated in Table 7. The analytic example, using the 2022

Sample Adult data file and the Sample Adult imputed income file, will examine the effect of the variables RATCAT\_A (the ratio of family income to the poverty threshold [*imputed*]), and WELLCHK (last doctor's visit was a wellness visit, recoded from WELLNESS\_A) on HEALTH (having good-to-excellent health, recoded from PHSTAT\_A).

### Example using SAS

In SAS, analysis of multiple imputed data is conducted in two stages:

1. Analysis: each of the M imputed datasets is analyzed separately using any method that would have been selected had there been a single complete dataset. This includes analytical procedure in SAS, such as PROC GLM, PROC MIXED, PROC LOGISTIC, PROC FREQ, etc. In SAS, analysis of multiply imputed data is invoked with a "BY \_IMPUTATION\_" statement, to indicate that the same analysis is performed within each of the imputed datasets. Users need to rename the NHIS imputation number identifier IMPNUM\_A to \_IMPUTATION.
2. Pooling: analysis results from M imputed datasets obtained from step 1 are combined into one overall result. This step can be carried out using SAS PROC MIANALYZE.

Import data files into SAS. See SAS input statements provided on the 2022 NHIS Data Release page. This example uses the libname 'NHIS.'

*/\*The sample code below illustrates renaming the IMPNUM\_A variable to \_IMPUTATION\_ for analyses in a new SAS dataset. \*/*

```
DATA IMPINC;
SET NHIS.ADULTINC22;
RENAME IMPNUM_A= _IMPUTATION_ ; *SAS identifies imputed datasets by imputation_ ; RUN;
```

*/\* Next, merge the Sample Adult file and Sample adult imputed Income file. Data files must be sorted by the common ID before they can be merged\*/*

```
PROC SORT DATA= IMPINC; BY HHX;
RUN;
```

```
PROC SORT DATA=NHIS.ADULT22 OUT=ADULT22; BY HHX;
RUN;
```

```
DATA NHIS22;
MERGE ADULT22 (IN=A) IMPINC; *Merging the imputed income and the main dataset;
BY HHX;
IF A;
RUN;
```

*/\* Sort the new dataset by imputation prior to analysis. Otherwise, your analyses will only show the first category of each variable for imputed analyses \*/*

```
PROC SORT DATA= NHIS22;
BY _IMPUTATION_;
```



```
RUN;
```

```
/*The survey analytic procedure (PROC SURVEYLOGISTIC) is used to account for the complex sampling design of NHIS. */
```

```
/*WELLCHK is a recode from WELLNESS_A and HEALTH is a recode from PHSTAT_A*/
```

```
PROC SURVEYLOGISTIC DATA=NHIS22;
```

```
STRATUM PSTRAT; CLUSTER PPSU;
```

```
WEIGHT WTFA_A;
```

```
CLASS RATCAT_A WELLCHK (DESC);
```

```
MODEL HEALTH(EVENT='1') = RATCAT_A WELLCHK;
```

```
ODS OUTPUT PARAMETERESTIMATES=lgsparms ODDSRATIOS=lgsodds; BY _IMPUTATION_;
```

```
RUN;
```

```
/*The ODS datasets from the code above will contain a set of estimates for each imputed dataset identified by the variable _imputation_ included in each of them.
```

```
The MIANALYZE procedure combines the results of the analyses of imputed data and generates valid statistical inferences.*/
```

```
PROC MIANALYZE PARMS(CLASSVAR=CLASSVAL)=lgsparms; CLASS RATCAT_A WELLCHK;
```

```
MODELEFFECTS RATCAT_A WELLCHK;
```

```
ODS OUTPUT PARAMETERESTIMATES=mian_lgsparms; *Combines the results of previous analyses;
```

```
RUN;
```

```
/*The mian_lgsparms output shows parameters from the pooled imputed datasets.*/
```

### *Example using SAS-callable SUDAAN*

SUDAAN reads in separate imputed datasets. To conduct analyses in SAS-callable SUDAAN, the following steps are taken:

1. Separate the multiply imputed NHIS.ADULTINC22 SAS data set into 10 individual imputed income datasets impinc1-impinc10.
2. Merge each imputed dataset with the main NHIS data file.

Import data files into SAS. See SAS input statements provided on the 2022 NHIS Data Release page. This example uses the libname 'NHIS.'

```
PROC SORT DATA=nhis.ADULT22; *Sorting by HHX; BY HHX;
```

```
RUN;
```

```
/* The following macro creates 10 separate imputation datasets from the NHIS.ADULTINC21 multiply imputed file. It then sorts them by the merge variable HHX and merges each with the NHIS.ADULT22 dataset. Ultimately 10 separate datasets are created for imputed analyses in SUDAAN */
```

```

%MACRO SEPARATE;
%DO I= 1 %TO 10; *Instructs SAS to do the procedure for 10 iterations;
DATA IMPINC&I; *CREATING 10 SEPARATE IMPUTED DATASETS;
SET NHIS.ADULTINC22;
WHERE IMPNUM_A= &I;
RUN;

PROC SORT DATA= IMPINC&I; *Sorting the 10 imputed datasets by HHX;
BY HHX;
RUN;
DATA NHIS22_&I; *Creating 10 separate analytic datasets;
MERGE NHIS.ADULT22 IMPINC&I;
BY HHX;

/* SUDAAN requires analytic datasets be sorted by the design/nest variables. These variables are PSTRAT
and PPSU on the NHIS. */
PROC SORT DATA= NHIS22_&I; *SORTING THE IMPUTED DATASETS BY DESIGN VARIABLES;
BY PSTRAT PPSU;
RUN;

%END;
%MEND;
%SEPARATE;

/*In SUDAAN, the option MI_COUNT indicates use of multiple imputed datasets.
The associated numeral indicates the number of imputed datasets to be used in the analysis. Note that the
data name used after the DATA= is the name of the first imputed dataset. This dataset name ends with
the number 1. Variable HEALTH in this example is coded 0/1 */

PROC RLOGIST DATA = NHIS22_1 FILETYPE=SAS DESIGN=WR MI_COUNT=10 ;
NEST PSTRAT PPSU / MISSUNIT;
WEIGHT WTFA_A;

SUBGROUP WELLCHK ;
LEVELS 2 ;

REFLEVEL WELLCHK=1 ;

MODEL HEALTH= RATCAT_A WELLCHK ;
EFFECTS WELLCHK= (2 -1)/EXP NAME="WELLCHK Yes vs No" ;

PRINT / betas=default risk=default tests=default expcntrst=default t_betafmt=f7.2 waldffmt=f8.2
dffmt=f10.0 orfmt=f5.2 loworfmt=f5.2 uporfmt=f5.2 exp_cntrstfmt=f13.2 low_cntrstfmt=f5.2
up_cntrstfmt=f5.2;
SETENV COLWIDTH=15 DECWIDTH=4 LABWIDTH=25 COLSPCE=1 TOPMGN=0;
RLABEL HEALTH="In good-excellent health";
RTITLE "Using LOGISTIC to Model good-excellent health";
RUN;

```

### Example using STATA

Import data files into Stata. See the sample Stata *.do* statements provided on the 2021 NHIS Data Release page.

```
*Change working directory to location of data files cd c:\nhis2022\  
  
use adult22  
sort hhx  
save nhis22  
  
use adultinc22  
sort hhx  
save impinc  
  
use nhis22  
merge 1:m hhx using "impinc"  
append using "nhis22"  
  
* Rename the NHIS imputation number identifier impnum_a to _mi_m  
  
replace impnum_a=0 if impnum_a==.  
save nhis22_mi, replace  
  
*set data to mi svyset  
  
mi import flong, m(impnum_a) id(hhx)  
  
*set dataset to multiple imputation for survey  
  
mi svyset [pweight=wtfa_a], strat(pstrat) psu(ppsu) singleunit(centered)  
save nhis22_mi_dat  
  
*mi describe will list the registration status of the variables.  
*mi varying will report the varying*and super-varying variables.  
*Verify that all varying variables are registered as imputed or passive.  
  
mi describe mi varying  
* Logistic regression. The dependent variable, health, is coded as 0 or 1.  
mi estimate: svy: logistic health ratcat_a wellchk  
*odds ratios  
mi estimate, or: svy: logistic health ratcat_a wellchk
```

### Example using R

```
#Import the data and rename the variable  
library(dplyr)  
IMPINC <- ADULTINC22 %>% rename(IMPUTATION = IMPNUM_A)
```

```
#Recode Variables
ADULT22$HEALTH <- ifelse(ADULT22$PHSTAT_A %in% c(1, 2, 3), 1,
  ifelse(ADULT22$PHSTAT_A %in% c(3, 4), 0, NA))

ADULT22$WELLCHK <- ifelse(ADULT22$WELLNESS_A %in% c(1,2), ADULT22$WELLNESS_A, NA)

#Sort the data by HHX
IMPINC <- IMPINC[order(IMPINC$HHX),]
ADULT22 <- ADULT22[order(ADULT22$HHX),]

#Merge the data by HHX
NHIS22 <- merge(ADULT22, IMPINC, by = "HHX", all.x = TRUE)

#Sort the data by IMPUTATION
NHIS22 <- NHIS22[order(NHIS22$IMPUTATION),]
```

### *Appending Sample Adult and Sample Child Files*

The 2022 Sample Adult and Sample Child can be appended to one another (i.e., add observations from different persons) to facilitate the analysis of measures that are common to both adults and children. An example of the need to combine observations or concatenate the Sample Adult and Sample Child files would be if the user is interested in generating an estimate of the U.S. civilian noninstitutional population of both children and adults or any subset of age ranges that includes both children and adults (e.g., ages 0 to 64 years).

To do so, data from the Sample Adult file and the Sample Child file should have comparable measures available. Since the names of the Sample Adult variables end in “\_A” and the names of the Sample Child variables end in “\_C”, comparable measures should be renamed to a common variable name. The Sample Adult and Sample Child weights should also be renamed to have the same variable name. The variance estimation variables have the same name for both the Sample Adult and Sample Child files, and no additional recoding and renaming is needed.

The following code illustrates the concatenation of the Sample Adult and Sample Child files for the purpose of generating an estimate of the U.S. civilian noninstitutional population who are uninsured, by age. The example code illustrates keeping variables of interest for the analysis and recoding them in order to generate the estimate of interest. The code does not include analytic procedures for generating the estimate. The variables used in this example are illustrated in Table 8.

**Table 8. Variables in example concatenating the Sample Adult file and the Sample Child**

Variable description	Variable name in the Sample Adult file	Variable name in the Sample Child file	Variable name in new file
Indicates person is the Sample Adult; Indicates person is the Sample Child	HHSTAT_A	HHSTAT_C	HHSTAT_A and HHSTAT_C
Pseudo-stratum for public-use file variance estimation	PSTRAT	PSTRAT	PSTRAT
Pseudo-PSU for public-use file variance estimation	PPSU	PPSU	PPSU
Weight - Final Annual	WTFA_A	WTFA_C	WTFA_NEW
Coverage status as used in Health United States	NOTCOV_A	NOTCOV_C	NOTCOV
Age of Sample Adult (top coded); Age of Sample Child	AGEP_A	AGEP_C	AGE

*Example using SAS*

Create a new file with all Sample Adult and all Sample Child records. Keep variance and sample weights, common variables in both files and new recodes combining key variables.

```

DATA PERSON22;
SET ADULT22 CHILD22;
KEEP PSTRAT PPSU WTFA_A WTFA_C HHSTAT_A HHSTAT_C NOTCOV_A AGEP_A NOTCOV_C AGEP_C
WTFA_NEW NOTCOV AGE;

/* recodes*/
IF HHSTAT_A=1 THEN DO;
WTFA_NEW=WTFA_A;
NOTCOV=NOTCOV_A;
AGE = AGEP_A;
END;

IF HHSTAT_C=1 THEN DO;
WTFA_NEW=WTFA_C;
NOTCOV=NOTCOV_C;
AGE = AGEP_C;
END;.
RUN;

```

*Example using STATA*

Create new separate files for the Sample Adult and Sample Child with the variables for variance, sample weight, and common variables of interest. Combine files and recode combining key variables.

```

use child22
keep NOTCOV_C HHSTAT_C ppsu pstrat WTFA_C AGEP_C
save childvars

use adult22
keep NOTCOV_A HHSTAT_A ppsu pstrat WTFA_A AGEP_A
save adultvars

append using childvars

*Recodes gen notcov=.
replace notcov=1 if NOTCOV_C==1 | NOTCOV_A==1
replace notcov=2 if NOTCOV_C==2 | NOTCOV_A==2

gen age=.
replace age=AGEP_C if HHSTAT_C==1
replace age=AGEP_A if HHSTAT_A==1

gen WTFA_new=.
replace WTFA_new= WTFA_C if HHSTAT_C==1
replace WTFA_new= WTFA_A if HHSTAT_A==1
save vars_child_adult

```

### Example using R

```

# read in data from ADULT22 and CHILD22
PERSON22 <- merge(ADULT22, CHILD22, all=TRUE)

# keep the desired variables
PERSON22 <- subset(PERSON22, select = c("PSTRAT", "PPSU", "WTFA_A", "WTFA_C",
    "HHSTAT_A", "HHSTAT_C", "NOTCOV_A",
    "AGEP_A", "NOTCOV_C", "AGEP_C"))

# create new variables based on the HHSTAT_A and HHSTAT_C variables
PERSON22$NOTCOV <- ifelse(PERSON22$HHSTAT_A == 1, PERSON22$NOTCOV_A, PERSON22$NOTCOV_C)
PERSON22$AGE <- ifelse(PERSON22$HHSTAT_A == 1, PERSON22$AGEP_A, PERSON22$AGEP_C)
PERSON22$WTFA_NEW <- ifelse(PERSON22$HHSTAT_A == 1, PERSON22$WTFA_A, PERSON22$WTFA_C)

# keep the desired variables
PERSON22 <- subset(PERSON22, select = c("PSTRAT", "PPSU", "WTFA_NEW", "NOTCOV", "AGE"))

```

### *Variance Estimation for Subsetted Data Analysis*

Frequently, analyses using NHIS data are restricted to specific population subgroups (e.g., persons aged 65 and older). NCHS recommends that subpopulation analyses be carried out using the full data file and the SUBPOPN statement in SUDAAN, or an equivalent procedure with another complex design variance estimation software package.

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, often the sample design structure available to the software is incomplete; subsetting data deletes important design information needed for variance estimation.

Analyses of large NHIS subgroups usually produce reliable estimates, but analyses of small subgroups may yield unreliable estimates, as indicated by their larger variances. The analyst should pay 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.

Below are examples for subsetting NHIS data using SUDAAN, Stata, SPSS, SAS, and R software packages for illustrative purposes. The following code is to subset the second category for each variable RACEALLP\_A and SEX\_A, which happens to be the value “2” in both cases in this example. These are Sample Adult variables for race and sex where RACEALLP\_A=2 is Black or African American only and SEX\_A=2 is female.

### *Example using SUDAAN*

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.

#### Strategy 1 (recommended)

Use the SUBPOPN statement with the SUDAAN method described above for the full Sample Adult dataset:

```
PROC .DESIGN = WR;
  NEST PSTRAT PPSU;
  WEIGHT WTFA _SA;
  SUBGROUP (variable names);
  LEVELS ... ;
  SUBPOPN RACEALLP_A=2 & SEX_A=2 / NAME="Analysis of Black or African American women;"
```

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

Strategy 2 (not recommended, except when Strategy 1 is infeasible)

Use the MISSUNIT option on the NEST statement with the method described above for subsetted data:  
NEST PSTRAT PPSU / 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.

### *Example using STATA*

```
Stata svy
Add SUBPOP to the SVY statement, e.g.:
svy, subpop( raceallp_a==2 & sex_a==2 ): mean <name of variable to be analyzed>
```

### *Example using SPSS*

```
SPSS csdescriptives or cstabulate
One must first define an indicator variable, e.g.:
DO IF (RACEALLP_A EQ 2 AND SEX_A EQ 2).
COMPUTE SUBGRP=1.
ELSE.
COMPUTE SUBGRP=0.
END IF.
```

And then refer to the indicator variable in csdescriptives or cstabulate, e.g.: CSDESCRIPTIVES (or CSTABULATE)  
/SUBPOP TABLE=SUBGRP

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

### *Example using SAS*

```
SAS proc surveymeans or surveyfreq
One must first define an indicator variable, e.g.:
IF RACEALLP_A=2 & SEX_A=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>;
```

This will produce tables for all values of the SUBGRP variable. As with SPSS, it is very important that the indicator variable is defined for all data records. Otherwise an invalid result can occur.

### *Example using R*

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 the syntax that follows specifies the subgroup of interest without using an equality test.

```
# subset for raceallp_a=2 & sex_a=2 without using equal signs  
subgrp <- subset(nhissvy, RACEALLP_A>1 & RACEALLP_A<3 & SEX_A>1svymean(~<name of variable to be  
analyzed>,design=subgrp)
```

Note that users may want to recode variables such that missing values (which have numeric codes greater than 1) are not treated as real values. For example, `sex>1` would include missing codes 7, 8 and 9 (don’t know, refused, not ascertained respectively)

## Questionnaire and Codebook Section Acronyms

Table 9. Acronym definition of sections in the questionnaire and codebooks: 2022 NHIS

Section	Section Description
ABH	Access Barriers to Care
ADO	Age of Disability Onset
ALC	Alcohol Use
ANX	Anxiety
AQS	Advice to Quit Smoking
AST	Asthma
BEH	Behavior
BLY	Bullying
BMI	Current pregnancy, height, weight
BMI	Height and Weight (Sample Child)
BSC	Baby Pediatric Symptom Checklist
CAN	Cancer
CFS	Chronic Fatigue Section
CHL	Cholesterol
CIV	Civic Engagement
CIG	Cigarettes and E-cigarettes w/ Cig. History and Smoking Quit Methods
CIH	Complementary and Integrative Health
COG	Cognition
COM	Communication
CON	Other Chronic Conditions
CPA	Content of Care - Physical Activity
CVC	Cardiovascular Conditions
CVL	Long-COVID
CVV	COVID-19 vaccination
DEP	Depression
DIB	Diabetes
DLD	Developmental and Learning Disabilities
DNC	Dental Care
DNU	Diet and Nutrition
EMP	Employment
EPI	Epilepsy
FDS	Food Security
FEM	Employment of family members
FGE	Fatigue
FLG	Flags
FOO	Food Related Programs
FRT	Front
GAD	GAD-7 Anxiety

**Table 9 continued. Acronym definition of sections in the questionnaire and codebooks: 2022 NHIS**

<b>Section</b>	<b>Section Description</b>
<b>GEN</b>	General
<b>GNI</b>	Gender Identity
<b>HEA</b>	Hearing
<b>HHC</b>	Household Composition
<b>HIS</b>	Health Status
<b>HIT</b>	Internet access and health information technology
<b>HOU</b>	Housing
<b>HPV</b>	HPV Vaccination
<b>HYP</b>	Hypertension
<b>IDN</b>	Identifier
<b>IMM</b>	Immunization
<b>IMS</b>	Immunization with supplements
<b>INC</b>	Family Income
<b>INS</b>	Health Insurance
<b>ISN</b>	Immunosuppression
<b>LNK</b>	Linkage
<b>LS1</b>	Satisfaction with Life
<b>LSF</b>	Life Satisfaction
<b>MAR</b>	Marital Status
<b>MHC</b>	Mental Health Care
<b>MOB</b>	Mobility
<b>NAT</b>	Nativity
<b>NHC</b>	Neighborhood Characteristics
<b>ORN</b>	Sexual Orientation
<b>OTB</b>	Other Tobacco
<b>PAR</b>	Parent Demographics
<b>PAY</b>	Difficulty Paying for Health Care
<b>PHQ</b>	PHQ-8 Depression
<b>PHY</b>	Physical Activity
<b>PMD</b>	Prescription Medication
<b>PTC</b>	Physical and other therapeutic care
<b>SCH</b>	Schooling
<b>SCR</b>	Screen time
<b>SDQ</b>	Child Mental Health - SDQ Questionnaire
<b>SLE</b>	Stressful Life Events
<b>SLP</b>	Sleep
<b>SOC</b>	Social Functioning
<b>SOS</b>	Social Support
<b>TBH</b>	Transportation Barrier to Care
<b>TBI</b>	Concussions - lifetime

**Table 9 continued. Acronym definition of sections in the questionnaire and codebooks: 2022 NHIS**

<b>Section</b>	<b>Section Description</b>
<b>TEL</b>	Telephone Use
<b>TLH</b>	Telehealth
<b>UCF</b>	Unit Control File
<b>UPP</b>	Self-care and Upper Body (Sample Child)
<b>UPP</b>	Upper Body, Motor skills and self-care (Sample Adult)
<b>UTZ</b>	Utilization
<b>VET</b>	Veterans Status
<b>VFY</b>	Verification
<b>VIS</b>	Vision
<b>VOL</b>	Volunteer Activities
<b>VSL</b>	Voice, Swallowing, Speech, and Language Communication Disorders
<b>WLK</b>	Walking

## Sample Adult's Health

### I. Health Status and Conditions

#### Annual Core

Several sections throughout the Sample Adult module measure the health of U.S. adults. Sample Adults were asked to self-report their height and weight, self-perceived health status, current pregnancy status for females aged 18–49 years, and whether a doctor or other health care professional had told them that they had series of selected conditions. Estimates derived from questions that ask about specific health conditions diagnosed by a doctor or health care professional may underestimate the true burden of these conditions in the population due to the undiagnosed status of the condition during its detectable pre-clinical and clinical phase and from reporting bias. For a list of health conditions measured in the annual core, by questionnaire section and reference periods of its respective questions, see Table 10. Sample Adults were also asked about the intake of medication to treat diabetes, hypertension and high cholesterol, type of diabetes, and visits to an emergency room due to asthma, if ever diagnosed with these respective conditions. Age of diagnosis was collected from Sample Adults ever diagnosed with diabetes and cancers.

**Table 10. Annual core content measures of health conditions about the Sample Adult, by questionnaire section and reference periods.**

Topic	Section	Reference period in available questions
Angina pectoris	CVC	Ever
Anxiety disorder	CON	Ever
Arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia	CON	Ever
Asthma	AST	Ever; Past 12 months; Current
Cancer and cancer kind	CAN	Ever
Chronic Obstructive Pulmonary Disease	CON	Ever
Coronary heart disease	CVC	Ever
Dementia, including Alzheimer's disease	CON	Ever
Depression	CON	Ever
Diabetes	DIB	Ever
Gestational diabetes	DIB	Ever
Health status	HIS	Current
Height and weight	BMI	Current; If pregnant, before pregnancy
Hyperlipidemia	CHL	Ever; Past 12 months
Hypertension	HYP	Ever; Past 12 months
Myocardial infarction	CVC	Ever
Pre-diabetes	DIB	Ever
Pregnancy status	BMI	Current
Stroke	CVC	Ever

### *Rotating Content*

**Fatigue symptoms.** Questions on fatigue in the FGE section, previously fielded in 2020, are part of the rotating content fielded every two years. Sample adults were asked how often they feel very tired or exhausted in the past three months, how long the feeling lasted, and the level of tiredness they felt. The questions can also be used as part of the fatigue domain in the WG extended set of disability identifiers and can be analyzed as part of the WG Extended Set on Functioning (see Table 11).

### *Emerging Content*

**COVID-19 infection.** Since July 2020, the Sample Adult module includes questions about positive COVID-19 diagnosis and testing and perceived symptom severity. In 2022, two additional questions were added to measure Long COVID or post-COVID conditions (PCC). Questions about whether the symptoms lasted for three or more months and whether they currently have symptoms were asked of adults who experience mild to severe COVID-19 symptoms, or the Sample Adult respondent refused or didn't know the symptom severity. According to the Department of Health and Human Services, Long COVID is defined as signs, symptoms and conditions that develop or continue after initial COVID-19 infection, are present after four weeks or more after the initial infection phrase, may be multi systemic, relapsing–remitting and progressing over time (U.S. Department of Health and Human Services, 2022).

These questions about COVID-19 are in the CVL section in 2022. In 2020–2021, the questions on COVID–19 diagnosis and testing were fielded in the CVD section, and testing was asked as a two-part-question.

### *Sponsored Content*

**Chronic Fatigue.** NCEZID sponsored two questions in the CFS section to provide national estimates for lifetime and current prevalence of chronic fatigue syndrome. Sample adults were asked about whether a doctor or other health professional had told them that they had chronic fatigue syndrome (CFS) or Myalgic Encephalomyelitis (ME), and if they currently have it. In 2021, these questions were included in the RCN section. There are few national population-based prevalence estimates of CFS/ME. Additionally, the COVID-19 pandemic has increased the need for national prevalence estimates because post-infection CFS/ME symptoms have been reported to be associated with COVID-19 (Bansal, Gubbi and Koch, 2022). Estimates of lifetime and current CFS/ME can provide data to assist in planning for the care of individuals living with this condition.

**Diabetes (insulin).** NIH/NIDDK and NCCDPHP sponsored three questions in the Sample Adult DIB section. Sample Adults who reported having been told by a doctor or health professional that they had diabetes and are now taking insulin were asked how long after diagnosis they began taking insulin, whether it was discontinued for more than 6 months after initiating it, and whether this occurred during the first year of diabetes diagnosis.

**Epilepsy.** NCCDPHP sponsored four questions on epilepsy in the Sample Adult EPI section. Sample Adults were asked whether they had ever been told by a doctor or health professional that they had a seizure disorder or epilepsy, and those diagnosed with the disorder were also asked, whether they were currently taking antiseizure medication, the number of seizures in the past year, and whether they saw a neurologist or epilepsy specialist in the past year. These questions are fielded during 2021-2022 to provide national and subpopulation estimates of epilepsy prevalence.

**Immunosuppression.** In 2022, NCIRD sponsored two questions about immunosuppression status in the Sample Adult ISN section. Sample Adults were asked whether a doctor or other health professional had told them that their prescription medication or any medical treatments in the past 12 months would weaken the immune system and if they currently had a condition that weakens the immune system. These questions began fielding in July 2020 as emerging content in response to the COVID-19 pandemic. Persons with an immunocompromised state are more likely to get severely ill from COVID-19 (CDC, 2023)

**Communication Disorders.** In 2022, NIDCD sponsored content in the VSL section related to sensory or communication disorders to help track [Healthy People 2030 Objectives](#) and to update national estimates. Sample Adults were asked about difficulty in the past 12 months with voice (i.e., hoarse, strained voice, or trouble speaking loud enough), swallowing, speech (i.e., stuttering, trouble pronouncing words), and language (i.e., using, learning, or understanding words). For each of the communication disorder experienced, follow-up questions asked for the number of days they had this and the level of difficulty experienced. An additional question asked those with a speech problem if this included stuttering, trouble pronouncing words correctly, or both. Sample adults who experienced any of the communication problems were also asked if these difficulties followed a brain injury or stroke, if they had seen a speech-language pathologist or specialist, and if the disorder had improved or gotten worse in the past 12 months. Those who had seen a speech-language pathologist or specialist were also asked how the health care services received had affected their lives. Similar content is also available for the Sample Child.

## Major Recodes

**Age of diabetes diagnosis.** The age when first told by a doctor or health professional that the Sample Adult had diabetes is top-coded at age 85.

**Age of cancer diagnosis.** The age when first told by a doctor or health professional that the Sample Adult had a specific kind of cancer is available for 27 of the 29 different kinds of cancers and 'other' kind of cancer collected in NHIS. To protect confidentiality, Sample Adults mentioning kidney or testicular cancer were suppressed along with the age of diagnosis for these specific cancers. The age of cancer diagnosis for those mentioning kidney and testicular cancer are included in the recode for age of diagnosis for 'other' kind of cancer. The age of cancer diagnosis was also calculated for combined cancer types colorectal (which combines colon and rectal cancer) and for head and neck cancers (which combines larynx-tracheal, mouth/tongue/lip, and throat (pharyngeal) cancers). The youngest age provided for the respective cancer type was assigned for Sample Adults with combined cancers, those who identified the same kind of cancer as the second or third kind of cancer, or mentioned 'other kind of cancer' more than once. Sample Adults who reported 'don't know' or 'refused' when asked about the type of cancer but provided an age of diagnoses for the unidentified cancer are included in the variable for age of cancer diagnosis for 'other kind of cancer.' Variables for age of cancer diagnosis are top-

coded at age 85 for all cancers, and bottom-coded at age 18 for cancers of the breast, cervix, prostate, and skin (nonmelanoma, melanoma and unknown type).

**Body Mass Index (BMI).** A categorical measure of BMI was created using unrestricted height and weight values which contain the greater range of height and weight values than are available on the public-use file. BMI was calculated using the formula:  $BMI = \text{kilograms} / \text{meters}^2$  where 1 kilogram = 2.20462 pounds and 1 meter = 39.37008 inches. The categorical measure of BMI was classified as follows: underweight is  $BMI < 18.5$ ; healthy weight is  $BMI 18.5$  to  $< 25$ ; overweight is  $BMI \geq 25$  to  $< 30$ ; and obese is  $BMI \geq 30$ . Sample Adults who answered don't know or refused for height or weight, reported values for either height or weight that were outside the limits for public data release, or for whom height or weight values for public release were set to 96 and 996 due to missing sex information were coded as 99 in the categorical BMI measure.

**Cancer types.** Sample Adults could name up to three kinds of cancer. Twenty-nine kinds of cancer could be identified in NHIS, with the option for 'other' not listed. Responses from the up to three kinds of cancers were recoded during editing into "mentioned"/ "not mentioned" variables for each cancer type. Due to confidentiality concerns, recodes that identified whether kidney cancer or testicular cancer were mentioned were suppressed, and Sample Adults reporting these cancers were included in the recode 'other cancer type' mentioned. Recodes were also created that combine specific kinds of cancers available in the public-use file. Sample Adults reporting colon or rectal cancer were combined into a separate variable indicating whether colorectal cancer was "mentioned"/ "not mentioned." Similarly, Sample Adults reporting larynx-tracheal, mouth/tongue/lip, or throat (pharyngeal) cancer were recoded into a separate variable indicating whether a head and neck cancer was mentioned/not mentioned. Sample Adults who reported having had cancer but reported 'don't know' or 'refused' when asked about the type of cancer are recoded as "don't know" or "refused" in these recodes, respectively. Respondents who reported a combination of "don't know" and "refused" to type of cancers are recoded as "don't know" in these recodes.

**Height and Weight.** Sample Adults had the option to report height and weight using the U.S. customary system (pounds: feet and inches) or the metric system (kilograms: meters and centimeters). Metric responses on height and weight were converted into the U.S. Customary system using the following conversion scale: 1 meter=39.37008 inches and 1 kilogram=2.20462 pounds (lbs.). The public-use height variable reflects total height in inches (e.g., 65" is 5' 5"), with height ranges 63–76 inches for men and 59–70 inches for women. The public-use weight variable reflects total pounds rounded to whole integers with weight ranges 126–299 lbs. for men and 100–274 lbs. for women. Pregnant women aged 18–49 years were asked to report weight before pregnancy. Sample Adults who reported values outside the public-use limits for either height or weight had data for both variables recoded to "96" or "996" ("Not available") on the public-use data file to protect the confidentiality of those who might be identifiable by their unusual physical characteristics. In addition, due to the sex-specific height and weight limits for public data release, all Sample Adults for whom sex was answered as don't know or refused were also coded as 96 for height and 996 for weight. Don't know and refused responses to either height or weight were retained in the height and weight public-use recodes.

**Missed workdays due to asthma.** The number of days that Sample Adults with current asthma were unable to work or get work done around the house in the past 12 months because of their asthma were top-coded at 30 or more days.

**Number of cancers.** A summary recode was created that indicates the number of kinds of cancer mentioned by the Sample Adult. Sample Adults who were never told that they had cancer were assigned the value 0. The number of kinds of cancers were derived from responses to the type of cancer told to have and whether they had a second, third and other (additional) types of cancer (assigned values 1–4, respectively, where 4 indicates four or more). This recode reflects the number of cancers mentioned even when the same kind of cancer was mentioned more than once.



**Number of seizures.** The response categories for the number of seizures in the past 12 months (0, 1, 2-3, 4-10 and more than 10) was top-coded at “4 or more” seizures, and values 7 (refused), 8 (not ascertained), and 9 (don’t know) were retained.

**Years since diabetes diagnosis.** This recode is the calculation of the Sample Adult’s current age (in years and top-coded at 85) minus the age (in years) when first told to have diabetes. This difference may be underestimated for Sample Adults with a top-coded age. The highest number of years since diabetes diagnosis available in the public data release is 84 years. Due to confidentiality concerns, Sample Adults aged 85 and older who were diagnosed at age 85 and over, are coded as 96. This recode updates the recode available in previous years for years since diabetes diagnoses.

## II. Functioning and Disability

### *Annual Core*

**Functioning and Disability.** The questions on functioning and disability found in sections VIS, HEA, MOB, COM, COG, UPP, ANX, DEP, PAI, and FGE of the Sample Adult module are part of sets of international standard measures developed, tested and endorsed by the Washington Group on Disability Statistics (WG). The WG is a city group established in 2001 under the United Nations Statistical Commission to address the need for population-based measures of disability by promoting and coordinating international cooperation in the area of health statistics focusing on disability data collection tools suitable for censuses and national surveys. The major objective is to provide necessary information on disability that is comparable throughout the world by identifying individuals with functional limitations in basic actions, regardless of nationality or culture. The questions reflect advances in the conceptualization of disability and use the World Health Organization’s International Classification of Functioning, Disability, and Health (ICF) as a conceptual framework. The intended use of these questions is to describe the functional status of adults and, when used with other questions on the survey, to evaluate whether adults with functional limitations have achieved similar levels of participation and inclusion as adults without functional limitations. These questions do not capture all aspects of difficulty in functioning, but rather focus on domains of functioning that are likely to identify the majority of adults at risk of participation restrictions in an unaccommodating environment.

The questions included for Sample Adults are from the WG Extended Set on Functioning (WG-ES). Two additional question sets developed by the WG are subsets of the WG-ES – the WG Short Set on Functioning (WG-SS) and the WG Short Set on Functioning – Enhanced (WG-SS Enhanced), comprised of 6 and 12 questions, respectively. For a list of questions included in each set and their respective questionnaire sections, see Table 11.

Questions ask about the Sample Adults’ level of difficulty (no difficulty, some difficulty, a lot of difficulty, or cannot do at all) in basic domains of functioning including seeing, hearing, mobility, communication, cognition, self-care, and upper body functioning and about the frequency and intensity of experiencing anxiety and depression, pain, and fatigue. In addition to questions about level of difficulty, several of the domains have questions to provide information on the use of accommodations. More information on the Washington Group and the question sets may be obtained by request to the WG Secretariat at [WG\\_Secretariat@cdc.gov](mailto:WG_Secretariat@cdc.gov) or found on the WG website at: <http://www.washingtongroup-disability.com/>.

**Table 11. Functional limitations included in the Sample Adult module, by topic, section and Washington Group Question Set.**

Topic	Section	WG-SS	WG-SS Enhanced	WG-ES
Wear glasses or contacts	VIS			√
Have difficulty seeing	VIS	√	√	√
Use a hearing aid	HEA			√
How often use hearing aid	HEA			√
Have difficulty hearing	HEA	√	√	√
Difficulty walking or climbing steps	MOB	√	√	√
Use equipment or receive help for getting around	MOB			√
Type of equipment	MOB			√
Difficulty walking 100 yards	MOB			√
Difficulty walking a third of mile	MOB			√
Difficulty walking up or down 12 steps	MOB			√
Difficulty communicating	COM	√	√	√
Difficulty remembering or concentrating	COG	√	√	√
Difficulty remembering, concentrating, or both	COG			√
How often have difficulty remembering	COG			√
Difficulty remembering few things, a lot, everything	COG			√
Difficulty with self-care	UPP	√	√	√
Difficulty raising a 2-liter bottle from waist to eye level	UPP		√	√
Difficulty using hands and fingers	UPP		√	√
How often feel worried, nervous, or anxious	ANX		√	√
Take prescription medication for these feelings	ANX			√
Level of feelings of worried, nervous, anxious	ANX		√	√
How often feel depressed	DEP		√	√
Take prescription medication for these feelings	DEP			√
Level of feelings of depressed	DEP		√	√
How often have pain	PAI			√
How much pain	PAI			√
How often felt very tired or exhausted*	FGE			√
How long tired or exhausted feelings last*	FGE			√
Level of tiredness or exhaustion**	FGE			√

\*Questions about fatigue are part of the rotating core content, rotating every two years starting in 2020.  
NOTE: WG-SS is WG Short Set on Functioning, WG-SS Enhanced is WG Short Set on Functioning – Enhanced, and WG-ES is WG Extended Set on Functioning.

The WG questions can be analyzed separately, by domain, or combined across domains. A disability status indicator is available to data users (See Major Recodes below) that identifies Sample Adults who are at greater risk than the general population for experiencing restrictions in participation because of difficulties doing certain universal, basic actions. This recode classifies Sample Adults with disability as those reporting “have a lot of difficulty” or “cannot do at all” for at least one of the six domains included in the WG Short Set on Functioning. Use of the functioning and disability data should be tailored to meet the needs of the analysis. Other disability indicators can be created from the WG Extended Set on Functioning and the WG Short Set – Enhanced. Consult

the WG website for guidance on the creation of these indicators. Changing the threshold for either the number of domains the respondent identifies having difficulty with or the degree of difficulty can create different identifiers that will capture different populations. For example, a recode that includes respondents who have “some difficulty” with any of the domains will capture a greater proportion of the population than a recode limited to include only those who report they “cannot do at all” to any of the domains. In this example, the functional abilities of the larger group will be much more heterogeneous than that of the smaller group. Analytic guidelines, including recommended disability identifiers, written for each of the WG questions sets may be obtained from the WG website: <https://www.washingtongroup-disability.com/analysis/analysis-overview/>.

A cognitive testing report is available for selected WG questions at <https://www.cdc.gov/QBank/Report.aspx?1216>. For a report that examines differences in survey responses of disability between the set of disability questions from Short Set on Functioning (WG–SS) and set of disability questions developed for the American Community Survey, see, <https://www.cdc.gov/nchs/data/nhsr/nhsr161-508.pdf>.

**Participation.** The SOC section in the Sample Adult module contains three questions about difficulty with participation in everyday life activities that are not part of the WG questions on functioning. These questions ask the Sample Adult whether, because of a physical, mental, or emotional condition, they are limited in the kind or amount of work they can perform, have difficulty doing errands alone, and have difficulty participating in social activities. The degree of difficulty (none, some, a lot, cannot do at all) is captured for the errands and social activities questions. Analysts can analyze each question separately or can combine the questions into an indicator to meet the needs of their analysis.

### *Sponsored Content*

The Administration for Community Living (ACL) sponsored a question in the ADO section. Sample adults reporting having "a lot of difficulty" or "cannot do at all" to the any one of five functioning questions (walking or climbing stairs, communicating, remembering or concentrating, self-care, or doing errands alone) were asked whether the difficulty began before age 22. The purpose of the question was to assist analysts who wish to identify adults with intellectual or developmental disability.

### *Major Recodes*

**Disability status composite indicator, age 18 and older.** An indicator of disability status based on the WG Short Set on Functioning that identifies Sample Adults who are at greater risk than the general population for experiencing restrictions in participation because of difficulties doing certain universal, basic actions. This recode classifies Sample Adults with disability as those reporting “a lot of difficulty” or “cannot do at all” for at least one of six domains of functioning: seeing (even if wearing glasses), hearing (even if wearing hearing aids), mobility (walking or climbing stairs), communication (understanding or being understood by others), cognition (remembering or concentrating), and self-care (such as washing all over or dressing). The remaining Sample Adults, that is those who responded "some difficulty" or “no difficulty” to at least one domain (and did not report “a lot of difficulty” or “cannot do at all” for any of the six domains of functioning) are classified as without disability. Those responding "don't know" or "refused" to all six questions are excluded.

### III. Health Care Access and Health Service Utilization

#### Annual Core

Several sections throughout the Sample Adult module measure access to and use of health services, as well as affordability of care. For a list of measures on these topics asked in the annual core, by questionnaire section and reference period of its respective questions, see Table 12. Similar content is also available for the Sample Child.

**Table 12: Annual core content measures of health care access, service use, and affordability of care asked of the Sample Adult, by questionnaire section and reference periods.**

Measure (section)	Reference period
<b>Immunizations (IMS)</b>	
Flu vaccine	Last 12 months, Month and Year
Pneumonia vaccine	Ever
Number of pneumonia vaccines received	Ever
<b>Medical Care (UTZ)</b>	
Saw a doctor for medical care	Last time interval
Medical and wellness visit combined	---
Wellness visit	Last time interval
Usual place to go for medical care	Current
Kind of place for medical care	Current
Number of urgent care visits	Last 12 months
Number of emergency department visits	Last 12 months
Any overnight hospitalization	Last 12 months
Delayed medical care due to cost	Last 12 months
Needed but did not get medical care due to cost	Last 12 months
<b>Mental Health Care (MHC)</b>	
Took medication for emotions/mental health	Last 12 months
Received therapy or counseling	Last 12 months; current
Delayed getting therapy/counseling due to cost	Last 12 months
Needed but did not get therapy/counseling due to cost	Last 12 months
<b>Prescription Medication (PMD)</b>	
Took prescription medication	Last 12 months
Skipped doses to save money	Last 12 months
Took less medication to save money	Last 12 months
Delayed filling prescription to save money	Last 12 months
Needed but did not get prescription due to cost	Last 12 months
<b>Problems Paying Medical Bills (PAY)</b>	
Anyone in family having problems paying medical bills*	Last 12 months
Have bills unable to pay at all*	Current
Level of worry about paying medical bills if sick/accident	Current
*These are family-level replicate questions asked once per family.	

### Rotating Content

Additional content about the use of selected health services is included in 2022–2023, as part of rotating content previously fielded in 2019–2020. These questions ask Sample Adults about receiving dental, eye, home, and physical or other therapeutic care, and about affordability of dental care. Table 13 lists rotating core measures of health care use, and affordability, by questionnaire section and reference period of its respective questions.

**Table 13. Measures of health care access, service use, and affordability of care asked of the Sample Adult in the rotating core, by questionnaire section and reference periods: 2022–2023 NHIS**

Measure (Section)	Reference period
<b>Dental Care (DNC)</b>	
Received a dental cleaning/exam	Last time interval
Delayed dental care due to cost	Last 12 months
Needed but did not get dental care due to cost	Last 12 months
<b>Physical and Other Therapeutic/Specialist Care (PTC)</b>	
Received eye exam from eye specialist	Last 12 months
Received physical/speech/rehabilitative/occupational therapy	Last 12 months
Received home care	Last 12 months

### Emerging Content

**Telemedicine.** The TLH section included a question about the use of telemedicine that asked whether the Sample Adult had an appointment with a doctor, nurse or other health professional by video or phone during the past 12 months. This question was first fielded in July 2020 for both Sample Adults and Sample Children as emerging content in response to the COVID-19 pandemic. From July 2020 through December 2021 this question was fielded in the ACC section, and moved to its own section, TLH, in 2022.

### Sponsored Content

**Complementary and Integrative Health.** Complementary and Integrative Health (CIH) questions were sponsored by NIH/NCCIH. Complementary and integrative health encompass the use of non-mainstream medical approaches together with conventional medicine (complementary), or in a coordinated way with conventional medicine (integrative) ([National Center for Complementary and Integrative Health](#)).

Questions about the use of chiropractic, acupuncture, massage, naturopathy, art or music therapy, meditation, guided imagery or progressive relaxation, and yoga in the past 12 months were asked of Sample Adults aged 18 and older. For each of the therapies or practices, with the exception of art or music therapy, two follow-up questions asked whether the respondent had used it to treat or manage physical pain or to restore their overall health. Additionally, adults that practiced yoga were also asked if their yoga practice included breathing exercises or meditation. These questions were also asked of Sample Children 4–17. Similar questions about the use of CIH in adults were previously fielded in the 2017 NHIS as part of the Adult Complementary Health (ACH) supplement.

**Immunization.** The IMS section includes several questions sponsored by NCIRD regarding CDC recommended vaccinations for flu, shingles, tetanus, HPV, and COVID-19 (See Table 14).

In December 2020, the FDA issued an emergency use authorization for the first COVID-19 vaccine, administered as two doses three weeks apart, for people aged 16 and older (FDA, 2023). Since then, a number of vaccines and booster doses have been approved or authorized by the FDA (FDA, 2023). In 2022, the question wording or response options of some of the COVID-19 vaccination questions were updated from the previous year to reflect changes in the availability of vaccines. The Sample Adult was asked whether they had received at least one dose of a COVID-19 vaccine, the number of vaccinations they received (up to 4 or more), the dates (month and year) of the most recent and next most recent vaccinations, and the brand of first vaccine received. Starting in quarter 4 of 2022, a revision to add an answer category was added to the question on vaccine brand, which resulted in a variable name change (SHOTTYPE\_A to SHOTTYPE1\_A) and a recode to consolidate answers from quarters 1-4 (See major recodes, below). In addition, while in 2021 the month and year of COVID-19 vaccines were edited to correct inconsistent timing of reported dates, dates of vaccination were not edited in 2022. Original responses are provided in the data file, which may include dates out of order or dates that are inconsistent with vaccines availability.

**Table 14. Sponsored content about immunization measures asked of Sample Adults, by population, reference period, and additional information collected: 2021 NHIS.**

Vaccine	Population	Reference Period	Additional information collected
<b>COVID-19*</b>	Adults 18 years and over	Ever Two most recent	Number of vaccines Vaccine brand
<b>Flu</b>	Female adults aged 18–49 or age unknown	Before/during/after a current or recent pregnancy (see flu vaccination timing, below)	Not applicable
<b>Shingles (Any vaccine or Shingrix)</b>	Adults aged 50 and over or adults aged 18 and over who have a weakened immune system	Ever (any and Shingrix) Most recent (any)	Number of Shingrix vaccines First or second dose of Shingrix in 2020-2021
<b>Tetanus</b>	Adults 18 years and over	During pregnancy, or last 10 years	Vaccine included pertussis vaccine (Tdap)
<b>Human Papilloma Virus (HPV)*</b>	Adults 18-64 years old	Ever	Age of first vaccination
*Similar content is available for Sample Children.			

In 2022, there were changes to the shingles vaccination question from the previous year. In 2022, the questions asked Sample Adults about ever receiving and the year of the most recent shingles vaccination (which could include Zostavax or Shingrix shingles vaccine), with additional questions focusing exclusively on Shingrix vaccines. Since November 2020, the Zostavax shingles vaccine is no longer available in the United States, and questions that specifically asked about Zostavax were thus discontinued. In addition, in 2022, the universe of the shingles questions was expanded from adults aged 50 and over to also include adults aged 18 and over who have a weakened immune system.

NCIRD also sponsored a question for female Sample Adults between 18 and 49 years old (or whose age was not known) that asked if they had a live birth in the past 12 months, and two questions for Sample Adults aged 18 and over about work or volunteer activities in health care settings. The purpose of these questions is to permit analysis of vaccine uptake by pregnant women and health care workers.

Flu vaccination timing in relation to pregnancy status was determined as follows:

Female Sample Adults between 18 and 49 years old (or whose age was not known) who reported that they were currently pregnant (asked previously in the BMI section), and who had received a flu vaccine in the past 12 months, and were interviewed from January through March or from August through December were asked: “Did you get a flu vaccination before or during your current pregnancy?” Female Sample Adults 18–49 years (or whose age is not known) and reported that they were not currently pregnant (or pregnancy status is not known), or those who were currently pregnant and were interviewed between April through July, were asked about pregnancy status during August through March as follows: since August 1<sup>st</sup> of last year if interviewed between January through March; from August of last year through March of current year if interviewed April through July; and since August 1 of current year if interviewed August through December. Those who reported to be currently pregnant or had a recent pregnancy during August through March, and who had received a flu vaccine in the past 12 months were asked: “Earlier you said you were pregnant sometime [since August 1<sup>st</sup>, {prior year}/from August {prior year} through March {current year}/since August 1<sup>st</sup>, {current year}] Did you get a flu vaccination before, during, or after your pregnancy?”

### Major Recodes

**Year of receipt of shingles vaccine.** A recodes was created with a lower limit of 2006 for the year of the most recent shingles vaccine. Sample Adults reporting years earlier than 2006 were assigned the value 9996.

**Brand of first COVID-19 vaccine** A recode was created to consolidate responses regarding brand of first COVID-19 vaccine received asked in quarters 1–3 of 2022 using SHOTTYPE\_A and asked in quarter 4 of 2022 using SHOTTYPE1\_A.

## IV. Health-Related Behaviors

### Annual Core

Sample Adults were asked about cigarette smoking and the use of electronic cigarettes in the CIG section. All adults were asked if they had smoked at least 100 cigarettes in their entire life. Those who had done so were asked whether they now smoke cigarettes every day, some days, or not at all. Every day smokers were asked about the number of cigarettes smoked every day, while someday smokers were asked the number of days that they smoked in the past 30 days and the average number of cigarettes smoked on those days. All Sample Adults were also asked about the use of electronic cigarettes or other electronic vaping products, even one time in their entire life, and if so, whether they now use this product every day, some days, or not at all.

### Rotating Core

In 2022, there are questions about alcohol use in the ALC section, physical activity in the PHY and CPA sections, sleep in the SLP section, and smoking history in the CIG section. This health-related behavior content rotates every other year and it previously fielded in 2020, see Table 15 for a list of rotating core health-related measures.

**Table 15. Health-related behaviors asked of the Sample Adult in the rotating core, by questionnaire section and reference periods: 2022 NHIS**

Measure (Section)	Reference period
<b>Alcohol Use (ALC)</b>	
One or more drinks	Ever, Past 12 months, in any one year, 30 days
Drinking frequency	Past 12 months
Average number of drinks	Past 12 months
Heavy and binge drinking	Past 12 months, 30 days
Advised to stop	Past 12 months
<b>Smoking History (CIG)</b>	
Initiation age	Since age 6
Cessation	Past 12 months, up to 70 years
Advised about ways to quit	Past 12 months
<b>Physical Activity (PHY)</b>	
Moderate intensity	Current
Vigorous intensity	Current
Muscle strengthening	Current
<b>Content of Care (CPA)</b>	
Advised to increase physical activity	Past 12 months
<b>Sleep (SLP)</b>	
Duration	Current
Quality	30 days
Medication	30 days



**Alcohol.** The questions in the ALC section serve to determine the frequency and quantity of drinking and whether a doctor or health professional had advised the Sample Adult to stop or cut down on their drinking. Sample Adults were asked if they ever had at least 1 drink of alcohol, excluding sips and small tastes, in their entire life, with examples of types of alcoholic beverages provided. Those who had not consumed alcohol in the past year were asked whether they have had at least 12 alcoholic drinks in any one year. Those who had consumed alcohol in the past year were asked about their alcohol drinking frequency and average quantity per drinking day in the past 12 months. Heavy drinking was defined as consuming 4 or more alcoholic drinks a day for women, and 5 or more drinks for men (and for Sample Adults who refused to identify as male or female or answered don't know to the question on sex). Binge drinking was defined as consuming 4 or more drinks for women and 5 or more drinks for men, on one occasion—or roughly two hours. Sample Adults who on average met the criteria for heavy drinking in the past 12 months were asked about any drinking in the past 30 days and binge drinking in the past 30 days. Sample Adults who on average were not heavy drinkers were asked about any heavy drinking in the past year, and binge drinking in the past 30 days.

A question about receiving advice from a doctor, dentist, or other health professional to stop or reduce drinking was asked of all Sample Adults who drank in the past 12 months, and it was not restricted to those who had also seen a doctor or other health professional about their health in the past year as done in 2020, thus, a new variable name is used in 2022.

**Smoking history.** Questions about smoking history in the CIG section were asked of Sample Adults who ever smoked 100 cigarettes, and included age when they first started smoking regularly, whether every day or someday smokers had stopped smoking for more than one day in the past 12 months because they were trying to quit, and time since they quit among former smokers.

A question also asked current cigarette smokers and former smokers who quit smoking in the past 12 months whether they had been advised by a doctor, dentist, or other health professional to quit cigarette smoking. This question was no longer restricted to those who had seen a health professional in the past 12 months, as fielded in 2020, and therefore uses a new variable name in 2022. Note that a separate sponsored question asked about receiving advice from a health care professional to quit any tobacco-related product.

**Physical activity.** The PHY section asked Sample Adults about physical activity they engaged in during their leisure time, including exercise, sports, or physically active hobbies. These questions are designed to assess compliance with the 2018 Department of Health and Human Services' Physical Activity Guidelines for Americans, which recommend that adults complete at least 150 minutes to 300 minutes of moderate-intensity activity, or 75 minutes to 150 minutes of vigorous-intensity aerobic activity per week, as well as moderate or greater intensity muscle strengthening activities on two or more days a week (U.S. Department of Health and Human Services, 2018). Moderate-intensity activities cause moderate increases in breathing or heart rate while vigorous-intensity activities cause large increases in breathing or heart rate. Muscle strengthening activities include exercises such as sit-ups, push-ups, or lifting weights. Sample Adults were asked in separate questions about the frequency and duration of doing moderate and vigorous physical activities, as well as the frequency of doing strengthening activities.

**Content of care.** A question in the CPA section asked Sample Adults whether they have been advised by a doctor or health professional to increase their amount of physical activity or exercise. In 2020, this question was fielded in the DPV section.

**Sleep.** Sample adults were asked in the SLP section about their average number of hours of sleep in a 24-hour period as well as how often during the past 30 days they woke up feeling rested, had trouble falling asleep, had trouble staying asleep, and took medications to help them sleep.

### *Sponsored Content*

**Other Tobacco Products.** The FDA sponsored seven questions about the use of cigars, pipes and smokeless tobacco products in the OTB section. In separate questions, all Sample Adults were asked about whether they had ever smoked a cigar, smoked a pipe filled with tobacco, or used smokeless tobacco products. Questions included examples and descriptions of these products. Those who said “yes” to each respective tobacco product were asked whether they now use it every day, some days or not at all. Sample Adults who had ever smoked any type of cigar were also asked about the number of days they smoked cigars in the past 30 days.

**Menthol Cigarettes.** The FDA also sponsored two questions in the CIG section about the use of menthol cigarettes, non-menthol, or no usual type of cigarettes among current and former smokers. Former smokers were asked about the usual type smoked in the 12 months prior to quitting cigarettes, and current smokers were asked current usual use.

**Smoking Cessation Methods.** NCI and NCCDPHP sponsored several questions in the CIG section about smoking cessation methods related to the use of nicotine replacement, nicotine-dopamine interference medication, behavioral or counseling approaches, and transition to electronic cigarettes. Former smokers who stopped smoking in the past two years were asked about cessation methods used when they stopped smoking completely, and smokers who tried to quit smoking in the past 12 months were asked about cessation methods used when they tried to quit smoking. A question also asked current smokers whether they would like to completely quit smoking cigarettes.

**Advised to Quit Tobacco Products.** NCI and NCCDPHP also sponsored a question in the AQS section that asked current smokers, former smokers that quit in the past 12 months, current users of e-cigarettes, cigars, pipes, or smokeless tobacco, whether they had been advised by a doctor, dentist, or other health professional in the past 12 months to quit smoking or to quit using other kinds of tobacco.

For additional information about the historical context of tobacco use questions in NHIS, see

<https://www.cdc.gov/nchs/nhis/tobacco.htm>

**Diet and Nutrition.** In 2022, questions about diet and nutrition in the DNU section were sponsored by NCCDPHP/NCI. Studies related to diet, nutrition, and incidence of cancer identified fruit and vegetable intake as a protective factor against cancer, while other items (e.g., sweetened drinks, potatoes, pizza) were associated with increased risk of cancer (Key et al., 2004). Sample adults were asked about the quantity and frequency of consumption of 14 items within three domains: 1) fruits (100% fruit juice, fruit); 2) sugar sweetened beverages (SSB) (soda, coffee and tea, sports and energy drinks, sweetened fruit drinks); and 3) vegetables (salad, fried potatoes, other kinds of potatoes, beans, other vegetables, salsa, tomato sauces, including in pizza). Similar content was previously fielded in the 2010 and 2015 NHIS.

### *Major Recodes*

**Cigarette smoking status.** Sample Adults were classified in terms of their lifetime and current cigarette smoking status. The recode includes the following categories: current every day smoker; current someday smoker; former smoker; never smoker; smoker, current status unknown; and unknown if ever smoked. Former smoker is defined as a person who has smoked at least 100 cigarettes in their entire life and currently does not smoke at all. Never smoker is a person who has never smoked any cigarettes or has smoked less than 100 cigarettes in their entire life.

“Smoker, current status unknown” is defined as a person who has smoked 100 cigarettes in their entire life and the question about current smoking practices was answered as don’t know or refused, or it was not ascertained. “Unknown if ever smoked” includes those whose response to ever having smoked at least 100 cigarettes in their entire life was answered as don’t know or refused, or it was not ascertained.

**Drinking status.** Sample Adults were classified in terms of their lifetime and current alcohol drinking status. This recode follows similar categories used in prior years of the NHIS; however, some of the alcohol questions in the redesigned NHIS have been revised and are administered in a different order, and Sample Adults who did not indicate their sex and reported more than 7 and up to 14 drinks per week were classified as ‘current unknown’ because they could not be differentiated between the ‘moderate’ or the ‘heavier’ drinking categories. The drinking status categories are as follows:

- Lifetime abstainer:** a person who has not had at least one alcoholic drink in his or her entire life
- Former infrequent:** a person who ever had at least one alcoholic drink but never as many as 12 in any one year and none in the past year
- Former regular:** a person who has had at least 12 alcoholic drinks in any one year but did not drink in the past year
- Former, unknown:** a person who had at least one alcoholic drink in their lifetime, did not drink in the past year, and their response to having at least 12 alcoholic drinks in any one year was refused or don’t know, or it was not ascertained
- Current infrequent:** a person who has had up to 11 alcoholic drinks in the past year.
- Current light:** a person who has had an average of up to 3 drinks per week in the past year.
- Current moderate:** men who had more than 3 drinks per week up to 14 drinks per week in the past year, or women who had more than 3 drinks per week up to 7 drinks per week in the past year; or sex is unknown or refused and person had more than 3 drinks per week up to 7 drinks per week
- Current heavier:** men who had more than 14 drinks per week on average in past year, or women who had more than 7 drinks per week in the past year; or sex is unknown or refused and person had more than 14 drinks per week on average in past year.
- Current, unknown:** a person who drank more than 7 and up to 14 drinks per week and sex is unknown or refused, therefore cannot be classified as either moderate or heavier current drinker; those who drank in the past year but whose average amount was unknown, refused, not ascertained, or inconsistent due to a response of zero drinks for average amount; and those whose average alcohol consumption last year was known but the frequency was unknown, refused, or inconsistent with the time period, i.e. responses of greater than 7 days a week or greater than 31 days a month.
- Drinking status unknown:** this includes those who had an unknown or implausible drinking frequency (> 7 days/week or >31 days/month) in the past year and whose average amount was also unknown, refused, inconsistent (zero average drinks on days drank), or not ascertained; those who refused to answer their drinking frequency in the past year or it was not ascertained and as a result, they were not asked the remaining alcohol questions; and those whose response to ever having at least 1 drink in their lifetime was refused or don’t know, or it was not ascertained.

**Duration of physical activity.** Sample Adults could report their length of moderate and vigorous physical activity in either minutes or hours. A recode was created to standardize duration by converting all values reported in hours to minutes.

**Electronic cigarette use status.** Sample Adults were classified in terms of their ever and current electronic cigarette use. The recode includes the following categories: current e-cigarette user; used e-cigarette, not current user; never e-cigarette user; e-cigarette user, current status unknown; and unknown if ever used e-cigarettes. Current e-

cigarette user is defined as a person who uses electronic cigarettes every day or some days. Not current user is defined as a person who has ever used an electronic cigarette even one time in their entire life and who currently does not use them at all. Never user is defined as a person who has never used electronic cigarette. User, current status unknown is defined as person who has used an electronic cigarette even one time in their entire life and the question about current use was answered as don't know or refused, or it was not ascertained. Unknown if ever used electronic cigarettes includes those whose response to ever having used an electronic cigarette in their entire life was answered as don't know or refused, or it was not ascertained.

**Frequency of drinking alcohol.** Sample Adults could answer the question about frequency of drinking alcohol in days per week, per month, or per year. Two recodes were created to standardize these frequencies. In the first recode, all drinking frequencies were converted to days per week by dividing monthly values by 4.33 and dividing yearly values by 52. A value of 0 on this recode indicates less than one day a week. The second recode converted frequencies to days per year by multiplying weekly frequencies by 52 and monthly frequencies by 12.

**Frequency of physical activity.** Sample Adults could answer questions about frequency of moderate, vigorous, and strengthening physical activity in times per day, per week, per month, or per year. Those who reported physical activity frequency in times per day, per month, or per year were converted to times per week to standardize values.

**Heavy drinking.** This recode classified Sample Adults who drank in the past 12 months into whether they engaged in heavy drinking in the past 12 months, defined as 4 or more drinks per day among women, and 5 or more drinks among men, and among those whose sex is refused or don't know. The recode is based on responses from the questions on the average number of drinks on the days drank in the past 12 months, any heavy drinking in the past 12 months, and any binge drinking in the past 30 days. Individuals who responded yes to binge drinking in the past 30 days are classified as yes in this recode even if they did not answer yes to heavy drinking in the past 12 months.

**Meets aerobic activity guidelines.** Sample Adults were classified based on whether their physical activity met the 2018 Physical Activity Guidelines for aerobic activity (U.S. Department of Health and Human Services, 2018) and categorized as either inactive, insufficiently active, or sufficiently active. Inactive is defined as zero minutes of moderate and vigorous activity per week. Insufficiently active is defined as less than 150 minutes of total activity per week, including moderate- and vigorous-intensity activity, with each minute of vigorous activity counting as two minutes. Sufficiently active is defined as 150 minutes or more of moderate and vigorous physical activity a week.

**Meets aerobic activity or strengthening guidelines.** Sample Adults were classified based on whether their physical activity met the 2018 Physical Activity Guidelines for aerobic activity or the guideline for strengthening activity, defined as engaging in muscle strengthening physical activity at least two times a week. Individuals were categorized as meeting neither guideline (aerobic activity nor strengthening activity); meeting only the aerobic activity guideline; meeting only the strengthening activity guideline; or meeting both guidelines. Sample Adults whose activity was known on only one guideline but unknown on the other were not classified in this recode.

**Years since quitting cigarettes.** Former smokers could answer the question about time since quitting smoking in days, weeks, months, or years. Those who reported time since quitting as number of days, weeks, or months were converted to number of years, where zero years indicates less than 365 days ago, less than 52 weeks ago, or less than 12 months ago. Time intervals exceeding these values were classified into the respective number of years. Due to confidentiality concerns regarding the Sample Adult's age, number of years since quitting is top-coded at 70 years. The variables used for this recode are available in the public-use file, and top-coded at 70 years.

**Corrections to cigarette questions.** Due to an instrument error in January 2022, respondents who were designated to receive the question were not asked about (1) whether they would like to quit smoking completely (CIG) and (2) whether a health professional advised them to quit smoking or tobacco products (AQS). In the public-use data file, these missed respondents are assigned the value 6. The error was corrected in February.

**Corrections to physical activity questions.** Recodes were created to correct reports of extreme frequency and duration for physical activity in the PHY section. Frequencies of moderate, vigorous, and strengthening activity greater than four times a day, 28 times a week, and 120 times a month were considered extreme. Recodes of the original source variables were created that assign these extreme values to 5 and 9995 for the time interval and number question, respectively. Reports of moderate and vigorous activity duration greater than 720 minutes or greater than 12 hours at a time were considered extreme. Recodes of the original source variables were created that assign these extreme values to 6 and 996 for the time interval and number question, respectively.

## V. Health Promotion

### *Rotating Core*

In 2022, questions about walking for transportation and leisure are included in the WLK section. These questions are part of the rotating content previously fielded in 2020. The section provides data to generate population estimates and track progress towards national health objectives like Healthy People 2030 related to physical activity and the built environment. Estimates on walking can inform new policies or interventions to promote walking activity, especially among vulnerable populations. Sample Adults were asked about walks for at least 10 minutes in the past 7 days for transportation, defined as walking they might have done to travel to and from work, to do errands, or to go from place to place, and for leisure, defined as walk for fun, relaxation, exercise, or to walk the dog. Sample Adults were also asked about the frequency (the number of days and times in a day walked) and duration of those walks.

### *Major Recodes*

**Walking for transportation.** Sample Adults who walked for transportation at least one time in the past 7 days were asked on average, how many times per day they walked for transportation and on average, how long those walks took. Respondents could choose a count of up to 995 minutes or hours. For ease of calculation and to limit the publication of extreme values, a recode was created for the average length of walks for transportation. This recode has a continuous count of 1–180 minutes, and a top-coded at 181 minutes for all walks with a duration over 3 hours. Users can convert these times to hours by dividing by 60.

**Walking for leisure.** Sample Adults who walked for leisure at least one time in the past 7 days were asked on average, how many times per day they walked for leisure and on average, how long those walks took. Respondents could choose a count of up to 995 minutes or hours. For ease of calculation and to limit the publication of extreme values, a recode was created for the average length of walks for leisure. This recode had a continuous count of 1–180 minutes, and a top-coded at 181 minutes for all walks with a duration over 3 hours. Users can convert these times to hours by dividing by 60.

## VI. Mental Health

### *Rotating Core*

Rotating every three years (last asked in 2019) are two mental health scales measuring symptoms of generalized anxiety disorder (GAD) and depression.

Symptoms of generalized anxiety disorder are measured using the 7-item Generalized Anxiety Disorder scale (GAD-7) (Spitzer et al., 2006). This is a brief scale to screen for GAD symptoms in the past two weeks and assess its severity in clinical settings and the general population (Rutter and Brown 2016, Löwe et al., 2008). GAD-7 has also been found to have moderately good operating characteristics for three other anxiety disorders – panic disorder, social anxiety disorder, and post-traumatic stress disorder (Kroenke et al., 2007). The GAD-7 was developed based on the most correlated items with a 13-item scale that included 9 items from the criteria for GAD in the Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition (DSM-IV) and 4 items on the basis of review of existing anxiety scales (Spitzer et al., 2006).

Symptoms of depression are measured using the 8-item Patient Health Questionnaire depression scale (PHQ-8) (Kroenke et al., 2009). This is a valid diagnostic and severity measure for current depressive disorders derived using the nine-item criteria for depressive disorders in the DSM-IV. Symptoms of depression are asked in reference to the past two weeks. PHQ-8 is used in clinical settings and in population-based studies to screen for symptoms indicative of a possible clinically significant depression and assess severity of depressive disorders (Kroenke et al., 2009). PHQ-8 is an abbreviated version of the nine-item PHQ-9 scale (Kroenke et al., 2001) that excludes the question about thoughts of death and self-injury, an indicator of possible suicide risk.

The set of questions in the GAD-7 and PHQ-8, available in the GAD and PHQ sections of the questionnaire, ask Sample Adults to assess how often they have been bothered over the last 2 weeks by a set of specific symptoms. Response options to the GAD-7 and PHQ-8 questions are the same: 1) not at all, 2) several days, 3) more than half the days and 4) nearly every day. These response categories correspond to 0 to 3 points, respectively, when scoring each question. The points for each question are summed to produce a total score between 0 and 21 in GAD-7 and between 0 and 24 for PHQ-8. Total scores may be used in analysis, and pre-determined cut-off values of the total score are available for assessing severity of symptoms (Kroenke et al., 2010). Variables classifying the total score into four categories of severity of symptoms of GAD and depression (i.e., none or minimal, mild, moderate, and severe) are available to data users (see major recodes, below). The recode for severe symptoms of depression is based on a PHQ-8 total score of 15 or greater. The PHQ-8 total score can also be classified into five categories of severity, where those scoring 15 or greater may be further categorized into moderately severe (point values 15 to 19) and severe symptoms (point values 20 to 24) (Kroenke et al., 2009). For both GAD-7 and PHQ-8, a total score of 10 or greater is indicative of a possible clinically significant condition.

Some other surveys do not include all 8 questions of the PHQ-8 or 7 questions of the GAD-7, but instead include only the first two items of each scale. These abbreviated measures are known as the PHQ-2 and GAD-2, respectively (Kroenke, et al., 2003; Kroenke et al., 2007). PHQ-2 and GAD-2 scores can be calculated from the NHIS for comparison with other surveys. The score of the two questions per scale are summed to produce a total score between 0 and 6. A score of 3 points or greater identifies possible cases for major depressive disorder based on the PHQ-2 and for generalized anxiety disorder based on the GAD-2.

### *Sponsored Content*

**Life Satisfaction.** NCCDPHP and ODP sponsored a question about life satisfaction in the LSF section. Sample adults were asked how satisfied they were with their life using response options of very satisfied, satisfied, dissatisfied, and very dissatisfied. Life satisfaction is an aspect of general well-being that represents people's subjective assessments of their lives overall and reflects factors such as health, work, and income, among others (OECD, 2013). The American Psychological Association defines life satisfaction as "the extent to which a person finds life rich, meaningful, full, or of high quality" (APA, 2022). Measures of life satisfaction, including the items used in NHIS, have been shown to be valid and reliable (Cheung and Lucas, 2014; OECD, 2013). Life satisfaction will be used as a Healthy People 2030 Overall Health and Well-being Measure to estimate and track adult well-being at the national level (Ochiai et al., 2021).

This question was also fielded in 2021, as part of an experiment to examine alternative response options and question placement in the survey. An internal unpublished analysis found that the 4-point scale retained in the 2022 NHIS was found to take less time to administer and had less measurement error than the 11-point scale, while there were no contextual effects from question placement or differences in item nonresponse.

### Major Recodes

**Severity of GAD symptoms.** Severity of GAD symptoms were categorized into no or minimal symptoms, mild symptoms, moderate symptoms, and severe symptoms for Sample Adults with responses to at least 6 of the 7 questions, using the cut-offs of the total score at <5, <10, <15, and 15 or greater. Instances when one question was refused or answered as don't know, or it was not ascertained, the average of the 6 ascertained answers (with point values 0-3) was calculated and added to the total score from the six ascertained questions. Total scores with decimal values were not rounded-up to the next whole integer for determining category of severity. Sample adults missing responses to two or more questions were categorized as having unknown severity of symptoms and coded as 8.

**Severity of depression symptoms.** Severity of depression symptoms were categorized into no or minimal symptoms, mild symptoms, moderate symptoms, and severe symptoms for Sample Adults with responses to at least 7 of the 8 questions, using the cut-offs of the total score at <5, <10, <15, and 15 or greater. Instances when one question was refused or answered as don't know, or it was not ascertained, the average of the 7 ascertained answers (with point values 0-3) was calculated and added to the total score from the seven ascertained questions. Total scores with decimal values were not rounded-up to the next whole integer for determining category of severity. Sample adults missing responses to two or more questions were categorized as having unknown severity of symptoms and coded as 8.

## VII. Barriers to care

### *Emerging Content*

**Nonfinancial barriers.** Questions about nonfinancial access barriers to medical care were included in the ABH section as emerging content. Sample adults were asked whether during the past 12 months they had delayed or forgone medical care because of the following reasons: 1) difficulty finding a medical facility that would accept their health insurance; 2) length of time to get to medical facility was too long from home or work; 3) an appointment wasn't available when needed it; 4) couldn't get to the medical facility when open; 5) were too busy with work or other commitments. A question about transportation barriers was fielded in the TBH section as emerging content. Sample adults were asked whether lack of reliable transportation in the past 12 months kept them from medical appointments, meetings, work, or from getting things needed for daily living.

The data are intended to produce national prevalence estimates of nonfinancial access barriers to care, consistent with efforts to inform about equity among underserved communities (EO 13985, 2021). Nonfinancial barriers or access to medical care have been described as dimensions or domains in accessibility, availability, accommodation, affordability, and acceptability (Penchansky and Thomas, 1981). Nonfinancial barriers for delaying or forgone health care have been associated with increased mortality, poorer health outcomes, and increased health expenditures (Prentice and Pizer, 2007; Heaps, Abramsohn, Skillen, 2021).

**Internet access and health care.** Five questions that asked Sample Adults about their internet access and use of the internet for health and medical purposes were added to the NHIS in July 2022 in the HIT section as emerging content. These questions aim to capture data on individuals' use of the internet to access health information, communicate with healthcare providers, and obtain medical test results. Internet access may enable access to telemedicine, promote health literacy, and support public health prevention (Benda et al., 2020; Early and Hernandez, 2021), and may have a positive impact on mental well-being, physical activity, and social isolation. (Kearns and Whitley, 2019, Duplaga and Szulc, 2019; Khalaila and Vitman-Schorr, 2018).

## VIII. Community engagement

### *Emerging Content*

Questions that asked Sample Adults about their participation in local elections and public meetings, and volunteer activities over the past 12 months were added to the NHIS in July 2022 in the CIV and VOL sections as emerging content. Higher levels of civic engagement, as measured by voter participation, volunteering, and community organizing, have been linked to higher levels of well-being and self-reported physical health as well as lower levels of chronic disease and depression. (Danso, 2017; Nelson et al., 2019; Stopka et al., 2022).



## Sample Child's Health

### I. Health status and conditions

#### *Annual Core*

The Sample Child module includes questions about the health status of the child and whether a doctor or other health care professional had diagnosed the child with asthma, diabetes, selected developmental conditions, and whether a representative from a school or a health professional had stated that the child had a learning disability.

The health status of Sample Children aged 0–17 years was asked in the Sample Child HIS section and assessed whether the child's health is generally excellent, very good, good, fair, or poor.

Questions regarding asthma were asked of children aged 0–17 years in the Sample Child AST section and measure the following: whether a doctor or other health care professional ever told that the Sample Child had asthma; still has asthma; had an episode of asthma or an asthma attack during the past 12 months; and had to visit an emergency room or urgent care center because of asthma during the past 12 months.

Questions regarding diabetes were asked of children aged 0–17 years in the Sample Child DIB section and measure whether a doctor or other health care professional ever told that Sample Child had prediabetes or borderline diabetes, and diabetes.

Questions regarding development conditions were asked of children aged 0–17 or aged 2–17 years in the Sample Child DLD section. Sample children aged 2–17 years were asked in separate questions whether a doctor or other health professional had ever told that the child had Attention Deficit/Hyperactivity Disorder (ADHD) or Attention-Deficit Disorder (ADD), and autism, Asperger's disorder, pervasive developmental disorder, or autism spectrum disorder. Sample children aged 0–17 years were asked in separate questions whether a doctor or other health professional had ever told that the child had an intellectual disability and any other developmental delay. For each condition, a follow up question asked whether the Sample Child currently had this developmental condition.

The DLD section also included two questions for Sample Children aged 2–17 years that asked whether a representative from a school or a health professional ever told that the child had a learning disability and currently had a learning disability.

#### *Rotating Core*

The Body Mass Index (BMI) section includes a set of questions rotating every other year for Sample Children aged 10–17 that ask how tall the child is without shoes in feet and inches or meters and centimeters and how much does the child weigh now in pounds or kilograms.

## Emerging Content

**COVID-19 infection.** Since July 2020, the Sample Child module includes questions about positive COVID-19 diagnosis and testing and perceived symptom severity. In 2022, two additional questions were added to measure Long COVID or post-COVID conditions (PCC). Questions about whether the symptoms lasted for three or more months and currently have symptoms were asked of children who experience mild to severe COVID-19 symptoms, or the Sample Child respondent refused or didn't know the symptom severity. According to the Department of Health and Human Services, Long COVID is defined as signs, symptoms and conditions that develop or continue after initial COVID-19 infection, are present after four weeks or more after the initial infection phase, may be multi-systemic, relapsing–remitting and progressing over time (Department of Health and Human Services, 2022).

The questions about COVID-19 are in the CVL section in 2022. In 2020–2021, the questions on COVID–19 diagnosis and testing were fielded in the CVD section, and testing was asked as a two-part- question.

## Major Recodes

**Body Mass Index (BMI).** A categorical measure of BMI was created using unrestricted height and weight values which contain the greater range of height and weight values than are available on the public-use file. BMI was calculated using the formula:  $BMI = \text{kilograms} / \text{meters}^2$  where 1 kilogram = 2.20462 pounds and 1 meter = 39.37008 inches. For children and teens, the classification of underweight, healthy weight, overweight and obese is age and sex specific. Values for BMI for age were categorized according to the National Center for Health Statistics' Data Table of BMI-for-age Charts. Table 16 shows the cutoff values for BMI for age for children aged 10 to 17 years by sex. Up to 5 decimal points were used for classifying Sample Children into the respective BMI categories. Sample children with the responses "don't know" or "refused" for height or weight, or those who had values for either height or weight that were outside the limits for public data release, or for whom height or weight values for public release were set to 96 and 996 due to missing sex information were coded as 9 in the categorical BMI measure. For additional information about children and teens BMI, see [https://www.cdc.gov/healthyweight/assessing/bmi/childrens\\_bmi/about\\_childrens\\_bmi.html](https://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html).

**Height and Weight.** The Sample Child respondent had the option to report height and weight using the U.S. customary system (pounds: feet and inches) or the metric system (kilograms: meters and centimeters). Metric responses on height and weight were converted into the U.S. Customary system using the following conversion scale: 1 meter=39.37008 inches and 1 kilogram=2.20462 pounds (lbs.). The public-use height variable for ages 10–17 years reflects total height in inches (e.g., 53" is 4' 5"), and can range from 48 to 76 inches for boys and from 41 to 70 inches for girls. In 2022, the lowest height is 43 inches. The public-use weight variable for ages 10 – 17 years reflects total pounds rounded to whole integers, and it can range from 50 to 280 lbs. for boys and from 46 to 230 lbs. for girls. Values reported for boys and girls outside the public-use limits for either height or weight had data for both variables recoded to "96" or "996" ("Not available") on the public-use data file to protect the confidentiality of those who might be identifiable by their unusual physical characteristics. In addition, due to the sex-specific height and weight limits by age for public data release, all Sample Children for whom sex was answered as don't know or refused were also coded as 96 for height and 996 for weight. Don't know and refused responses were also recoded to "96" for height and "996" for weight.

**Missed days due to asthma.** A recode was created of the number of missed days to daycare or school due to asthma during the past 12 months, top-coded at 10 or more days.

Table 16. Body Mass Index (BMI) for age values for children aged 10 to 17 years old, by sex.

Age and Sex	Underweight	Healthy weight	Overweight	Obese
<b>Age 10</b>				
Boys	<14.21866	14.21866 to <19.39041	19.39041 to <22.15409	>=22.15409
Girls	<14.03535	14.03535 to <19.984	19.984 to <22.98258	>=22.98258
<b>Age 11</b>				
Boys	<14.56001	14.56001 to <20.19667	20.19667 to <23.21358	>=23.21358
Girls	<14.4029	14.4029 to <20.86984	20.86984 to <24.14141	>=24.14141
<b>Age 12</b>				
Boys	< 14.97745	14.97745 to < 21.02386	21.02386 to < 24.22985	>= 24.22985
Girls	<14.83262	14.83262 to <21.74263	21.74263 to <25.25564	>=25.25564
<b>Age 13</b>				
Boys	< 15.45918	15.45918 to < 21.85104	21.85104 to < 25.17811	>= 25.17811
Girls	<15.30749	15.30749 to <22.57506	22.57506 to <26.2988	>=26.2988
<b>Age 14</b>				
Boys	< 15.99065	15.99065 to < 22.66325	22.66325 to < 26.04662	>= 26.04662
Girls	<15.80753	15.80753 to <23.34689	23.34689 to <27.25597	>=27.25597
<b>Age 15</b>				
Boys	< 16.55481	16.55481 to < 23.45117	23.45117 to < 26.83688	>= 26.83688
Girls	<16.30974	16.30974 to <24.04503	24.04503 to <28.12369	>=28.12369
<b>Age 16</b>				
Boys	< 17.1325	17.1325 to < 24.21087	24.21087 to < 27.56393	>= 27.56393
Girls	<16.78787	16.78787 to <24.66372	24.66372 to <28.90981	>=28.90981
<b>Age 17</b>				
Boys	< 17.70284	17.70284 to < 24.94362	24.94362 to < 28.25676	>= 28.25676
Girls	<17.21234	17.21234 to <25.20482	25.20482 to <29.6335	>=29.6335

Source. National Center for Health Statistics. Data Table of BMI-for-age Charts, available at [https://www.cdc.gov/growthcharts/html\\_charts/bmiagerev.htm](https://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm).

## II. Functioning and Disability

### *Annual Core*

The questions on functioning and disability found in sections VIS, HEA, MOB, UPP, COM, COG, ANX, DEP, and BEH of the Sample Child module compose the Child Functioning Module (CFM) that is part of a set of international standard measures developed, tested and endorsed by the Washington Group on Disability Statistics (WG). The WG is a city group established in 2001 under the United Nations Statistical Commission to address the need for population-based measures of disability by promoting and coordinating international cooperation in the area of health statistics focusing on disability data collection tools suitable for censuses and national surveys. The major objective is to provide necessary information on disability that is comparable throughout the world by identifying individuals with functional limitations in basic actions, regardless of nationality or culture. The questions reflect advances in the conceptualization of disability and use the World Health Organization's International Classification of Functioning, Disability, and Health (ICF) as a conceptual framework. The CFM was developed jointly with the United Nations Children's Fund (UNICEF). There are questions about children 2–4 years of age (CFM 2–4) and questions about children 5–17 years of age (CFM 5–17). The intended use of these questions is to describe the functional status of children and, when used with other questions on the survey, to evaluate whether children with functional limitations have achieved similar levels of participation and inclusion as children without functional limitations. These questions do not capture all aspects of difficulty in functioning, but rather focus on domains of functioning that are likely to identify the majority of children at risk of participation restrictions in an unaccommodating environment.

Questions ask about the Sample Child's level of difficulty (no difficulty, some difficulty, a lot of difficulty, or cannot do at all) in basic domains of functioning including seeing, hearing, mobility, dexterity, self-care, communication, cognition, playing, learning, relationships, and behavior and about the frequency of experiencing anxiety and depression as well as kicking/biting/hitting others. The CFM 2–4 and CFM 5–17 are designed to include domains of functioning relevant for each age group. For a list of questions asked in each set, and the associated functioning domain, questionnaire section, and respective age range, see Table 17. In addition, several of the domains have questions to provide information on the use of accommodations. Questions about the use of equipment or assistance with walking were asked about children who had difficulty walking both with and without equipment or assistance. More information may be obtained by request to the WG Secretariat at [WG\\_Secretariat@cdc.gov](mailto:WG_Secretariat@cdc.gov) or found on the WG website at: <http://www.washingtongroup-disability.com/>.

The WG questions can be analyzed separately, by domain, or combined across domains. CFM disability status indicators are available to data users (see Major Recodes below). The disability composite indicators for Sample Children aged 2–4 and 5–17 identify children who are at greater risk than the general population for experiencing restrictions in participation because of difficulties doing certain universal, basic actions. Consult the WG website for guidance on the creation of these indicators. Changing the threshold for either the number of domains the respondent identifies the child having difficulty with or the child's degree of difficulty can create different identifiers that will capture different populations. For example, a recode that includes Sample Children with the response "some difficulty" to any of the domains will capture a greater proportion of the population than a recode that includes only Sample Children with the response "cannot do at all" to any of the domains. In this example, the functional abilities of the larger group will be much more heterogeneous than that of the smaller group. Analytic guidelines written for each of the CFM questions sets, including recommended disability identifiers, may be obtained from the WG website.

**Table 17. Annual core content measures of functional limitations measured in the Sample Child module, by domain, question topic, questionnaire section and age range of question**

Domain	Question topic	Section	Age range for question
Seeing	Wear glasses or contacts	VIS	2–17
Seeing	Have difficulty seeing (with glasses, if worn)	VIS	2–17
Hearing	Use a hearing aid	HEA	2–17
Hearing	Have difficulty hearing sounds (with hearing aid, if used)	HEA	2–17
Mobility	Use equipment or assistance for walking	MOB	2–17
Mobility	Difficulty walking	MOB	2–4
Mobility	Difficulty walking 100 yards	MOB	5–17
Mobility	Difficulty walking a third of a mile	MOB	5–17
Dexterity	Difficulty picking up small objects	UPP	2–4
Self-care	Difficulty with self-care	UPP	5–17
Communication	Difficulty understanding you	COM	2–4
Communication	Difficulty understanding Sample Child	COM	2–4
Communication	Difficulty being understood by people inside of household	COM	5–17
Communication	Difficulty being understood by people outside of household	COM	5–17
Learning	Difficulty learning things	COG	2–17
Cognition	Difficulty remembering things	COG	5–17
Affect	How often seem very anxious, nervous, or worried	ANX	5–17
Affect	How often seem very sad or depressed	DEP	5–17
Playing	Difficulty playing	BEH	2–4
Behavior	Kick, bite, or hit other children or adults	BEH	2–4
Behavior	Difficulty controlling behavior	BEH	5–17
Cognition	Difficulty concentrating	BEH	5–17
Behavior	Difficulty accepting changes in routine	BEH	5–17
Relationships	Difficulty making friends	BEH	5–17

### Major Recodes

**Disability status composite indicator, age 2–4.** An indicator of disability that captures Sample Children aged 2–4 who are at greater risk than the general population for experiencing restrictions in participation because of difficulties doing certain universal, basic actions. This recode classifies children with disability as those with the responses “a lot of difficulty” or “cannot do at all” for at least one of the questions asking about the Sample

Child’s difficulty seeing, hearing, walking, dexterity, communication, learning, and playing, or with the response “cannot do at all” to the question about controlling behavior. The remaining Sample Children, with the responses “some difficulty” or “no difficulty” to at least one question (and who do not have the responses “a lot of difficulty” or “cannot do at all” for any of the questions), are classified as without disability. Sample Children with a response of “don’t know” or “refused” to all questions are excluded.

**Disability status composite indicator, age 5–17.** An indicator of disability that captures Sample Children aged 5–17 who are at greater risk than the general population for experiencing restrictions in participation because of difficulties doing certain universal, basic actions. This recode classifies children with disability as those with the responses “a lot of difficulty” or “cannot do at all” for at least one of the questions asking about the Sample Child’s difficulty seeing, hearing, walking, self-care, communication, learning, remembering, concentrating, accepting change, controlling behavior, making friends or the response “daily” to questions asking how often the Sample Child feels anxious, nervous, or worried or feels depressed. The remaining Sample Children, with the responses “some difficulty” or “no difficulty” to at least one question (and who do not have the responses “a lot of difficulty” or “cannot do at all” or “daily” for any of the questions), are classified as without disability. Sample children with a response of “don’t know” or “refused” to all questions are excluded.

### III. Health Care Access and Health Service Utilization

#### Annual Core

Several sections throughout the Sample Child module measure access to and use of health services, as well as affordability of care. Similar content is also available for Sample Adults. For a list of measures on these topics asked in the annual core, by questionnaire section and reference period of its respective questions, see Table 18.

**Table 18. Annual core content measures of health care access, service use, and affordability of care asked of the Sample Child, by questionnaire section and reference periods.**

Measure (Section)	Reference period
<b>Immunizations (IMM)</b>	
Flu vaccine (any, number of vaccines up to 2)	Last 12 months, Month and Year
<b>Medical Care (UTZ)</b>	
Saw a doctor for medical care	Last time interval
Medical and wellness visit combined	---
Wellness visit	Last time interval
Usual place to go for medical care	Current
Kind of place for medical care	Current
Number of urgent care visits	Last 12 months
Number of emergency department visits	Last 12 months
Any overnight hospitalization	Last 12 months
Delayed medical care due to cost	Last 12 months
Needed but did not get medical care due to cost	Last 12 months
<b>Prescription Medication (PMD)</b>	
Took prescription medication	Last 12 months
Delayed filling prescription to save money	Last 12 months
Needed but did not get prescription due to cost	Last 12 months

**Table 18 continued. Annual core content measures of health care access, service use, and affordability of care asked of the Sample Child, by questionnaire section and reference periods.**

Measure (Section)	Reference period
<b>Problems Paying Medical Bills (PAY)</b>	
Anyone in family having problems paying medical bills*	Last 12 months
Have bills unable to pay at all*	Current
Level of worry about paying medical bills if sick/accident	Current
*These are family-level replicate questions asked once per family	

### Rotating Core

Additional content about the use of selected health services is included in 2022–2023, as part of rotating content previously fielded in 2019–2020. These questions ask Sample Children about receiving dental, mental health, eye, home, and physical or other therapeutic care, and about affordability of dental and mental care. Part of the content on mental health (i.e., receiving therapy and taking medication for mental health) was also fielded in 2021 to help monitor the use of mental health services in children during the COVID-19 pandemic. Table 19 lists rotating core measures, by questionnaire section and reference period of its respective questions.

**Table 19. Measures of health care access, service use, and affordability of care asked of the Sample Child in the rotating core, by questionnaire section and reference periods: 2022–2023 NHIS.**

Measure	Reference period
<b>Dental Care (DNC)</b>	
Received a dental cleaning/exam	Last time interval
Delayed dental care due to cost	Last 12 months
Needed but did not get dental care due to cost	Last 12 months
<b>Mental Health Care (MHC)</b>	
Took medication for emotions/mental health	Last 12 months
Received therapy or counseling from mental health professional	Last 12 months
Delayed getting therapy/counseling due to cost	Last 12 months
Needed but did not get therapy/counseling due to cost	Last 12 months
<b>Physical and Other Therapeutic/Specialist Care (PTC)</b>	
Received eye exam from eye specialist	Last 12 months
Received physical/speech/rehabilitative/occupational therapy	Last 12 months
Received home care	Last 12 months

### Emerging Content

**Teens.** In 2021–2022, the Sample Child UTZ section includes questions which asked whether Sample Children aged 12–17 had time during their last medical visit or their last wellness visit to speak to the health care professional privately. An additional question asked if there was one or more than one person that the Sample Child respondent thought of as the Sample Child’s personal health care professional. These questions assess whether the adolescent may have received services or advice that the parent or guardian may not be aware of, and whether the adolescent had the opportunity to discuss their health care with their provider as part of a transition into adulthood.

**Telemedicine.** The TLH section included a question about the use of telemedicine that asked whether the Sample Child had an appointment with a doctor, nurse or other health professional by video or phone during the past 12 months. This question was first fielded in July 2020 for both Sample Adults and Sample Children as emerging content in response to the COVID-19 pandemic. From July 2020 through December 2021 this question was fielded in the ACC section, and moved to its own section, TLH, in 2022.

### *Sponsored Content*

**Complementary and Integrative Health.** Complementary and Integrative Health (CIH) questions were sponsored by NIH/NCCIH. Complementary and integrative health encompass the use of non-mainstream medical approaches together with conventional medicine (complementary), or in a coordinated way with conventional medicine (integrative) ([National Center for Complementary and Integrative Health](#)).

Questions about the use of chiropractic, acupuncture, massage, naturopathy, art or music therapy, meditation, guided imagery or progressive relaxation, and yoga in the past 12 months were asked of Sample Children aged 4–17 years. For each of the therapies or practices, with the exception of art or music therapy, two follow-up questions asked whether the child had used it to treat or manage physical pain or to restore their overall health. Additionally, children that practiced yoga were also asked if their yoga practice included breathing exercises or meditation. These questions were also asked of Sample Adults. Similar questions about the use of CIH in children were previously fielded in the 2017 NHIS as part of the Child Complementary Health (CCH) supplement.

**COVID-19 vaccination.** Beginning in July 2021, questions about COVID-19 vaccination sponsored by NCIRD were added to the Sample Child module in the CVV section for children aged 12-17 years. At the start of 2022, these questions were asked for children aged 5-17, and starting in August 2022 they were asked for children of all ages. In December 2020, the FDA issued an emergency use authorization for the first COVID-19 vaccine, administered as two doses three weeks apart, for people aged 16 and older (FDA, 2023). This emergency use authorization was expanded to include children aged 12-15 years in May 2021 (FDA, 2023), to children aged 5-11 in October 2021, and to children aged 6 months and older in June 2022 (FDA, 2023). The Sample Child respondent was asked whether the child has had a COVID-19 vaccination, the number of vaccinations the child has received, the dates (month and year) of the most recent and next most recent vaccinations, and the brand of the first vaccine received. Similar content was collected from Sample Adults starting in January 2021. While in 2021 the month and year of COVID-19 vaccines were edited to correct inconsistent timing of reported dates, for 2022 data, the original responses are provided without any editing for reasonableness (dates may be out of order or before vaccines were available).

**HPV Vaccination.** NCIRD sponsored two questions on human papillomavirus (HPV) vaccination in the HPV section of the Sample Child module. The questions asked whether the Sample Child aged 9–17 had been vaccinated for HPV and the age when the child received the first shot. Similar content was collected from Sample Adults in the IMS section. To prevent most HPV-attributable cancers, CDC recommends a 2-shot series at ages 11–12 years, although the vaccination series can start when children are nine years old. A third shot is also recommended for children with a weakened immune system and those starting vaccination at 15 years or older.



## IV. Behavioral and Mental Health

### *Annual Core*

The Baby Pediatric Symptom Checklist (BPSC) is a 12-item validated screening tool used for assessing social and emotional difficulties among children aged 0–17 months (Sheldrick, 2013). The BPSC is one component of the larger Survey of Well-being of Young Children (SWYC), a screening instrument designed for use in a clinical setting, such as a pediatric primary care. Information about SWYC is available at <https://www.floatinghospital.org/The-Survey-of-Wellbeing-of-Young-Children/Overview.aspx>

In the BSC section of the Sample Child module, parents or adults knowledgeable and responsible for the child’s health rated a series of behaviors related to irritability, inflexibility, and difficulty with routines that may be used to identify risk for social and emotional difficulties. Each item can be rated as “not at all”; “somewhat”; or “very much” and responses are assigned point values of 0, 1 and 2 respectively. Items with missing responses will count as 0 points. The BPSC is constructed of three subscales (irritability, inflexibility and difficulty with routines) and each subscale is composed of 4 items. Any summed scale of three or more on any of the three subscales indicates that a child is at risk and in practice will prompt further evaluation with a health care professional. For analysis, users may sum responses to operationalize risk as a continuous variable. As thresholds have not yet been developed to operationalize a categorical variable for the general population, no cut-off scores have been provided. Although the BPSC instrument was developed for children under age 18 months, questions were fielded among Sample Children aged 0–23 months. Further, original validation of the BPSC was confirmed using NHIS data (Zablotsky, Black, Sheldrick, Perrin, Blumberg, 2022)

### *Rotating Core*

The Strengths and Difficulties Questionnaire in the SDQ section is a series of questions designed to be used as a screening tool for clinical assessment of mental disorders and for epidemiological research of emotional and behavioral problems in children and adolescents aged 4-17 years. The parent respondent version of the SDQ was used and this instrument is copyrighted by Robert Goodman, Ph.D., FRCPSYCH, MRCP (Goodman, 1997; Goodman 2009).

Parents or adults knowledgeable and responsible for the child’s health were asked to answer, on behalf of the Sample Child, 25 questions that assess emotional symptoms, conduct problems, hyperactive behavior, peer relationships and prosocial behaviors. The instrument is intended to be scored by summing items within each subscale, and four of the five subscales can be summed to create a total score which may be used in analysis as a continuous variable. Higher scores to items assessing emotional symptoms, conduct problems, hyperactive behavior, and peer relationships indicate more problematic attributes. See Table 12 for a list of SDQ subscales by item and scoring values. Recodes of summary scores are available to data users (see Major Recodes below). Additionally, the SDQ section houses questions that collect information about the duration and impact of symptoms, which can be useful for assessing the child’s problem, the impact that the problem has on the child and his/her family. Further information about the SDQ, is available at <http://www.sdqinfo.com>.

### Emerging Content

**Bullying.** Since 2021, bullying has been included as an emerging topic in the Sample Child module. Bullying victimization is associated with negative health outcomes including increased suicide attempts, depressive symptoms and lower school performance and school attachment (Schneider et al., 2012). Three questions about the frequency of being bullied, recent electronic bullying and frequency of bully perpetration were asked in the BLY section. Bullying questions were only asked among adolescents, or Sample Children aged 12-17. Bullying measures are adapted from the Youth Risk Behavior Surveillance System (YRBSS)

**Life Satisfaction.** Since 2021, a question about life satisfaction was included in the LS1 section of the Sample Child module for Sample Children aged 12–17 years. The Sample Child respondent was asked to rate how they thought the child felt about their life as a whole these days, using a scale of 0 to 10 where 0 means "very dissatisfied" and 10 means "very satisfied." This question was previously sponsored content, but it was retained in 2022 as emerging content.

**Social Support.** Since 2021, the receipt of social support has been included as an emerging topic in the Sample Child module. Presence of a positive adult mentor in a child or adolescent's life is associated with the ability to flourish, complete high school, attend college, reduced problem behaviors (e.g., gang membership), improved psychological well-being and health (DuBois and Silverthorn, 2005). Two questions related to the frequency of receiving social and emotional support and having an adult in their life that makes a difference were included in the SOS section. Receipt of social and emotional support was derived from the Sample Adult questionnaire, while the adult that makes a difference question was derived from the [National Survey of Children's Health](#). Social support questions were only asked among adolescents, or Sample Children aged 12-17.

### Major Recodes

**SDQ subscales.** Summary scores of each of the five SDQ subscales (emotional, conduct, hyperactive, peer relationships and prosocial), have been calculated and provided as a continuous variable ranging from 0-10 with higher scores indicating more problematic attributes for emotional, conduct, hyperactive, and peer relationship subscales and lower scores indicating more problematic attributes for the prosocial subscale. The five subscales can be analyzed separately to look at specific psychological problems. In order to calculate the SDQ subscales, variables with original answer codes of 1, 2, 3 in the instrument which correspond to point values of 0, 1, 2 (when question is worded negatively) or 2, 1, 0 (when question is worded positively) were summed. See Table 20 for subscales with items and point values. SDQ subscale recodes were assigned a code of 88 if more than two items needed to calculate the subscale were answered as "don't know," "refused," or were not ascertained.

**SDQ total.** The SDQ total score is based on the sum of the first four subscales (emotional, conduct, hyperactive, and peer relationships). This recode is a continuous variable ranging from 0-40 with higher scores indicating more problematic attributes. For the SDQ total score, a code of 88 has been assigned if more than one subscale score is missing.

**SDQ impact score.** The SDQ impact score is based on the sum of the impact questions, which are asked only among Sample Children 4-17 years who had responded as having had minor, definite or severe difficulties with emotions,

concentration, behavior, or being able to get along with other people. This recode is a continuous variable ranging from 0–10, with higher scores indicating that difficulties have a greater impact on the child’s and or family’s life. For the SDQ impact total score, a code of 88 is assigned if more than one item is missing.

**Table 20. Strengths and Difficulties Questionnaire subscales, items in each subscale, and point values for scoring**

<b>Subscale and Question Topic</b>	<b>Not true</b>	<b>Somewhat true</b>	<b>Definitely true</b>
<b>EMOTIONAL SUBSCALE</b>			
Often complains of headaches, stomach-aches, or sickness	0	1	2
Many worries, often seems worried	0	1	2
Often unhappy, depressed, or tearful	0	1	2
Nervous or clingy in new situations	0	1	2
Many fears, easily scared	0	1	2
<b>CONDUCT SUBSCALE</b>			
Often has temper tantrums or a hot temper	0	1	2
Generally obedient, usually does what parents want	2	1	0
Often fights with other children or bullies	0	1	2
Often lies or cheats	0	1	2
Steals from home, school, or elsewhere	0	1	2
<b>HYPERACTIVITY SUBSCALE</b>			
Restless, overactive, cannot stay still for long	0	1	2
Constantly fidgeting or squirming	0	1	2
Easily distracted, concentration wanders	0	1	2
Thinks things out before acting	2	1	0
Sees task through to the end, good attention span	2	1	0
<b>PEER RELATIONSHIP SUBSCALE</b>			
Rather solitary, tends to play alone	0	1	2
Has at least one good friend	2	1	0
Generally liked by other children	2	1	0
Picked on or bullied by other children	0	1	2
Gets on better with adults than other children	0	1	2
<b>PROSOCIAL SUBSCALE</b>			
Considerate of other people’s feelings	0	1	2
Shares readily with other children	0	1	2
Helpful if someone is hurt, upset, or feeling ill	0	1	2
Kind to younger children	0	1	2
Often volunteers to help others	0	1	2

Source: Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, and allied disciplines.* 38(5):581-6. 1997.

**Prosocial subscale.** A fifth subscale that describes children’s positive behaviors. These items are excluded from the overall SDQ score when using the SDQ to identify children who have or are at increased risk of having psychological disorders.

## V. Stressful Life Events

### *Rotating Core*

Stressful life events have been associated with early and lifelong health conditions that may extend into adulthood, including diabetes, mental health disorders, and cardiovascular problems (Felitti et al., 1998). Continuing from 2021 and previously fielded in 2019, the SLE section includes four rotating core questions that asked the Sample Child respondent whether the child has witnessed or experienced neighborhood violence, lived with someone with a mental illness, lived with someone with a drug or alcohol problem, and had a parent who was incarcerated after the child was born. The questions were adapted from the Adverse Childhood Experience (ACE) questions that have previously been fielded in the [National Survey of Children’s Health](#).

Since 2021, four additional questions are included in the SLE section as either emerging or sponsored content. The combined set of eight questions are intended to create a more comprehensive picture of potential stressful life events during the childhood years while providing compatible measures for an ongoing online survey of adolescents whose parents completed the NHIS Sample Child interview.

Given the sensitive nature of these questions, an introductory paragraph precedes the first question and informs the respondent that these events can happen in any family and that any question can be skipped. It is possible that some respondents may choose to skip some of these questions, particularly if other individuals in the household are present at the time of the interview

### *Emerging Content*

The Sample Child respondent was asked whether the child had ever been treated or judged unfairly because of their race or ethnic group, and among Sample Children aged 12-17 years, whether the child had been treated or judged unfairly because of their sexual orientation or gender identity. Both measures were also asked on the NSCH beginning in 2020.

### *Sponsored Content*

NCIPC sponsored two stressful life event questions. The Sample Child respondent was asked whether the child had ever lived with a parent or adult who was verbally abusive, and whether the child ever experienced unmet basic needs such as not having enough to eat, not going to a doctor when was sick, or not having a safe place to stay. These questions were adapted from the Optional Questions List for YRBSS and the ACE Module in the Behavioral Risk Factor Surveillance System.

## VI. Health-Related Behaviors

### *Rotating Core*

In 2022 and rotating once every two years since 2020 is content about children's use of electronic devices, sleep quality and physical activity.

Screen time in the SCR section includes a single question asked of Sample Children aged 0–17 years assessing whether they spend more than 2 hours a day on electronic devices.

Sleep content in the SLP section includes typical routines and sleep behaviors among Sample Children aged 2–17 years such as frequency of waking up well-rested, difficulty getting out of bed in the morning, being tired during the day, falling asleep or taking naps during the day, and having routine bed and wake times.

Physical activity content in the PHY section asked of Sample Children aged 6–17 years collects information about participation in sports teams, and gym class, and the frequency (i.e., how often) Sample Children participate in the following activities: exercise or sports, strength training activities, walking at least 10 minutes, bicycling at least 10 minutes.

## VII. Injuries

### *Emerging Content*

Content about head injuries or concussions experienced by Sample Children was included in the TBI section as emerging content during 2020–2022. The Sample Child respondent was asked to think about all head injuries, for example, from playing sports, car accidents, falls, or being hit by something or someone. The questions ask about head injuries that may have occurred anytime in Sample Child's life, ever experiencing concussion symptoms as a result of that head injury (e.g., gap in memory, headaches, vomiting, blurred vision, or changes in mood or behavior), receiving assessment for a concussion from a health care professional, and a diagnosis of a concussion from a health care professional.

## VIII. Health Promotion

### *Rotating Core*

In 2022 and rotating once every two years since 2020 is content about neighborhood characteristics of Sample Children aged 6–17 years. The questions are intended to provide information about the environment of the child which may create barriers for physical activity (Franzini, et al., 2010). Questions in the NHC section ask about the presence of neighborhood amenities favorable to outdoor activities and perception of neighborhood safety to engage in outdoor activities:

- Roads, sidewalks, paths, or trails for walking or biking on
- Parks or playgrounds in proximity of walking or biking to
- Traffic causing safety concerns for walking or biking
- Crime causing safety concerns for walking or biking

## Health Insurance

### *Annual Core*

The health insurance sections (INS) of the Sample Adult and Sample Child modules have a full range of items addressing health insurance such as coverage status, sources of coverage, characteristics of coverage, and reasons for no coverage. The flow and content of the questions pertaining to health insurance programs covered in the INS sections are similar to questions covered in the 1997–2018 NHIS Family Core. The main difference starting in 2019 and continuing through 2022 is that instead of asking health insurance for all family or household members, one adult and one child (if present) are selected from each household to receive these questions. The Sample Adult and Sample Child receive a similar set of questions with a few exceptions that will be outlined below.

### *Health Insurance Coverage Status*

An individual is considered currently insured if they currently have coverage through private health insurance, Medicare, Medicaid, Children’s Health Insurance Program (CHIP), military (TRICARE, Veterans Administration (VA), and CHAMP-VA), other state-sponsored health plans, or other government program. Individuals without any of the aforementioned types of coverages, with only Indian Health Service coverage, or a non-comprehensive plan that covers only dental, vision, or prescription drugs are considered uninsured.

For ease of analysis two recodes are available, NOTCOV\_A (on the Sample Adult file) and NOTCOV\_C (on the Sample Child file) that reflect this definition of noncoverage as used in *Health, United States* (in which persons with *only* Indian Health Service coverage or a single service plan that covers only dental, vision, or prescription drugs are considered uninsured).

### *Sources of Coverage*

Sample Adult and Sample Child respondents could identify one or more sources of medical care coverage, and single service plans were asked as separate questions. The following sources of healthcare coverage were collected in the interview:

- Private health insurance: Coverage obtained through employment or directly purchased (including Medigap plans)
- Medicare: The federal health insurance program for adults who are 65 and older, certain younger people with disabilities, and people with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant, sometimes called ESRD)
- Medicaid: A joint federal and state program that provide free or low-cost healthcare coverage to Americans, including some low-income people, families and children, pregnant women, the elderly, and people with disabilities

- CHIP (Children’s Health Insurance Program): A joint federal and state program that provides low- cost health coverage to children in families that earn above the income threshold to qualify for Medicaid
- Military: TRICARE, VA or CHAMP-VA
- Other state-sponsored health plans
- Other government program
- Indian Health Service: A part of the federal government that delivers direct medical and public health services to federally recognized Native American Tribes and Alaska Native people
- Single service dental plan
- Single service vision plan
- Single service prescription drug plan

### *Characteristics of Coverage*

For all coverage types, except for the Indian Health Service, additional follow-up questions specific to the type of coverage are asked. Some of these characteristics of coverage are broad and are relevant to more than one coverage type, whereas other characteristics are very specific to a particular type of coverage. In summary, the INS section includes detailed follow-up questions in the following areas:

- Health Insurance Marketplace, state exchanges or Healthcare.gov as to how coverage (private, Medicaid, CHIP, state-sponsored plans, other government programs) was obtained
- Enrollment in a high deductible health plan (private, Medicaid, CHIP, state-sponsored plans, or other government programs)
- Premium (private, Medicaid, CHIP, state-sponsored plans, or other government programs)
- Characteristics of private insurance (up to two plans per person)
  - exchange-based
  - policyholder
  - coverage of individuals other than the policyholder
  - relationship to the policyholder (only Sample Adult)
  - how plan was obtained (e.g., work, directly purchased)
  - who pays for the plan (e.g., self or family, work)
  - annual amount of premium paid by individual or family
  - prescription drug benefit
  - dental benefit
  - vision benefit
  - health savings account

- Characteristics of Medicare
  - Medicare Parts (A, B, C, D)
  - Medicare Advantage plan or Medicare managed care
  - Dental care coverage among those with Medicare Advantage plans (beginning 2022)
  - Vision care coverage among those with Medicare Advantage plans (beginning 2022)
  
- Types of military healthcare
  - VA (only Sample Adult)
  - TRICARE
  - CHAMP-VA

### *Continuity of Coverage*

For persons with coverage, they were asked if there was any time in the past 12 months when they did not have coverage. If they answered “yes,” they were asked for how many months they did not have coverage. For persons who were currently uninsured, they were asked when the last time was that they had coverage. If it was less than a year, they were asked for how many months they did not have coverage.

### *Reasons for No Coverage*

There are two sets of questions concerning the reasons for not having healthcare coverage. The first set focuses on reasons for no longer being enrolled in health coverage among those individuals who currently lacked coverage for less than three years. Reasons measured included the following: policyholder retired; lost a job or changed employers; a deadline was missed for signing up or paying for coverage; became ineligible due to age or leaving school; the cost of the coverage increased; and had Medicaid or other public coverage but were no longer eligible. The second set of questions focuses on reasons for not obtaining health coverage among all individual who currently did not have healthcare coverage. Reasons measured included the following: currently uninsured because coverage is not affordable; do not need or want coverage; not eligible for coverage; process of signing up is too difficult or confusing; cannot find a plan that meets needs; has applied for coverage but it has not started yet; and other reason. Based on coding open-ended responses, two additional categories are available as recoded variables. These additional reasons captured were retired, lost a job or changed employers and missing a deadline for signing up for coverage.

### *Replicate*

To reduce respondent burden, under certain circumstances a family who shared the same private plans were only asked about detailed characteristics of shared plans once, either in the Sample Adult interview or Sample Child interview, whichever occurred first. To be eligible for this replicate, the Sample Adult and Sample Child must be from the same family, the private plan must cover more than one person, the private plan must have a “valid” plan name, (i.e. it cannot have a refused or not known as the name of the plan), and the plan has to have information as to either where the plan was obtained or who pays for it, (i.e. it cannot have refused or not known to either of these fields). In addition, if either the Sample Adult or Sample Child indicated that they have Medigap coverage through HIKIND03\_A or HIKIND03\_C, they were not eligible for the replicate. Families may share up to two private plans. It is important to note that due to confidentiality concerns the HIKIND03\_C is not



available on the Sample Child file. Responses to HIKIND03\_C have been combined with responses to HIKIND01\_C and are made available in HIKIND01R\_C.

### *Processing Health Insurance Responses*

The INS sections use responses to follow-up questions to evaluate the reliability of the reported health insurance coverage and to adjudicate conflicting information. For many survey respondents, health insurance is a complex topic and some inconsistencies in survey responses are expected. If the responses to follow-up questions are inconsistent with the original health insurance coverage indicated, the original responses are edited. As a result, a portion of the Sample Adults and Sample Children are reassigned to a different type of coverage or reclassified from insured to uninsured (or vice versa). Conversely, follow-up responses in agreement with the original health insurance response are not edited and are included in the recodes. Therefore, it is best to use the recodes created, and listed in Table 21 below, for specific types of healthcare coverage and noncoverage because of the complicated editing process that takes place in the INS sections.

**Table 21. Annual core content of health insurance recoded variables for the Sample Adult and Sample Child**

Type of health insurance coverage	Sample Adult file	Sample Child file
Private health plans	PRIVATE_A	PRIVATE_C
Medicare	MEDICARE_A	OTHGOVR_C*
Medicaid	MEDICAID_A	MEDICAID_C
Children's Health Insurance Program (CHIP)	CHIP_A	CHIP_C
Military health plans	MILITARY_A	MILITARY_C
Indian Health Service	IHS_A	IHS_C
Other government programs	OTHGOV_A	OTHGOVR_C*
State-sponsored health plans	OTHPUB_A	OTHPUB_C
Uninsured	NOTCOV_A	NOTCOV_C
*This recode combines Sample Children covered by Medicare, other government programs or both.		

## Characteristics about the Sample Adult and Sample Child

### Annual Core

#### Sex, Age, Hispanic origin and Race

The NHIS collects information across different modules and sections that describe the Sample Adult and Sample Child's sex, age, Hispanic origin, and race. Table 22 lists the variables that summarized the final public-use variables on sex, age, Hispanic origin, and race available for them. These variables are described in the HHC section of the Codebooks.

**Table 22. Annual core content on public-use variables for sex, age, Hispanic origin, and race variables**

Description	Sample Adult variable	Sample child variable
<b>Sex</b>	SEX_A	SEX_C
<b>Age*</b>	AGEP_A	AGEP_C
<b>Hispanic origin</b>	HISP_A	HISP_C
<b>Hispanic group detail</b>	HISDETP_A	HISPDETP_C
<b>Single and multiple race groups</b>	RACEALLP_A	RACEALLP_C
<b>Single and multiple race groups combined with Hispanic origin</b>	HISPALLP_A	HISPALLP_C
<p>*During rostering, household respondents who refused or don't know the age of a household member are then asked a set of age range questions about the household member. The public-use data file includes information from the household respondent on whether the person selected to be the Sample Adult is under 65 years or age or 65 or older (AGE65). In the Sample Adult and Sample Child modules, the respondent can then provide the missing age, correct it, or also decline to provide this information. AGEP_A and AGEP_C are based on the final age information.</p>		

Starting in 2019, responses of “refused” or “don't know” to the sex and age questions are allowed. To preserve confidentiality, the ages of adults aged over 85 years are top-coded at 85 on the Sample Adult public-use data file.

In accordance with the Office of Management and Budget's Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (OMB, 1997) for the collection of ethnicity and race 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/Mexican American/Chicano, Central American, South American, Puerto Rican, Cuban, Dominican, or other Hispanic origin. Race is based on the Sample Adult's description of his or her own racial and ethnic identity, and an adult knowledgeable and responsible for the child's health provides this information for the Sample Child. More than one race can be reported for a Sample Adult and Sample Child.

Public-use race and Hispanic origin variables are imputed when unknown using hot deck imputation methods (Andridge and Little, 2010) where donor information was provided from either within the household or the local geographic area. Less than 1% of NHIS cases have imputed values for race or Hispanic origin. Race was not

imputed for Hispanic persons who did not identify with any of the OMB race categories and thus race was coded as ‘not ascertained’ for these persons. Imputation was implemented to determine whether the person was of any Hispanic or Latino origin but not to determine the country or geography of origin.

The confidentiality of respondents and their families could be compromised if there were extensive details available about the Sample Adult, Sample Adult’s spouse or partner, Sample Children and their parents, and other family members. Since 2019, detailed information about race and ethnicity that was previously available in public-use files is suppressed as more detailed geographic information (e.g., region and urbanization level) is included in the public use file. The following four single-race categories are available for Sample Adult and Sample Children in the public-use files: 1) White; 2) Black or African American; 3) Asian; and 4) American Indian or Alaska Native (AIAN). The only multiple race category available in the public-use files is AIAN and another race. Sample Adult and Sample Child respondents indicating a single race other than the four mentioned or reporting more than one race, other than including AIAN, were combined into the “other single and multiple races” category. Detailed Hispanic origin information available in the public-use files includes Mexican and Mexican American only.

Estimates, as three-year averages, of selected health topics for adults by detailed race and ethnicity are available at the NHIS website, see [Interactive Summary Health Statistics for Adults, by Detailed Race and Ethnicity](#).

For additional information about the historical context of race and ethnicity data collection in the NHIS, including editing, references for NHIS race and Hispanic origin, and background documents related to race and ethnicity data collection in federal data systems, see Appendix II of the 2012 NHIS *Survey Description*.

### *Nativity*

Information on place of birth (U.S. state or territory, or outside of the U.S.) and citizenship status is collected in the NAT section of the Sample Adults and Sample Child modules. To protect confidentiality, information on U.S. state or territory of birth and detailed citizenship status is not available in the public-use data files. Information about whether the Sample Adult or Sample Child was born in the U.S., and for those born outside the U.S., whether the Sample Adult or Sample Child is a U.S. citizen and years in the U.S. (categorical variable) is available in the public-use files.

### *Schooling*

The schooling sections (SCH) of the Sample Adult and Sample Child modules contain questions to determine the number of school days missed during the past 12 months. All Sample Children under aged 5–17 years, and Sample Adults aged 18 and over who were enrolled in or attending school at the time of interview were asked about the number of school days missed because of illness, injury, or disability during the 12 months prior to the interview. Persons responding for the Sample Children aged 0–17 years were asked if the child ever had a special education or early intervention plan, such as an Individualized Education Plan (IEP), or an Individualized Family Service Plan (IFSP). Those who responded “yes” were then asked if the child currently (sometime in the past school year) a special education or early intervention plan, and if he/she received these services to help with their emotions, concentration, behavior, or mental health.

## Education

Information on educational attainment, measured as the highest level of school or highest degree completed is asked about all adults in the household during the household roster module. The variable EDUCP\_A describes the highest educational attainment of the Sample Adult, top-coded to combine professional school degree (e.g., MD, DDS, DVM, JD) with doctoral degree (e.g., PhD, EdD). Additionally, similar recodes have been created that describe the level of educational attainment for the spouse or partner of the Sample Adult (described under “Characteristics about the Spouse or Partner of the Sample Adult”), and the parents of the Sample Child (described under “Characteristics about the Sample Child’s Parents”). Persons responding for the Sample Children are not asked about highest school grade or education the Sample Child has completed.

## Employment

The Sample Adult EMP section contains information regarding the Sample Adult’s work status in the week before the interview, main reason for not working for pay, when they last worked for pay, and among those who worked for pay in the past year, usual number of hours worked, work benefits, and days of sick leave taken.

The variable names for employment questions were revised in 2021 due to changes in question order and universe. Beginning in 2021, the annual content of the EMP section fielded during 2019-2020 was restructured to ask for the reason for not working for pay last week and the last time the Sample Adult worked for pay earlier in the set of questions. This change informed the universe of subsequent questions.

Sample Adults who did seasonal or contract work but did not work for pay last week were asked about the last time they worked for pay (and like in 2019-2020, it was also asked of those who did not work last week because they were retired, unable to work for health reasons, taking care of the house/family, going to school, or for some other reason, or refused or answered don’t know). Sample Adults who worked anytime in the past year (for pay or not for pay at family business) were asked about typical total number of hours worked at all job or businesses, and whether they work 35 or more hours was only asked as a follow-up question to those who refused or didn’t know how many hours they usually work or worked (both questions previously asked of those who only worked for pay last week, and hours worked were unknown or less than 35 hours). Lastly, Sample Adults who worked anytime in the past year were asked about job-sponsored health coverage, paid sick leave, and workdays missed due to illness or injury (previously asked of those working at or on temporary leave last week, performed seasonal/contract work regardless of when they last worked for pay, and those who work, but not for pay at a family business).

Sample Adults were first asked whether they worked for pay at a job or business last week; if not, they were asked if they had a job or business last week, but were temporarily absent due to illness, vacation, or family or maternity leave. Those who were working or temporarily absent from work were then asked how many hours in total they usually work at all jobs or businesses. Sample Adults who refused or did not know how many hours they usually work were asked if they usually work 35 or more hours per week in total at all their jobs or businesses.

Sample Adults who were not working last week or were not temporarily absent from a job or business last week were asked the main reason they were not working for pay at a job or business last week. Sample Adults who were unemployed, laid off, looking for work, performing seasonal or contract work, retired, unable to work for health reasons/disabled, taking care of the house or family, going to school, or had some other reason, as well as refused or don’t know responses were then asked when was the last time they worked for pay at a job or business, even if only for a few days.

Employed Sample Adults – those who were working last week, temporarily absent last week, performed seasonal or contract work in the past 12 months, or were working, but not for pay – and all other Sample Adults working within the past 12 months were asked whether paid sick leave is/was available and whether health insurance is/was offered through their workplace, as well as how many days of work they missed because of illness, injury or disability in the past year.

### *Employment Recodes*

Due to changes in the question order and universe in the EMP section starting in 2021, employment recode names have been updated.

**Employment status.** To ensure that the variables in this section are internally consistent with one another, a summary recode, EMPWRKLSW1\_A, identifies all employed Sample Adults, including those who were working during the last week; those with a job or business but temporarily absent the last week; those performing seasonal or contract work in the past 12 months; and those doing unpaid work. This recode updates EMPWRKLSW\_A from the 2019-2020 NHIS in that it now only includes those who reported seasonal/contract work if they worked in the past 12 months.

**Hours Worked.** Recode EMPWKHRS3\_A provides the hours worked last week top-coded at 95 for confidentiality. Recode EMPWRKFT1\_A identifies all employed Sample Adults who usually worked 35 or more hours per week. The two recodes about hours worked update those from 2019-2020 in that these recodes only include seasonal or contract workers if they worked in the past 12 months, and these seasonal/contract workers along with those who worked but not for pay at a family business are no longer coded as not ascertained because they are now asked about their work hours.

**Last worked.** EMPLSTWOR1\_A indicates the last time that Sample Adults who were not working in the last week or temporarily absent from their usual job or business had worked for pay, and Sample Adults performing unpaid work are coded as “8” or “not ascertained” on this recode. This recode differs from 2019-2020 in that those performing seasonal/contract work are no longer coded as not ascertained because they now provide information on when they last worked.

**Workdays missed.** To protect confidentiality, information on the number of workdays missed due to health reasons is top-coded at 130 days, EMPDYSMSS3\_A. This recode differs from 2019-2020 in that it no longer includes all seasonal/contract workers in the variable universe but only seasonal/contract workers who worked in the past 12 months.

### *Marital Status*

The MAR section contains information on the marital status of the Sample Adult. Sample Adults are first asked if they are “now married, living with a partner together as an unmarried couple, or neither.” Sample Adults who responded they are married are asked if their spouse lives in the same residence. If not living in the same residence, they are asked if this is because the Sample Adult and their spouse are legally separated. Sample Adults who are married or living with a partner are asked to verify the sex of their spouse or partner that was obtained during rostering. Sample Adults whose answer to the initial marriage and cohabitation question was other than “married,” are asked if they have ever been married. Sample Adults who are currently living with a partner and have been married are asked their current legal marital status – that is, whether they are currently

married, widowed, divorced, or separated. Sample Adults who are neither married nor living with a partner but have been married are asked if they are now widowed, divorced, or separated.

Additional information about the spouse or partner of the Sample Adult is described under “Characteristics about the Spouse or Partner of the Sample Adult.”

### *Parental Status*

The MAR section also includes recodes which describe whether the Sample Adult is a parent of a child residing in the family (PARSTAT\_A) and the Sample Adult’s relationship to Sample Child (SAPARENTSC\_A).

### *Sexual Orientation*

Sample Adults were asked about their sexual orientation. This question was asked before determining marital status of the Sample Adult and sex of the spouse or partner living in the household. Sexual orientation was not asked about the Sample Child.

### *Proxy Status*

Generally, Sample Adults provide information for themselves during the Sample Adult interview. However, in a small number of cases, proxy responses are allowed if the Sample Adult had a physical or mental condition that prevented them from responding. The variable PROXY\_A indicates those cases for which a proxy respondent provided the information.

### *Veteran Status*

The VET section contains information about the Sample Adult’s military veteran status and use of Veteran Affairs (VA) services. Specifically, it includes information on whether the Sample Adult ever served in the U.S. Armed Forces, military Reserves, or National Guard, if served in active duty for training or in a combat setting or humanitarian peace-keeping mission, and whether has received a service-connected disability rating.

Information about use of VA services focused on whether the Sample Adult sought care from at VA Hospital or other VA-affiliated facility in the past 12 months, and whether they have ever enrolled in or used VA healthcare.

## *Emerging Content*

### *Gender identity*

Experimental questions about gender identity were included in the GNI section of the Sample Adult module to examine alternative approaches for measuring gender identity and improve the measurement of sex and gender across federal surveys. Data from this section are not included in the public-use file and are available through the [Research Data Center](#).

The GNI questions were a three-part series that asked about:

1. Sex assigned at birth (using the binary responses of male and female).
- 2a. How they currently describe themselves using two sets of closed-ended responses:
  - I. male, female, transgender, with the option for none of the these
  - II. man, women, or some other way.
- 2b. A follow-up question to document a verbatim response for how they describe themselves among those answering none of these or some other way.
3. A confirmation question for discordant responses between the first two questions.

Sample Adults were randomized to one of four groups using a 2x2 experimental factorial design to examine the impact of varying: 1) question order for sex at birth and how they currently describe themselves; and 2) response options for currently described gender. For more information about measuring gender identity, see [Measuring Sexual Orientation and Gender Identity Research Group | FCSM.gov](#), [DPCPSI - Sexual & Gender Minority Research Office > Reports \(nih.gov\)](#), [Measuring Sex, Gender Identity, and Sexual Orientation |The National Academies Press](#).

## Characteristics about the Parents of the Sample Child

### *Annual Core*

Starting in 2019, the PAR section includes variables and recodes describing the demographic characteristics of parents residing with the Sample Child, type of relationship (biological, adoptive, step, foster) between the child and his or her parent(s), each parent's current and legal marital status, and information on nativity (whether the parent was born in the U.S. or a U.S. territory).

Detailed information on the type of relationship between Sample Child and parent is suppressed in the public-use data due to disclosure risks. Specifically, adoptive and biological children are combined in a single category, and foster children cannot be identified. In addition, while the redesigned NHIS allows for the collection of information on up to four parents as long as they all reside in the same household as the child, only information on the first two are released in the public-use file. Instances where a Sample Child had three or four residential parents are rare. Detailed information for the parents residing with the Sample Child is available through the NCHS RDC. For a list of restricted variables, see the PAR section of the Sample Child Codebook for restricted variables.

Several recodes are available in the PAR section of the public-use data that describe the demographic characteristics of up to two parents residing with the Sample Child. The information for these recodes was obtained from questions asked in various sections of the Sample Child module (i.e., PAR, GEN, HHC). Table 23 lists the available recodes, which include parental sex, age (bottom-coded at 20 and top-coded at 65), education of the parent with the highest educational attainment (top-coded to combine professional school degree and doctoral degree), employment status (including full- versus part-time work), current marital status, legal marital status, and whether the Sample Child's parents are of the same sex or of opposite sex.

To protect confidentiality, detailed information about race and ethnicity of the Sample Child's resident parents is suppressed. To assist data users interested in studying interracial families, the PAR section also includes four recodes to indicate (1) whether the Sample Child's race is the same as all parents in the household, (2) whether the Sample Child and all the parents in the household are of same Hispanic or Latino origin category (i.e., Hispanic, non-Hispanic), (3) whether the Sample Child's parents are of the same race to each other, and (4) whether the Sample Child's parents are of the same Hispanic or Latino origin category to each other. These recodes are "yes" and "no" answers. The "yes" response means that the Sample Child and all resident parents are in the same racial category, or the Sample Child and all resident parents are in the same Hispanic or Latino origin category. A "no" response means that either the Sample Child and at least one of the parents is of a different racial category, or that either the Sample Child and at least one of the parents are not of the same Hispanic or Latino origin category, respectively. Similarly, same race and same Hispanic or Latino origin category between parents is based on a common racial and Hispanic or Latino origin category. Same race is in reference to the racial categories available in the public-use file, that is White, Black or African American, Asian, AIAN, AIAN and another race, and all other single and multiple races. Same Hispanic or Latino origin category is based on whether the child and all parents, and whether all the parents, are of any Hispanic or Latino ethnicity (regardless of country or area of origin). Sample children residing with one parent in the household are categorized as missing (not in universe) in the variables categorizing whether the Sample Child's parents are of the same race to each other, and whether the Sample Child's parents are in the same Hispanic or Latino origin category.

Additional information about the family of the Sample Child is described under "Characteristics about the Family and Household of the Sample Adult and Sample Child."



**Table 23. Annual core content of public-use recodes in the PAR section of the Sample Child codebook describing demographic characteristics of the parents residing with the Sample Child.**

Description	Residential parent 1 variable	Residential parent 2 variable	Variable for both residential parents
Relationship type to Sample Child <sup>1</sup>	RELCHPARENTP1_C	RELCHPARENTP2_C	
Sex	PARSEX1_C	PARSEX2_C	
Age	PARAGETC1_C	PARAGETC2_C	
Education of the Sample Child's parent with the highest education			MAXPAREDUP_C
Working last week	PARWORK1_C	PARWORK2_C	
Working full-time last week (35+ hours)	PARWKFT1_C	PARWKFT2_C	
Current marital status	MARSTAT1_C	MARSTAT2_C	
Legal marital status	LEGMSTAT1_C	LEGMSTAT2_C	
Two parents of same or of opposite sex			PARSAMEOPP_C
Sample child's Hispanic ethnicity is the same as both parents			SCPARHISP_C
Sample child's race is the same as both parents			SCPARRAC_C
Sample child's parents are of the same Hispanic ethnicity			HISPPARSC_C
Sample child's parents are of the same race			RACPARSC_C

<sup>1</sup>In 2021, the variables for categorizing the type of relationship between the Sample Child and each residential parent were updated to more clearly identify when there was no residential parent or only one residential parent in the household. The variables of RELCHPARENTP1\_C and RELCHPARENTP2\_C replace 2019–2020 variables RELCHPARP1\_C and RELCHPARP2\_C.

NOTES. The Sample Child's parents are labeled as parent 1 and parent 2 based in the order that this information was provided by the respondent and entered by the interviewer during the interview for questions WHOPAR and WHOFOST. All information in the recodes about parent 1 (e.g., sex, age) are about the same person, and similarly for parent 2.

## Characteristics about the Spouse or Partner of the Sample Adult

### Annual Core

Starting in 2019, the MAR section includes several recodes describing the demographic characteristics of the spouse or partner living with the Sample Adult, if married or cohabiting. The information for these recodes was obtained from questions asked in various sections of the Sample Adult module (i.e., MAR, GEN, HHC). These recodes are shown in Table 24 and include the spouse or partner’s sex, age (bottom-coded at 20 and top-coded at 85 years), education (top-coded to combine professional school degree and doctoral degree), current work status, and whether the spouse or partner worked full-time (35 or more hours per week).

For data users interested in studying interracial families, recodes of the race and ethnicity of the spouse or partner in the household are available in reference to the Sample Adult’s race and ethnicity. These recodes are “yes” and “no” answers, where a “yes” response means that Sample Adult and the spouse or partner are in the same racial category, or the same Hispanic or Latino origin category (i.e., Hispanic, non-Hispanic), and a “no” response means that the Sample Adult and the spouse or partner are of a different racial category, or Hispanic or Latino origin category, respectively. Same race is in reference to the racial categories available in the public-use file, that is White, Black or African American, Asian, AIAN, AIAN and another race, and all other single and multiple races. Same Hispanic ethnicity is based on any Hispanic or Latino ethnicity (regardless of country or area of origin).

Additional information about the family of the Sample Adult is described under “Characteristics about the Family and Household of the Sample Adult and Sample Child.”

**Table 24. Annual core content of public-use recoded variables in the MAR section of the codebook describing demographic characteristics of the spouse or partner residing with the Sample Adult**

Description	Variable for spouse	Variable for cohabiting partner
Sex	SPOUSESEX_A	PRTNRSEX_A
Age	SPOUSAGETC_A	PRTNRAGETC_A
Hispanic ethnicity of Sample Adult and spouse/partner are the same	SASPPHISP_A	SASPPHISP_A
Race of Sample Adult and spouse/partner are the same	SASPPRACE_A	SASPPRACE_A
Education	SPOUSEDUCP_A	PRTNREDUCP_A
Working last week	SPOUSWRK_A	PRTNRWRK_A
Working full-time last week (35+ hours)	SPOUSWKFT_A	PRTNRWKFT_A

## Characteristics about the Family and Household of the Sample Adult and Sample Child

### Annual Core

#### Family and Household Composition variable

Table 25 provides a list of various recodes included on the public-use Sample Adult and Sample Child data files that describe the families and households in which the Sample Adult and Sample Child live. Flag variables indicate source respondent for each module and whether the sample unit was a single or multiple family household. Also listed are recodes that indicate the level of education attained by the adult with the highest education in the Sample Adult's and Sample Child's family. In addition, several top-coded counters provide the number of family members in various age groups in the household.

**Table 25. Annual core content of public-use recodes of family and household composition and counter variables available for Sample Adult and Sample Child**

Description	Codebook section	Variable name for Sample Adult's family	Variable name for Sample Child's family
Sample Adult is the household respondent or the proxy who lives in the household	FLG	HHRESPSA_FLG	
Sample child respondent is the household respondent	FLG		HHRESPSC_FLG
Number of adults in the Sample Adult's/Sample Child's family (top-coded)	FAM	PCNTADLT_A	PCNTADLT_C
Number of children in the Sample Adult's/Sample Child's family (top-coded)	FAM	PCNTKIDS_A	PCNTKIDS_C
Indicator for at least one person is 65 and over in the Sample Adult's/Sample Child's family	FAM	OVER65FLG_A	OVER65FLG_A
Education of the adult with the highest education in the Sample Adult's/Sample Child's family (top-coded)	FAM	MAXEDUCP_A	MAXEDUCP_C
Flag indicating Sample Adult/Sample Child lives in households containing more than one family	FAM	MLTFAMFLG_A	MLTFAMFLG_C
Counters of persons aged 0–17 years in the Sample Adult's/Sample Child's households (top-coded)	HHC	PCNTLT18TC	PCNTLT18TC
Counters of persons aged 18 years and older in the Sample Adult's/Sample Child's households (top-coded)	HHC	PCNT18UPTC	PCNT18UPTC

### Family Employment

The family employment section (FEM) contains employment information for all related adults in the Sample Adult's and Sample Child's families. To reduce respondent burden, these questions were asked once per family. Responses to these questions were used to create recodes that list the number of persons aged 18 and older who work and who work full-time (i.e., 35 or more hours per week) in the Sample Adult and Sample Child family. These recodes are top-coded for confidentiality, see Table 26 a list of family employment variables available for public-use.

**Table 26. Annual core content of employment counter variables available in the Sample Adult or Sample Child files.**

Description	Module, section	Variable Name	Notes
<b>Number of adults in the Sample Adult's family who are working</b>	Sample Adult, FEM	PCNTADTWKP_A	Top-coded at 3.
<b>Number of adults in the Sample Adult's family who are working full-time (35 or more hours per week)</b>	Sample Adult, FEM	PCNTADTWFP_A	Top-coded at 3.
<b>Number of adults in the Sample Child's family who are working</b>	Sample Child, FEM	PCNTADTWKP_C	Top-coded at 3.
<b>Number of adults in the Sample Child's family who are working full-time (35 or more hours per week)</b>	Sample Child, FEM	PCNTADTWFP_C	Top-coded at 3.

### Family Income

The family income section (INC) contains information regarding a variety of income sources, as well as estimates of total combined family income. All questions are asked once per family, using the family-level-replicate interviewing approach. Respondents are told at the start of the family income section that all questions are seeking information about possible income sources in the previous calendar year, and the names of all family members (collected earlier) to consider when responding. No personal earnings information is collected as part of the redesigned questionnaire.

Respondents were asked whether anyone in the family received income from a variety of sources (e.g., wages, salary from self-employment, social security, railroad retirement, government assistance). Respondents are also asked to report their "best estimate" of their family's total income (in dollars) from all sources for all family members living in the household before taxes in the last calendar year. 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 obtains 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 estimated total family income. These 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 asking respondents about the family's income relative to specific dollar values (i.e., \$75,000, \$100,000, and

\$150,000), these respondents were also asked about the family's income relative to the federal poverty threshold (100%, 138%, 200%, 250% and 400%) and take into account each family's size (collected earlier in the interview).

The poverty thresholds used in the questionnaire, and shown in Table 27, are intended to approximate the U.S. Census Bureau's weighted average poverty thresholds for 2021. Because these values were not available when the 2022 NHIS instrument was created, the poverty thresholds used in the 2022 instrument were derived by NCHS from the 2020 poverty thresholds by size of family and number of related children under 18 years, the average Consumer Price Index for all urban consumers (CPI-U) from 2020, the forecasted annual growth rate of the CPI-U for 2021, actual monthly CPI values (all consumers) for January-July 2021, and projected CPI values (all consumers) for August-December 2021.

**Table 27. Poverty thresholds used in the 2021 NHIS Instrument, by family size.**

Family Size	100% of the federal poverty level	138% of the federal poverty level	200% of the federal poverty level	250% of the federal poverty level	400% of the federal poverty level
1 person < 66 years	\$14,000	\$19,000	\$28,000	\$35,000	\$56,000
1 person ≥ 66 years	\$13,000	\$18,000	\$26,000	\$32,000	\$52,000
2 persons, both < 66	\$18,000	\$25,000	\$36,000	\$45,000	\$73,000
2 persons, 1 is ≥ 66	\$16,000	\$23,000	\$33,000	\$41,000	\$65,000
3 persons	\$21,000	\$30,000	\$43,000	\$54,000	\$86,000
4 persons	\$28,000	\$38,000	\$55,000	\$69,000	\$110,000
5 persons	\$33,000	\$45,000	\$65,000	\$82,000	\$131,000
6 persons	\$37,000	\$51,000	\$74,000	\$92,000	\$148,000
7 persons	\$42,000	\$58,000	\$84,000	\$105,000	\$168,000
8 persons	\$47,000	\$64,000	\$93,000	\$117,000	\$187,000
9 or more persons	\$56,000	\$78,000	\$112,000	\$140,000	\$225,000

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 the previous year was less than the applicable poverty threshold, or if the family's income was greater than or equal to that same poverty threshold.

In 2022, 76.6% of Sample Adults and 81.5% of Sample Child respondents provided their family income. Missing family income for Sample Adults (23.4%) and Sample children (18.5%) were imputed using family income bracket responses or other survey information. Missingness on family income is not completely at random and excluding observations with missing income information can result in biased analyses. Reported and imputed family income information are used to create poverty ratio (see, Recodes of Family Income and Imputed Family Income, below).

To protect confidentiality, continuous and categorical family income, reported and imputed, as well as the variables obtained from the income bracketing questions are not available on the NHIS public-use data files. See Appendix for availability of restricted income questions.

### Recodes of Family Income and Imputed Family Income

Missing data on family income and earnings in the NHIS are imputed using a multiple imputation methodology. Imputation is the process of replacing missing data with substituted values based on information collected from other observations in the dataset. Multiple imputation accounts for the extra variability due to imputation in statistical analyses.

Ten multiply imputed income data values are created for the Sample Adult and the Sample Child's families. Recent literature on multiple imputation analysis suggests that increasing the number of imputations (e.g., to 10 or higher) produces more precise estimates for a wide variety of analyses (van Buuren, 2012). The family income is imputed within the lower and upper bound when the income bracketing questions are answered.

Ten sets of top-coded continuous poverty ratios and grouped poverty ratios are available for the Sample Adult and Sample Child families. These recodes incorporate information from reported and imputed total family income and are included in the respective income files (Table 28) as a continuous ratio of total family income and family size relative to the poverty threshold and as a finite number of categories.

**Table 28. Top-coded variables for family poverty ratio, and related flags available in the Sample Adult and Sample Child files**

Description	Sample Adult file (adultinc22)	Sample Child file (childinc22)
<b>Top-coded poverty ratio</b>	POVRATTC_A	POVRATTC_C
<b>Grouped poverty ratio</b>	RATCAT_A	RATCAT_C
<b>Imputed income flag</b>	IMPINFLG_A	IMPINFLG_C
<b>Imputation number*</b>	IMPNUM_A	IMPNUM_C

\*This variable was named IMPNUM in 2019-2020.

In cases where the Sample Adult and Sample Child are in the same family, these corresponding values are identical. The poverty ratio variable is top-coded to the mean of the top 95<sup>th</sup> percentile of the 20 distinct imputations from the Sample Adult and Sample Child inhouse poverty ratio values (i.e., 10 poverty ratio values from the Sample Adult and 10 poverty ratio values from the Sample Child).

All 10 imputations are stacked in a single file with a variable (IMPNUM\_A or IMPNUM\_C) indicating the imputation number in each file. Stacking the 10 multiple imputation datasets into one allows for fewer steps in data preparation for analyses in SAS and Stata. (For sample code on how to use multiply imputation in analysis, see the section 'Merging Survey Data and Imputed Income Files.')

For the convenience of analyses that don't need or use multiple imputed data, the same variables, i.e., POVRATTC\_A and RATCAT\_A (for the Sample Adult's family) and analogously for the Sample Child family from a single imputation are also available in the Sample Adult and Sample Child files, respectively.

**Analysts are reminded that imputed income files for each year should be merged with the relevant data files for that year before concatenating data files for multiple years.**

For technical information about the imputation model, please refer to the "Imputed Income Technical Document" available with the annual file releases on the NHIS website, under "Using the NHIS."

### *Food Related Programs*

The food related programs (FOO) section includes three questions to ascertain past 12-month participation in the food assistance program Supplemental Nutrition Assistance Program (SNAP), free or reduced-cost breakfasts or lunches at school, and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC).

These are family-level replicate questions asked once per family. The universes for the questions in this section coincide with program eligibility. Sample Adult and Sample Child respondents were asked about SNAP assistance program participation. Sample Adult and Sample Child respondents living in families with females 12–55 years of age or children 0–5 years of age were asked about WIC program participation. Sample Adult and Sample Child respondents living in families with children between the ages of 5–17 were asked about free or reduced-cost breakfasts or lunches at school.

### *Housing*

The housing section (HOU) collects information on housing tenure, length of residence, and participation in Federal, State, or local government housing assistance programs among renters. To reduce respondent burden, these questions were asked once per family.

### *Region and urbanization level*

Geographical classification of the U.S. population is provided on the NHIS in two ways: region and urban-rural classification. In the geographical classification, states are grouped into four regions used by the U.S. Census

Bureau: Northeast (Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, and Pennsylvania), Midwest (Ohio, Illinois, Indiana, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Kansas, and Nebraska), South (Delaware, Maryland, District of Columbia, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Oklahoma, Arkansas, and Texas), and West (Washington, Oregon, California, Nevada, New Mexico, Arizona, Idaho, Utah, Colorado, Montana, Wyoming, Alaska, and Hawaii).

The urban-rural classification is based the 2013 NCHS Urban-Rural Classification Scheme for Counties, which groups U.S. counties and county-equivalent entities into six urban-rural categories: large central metro, large fringe metro, medium metro, small metro, micropolitan, and non-core (Ingram and Franco, 2014). The categorization included on the NHIS public-use files combine medium and small metropolitan areas into a single group and micropolitan and non-core areas into a single group to yield the following four categories: large central metro, large fringe metro, medium and small metro, and nonmetropolitan. Additional information on the development of this classification scheme can be found in “2013 NCHS Urban-Rural Classification Scheme for Counties” available at: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_166.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf).

### *Telephone Use*

The telephone use section (TEL) contains information regarding the availability of a working non-cellular telephone (land line) and of a cellular (wireless, mobile) telephone in the home, and for adults with both types of working telephones, whether the calls received are answered “all or almost all,” “some,” or “very few or none” on the cell phone. The purpose of the telephone questions is to track the percentages of wireless-only adults and of children living in homes with only wireless telephones. The telephone use data are in the Paradata file. Variable information is found in the TEL sections (adult and child) of the Paradata codebook.

### *Telephone Use Recodes*

Recodes that classify the telephone status of the household (HH\_PLASS) and the Sample Adult (AD\_PLASS) are included in the Paradata file. The response categories in the recodes correspond to the classification presented in the NCHS report “Wireless Substitution: Early Release of Estimates from the National Health Interview Survey” <https://www.cdc.gov/nchs/nhis/erwirelessubs.htm>.

### *Sponsored Content*

#### *Food Related Programs*

The FOO section included a question sponsored by the USDA about participation in the Supplemental Nutrition Assistance Program (SNAP) in the last month. This question was asked of Sample Adult and Sample Child respondents who reported receiving SNAP benefits in the last 12 months. To reduce respondent burden, this question was asked once per family.



## Family Food Security

Family food security refers to access at all times to enough food for active, healthy lives. The food security questions in the FDS sections of the Sample Adult and Sample Child modules are sponsored by the USDA and consist of the same 10 questions measuring food security status of families in the past 30 days. The first three questions asked about being worried that food would not last until there is money to buy more, food bought did not last and didn't have money to buy more, and not being able to afford to eat balanced meals. The next set of statements asked about cutting or skipping meals, eating less than should, being hungry but not eating, losing weight, and not eating for a whole day because there wasn't enough money to buy food. Respondents were also asked for the number of days that a meal was skipped, and the number of days that they did not eat for a whole day. During 2011–2018, the food security questions were administered as part of the Family Component at the beginning of the interview. Starting in 2019, the food security questions are administered once per family in the later portion of the Sample Adult and Sample Child modules using the family-level replicate interviewing approach. Responses to food security questions can be combined to create a raw food security score and categories for degree of food insecurity (see family food security recodes). The purpose of the questions is to examine the relationship between health and food insecurity. For more information about the USDA's food security research and standard procedures for measuring food insecurity and hunger in the United States, see <https://www.fns.usda.gov/sites/default/files/FSGuide.pdf>.

## Family Food Security Recodes

The ten questions in the Food Security (FDS) section were used to determine a 3 and 4 level scale of food security status of adults and children as recommended by the USDA Economic Research Service. The food security status classification variables were derived from a raw food security score created to represent the number of affirmative responses 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. Each affirmative response has a score of 1 for a total score ranging from 0 to 10. Not all ten questions were asked of all respondents. Respondents who answer "never true" to the first three questions (with a score of zero for each question) are determined to be food secure and are not asked additional questions. Subsequently, only those with an affirmative answer to questions that specify because there wasn't enough money for food, they cut the size or skipped meals, ate less than they should, were hungry but didn't eat, or lost weight, were asked about not eating for a whole day, and the number of days that occurred. Respondents who answered "don't know" or "refused" or whose answers were not ascertained to the first three questions in the food security set are classified as not ascertained food security status and coded as 8. Information from any affirmative response was summed to the raw score, including when respondents answered "don't know" or "refused" or whose answers were not ascertained for questions that came after the initial three food security questions within the set.

Two options for food security status classification variables were created: 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)

## NCHS Data Linkage Program

The Data Linkage Program at NCHS is a cross-cutting program housed in the Division of Analysis and Epidemiology (DAE) which aims to maximize the scientific value of the Center’s population-based surveys by linking NCHS survey data with data collected from vital and other administrative records. Linked data files enable researchers to augment information for major diseases, risk factors, and health services utilization, by linking exposures to outcomes and in some cases introducing a longitudinal component to survey data.

For more information on the NCHS Data Linkage Program, please visit the NCHS Data Linkage website: <https://www.cdc.gov/nchs/data-linkage/index.htm>

### Data Sources Linked

The Data Linkage Program currently links NHIS data with:

- National Death Index (NDI) death certificate data, including cause of death (<https://www.cdc.gov/nchs/data-linkage/mortality.htm>)
- Centers for Medicare and Medicaid Services (CMS) enrollment and claims data for:
  - o Medicare (<https://www.cdc.gov/nchs/data-linkage/medicare.htm>)
  - o Medicaid/CHIP (<https://www.cdc.gov/nchs/data-linkage/medicaid.htm>)
- Department of Housing and Urban Development (HUD) administrative data from the largest rental housing assistance programs (<https://www.cdc.gov/nchs/data-linkage/hud.htm>)
- United States Renal Data System (USRDS) data on End Stage Renal Disease (ESRD) (<https://www.cdc.gov/nchs/data-linkage/esrd.htm>)
- Department of Veterans Affairs (VA) administrative data, including use of VA benefits (<https://www.cdc.gov/nchs/data-linkage/va.htm>)

For a list of linked surveys and administrative data by year, please see [https://www.cdc.gov/nchs/data/datalinkage/LinkageTable\\_1.pdf](https://www.cdc.gov/nchs/data/datalinkage/LinkageTable_1.pdf).

### Linkage Methodology

Only NHIS participants who have provided consent as well as the necessary personally identifiable information (PII) are considered *linkage-eligible*. *Linkage-eligibility* is distinct from program eligibility, which refers to whether a person meets eligibility criteria for a benefits program. *Linkage eligibility* refers to the potential ability to link data obtained from an NHIS participant to administrative data. Survey participants are informed of NCHS’ intent to conduct data linkage activities through a variety of procedures such as “advance letters,” participant brochures, and during the interview when verbal consent is requested. NHIS participants selected to be the Sample Adult or Sample Child (by proxy respondent) are asked for the last four digits of their Social Security Number (SSN) and Medicare Beneficiary Identifier (MBI) for participants aged 65 and older. Additionally, those who refused to provide the last four digits of their SSN or MBI are asked if they would consent to linkage based on their other identifying information. Only Sample Adult and Sample Child participants who provided the last four digits of SSN or MBI or provided consent for linkage without SSN or MBI are included in linkage activities for 2007 NHIS forward. Since 2010, approximately 90% of NHIS Sample Adult participants are linkage eligible.

Questions to determine linkage eligibility are collected in the LNK sections of the Sample Adult and Sample Child

interview, and associated variables are stored in the public-use Paradata file.

The individual-level linkages are conducted using both probabilistic and deterministic techniques. The algorithms rely on PII such as SSN, name, and date of birth. Please refer to the appropriate linkage documentation for further information on methodology and analytic considerations (for example for the linked NDI data, <https://www.cdc.gov/nchs/data-linkage/mortality-methods.htm>).

## Geocoded Data

Sample addresses from NHIS are also geocoded to standard Census geocoded areas. This enables researchers to merge contextual data (e.g., county level data, air quality data) with NHIS data. The geocodes are considered restricted-use and only available through the NCHS or Federal Statistical Research Data Centers (RDCs).

## Public-Use Linked Data

Public-use Linked Mortality Files (LMFs) are available containing a limited set of mortality variables for adult participants only. The public-use versions of the LMFs were subjected to data perturbation techniques to reduce the risk of participant re-identification. For select records, synthetic data were substituted for follow-up time or underlying cause of death. Information regarding vital status was not perturbed. For more information on the public-use LMFs, please visit the public-use LMF webpage: <https://www.cdc.gov/nchs/data-linkage/mortality-public.htm>

## Restricted-Use Linked Data

All other linked data files are restricted-use and available only through the NCHS or Federal Statistical RDCs. For more information about the restricted-use linked data, including the file contents, methods used for linkage and analytic consideration, follow the links provided for each of the following data linkages:

- National Death Index (NDI), Restricted-Use Linked Mortality Data (<https://www.cdc.gov/nchs/data-linkage/mortality-restricted.htm>)
- NCHS-CMS Medicare (<https://www.cdc.gov/nchs/data-linkage/medicare-restricted.htm>)
- NCHS-CMS Medicaid (<https://www.cdc.gov/nchs/data-linkage/medicaid-restricted.htm>)
- NCHS-HUD (<https://www.cdc.gov/nchs/data-linkage/hud-restricted.htm>)
- NCHS-USRDS ESRD (<https://www.cdc.gov/nchs/data-linkage/esrd-restricted.htm>)
- NCHS-VA (<https://www.cdc.gov/nchs/data-linkage/va-restricted.htm>)
- Geocoded data ([https://www.cdc.gov/rdc/geocodes/geowt\\_nhish.htm](https://www.cdc.gov/rdc/geocodes/geowt_nhish.htm))

For more information about accessing the restricted-use linked data, please visit the NCHS RDC website: <https://www.cdc.gov/rdc/index.htm>

---

## Medical Expenditure Panel Survey (MEPS)

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.

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: [https://meps.ahrq.gov/mepsweb/data\\_stats/more\\_info\\_download\\_data\\_files.jsp#hc-nhis](https://meps.ahrq.gov/mepsweb/data_stats/more_info_download_data_files.jsp#hc-nhis).

---

## References

- The American Association for Public Opinion Research (AAPOR). (2016). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (9th edition). AAPOR.
- American Psychological Association. (2022). Life satisfaction. APA Dictionary of Psychology. Accessed 5/9/2022. Available at <https://dictionary.apa.org/life-satisfaction>
- Andridge RR, Little RJ. (2010). A Review of Hot Deck Imputation for Survey Non-response. *Int Stat Rev*, 78(1):40-64.
- Bansal R, Gubbi S, Koch CA. (2022). COVID-19 and chronic fatigue syndrome: An endocrine perspective. *Clin. Transl. Endocrinol*; 27:100284.
- Benda NC, Veinot TC, Sieck CJ, et al. (2020) Broadband Internet Access Is a Social Determinant of Health! *Am J Public Health*, 110(8): 1123–1125.
- Cheung F., Lucas RE. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Qual Life Res*, 23(10), 2809-2818.
- Danso K. 2017. "Immigrant health disparities: Does neighborliness improve health?" *J Sociol Soc Welf*, 44(3): 75-94.
- DuBois DL., Silverthorn N. (2005). Natural Mentoring Relationships and Adolescent Health: Evidence from a National Study, *Am J Public Health*, 95(3), 518-524.
- Duplaga M., Szulc K. 2019. "The association of Internet use with wellbeing, mental health and health behaviors of persons with disabilities." *Int J Environ Res Public Health*, 16(18), 3252.
- Early J, Hernandez A. Digital Disenfranchisement and COVID-19: Broadband Internet Access as a Social Determinant of Health. *Health Promot Pract*. 2021;22(5):605-610.
- Executive Order 13985 on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government 86 FR 7009 7009-7013 (2021).
- Felitti VJ., Anda RF., Nordenberg D., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) study. *Am J Prev Med*, 14:245-258.
- Franzini, L., Taylor, W., Elliott, M.N., et al. (2010). Neighborhood characteristics favorable to outdoor physical activity: Disparities by socioeconomic and racial/ethnic composition, *Health & Place*, 16(2), 267-274.
- Goodman A, Goodman R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J. Am. Acad. Child Adolesc. Psychiatry*, 48:400-3.
- Goodman R. (1997). The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry*, 38(5):581-6.
- Heaps W, Abramsohn E, Skillen E. Public Transportation in the US: A Driver of Health and Equity. *Health Affairs Health Policy Brief*. 2021.

- Ingram DD, FrancoSJ. (2014). 2013 NCHS urban–rural classification scheme for counties. National Center for Health Statistics, Vital Health Stat, 2(166).
- Kearns A, Whitley E. 2019. “Associations of Internet access with social integration, wellbeing and physical activity among adults in deprived communities: Evidence from a household survey.” BMC Public Health, 19:860.
- Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. (2004). Diet, nutrition and the prevention of cancer. Public Health Nutr;7(1a):187-200.
- Khalaila R, Vitman-Schorr A. 2018. “Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: Mediating and moderating effects.” Qual Life Res, 27: 479-489.
- Korn, E.L., Graubard, B.I. (1999). Analysis of Health Surveys, John Wiley & Sons.
- Kroenke K, Spitzer RL, Williams JBW. (2001). The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med, 16: 606–613.
- Kroenke K, Spitzer RL, Williams JB. (2003). The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. Med Care, 41:1284-92.
- Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann. Intern. Med, 146(5):317-325.
- Kroenke K, Spitzer RL, Williams JB, Löwe B. (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry, 32(4):345-359.
- Kroenke K, Strine TW, Spitzer RL, Williams JB, et al. (2009) The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord, 114, 163–173.
- Löwe B, Decker O, Müller S, Brähler E, et al. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care. 46, 266–274.
- Nelson C, Sloan J, and Chandra A. 2019. Examining civic engagement links to health: Findings from the literature and implications for a culture of health. Santa Monica, CA: RAND Corporation.
- Ochiai E, Kigenyi T, Sondik E, et al. (2021). Healthy People 2030 Leading Health Indicators and Overall Health and Well-being Measures: Opportunities to Assess and Improve the Health and Well-being of the Nation. JPHMP, 27(1), S235-S41.
- Olson, K. (2013). Paradata for nonresponse adjustment. Ann Am Acad Pol Soc Sci, 645(1):142-170.
- OECD (2013), OECD Guidelines on Measuring Subjective Well-being, OECD Publishing, Paris, <https://doi.org/10.1787/9789264191655-en>.
- Penchansky R., Thomas J.W. (1981). The Concept of Access: Definition and Relationship to Consumer Satisfaction. Med Care, 19, 127-140.
- Prentice JC, Pizer SD. Delayed access to health care and mortality. Health Serv. Res, 2007;42(2):644-662.

- Rutter LA, Brown TA. (2017). Psychometric Properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in Outpatients with Anxiety and Mood Disorders. *J Psychopathol Behav Assess*, 39(1):140-146.
- Schneider SK, O'Donnell L, Stueve A, Coulter RWS. (2012). Cyberbullying, School Bullying, and Psychological Distress: A regional Census of High School Students. *Am J Public Health*, 102(1):171-177.
- Sheldrick RC, Henson BS, Neger EN, et al. (2013). The Baby Pediatric Symptom Checklist: Development and initial validation of a new social/emotional screening instrument for very young children. *Acad Pediatr*, 13:72- 80.
- Spitzer RL, Kroenke K, Williams JB, Löwe B. (2006). Brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. intern. med*, 166:1092-7.
- Stopka TJ, Feng W, Corlin L, King E, Mistry J, Mansfield W, Wang Y, Levine P, and Allen JD. 2022. "Assessing equity in health, wealth, and civic engagement: A nationally representative survey, 2020." *Int. J. Equity Health*, 21(1):12.
- Stussman BJ, Taylor BL, Riddick H. (2003). Partials and Break-offs in the National Health Interview Survey, 2002. Paper presented at the 2003 Federal Committee on Statistical Methodology Research Conference, Arlington, Virginia.
- Toth, D. (2019). *rpms: An R Package for Modeling Survey Data with Regression Trees*. Available at: [https://mran.microsoft.com/snapshot/2018-08-19/web/packages/rpms/vignettes/rpms\\_2018\\_01\\_22.pdf](https://mran.microsoft.com/snapshot/2018-08-19/web/packages/rpms/vignettes/rpms_2018_01_22.pdf)
- U.S. Food & Drug Administration. 2023. COVID-19 Vaccines. Available at: <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines>.
- U.S. Department of Health and Human Services. (2018). *Physical Activity Guidelines for Americans*, 2nd edition. Washington, DC: U.S. Department of Health and Human Services. Available from: [https://health.gov/sites/default/files/2019-09/Physical\\_Activity\\_Guidelines\\_2nd\\_edition.pdf](https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf).
- U.S. Department of Health and Human Services, Office of the Assistant Secretary for Health. (2022). *National Research Action Plan on Long COVID*, 200 Independence Ave SW, Washington, DC 20201 Available at <https://www.covid.gov/assets/files/National-Research-Action-Plan-on-Long-COVID-08012022.pdf>
- Valliant R, Dever JA, Kreuter F. (2018). *Practical tools for designing and weighting survey samples*. New York: Springer.
- van Buuren S. (2012). *Flexible Imputation of Missing Data* (2nd Edition).
- Zablotsky B, Black LI, Sheldrick RC, Perrin EC, Blumberg SJ. (2022). Assessing the Validity of the Baby Pediatric Symptom Checklist Using a Nationally Representative Household Survey. *Acad Pediatr*, 6:S1876-2859(22)00559-9.



## Appendix: Availability of Health, Health Insurance and Selected Demographic Questions Not Included in the Sample Adult and Sample Child Public-use Files

Table. List of questions not available in the public-use files and availability through the Research Data Center (D), as a recode (R), or not available (NA): 2022 NHIS					
Module	Section Acronym	Questionnaire Variable	Description	Availability	Public-use Recode
Roster	HHC	EDUC	Highest level of education completed	D R	EDUCP_A MAXPAREDUP_C SPOUSEDUCP_A PRNREDUCP_A MAXEDUCP_A MAXEDUCP_C
Roster, Sample Adult, Sample Child	HHC, VFY	RACE, RACE_SP, RACE_VRBAT, NEWRACE_A, NATORG, NEWNATORG_A, HISPTYPE_A, HISPOTHER_A, HISPVRBAT_A, PITYPE_A, PIOTHER_A, PIVRBAT_A, ASIANTYPE_A, ASIANOTHER_A, ASIANVRBAT_A, RACEOTHER_A, RACEVRBAT_A, MLTRACE_A, NEWNATORG_C, HISPTYPE_C, HISPOTHER_C, HISPVRBAT_C, PITYPE_C, PIOTHER_C, PIVRBAT_C, ASIANTYPE_C, ASIANOTHER_C, ASIANVRBAT_C, RACEOTHER_C, RACEVRBAT_C, MLTRACE_C	Race and Ethnicity	D R	RACEALLP_A, HISPALLP_A, HISP_A, HISDETP_A, RACEALLP_C, HISPALLP_C, HISP_C, HISDETP_C, SASPPRACE_A, SASPPHISP_A, SCPARRAC_C, RACPARSC_C, SCPARHISP_C, HISPPARSC_C
Sample Adult	BMI	HEIGHTFT_A, HEIGHTIN_A, HEIGHTM_A, HEIGHTCM_A	Height	D R	HEIGHTTC_A, BMICAT_A
Sample Adult	BMI	WEIGHTLB_A, WEIGHTKG_A	Weight	D R	WEIGHTLBTC_A, BMICAT_A
Sample Child	BMI	HEIGHTFT_C, HEIGHTIN_C, HEIGHTM_C, HEIGHTCM_C	Height	D R	HEIGHTTC_C, BMICAT_C
Sample Child	BMI	WEIGHTLB_C, WEIGHTKG_C	Weight	D R	WEIGHTLBTC_C, BMICAT_C
Sample Adult	CAN	CANKIND1_A-CANKIND3_A	Kind of cancer	D R	BLADDCAN_A, BLOODCAN_A, BONECAN_A, BRAINCAN_A, BREASCAN_A, CERVICAN_A, COLONCAN_A, ESOPHCAN_A, GALLBCAN_A, LARYNCAN_A, LEUKECAN_A, LIVERCAN_A, LUNGCAN_A, LYMPHCAN_A, MELANCAN_A, MOUTHCAN_A, OVARYCAN_A, PANCRCAN_A, PROSTCAN_A, RECTUCAN_A, SKNMCAN_A, SKNNMCAN_A, SKNDKCAN_A, STOMACAN_A, THROACAN_A, THYROCAN_A, UTERUCAN_A, HDNCKCAN_A, COLRCCAN_A, OTHERCANP_A
Sample Adult	CAN	CANKIND1_A-CANKIND3_A	Kind of cancer (kidney, testicular)	D	
Sample Adult	CAN	CANAGE1_A-CANAGE3_A	Age of cancer diagnosis	D R	BLADDAGETC_A, BLOODAGETC_A, BONEAGETC_A, BRAINAGETC_A, BREASAGETC_A, CERVIAGETC_A, COLONAGETC_A, ESOPHAGETC_A, GALLBAGETC_A, LARYNAGETC_A, LEUKEAGETC_A, LIVERAGETC_A, LUNGAGETC_A, LYMPHAGETC_A, MELANAGETC_A, MOUTHAGETC_A, OVARYAGETC_A, PANCRAGETC_A, PROSTAGETC_A, RECTUAGETC_A, SKNMAGETC_A, SKNNMAGETC_A, SKNDKAGETC_A, STOMAAGETC_A, THROAGETC_A, THYROAGETC_A, UTERUAGETC_A, HDNCKAGETC_A, COLRCAGETC_A, OTHERAGETC_A

<b>Sample Adult</b>	CAN	CANAGE1_A-CANAGE3_A	Age of cancer diagnosis (kidney, testicular)	D	
<b>Sample Adult</b>	CAN	CANMORE_A	More than three kinds of cancer	D R	NUMCAN_A
<b>Sample Adult</b>	CIG	SMKQTN_A	Time since quitting smoking-number	D R	SMKQTNP_A, SMKQTY_A
<b>Sample Adult</b>	DIB	DIBAGE_A	Age of diabetes diagnosis	D R	DIBAGETC_A, DIFYRSTC1_A
<b>Sample Adult</b>	EMP	EMPLASTWK_A	Worked for pay last week	D R	EMPWRKLSW1_A
<b>Sample Adult</b>	EMP	EMPNOWRK_A	Temporarily absent from work last week	D R	EMPWRKLSW1_A
<b>Sample Adult</b>	EMP	EMPHOURS_A	Hours worked per week	D R	EMPWKHRS3_A, EMPWRKFT1_A
<b>Sample Adult</b>	EMP	EMPFULLTIM_A	Work 35+ hours per week	D R	EMPWRKFT1_A
<b>Sample Adult</b>	EMP	EMPWHENWRK_A	Last time worked	D R	EMPLSTWOR1_A
<b>Sample Adult</b>	EMP	EMPDAYMISS_A	Days missed work, past 12 months	D R	EMPDYSMSS3_A
<b>Sample Adult</b>	EPI	EPINUMSEZ_A	Number of seizures	D R	EPINUMSEZP_A
<b>Sample Adult</b>	FEM	FEMWORK_A	Employment status of other adults	D R	PCNTADTWRP_A
<b>Sample Adult</b>	FEM	FEMWKFT_A	Other adult family member works 35+ hours per week	D R	PCNTADTWFP_A
<b>Sample Child</b>	FEM	FEMWORK_C	Employment status of adults	D R	PCNTADTWRP_C, PCNTPARWKP_C
<b>Sample Child</b>	FEM	FEMWKFT_C	Adult family member works 35+ hours per week	D R	PCNTADTWFP_C, PCNTPARWFP_C
<b>Sample Adult</b>	IMS	SHINGYEAR_A	Year of most recent Shingrix vaccine	D R	SHINGYEARP_A
<b>Sample Adult</b>	INC	INCTOTAL_A, INC100FILL_A, INC250FILL_A, INC138FILL_A, INC75K_A, INC100K_A, INC138K_A, INC150K_A, INC250K_A, INC400K_A, INC100PCT_A, INC138PCT_A, INC200PCT_A, INC250PCT_A, INC400PCT_A	Total family income as poverty ratio	D R	POVRATTC_A, RATCAT_A
<b>Sample Child</b>	INC	INCTOTAL_C, INC100FILL_C, INC250FILL_C, INC138FILL_C, INC75K_C, INC100K_C, INC138K_C, INC150K_C, INC250K_C, INC400K_C, INC100PCT_C, INC138PCT_C, INC200PCT_C, INC250PCT_C, INC400PCT_C	Reported and imputed Sample Child family income	D R	POVRATTC_C, RATCAT_C
<b>Sample Adult</b>	INS	HIKIND_A	Kinds of health insurance	R	MEDICARE_A, MEDICAID_A, PRIVATE_A, CHIP_A, OTHPUB_A, OTHGOV_A, MILITARY_A, IHS_A, NOTCOV_A, COVER_A, COVER65_A, HIKIND01_A, HIKIND02_A, HIKIND03_A, HIKIND04_A, HIKIND05_A, HIKIND06_A, HIKIND07_A, HIKIND08_A, HIKIND09_A, HIKIND10_A
<b>Sample Adult</b>	INS	HICHANGE_A	Verification of insurance coverage	NA	
<b>Sample Adult</b>	INS	MCANAME_A	Verbatim response to name of Medicare Advantage or Medicare HMO plan	R	MCADVR_A
<b>Sample Adult</b>	INS	MACHMN_A	Verbatim response to name of Medicaid managed care plan	NA	
<b>Sample Adult</b>	INS	PLANNAME1_A	Adult shares child's plan 1	R	PRIVATE_A

Sample Adult	INS	POLHLDA1_A	Policyholder for adult who shares child's plan 1	R	POLHLD1_A, POLHLD2_A, PRPLCOV1_A, PRPLCOV2_A, PRPOLH1_A, PRPOLH2_A
Sample Adult	INS	PRPOLHP1_A	Relationship to policyholder for adult who shares child's plan 1	R	POLHLD1_A, POLHLD2_A, PRPLCOV1_A, PRPLCOV2_A, PRPOLH1_A, PRPOLH2_A
Sample Adult	INS	PLANNAME2_A	Adult shares child's plan 2	R	PRIVATE_A
Sample Adult	INS	POLHLDA2_A	Policyholder for adult who shares child's plan 2	R	POLHLD1_A, POLHLD2_A, PRPLCOV1_A, PRPLCOV2_A, PRPOLH1_A, PRPOLH2_A
Sample Adult	INS	PRPOLHP2_A	Relationship to policyholder for adult who shares child's plan 2	R	POLHLD1_A, POLHLD2_A, PRPLCOV1_A, PRPLCOV2_A, PRPOLH1_A, PRPOLH2_A
Sample Adult	INS	HIPNAM1_A	Verbatim response to name of Sample Adult's first private plan	R	EXCHPR1_A
Sample Adult	INS	MORPLAN_A	Any other plans	NA	
Sample Adult	INS	HIPNAM2_A	Verbatim response to name of Sample Adult's second private plan	R	EXCHPR2_A
Sample Adult	INS	POLHLD_A	Policyholder for private plan	R	POLHLD1_A, POLHLD2_A
Sample Adult	INS	PRPLCOV_A	Plan cover others	R	PRPLCOV1_A, PRPLCOV2_A
Sample Adult	INS	PRPOLH_A	Relationship to policyholder	R	PRPOLH1_A, PRPOLH2_A
Sample Adult	INS	PLNWRK_A	How adult's plan was obtained	R	PLNWRKR1_A, PLNWRKR2_A
Sample Adult	INS	PLNWKSP_A	Verbatim response to how plan was obtained	R	PLNWRKR1_A, PLNWRKR2_A
Sample Adult	INS	PLNEXCHG_A	Plan obtained through the Marketplace	R	PLNEXCHG2_A, PLNEXCHG1_A
Sample Adult	INS	PLNPAY_A	Who pays for this plan	R	PLN1PAY1_A, PLN1PAY2_A, PLN1PAY3_A, PLN1PAY4_A, PLN1PAY5_A, PLN1PAY6_A, PLN2PAY1_A, PLN2PAY2_A, PLN2PAY3_A, PLN2PAY4_A, PLN2PAY5_A, PLN2PAY6_A
Sample Adult	INS	HICOSTN_A, HICOSTT_A	Premium amount that family or adult pays for plan	R	HICOSTR1_A, HICOSTR2_A
Sample Adult	INS	PRDEDUC_A	Plan has a deductible	R	PRDEDUC1_A, PRDEDUC2_A
Sample Adult	INS	PRHDHP_A	Annual deductible	R	PRHDHP1_A, PRHDHP2_A
Sample Adult	INS	HSAHRA_A	Health savings account	R	HSAHRA1_A, HSAHRA2_A
Sample Adult	INS	PRRXCOV_A	Plan has prescription drug coverage	R	PRRXCOV1_A, PRPXCOV2_A
Sample Adult	INS	PRDNCOV_A	Plan has dental coverage	R	PRDNCOV1_A, PRDNCOV2_A
Sample Adult	INS	PRVSCOV_A	Plan has vision coverage	R	PRVSCOV1_A, PRVSCOV2_A
Sample Adult	INS	CHNAME_A	Verbatim response to name of Sample Adults' Children's Health Insurance Program (CHIP) plan	NA	
Sample Adult	INS	OPNAME_A	Verbatim response to name of Sample Adults' state-sponsored plan	R	PLEXCHOP_A
Sample Adult	INS	OGNAME_A	Verbatim response to name of Sample Adults' other government plan	R	PLEXCHOGR_A
Sample Adult	INS	MILSPC_A	Type of military related health care	R	MILSPC1_A, MILSPC1R_A, MILSPC2_A, MILSPC3_A

<b>Sample Adult</b>	INS	RSNHIOHSP_A	Verbatim response to reasons for not getting coverage	R	RSNHICOST_A, RSNHIWANT_A, RSNHIELIG_A, RSNHICONF_A, RSNHIMEET_A, RSNHIWAIT_A, RSNHIOH_A, RSNHIJOB_A, RSNHIMISS_A
<b>Sample Child</b>	INS	HIKIND_C	Kinds of health insurance	R	MEDICAID_C, PRIVATE_C, CHIP_C, OTHPUB_C, OTHGOVR_C, MILITARY_C, IHS_C, NOTCOV_C, COVER_C, HIKIND01R_C, HIKIND04_C, HIKIND05_C, HIKIND06_C, HIKIND07_C, HIKIND08_C, HIKIND09R_C, HIKIND10_C
<b>Sample Child</b>	INS	HICHANGE_C	Verification of insurance coverage	NA	
<b>Sample Child</b>	INS	MCANAME_C	Verbatim response to name of Medicare Advantage or Medicare HMO plan	D, R	
<b>Sample Child</b>	INS	MCPART_C	Type of Medicare coverage	D	
<b>Sample Child</b>	INS	MCCHOICE_C	Enrolled in Medicare Advantage Plan	D	
<b>Sample Child</b>	INS	MCHMO_C	Medicare HMO	D	
<b>Sample Child</b>	INS	MCDNCOV_C	Dental care coverage for those enrolled in Medicare Advantage plans	D	
<b>Sample Child</b>	INS	MCVSCOV_C	Vision care coverage for those enrolled in Medicare Advantage plans	D	
<b>Sample Child</b>	INS	MCPARTD_C	Medicare Part D	D	
<b>Sample Child</b>	INS	MACHMN_C	Verbatim response to name of Medicaid managed care plan	NA	
<b>Sample Child</b>	INS	PLANNAME1_C	Child shares adults' plan 1	R	PRIVATE_C
<b>Sample Child</b>	INS	POLHLDA1_C	Policyholder for child who shares adult's plan 1	R	POLHLD1_C, POLHLD2_C, PRPLCOV1_C, PRPLCOV2_C, PRPOLH1_C, PRPOLH2_C
<b>Sample Child</b>	INS	PLANNAME2_C	Child shares adults' plan 2	R	PRIVATE_C
<b>Sample Child</b>	INS	POLHLDA2_C	Policyholder for child who shares adult's plan 2	R	POLHLD1_C, POLHLD2_C, PRPLCOV1_C, PRPLCOV2_C, PRPOLH1_C, PRPOLH2_C
<b>Sample Child</b>	INS	HIPNAM1_C	Verbatim response to name of Sample Child's first private plan	R	EXCHPR1_C
<b>Sample Child</b>	INS	MORPLAN_C	Any other plans	NA	
<b>Sample Child</b>	INS	HIPNAM2_C	Verbatim response to name of Sample Child's second private plan	R	EXCHPR2_C
<b>Sample Child</b>	INS	POLHLD_C	Policyholder for private plan	R	POLHLD1_C, POLHLD2_C
<b>Sample Child</b>	INS	PRPLCOV_C	Plan cover others	R	PRPLCOV1_C, PRPLCOV2_C
<b>Sample Child</b>	INS	PLNWRK_C	How plan was obtained	R	PLNWRKR1_C, PLNWRKR2_C
<b>Sample Child</b>	INS	PLNWKSP_C	Verbatim response to how plan was obtained	R	PLNWRKR1_C, PLNWRKR2_C
<b>Sample Child</b>	INS	PLNEXCHG_C	Plan obtained through the Marketplace	R	PLNEXCHG2_C, PLNEXCHG1_C
<b>Sample Child</b>	INS	PLNPAY_C	Who pays for this plan	D, R	PLN1PAY1_C, PLN1PAY2_C, PLN1PAY3_C, PLN1PAY5_C, PLN1PAY6R_C, PLN2PAY1_C, PLN2PAY2_C, PLN2PAY3_C, PLN2PAY5_C, PLN2PAY6R_C

<b>Sample Child</b>	INS	HICOSTN_C, HICOSTT_C	Premium amount that family pays for plan	R	HICOSTR1_C, HICOSTR2_C
<b>Sample Child</b>	INS	PRDEDUC_C	Plan has a deductible	R	PRDEDUC1_C, PRDEDUC2_C
<b>Sample Child</b>	INS	PRHDHP_C	Annual deductible	R	PRHDHP1_C, PRHDHP2_C
<b>Sample Child</b>	INS	HSAHRA_C	Health savings account	R	HSAHRA1_C, HSAHRA2_C
<b>Sample Child</b>	INS	PRRXCOV_C	Plan has prescription drug coverage	R	PRRXCOV1_C, PRPXCOV2_C
<b>Sample Child</b>	INS	PRDNCOV_C	Plan has dental coverage	R	PRDNCOV1_C, PRDNCOV2_C
<b>Sample Child</b>	INS	PRVSCOV_C	Plan has vision coverage	R	PRVSCOV1_C, PRVSCOV2_C
<b>Sample Child</b>	INS	CHNAME_C	Verbatim response to name of Sample Child's Children's Health Insurance Program (CHIP) plan	NA	
<b>Sample Adult</b>	MAR	SPOUSWHO_A	Person number for spouse	D	
<b>Sample Adult</b>	MAR	SPOUSSEX_A, SPOUNEWSEX_A	Confirm spouse's sex; Correct spouse's sex	D R	SPOUSESEX_A
<b>Sample Adult</b>	MAR	PARTNERWHO_A	Person number for partner	D	
<b>Sample Adult</b>	MAR	PARTNERSEX_A, PARTNEWSEX_A	Confirm partner's sex; Correct partner's sex	D R	PRTNRSEX_A
<b>Sample Adult</b>	MAR	LEGALSTAT_A	Legal marital status	D	LEGMARSTAT_A
<b>Sample Adult</b>	MAR	WIDIVSEP_A	Widowed/Divorced/Separated	D R	MARSTAT_A
<b>Sample Adult</b>	NAT	CITIZEN_A	Citizen status	D R	CITZNSTP_A
<b>Sample Adult</b>	NAT	NATSTBORN_A	Place of birth	D	
<b>Sample Adult</b>	NAT	NATCTZN_A	How Sample Adult became U.S. citizen	D	
<b>Sample Child</b>	NAT	CITIZEN_C	Citizen status	D R	CITZNSTP_C
<b>Sample Child</b>	NAT	NATSTBORN_C	Place of birth	D	
<b>Sample Child</b>	NAT	NATCTZN_C	How Sample Child became U.S. citizen	D	
<b>Sample Child</b>	PAR	RELCHPAR_C	Biological/Adoptive/Step/Other to type of parent-child relationship	D R	RELCHPARENTP1_C-RELCHPARENTP21_C
<b>Sample Child</b>	PAR	MARITAL_C	Married/Living with partner as unmarried couple/neither to questions about marital status of Sample Child's parents	D R	MARITAL1_C-MARITAL2_C
<b>Sample Child</b>	PAR	SPOUSLIV_C	Spouse of first-fourth parent lives there	D	
<b>Sample Child</b>	PAR	SPOUSEP_C	First-fourth parents are legally separated	D	
<b>Sample Child</b>	PAR	SPOUSWHO_C	Person number of first-fourth residential parent's spouse	D	
<b>Sample Child</b>	PAR	SPOUSSEX_C; FIXSPOUSSEX_C_	Confirming sex of first-fourth parent's spouse; Correcting sex of first-fourth parent's spouse	D	
<b>Sample Child</b>	PAR	PARTNERWHO_C	Person number of first-fourth residential parent's unmarried partner	D	

<b>Sample Child</b>	PAR	PARTNERSEX_C; FIXPARTSEX_C	Confirming sex of first-fourth parent's unmarried partner; Correcting sex of first-fourth parent's unmarried partner	D	
<b>Sample Child</b>	PAR	EVRMARRIED_C, WIDIVSEP_C	Ever been married	D R	MARSTAT1_C-MARSTAT2_C
<b>Sample Child</b>	PAR	LEGALSTAT_C	Married/Widowed/Divorced/Separated to questions about legal marital status	D R	LEGMSTAT1_C-LEGMSTAT2_C
<b>Sample Child</b>	PAR	PARBORN_C	Sample child's first-second parent born in the US/US territory	D	
<b>Sample Child</b>	PAR	FOSTPAR	Sample Child currently in foster care	D	
<b>Sample Adult</b>	PHY	MODN_A	Moderate physical activity-number	D R	MODNR_A, MODFREQW_A
<b>Sample Adult</b>	PHY	MODTP_A	Moderate physical activity-time period	D R	MODTPR_A, MODFREQW_A
<b>Sample Adult</b>	PHY	MODLN_A	Length of moderate physical activity-number	D R	MODLNR_A, MODMIN_A
<b>Sample Adult</b>	PHY	MODLTP_A	Length of moderate physical activity-time period	D R	MODLTPR_A, MODMIN_A
<b>Sample Adult</b>	PHY	VIGN_A	Vigorous physical activity-number	D R	VIGNR_A, VIGFREQW_A
<b>Sample Adult</b>	PHY	VIGTP_A	Vigorous physical activity-time period	D R	VIGTPR_A, VIGFREQW_A
<b>Sample Adult</b>	PHY	VIGLN_A	Length of vigorous physical activity-number	D R	VIGLNR_A, VIGMIN_A
<b>Sample Adult</b>	PHY	VIGLTP_A	Length of vigorous physical activity-time period	D R	VIGLTPR_A, VIGMIN_A
<b>Sample Adult</b>	PHY	STRN_A	Strengthening physical activity-number	D R	STRNR_A, STRFREQW_A
<b>Sample Adult</b>	PHY	STRTP_A	Strengthening physical activity-time period	D R	STRTPR_A, STRFREQW_A
<b>Sample Adult</b>	SCH	SCHDYSMSS_A	Number of school days missed, past 12m	D R	SCHDYSSTC_A
<b>Sample Child</b>	SCH	SCHDYSMSS_C	Number of school days missed, past 12m	D R	SCHDYSSTC_C
<b>Sample Adult</b>	UTZ	EMERGE12M_A	Number of times visited hospital emergency room, past 12m	D R	EMERG12MTC_A
<b>Sample Adult</b>	UTZ	URGENT12M_A	Number of times visited urgent care, past 12m	D R	URGNT12MTC_A
<b>Sample Child</b>	UTZ	EMERGE12M_C	Number of times visited hospital emergency room, past 12m	D R	EMERG12MTC_C
<b>Sample Child</b>	UTZ	URGENT12M_C	Number of times visited urgent care, past 12m	D R	URGNT12MTC_C
<b>Sample Adult</b>	WLK	WLKTRANLGT_A WLKTRANTP_A	Walk transportation	D R	WLKTRANTC_A
<b>Sample Adult</b>	WLK	WLKLEISLGT_A WLKLEISTP_A	Walk leisure	D R	WLKLEISTC_A

NOTE: The Research Data Center (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: <https://www.cdc.gov/rdc/>.

Blank Page