**2022 National Health Interview Survey (NHIS)**
**Public Use Data Release**



# Paradata File 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 and NHIS Electronic Mail List

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

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

The website features links to downloadable public use data and documentation for the 2022 NHIS, as well as important information about any modifications or updates to the data and/or documentation. Published reports from previous years' surveys are also available, as are updates about future surveys and datasets.

For more information about the 2022 NHIS health data files, please see the 2022 NHIS *Survey Description* and supporting documentation for that release by choosing the "Using the NHIS" tab at the website above.

Data users are encouraged to join the NHIS Listserv, an electronic mail list. The Listserv is made up of over 3,000 NHIS data users located around the world who receive e-news about NHIS surveys (e.g., new releases of data or modifications to existing data), publications, workshops, and conferences. To join, scroll down to "Related Sites" on the NHIS website, and then click on "NHIS Listserv."

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

## Guidelines for Citation of Data Source

With the goal of mutual benefit, the National Center for Health Statistics (NCHS) requests that recipients of data files cooperate in certain actions related to their use. Any published material derived from the 2022 Paradata File should acknowledge NCHS as the original source. The suggested citation to appear at the bottom of all tables and graphs is as follows:

- Data Source: NCHS, National Health Interview Survey, 2022

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

- National Center for Health Statistics. Paradata File Description, National Health Interview Survey, 2022. Hyattsville, Maryland. 2022.

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

- 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. 2022.

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

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

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

# 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 [Pub. L. No. 115-435, 132 Stat. 5529 § 302]. 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.

# Introduction

This document describes the 2022 National Health Interview Survey (NHIS) Paradata File, a complementary file to the 2022 NHIS public use health data files.

The National Health Interview Survey is a multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), and is the principal source of information on the health of the civilian noninstitutionalized household population of the United States. The U.S. Census Bureau, under a contractual agreement, is the data collection agent for the National Health Interview Survey. NHIS data are collected through a personal household interview by Census Field Representatives (FRs), with some telephone follow-up permitted to complete the case when necessary. The NHIS has been conducted continuously since its beginning in 1957. Public use microdata files are released on an annual basis.

The NHIS Paradata File does not contain health-related data, as do the NHIS regular public use data files, but rather contains paradata, which are data 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 paradata come from a number of sources:

- The Contact History Instrument (CHI), a supplement 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), the location where the interview took place and reasons for partial interviews/breakoffs.

- Date and time variables from each module of the instrument (Household, Sample Adult and Sample Child). The date and time information is 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 health data release, which contains 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. These are referred to as *Type A* cases. In addition, data are provided on cases that were deemed out of scope, such as households with Armed Forces-only adults. The out-of-scope cases are referred to as *Type B* outcomes in this document. A more detailed explanation of the outcome codes on the Paradata File can be found in the section "General Information about the 2022 Data" later in this document.

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 NHIS regular public use health data file. More information about merging the regular health files and Paradata File can be found in Appendix I.

# Sample Design, Weighting, and Variance Estimation

## Sample Design

The sample design for the 2022 Paradata File is the same as that for the regular 2022 NHIS public use data files. This information can be found in the *Survey Description* accompanying the regular 2022 NHIS health data files release. This document is available under the "Using the NHIS" tab at the following location:

- [http://www.cdc.gov/nchs/nhis.htm](http://www.cdc.gov/nchs/nhis.htm).

## Weighting

The NHIS is designed to make inferences about the civilian noninstitutionalized population of the United States. This is accomplished by inflating the responses of each surveyed person or household by a national weight factor. The national weight permits calculation of approximately unbiased design-based estimates of target U.S. population totals.

A paradata weight variable (WTIA_PD) is included on the Paradata File. This weight reflects the probability of household selection and does not include nonresponse and post-stratification adjustments. It is the appropriate weight to use when analyzing all cases (responding, nonresponding, and out-of-scope) on the Paradata File (see below).

When using the Paradata File on a stand-alone basis, data users have the option of weighting or not weighting estimates. If the goal is to make population inferences, then the paradata weight should be used. For example, a data user may want to estimate the proportion of the household population that fails to participate in face-to-face interview surveys because of privacy concerns or time burdens. The paradata weight would be necessary to address this research question.

However, not every analysis involving the Paradata File will entail making population inferences. For example, a data user may want to describe the level of effort (number of contact attempts) expended by NHIS interviewers across a wide range of final case dispositions (complete interview, partial interview, noncontact, refusal, etc.). Use of WTIA_PD would not be necessary in this situation.

If data from the Paradata File are used in support of analyses of the 2022 NHIS health data files, the corresponding weights provided on the health data file should be used. For example, a data user may be interested in exploring the determinants of health care access and utilization among sample adults and would like to control for mode of interview. The necessary mode measures located on the Paradata File would be merged with the Sample Adult File, and the final sample adult weight (WTFA_A) would be used in the subsequent analysis.

## 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 additional information about using weights and variance variables in analysis, including examples in selected statistical software packages, see "Using weights and variance estimation variables" in the 2022 NHIS Survey Description.

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

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.

# General Information about the 2022 Data

The Paradata File contains 58,962 records. Of these records, 58,139 cases are considered "in scope," i.e., cases that are eligible for interview. A case with an outcome code of 201 is a fully completed interview, in which the Household Composition, Sample Adult, and Sample Child (if a child was present) modules were completed. A case with an outcome code of 203 is a sufficient partial interview, meaning that at least a sufficient portion of the either the Adult or Child Module was completed. Cases with an outcome code of 213, 215, 216, 217, 218, or 219 are in-scope cases that did not result in either a fully complete or sufficiently complete interview. These include refusal cases (218) and insufficient partial cases (215) where the Household Module was started, but the Adult or Child module was not completed to a sufficient degree. These are referred to as *Type A* or *nonresponding* cases. In addition, 823 cases are considered "out of scope" and thus ineligible for interview (also called Type B cases). Type B cases are households comprised of military-only adult and houeholds whose usual residence is elsewhere. *Type B* cases are included in the data file because, although they are not counted in the NHIS response rate, they represent households where some contact was made with person(s) at that address in order to determine eligibility for the interview. Table 1 shows the frequency distribution of interview outcome codes for the Paradata File.

**Table 1. Frequency distribution of outcome codes for 2022 National Health Interview Survey cases**

| Outcome code | Number of Households |
|---|---|
| **In-scope cases** | |
| Interview cases | |
| 201 – Fully complete case | 26,125 |
| 203 -  Sufficient partial case | 3,138 |
| Type A cases (nonresponding) | |
| 213 - Language problem | 261 |
| 215 - Insufficient partial case | 5,825 |
| 216 – No one home, repeated calls | 4,188 |
| 217 – Temporarily absent, no follow-up | 203 |
| 218 – Refusal case | 16,621 |
| 219 – Other Type A | 1,778 |
| **Out-of-scope cases** | |
| 299 - Type B: Occupied entirely by Armed Forces adults or occupied entirely by persons with usual residence elsewhere. | 823 |
| Total | 58,962 |

# 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 512b of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) require that confidentiality be maintained without exception. Violations of CIPSEA are a class E felony, punishable by imprisonment for not more than 5 years, a fine of not more than $250,000, or both. Strict procedures in survey operations and data dissemination are used by NCHS, its data collection contractors, and other agents to prevent disclosure of survey subjects' identities.

The risk of inadvertent disclosure of confidential information regarding individual respondents is higher when there exists a publicly released data set having detailed geography variables and a detailed and extensive set of survey observations. For this reason, the NHIS does not publicly release state identifiers and some other geographic variables, and the original design strata and 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.

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

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

# Information about the 2022 Paradata File Documentation

The 2022 NHIS Paradata File and associated documentation for 2022 are available on the NHIS website, https://www.cdc.gov/nchs/nhis/2022nhis.htm.

Information about the general 2022 NHIS data release is also available here.

Like the regular NHIS release of health data, the Paradata File includes the supporting documentation listed below.

- Summary Report
- Codebook Report
- Data in column-delimited ASCII format
- Data in comma separated Values (CSV) format
- SAS, STATA, and SPSS sample programs with input statements

## Summary Report

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 Report

The Codebook report combines all the detailed information for a variable with the unweighted frequencies (counts and percentages) found in the data. The Codebook report is a combination of the former variable layout and variable frequency reports from prior years. 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 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.

Recodes — A *recode* is a variable derived from the reordering, collapsing, or verbatim coding of another variable, such as the recode UNABLE1R found in the Paradata File. Alternatively, a recode may be constructed from two or more variables, such as the point in the interview period when the Household Composition Module was started (STRTPNT). If a particular variable was used in making recode variables, then those recode variables are listed as a cross reference.

The Paradata File contains a number of recodes related to the date and time variables, including the start point of the Household Composition Module in the interview period and the time of day when the Household Composition Module was started. For confidentiality reasons, many of the original continuous measures (i.e., number of days) have been recoded into ordinal, or ranked measures such as "Early," "Middle," or "Late." Further information about these specific recodes can be found in the "Measures of Time" section of the document.

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

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

Notes — *Notes* provide information that analysts need to know about a particular variable, such as assumptions, limitations, caveats, differences between instrument versions, or other important information. Analysts are encouraged to read the notes for variables of interest.

Universe — The *universe* refers to those respondents eligible to be asked a given question. For example, the universe for many variables on the Paradata file is "All Households." This means that all cases should have a valid value for the variable.

# Paradata File Variables

The Paradata File is a case-level file.  Each record in the file represents a unique household. Although multiple families may share one household, the household is the unit of measure. What follows are descriptions of conceptually grouped variables included on the Paradata File.

## Measures of Time

A series of time measures can be found on the Paradata File. The first set of time measures convey information on the start dates of main sections of the survey instrument and the completion date of the case. For confidentiality purposes, actual dates are not provided. Instead, original date variables have been recoded into categorical measures that anchor the case within the monthly NHIS interview period.[1] For example, STRTPNT indicates that the Household Composition Module of the survey instrument was started in the early, middle, or late portion of the monthly interview period. Similarly, ENDPNT indicates that the case was completed in the early, middle, or late portion of the monthly interview period. Among several possible uses, measures of this sort allow researchers to explore whether the quality of data (e.g., item nonresponse) collected varies by when in the interview period the case was started and completed.

The following measures correspond to the starting points, within the monthly interview period, of main sections of the survey instrument. Valid values are present on these measures for cases where the specific section or module of the survey instrument was started.

- STRTPNT indicates that the Household Composition Module of the instrument was started early in the interview period, in the middle of the interview period, or late in the interview period.

- SASTRPNT indicates that the sample adult module was started early in the interview period, in the middle of the interview period, or late in the interview period.

- SCSTRPNT indicates that the sample child module, if there was a sample child, was started early in the interview period, in the middle of the interview period, or late in the interview period.

The following measure identifies the point in the monthly interview period when the case was completed and a final disposition (or outcome) code was assigned.

- ENDPNT indicates that the case was completed and a final disposition (or outcome) code was assigned early in the interview period, in the middle of the interview period, or late in the interview period.

---

[1]In 2011, the NHIS transitioned to a monthly interview sample.  An interviewer received a new assignment of sample cases on the first of the month and had until the end of the month to complete the assignments.  Prior to 2011, there was a 17-day interview period to complete the assignment.

Four additional items capture the time of day when sections of the survey instrument were started. The measures were recoded into three broad time periods of morning (including overnight), afternoon, and evening. Valid values are present on these measures for cases where the specific section or module of the survey instrument was started.

The following are descriptions of the time of day measures:

- HHC_TOD indicates that the Household Composition Module was started in the morning, afternoon, or evening.

- SA_TOD indicates that the Sample Adult Module was started in the morning, afternoon, or evening.

- SC_TOD indicates that the Sample Child Module, if there was a Sample Child, was started in the morning, afternoon, or evening.

## Measures of Contactability

One of the conceptual groupings of measures on the Paradata File are those describing contactability, or the effort involved in making contact with a sample household. As with all measures on the file, this particular group may be utilized in many types of analyses, though conceptually they are particularly amenable to exploring survey contact and, conversely, noncontact processes.

The bulk of "noncontact" measures on the Paradata File are case-level summaries of CHI-based, interviewer-recorded explanations of noncontact outcomes. For each contact attempt that an interviewer makes on a sample household, she/he must report one of the following three outcomes: contact with a sample unit member, contact with a non-sample unit member, and noncontact. There are two sets of CHI-based noncontact measures based on mode of attempt.

- In-person noncontact measures: NCTP01R through NCTPR99R:

These dichotomous measures (0=never entered by interviewer across all in-person noncontacts, 1=entered by interviewer at 1 or more in-person noncontacts) describe noncontact outcomes in personal visit contact attempts. Examples include "no one home—no eligible person home" and "unable to reach/locked gate/buzzer entry." The interviewer could record up to eleven different reasons or explanations.

- Telephone noncontact measures: NCTL01R through NCTEL99R:

These dichotomous measures (0=never entered by interviewer across all telephone noncontacts, 1=entered by interviewer at 1 or more telephone noncontacts) describe noncontact outcomes in telephone contact attempts. Examples include "busy signal" and "disconnected." The interviewer could record up to eight different reasons or explanations.

For each contact attempt that an interviewer makes on a sample household, she/he must report one of the following three outcomes: contact with a sample unit member, contact with a non-sample unit member, and noncontact. There are two sets of CHI-based noncontact measures based on mode of attempt.

Four measures are included under this conceptual grouping: CTSTAT1, CTSTAT2, CTSTAT3 and UNABLE1R.

- CTSTAT1, CTSTAT2, and CTSTAT3 provide the number of contact attempts that resulted in contact with sample unit members, the number of contact attempts that resulted in contact with non-sample unit members, and the number of contact attempts that resulted in noncontact.

- UNABLE1R provides the number of times contact was made with a sample unit member but the interviewer was unable to conduct or complete an interview because the "eligible person was not available." This measure captures situations of sub-level noncontact -contact was made at the household level, but the household member needed to complete a particular section of the interview was not available. This largely pertains to the sample adult, followed by the sample child respondent.

There are several points to remember when working with these measures.

- CHI entries were not made for 268 cases.

- The CHI-based measures presented above were created by summing a particular entry across all contact attempts and then top-coding values greater than one. For example, if two personal visit contact attempts for a household resulted in contact with a sample unit member, but "eligible person was not available" was entered for both attempts, the summary CHI measure (UNABLE1R) would have a value of 1.

- These measures are based on interviewer self-reports. A corresponding amount of recall error is expected, especially when CHI entries were not made at the time of the contact attempt.

## Contact Strategies

As part of CHI, interviewers are instructed to record all strategies used on every contact attempt. Twelve possible strategies, including "none" and "other" categories, are available. The following is a list of each possible strategy and its description:

- Advance letter given (STRAT01R): Interviewers are instructed to ask the household member if she/he received the advance letter (mailed from the regional office several weeks before the interview period). If the household member did not receive the letter, the interviewer will provide the household member with a copy of the letter. The letter can also be left in situations of noncontact.

- Scheduled appointment (STRAT02R): If the respondent is unavailable at the time of contact, the interviewer may choose to schedule an appointment some other time during the interview week.

- Left note/appointment card (STRAT03R): If no contact is made at the time the interviewer visits the household, she/he may leave a personalized note or an appointment card.

- Left promotional packet/informational brochure (STRAT04R): If the householder is not available at the time of visit, the interviewer may give the householder informational materials to review before an interview is conducted. These materials may include brochures about how the NHIS data are used, a ruler, magnet, or other information related to the survey.

- Called household (STRAT05R): The interviewer may choose to call the household to make contact. This is usually done after a personal visit has been attempted.

- Left message on answering machine (STRAT06R): Interviewers may choose to leave a message on an answering machine for the householder.

- Waited for respondent (STRT11R): The interviewer may wait near the residence in the hopes that the interviewer will see the householder arrive.

- Checked with neighbors (STRAT12R): The interviewer may try to learn more about the best time to reach the householder by asking neighbors about the householder's schedule.

- Contacted other family members (STRAT13R): If the interviewer has difficulty contacting a respondent, she/he may choose to contact a relative of the respondent in order to help to locate the respondent.

- Contact property manager/doorman (STRT14R): The interviewer may need to contact the property manager or doorman if the community where the household is located has limited access.

- None (STRAT98R): No strategies were used on this contact attempt.

- Other-specify (STRAT99R): Some other strategy not listed was used on the contact attempt.

## Measures of Cooperation

Another conceptual grouping of variables found in the Paradata File is related to respondent cooperation. The more the respondent cooperates with the question and survey processes, the greater the chance for a quality interview. Several measures of cooperation come from the Contact History Instrument.

- Language concerns (LANG1R–LANG5R): These dichotomous coded measures (0=never entered, 1=entered at 1 or more contacts) describe contact outcomes in contact attempts, both in person and over the telephone. Some examples include "no household member able to translate" and "unable to find translator."

- Respondent reluctance (RELC01R–RELUC99R): These dichotomous coded measures (0=never entered, 1=entered at 1 or more contacts) represent contact outcomes in contact attempts, both in person and over the telephone. Some examples include "too busy" and "interview takes too much time."

- Unable to complete interview (UNABLE2R–UNABL99R): These dichotomous coded measures (0=never entered, 1=entered at 1 or more contacts) represent contact outcomes in contact attempts, both in person and over the telephone. Some examples include "inconvenient time" and "health problem." These variables could also be conceptually grouped with the partial/break-off variables from the file.

## Partials/Breakoffs

A series of questions to ascertain reasons for partially completed interviews (interviews that are not entirely complete) appear in the Back Section of the instrument. PARWHY is a variable that differentiates between the interview being stopped in progress (break-off) and being unable to reach a particular respondent or obtain some information on follow-up (partial, other than break-off).  We also collect the main reason the interview is not complete (BRKREASON).

For cases marked as a break-off in PARWHY, we identify the module where the break-off occurred. BRKHHC indicates the breakoff occurred in the Household Composition Module, BRKSA indicates the break-off occurred in the Sample Adult Module, and BRKSC indicates the break-off occurred in the Sample Child Module.

In addition, for non-completed cases that were not a break-off, we identify cases where the Household, Sample Adult and Sample Child Modules are not complete (NCOMHHC, NCOMSA, and NCOMSC).

## Mode Measures

The NHIS is primarily a personal visit survey. However, there are limited circumstances in which the interviewers are permitted to conduct an interview by telephone. Once an initial personal contact has been made, telephone contacts may be used to complete partial interviews, or to collect other missing parts of the interview. There are two sources of mode data on this file: data at the household level from the Back Section, and data from the summary Contact History File.

### Back Section Variables

At the completion of each interview, interviewers are asked a series of questions about the administration of the interview for the Sample Adult and Sample Child Modules. If the Sample Adult Module was fully or partially completed, they are asked: "Was all, some, or none of the Sample Adult interview conducted by telephone?" (INTMODESA). They are then asked for the main reason the Sample Adult interview was completed by phone (INTMODEWHYSA).

Similarly, If a Sample Child Module was fully or partially completed, they are asked "Was all, some, or none of the Sample Child interview conducted by telephone?" (INTMODESC). They are then asked for the main reason the Sample Child interview was completed by phone (INTMODEWHYSC).

### CHI Variables

For each contact attempt, the interviewer records whether the attempt was made by telephone or by personal visit.

- MODE_P indicates the number of contact attempts on a case that were made by personal visit. To ascertain the proportion of contact attempts made by personal visit, divide MODE_P by the total number of contact attempts (TOTCOUNT).

- MODE_T indicates the number of contact attempts on a case that were made by telephone. To ascertain the proportion of contact attempts made by telephone, divide MODE_T by the total number of contact attempts (TOTCOUNT).

## Case-Level Information

There is also a set of variables that give general case-level information about the record. For each case, the variable SRVY_YR (survey year) will have a value of "2022."

The variable OUTCOME1 represents the disposition of the case. If the Paradata File is linked to the health data files, it is important to remember that there is not a 1:1 match between the two. The Paradata File represents field data before cleaning. There are some 201 and 203 outcomes (fully complete and sufficiently complete interviews) on the Paradata File that were dropped from the final 2022 public use health files because the record was not deemed adequate for data release (for example, there may have been too many "Don't know" or "Refused" responses). The variables QUALITY_SA and QUALITY_SC flag records that have been removed from the health data files for quality reasons.

As mentioned before, the Paradata file also contains in-scope but not sufficiently complete cases (refusals, insufficient partial cases, etc.) and out-of-scope cases (usual residence elsewhere, all active-duty military households).

## Appendix I. Merging the Paradata File to Public Use Health Data Files

NHIS data files can be merged within a survey year. The purpose of merging data within the same data year is to either incorporate variables from different data files when persons are common to both files (i.e. merging by common factor). An example of the need to merge a data file is if the user is interested in merging the Paradata file with the Sample Adult file (or the Sample Child file) to explore the 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. Nearly all of the 201 or 203 outcomes on the Paradata File can be linked with the health data files. As mentioned previously, the Paradata File represents the field data before cleaning; thus, there are some cases, including 201 and 203 outcomes that are dropped before the creation of the public use health data files. These non-matched 201 and 203 outcomes, along with the Type A and Type B outcomes, will be excluded when the files are linked.

In 2022, the Sample Adult file has 27,651 records, the Paradata file has 58,962 records, and the Sample Child has 7,464 records.  If making population inferences of noninstitutionalized adults aged 18 or over using this linked file, the final annual weight on the Sample Adult Health File (WTFA_A) should be used.

*Example: Merging survey data and Paradata 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: Merging survey data and Paradata using STATA

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 */

# Appendix II. Changes From the 2021 Paradata File

*A General Note About Interviewing Procedures:*

After altering contact practices in the preceding years due to the COVID-19 pandemic, interviewers fully returned to regular survey interviewing procedures in 2022, whereby first contact attempts to households were made in person, with follow-up allowed by telephone.

For more information on NHIS data collection and interviewing procedures in 2022, please see the 2022 NHIS *Survey Description* for that release by choosing the "Using the NHIS" tab at the website below:

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