NCHS Rapid Surveys System

NATIONAL CENTER FOR HEALTH STATISTICS

# Survey Description

## Round 3: Data collected January-February 2024

Last revised July 10, 2024

CDC
U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

## Table of contents

## Tables

# 1. Introduction

In 2022, the National Center for Health Statistics (NCHS) launched the Rapid Surveys System (RSS) with three main goals:

1. to provide the Centers for Disease Control and Prevention (CDC) and partners with time-sensitive data of known quality about emerging and priority health concerns,
2. to use these data collections to increase NCHS's expertise and to evaluate the quality of public health estimates generated from commercially available probability-based online panels (hereafter referred to as online panels), and
3. to improve methods to appropriately communicate the fitness for use of public health estimates generated from online panels.

The RSS platform is designed to approximate national representation of the adult population of the United States in ways that many data collection approaches cannot. Specifically, the platform is designed to collect self-reported health data using two online panels, combining and weighting the resulting data. Online panels generally consist of persons who are recruited using statistical sampling and agree to participate in multiple surveys, typically in exchange for payment or prizes. These survey panels are designed to take advantage of the efficiencies in using online surveys, although other modes such as telephone can be used to improve data accuracy. The RSS incorporates multiple mechanisms to carefully evaluate the resulting survey data for their appropriateness for use in public health surveillance and research (e.g., hypothesis generating).

The RSS will be used multiple times per year to gather data about issues of importance to CDC and U.S. Department of Health and Human Services (HHS). In the first year (2023), there were two rounds of data collection. In future years, up to four rounds of data collection per year are anticipated. Each round of data will be shared with the public through dashboards, tables, infographics, and data files with accompanying codebooks. For the initial rounds, the data will be released approximately 6 months after the data collection period. Faster releases are expected in future cycles.

## 1.1 Motivation

Typically, NCHS data systems generate robust nationally representative statistics using methods that maximize relevance, accuracy, and reliability. For example, surveys such as the National Health Interview Survey (NHIS) allow for description of health outcomes among the U.S. population across time and for many demographic groups; however, such efforts require longer periods of time for data collection and processing. Although NCHS's gold standard sampling, interviewing, and post-processing strategies are pivotal for examining national annual trends in disease and behavioral risk factors and differences across demographic and geographic groups, they are less useful for responding to more immediate or "real-time" public health issues that may arise with

little warning or notice. The RSS can be used to collect time-sensitive data about emerging and priority health concerns for decision making by reducing the time needed for data collection and processing. Yet unlike many other quick-turnaround surveys and polls, the RSS will also provide decision makers with information on the quality of the estimates using methodological evaluations and ongoing quality assessments.

CDC uses other data sources to identify and track emerging public health threats, such as those associated with disease outbreaks. During the COVID-19 pandemic, the implications of *unknown data quality* from some public health surveillance approaches became clearer. For example, new surveys such as the U.S. Census Bureau Household Pulse Survey (https://www.census.gov/programs-surveys/household-pulse-survey.html) were launched without going through usual design evaluation. The quality of estimates from these surveys was not well understood; however, in the absence of better information, these estimates were disseminated to the public. In response to these challenges, the RSS was launched to develop a mechanism that facilitates collection of time-sensitive survey data using online panels with thorough and ongoing evaluation of data quality.

NCHS has been using online panels for methodological research since 2015 through the Research and Development Survey (RANDS); for more information please see https://www.cdc.gov/nchs/rands/index.htm. The RSS expands on this effort to produce estimates of health conditions and related topics using online panels. Because online panels do not meet the rigor of gold standard NCHS surveys, the RSS includes mechanisms to facilitate continuous quality improvement by supplementing these panels with intensive efforts to understand how well the estimates reflect the population as a whole and, depending on the sample size, topic, and analytic goals, how well estimates for specific subpopulations defined by demographics or socioeconomic characteristics can be generated. An important aspect of the RSS is a data dissemination strategy that communicates the strengths and limitations of data collected through the RSS's online panel approach, depending on the topic and sample, as compared to more traditional data collection methods. Although these assessments will provide users more specific information on the quality of the estimates each round, the RSS is intended to be suitable for:

- time-sensitive data needs for which existing resources are unavailable, not timely, or of insufficient or unknown quality;
- current public health attitudes or behaviors, because estimates from annual surveys might be outdated when available; and
- developmental work to improve concept measurement and inform future question design.

## 1.2 RSS samples

The RSS uses cross-sectional samples of adults in the United States from two online panels: NORC at the University of Chicago's (NORC's) AmeriSpeak Panel and Ipsos's KnowledgePanel. Each sample is surveyed over the same time period, using the same RSS survey instrument, so that the samples can be combined. This allows NCHS to produce statistics more rapidly than using data from traditional NCHS household surveys that rely on trained interviewers and face-to-face data collection. The sample design for the third round of RSS (RSS-3) is described in Section 3.1.

## 1.3 Survey components

Each round's questionnaire consists of four components:

1. basic demographic information on respondents to be used as covariates in analyses;
2. new, emerging, or supplemental content proposed by NCHS, other units within CDC, and other federal agencies;
3. questions used for calibrating the survey weights, or adjusting the weights to increase precision; and
4. additional content selected by NCHS to evaluate against relevant benchmarks, which provides an assessment of how well the weighting scheme performed.

Component 1 contains questions about demographic characteristics. Both AmeriSpeak and KnowledgePanel have existing, previously collected, demographic data for each panel member. These data are collected periodically from each panel member during their panel tenure, and therefore prior to the completion of the RSS questionnaire. These panel "profile" variables include sex, race, education, and other characteristics. Profile variables differ slightly between the two panel providers and are harmonized when combined for RSS data. Some demographic variables for component 1 come directly from an RSS questionnaire, while others come indirectly from the data respondents provide when initially joining the individual panels and periodically throughout their panel tenure.

Component 2 provides relevant, timely data on new, emerging, and priority health topics that are fit for use for decision making. Prior to each round of RSS, CDC divisions and other HHS agencies are invited to propose topics and questions to be added to that round. Sponsored topic areas must be consistent with at least one of the four RSS topic criteria:

1. **Time-sensitive data needs**
2. **Public health attitudes and behaviors** (e.g., opinions, beliefs, stated preferences, and hypotheticals)
3. **Developmental work** to improve concept measurement/questionnaire design
4. **Methodological studies** to compare, test, and develop approaches to data collection and analysis

Component 3 includes questions that are used for calibration of RSS weights to improve the precision of the resulting estimates and come from the NHIS, the American Community Survey (ACS), and other large federal surveys. RSS survey weights are calibrated so that the weighted distributions of calibration variables align with the source survey(s). In addition, new content may be added to both NHIS and an RSS questionnaire to evaluate its utility for calibration. Information about RSS-3 weight calibration can be found in Section 3.4.

Component 4 includes questions specifically used for benchmarking, evaluating, and communicating the utility of health estimates collected via RSS. Over time, the benchmarks may also be helpful in evaluating estimates from other available online panels. These questions come from NHIS or other surveys that have relatively stable estimates with known quality. Once RSS data are weighted and calibrated to NHIS data, the RSS data are compared with the benchmark data to evaluate the quality of the estimates. See Section 2.1 for information about RSS-3 benchmarking.

Each round of RSS data collection will emphasize different topics, and the content of each component will vary. Component 3 calibration questions will be relatively consistent across rounds, although they may change based on the results of methodological evaluations. Additionally, data (known as paradata) are collected about the survey mode (web or phone), length of time taken to complete the full questionnaire and timing of each section, whether the data came from AmeriSpeak or KnowledgePanel, and other data collection information. More information about the RSS-3 components can be found in Section 2. For details on the RSS-3 questionnaire, please see www.cdc.gov/nchs/data/rss/round3/questionnaire.pdf.

This document provides guidance and information for users of RSS-3 data and estimates. Section 2 provides an overview of RSS-3 content and methodology studies. Section 3 describes the survey methodology, including sample design, data collection, editing and imputation, sample weights, and confidentiality. Section 4 provides an overview of RSS-3 data files and guidance on variance estimation and data quality.

# 2. Round 3 overview

This section summarizes the content and methodology studies for RSS-3.

## 2.1 Round 3 content

As described in Section 1.1, RSS-3 includes four components: demographic characteristics, new emerging or supplemental topics, calibration variables, and benchmarking variables, which are described below. For a complete list of variables, please see the RSS-3 codebook (https://www.cdc.gov/nchs/rss/data.html).

**Demographic Questions.** Demographic data such as age, race, ethnicity, gender, marital status, education, income, and employment come from the panel profile variables. Because panel profile data were collected by the panel providers over a period of years, questions on marital status and employment were asked again using standardized questions in the RSS-3 questionnaire for comparison to the corresponding panel variables.

**New, Emerging, or Supplemental Topic Questions (from contributing agencies).** In RSS-3 there were four new emerging or supplemental topics:

1. Family health history and barriers to collecting family health history (Office of Genomics and Precision Public Health [OGPPH])
2. Genetic testing for hereditary forms of cancer and heart disease (OGPPH)
3. Employment status, work arrangements, leave availability, and employer insurance coverage (National Institute for Occupational Safety and Health)
4. Sexual health, communication with partners and healthcare providers about sexual health, sexual health services utilization, sources for sexual health information (National Center for HIV, Viral Hepatitis, STD, TB Prevention)

**Calibration Questions.** Questions included for calibration of RSS-3 weights, or for evaluation for use as calibration variables, include social and work limitations, internet use in general and for medical reasons, language used at home and in other settings, telephone use, civic engagement, employment, marital status, and ever diagnosed with high cholesterol. Several panel profile variables were also used for calibration: Census region, age, race, ethnicity, gender, education, income, and home ownership. See Section 3.4 for more information about the calibration variables used in RSS-3.

**Benchmarking Questions.** In RSS-3, all estimates used for benchmarking came from the 2023 NHIS. RSS-3 respondents were asked about self-reported health status; chronic conditions including hypertension, asthma, and cancer; social determinants including difficulty paying for food and food insecurity; access to and use of medical care including use of urgent care centers, emergency rooms, and hospitals; cigarette smoking and use of other nicotine products; and disability including difficulties in vision, hearing, walking or climbing steps, communicating, remembering or concentrating, and self-care. The weighted RSS-3 estimates for these benchmarking questions were compared with NHIS data from the third quarter of 2023 (https://www.cdc.gov/nchs/data/rss/round3/quality-profile.pdf). The full NHIS questionnaire can be found at https://www.cdc.gov/nchs/nhis/2023nhis.htm.

## 2.2 Round 3 methodological studies

The methodology studies in RSS-3 are described below.

### 2.2.1 Cognitive interviews

RSS-3 included accompanying cognitive interviews conducted by RTI International and the NCHS Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER). All questions in the RSS-3 survey were included, with probing related to the sponsored questions outlined in Section 2.1 of this document.

For RSS-3, cognitive interviews were conducted after the survey was fielded. Because of this, cognitive interviews should be understood as an examination of the RSS-3 items' construct validities, or how well a question captures the intended measurement, rather than as a method to evaluate question wording. The cognitive interviewing report, including a question-by-question analysis, will be available in Fall 2024 on Q-Bank (https://wwwn.cdc.gov/qbank) and the RSS Data Files and Documentation page (https://www.cdc.gov/nchs/rss/data.html). Data users should consult this report to understand what information the survey questions captured and to frame their own analysis of the RSS-3 data.

Cognitive interviews will be conducted in most rounds of RSS. Probes are planned for sponsored questions in future rounds but not for all questions. Information on cognitive interviewing can be found at https://www.cdc.gov/nchs/ccqder/evaluation/CognitiveInterviewing.htm.

# 3. Methodology

This section describes the procedures used to collect and prepare RSS-3 data, including sample design, data collection, weighting, editing and imputation, and protection of respondent confidentiality.

## 3.1 Sample design

As described in Section 1, RSS uses two online panels: Ipsos KnowledgePanel and NORC AmeriSpeak. These panels, described in Section 3.1.1, served as the sampling frames. One sample was selected from each frame as described in Section 3.1.2.

### 3.1.1 Sampling frames

These panels are based on probability samples of the population of U.S. households and are designed to serve as sampling frames for sample selection and production of national estimates. All members of both panels complete questions about demographics and household composition before participating in any surveys. Although both panels include individuals aged 13–17, these younger panelists are not included in RSS-3.

The KnowledgePanel has approximately 60,000 adults in 44,000 households. Panel members are primarily recruited using address-based sampling (ABS) methods. It is used to conduct surveys for researchers, government agencies (federal, state, and city), and commercial companies on a variety of topics including health, political, and

consumer studies. For additional information see https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel.

The AmeriSpeak Panel includes approximately 59,000 adults in 53,000 households. Panel members are recruited using ABS of U.S. households. It is used to conduct surveys for academic, commercial, nonprofit, and government organizations. For additional information see https://amerispeak.norc.org/us/en/amerispeak.html.

### 3.1.2 Sample selection

The target population of RSS-3 is all U.S. adults aged 18 and older. A sample was selected from each frame using similar designs, described below. In both cases, only one panel member per household was selected to participate.

**KnowledgePanel.** A probability sample was drawn from the KnowledgePanel using a probability proportional to size (PPS) sample selection method. The measure of size for the PPS selection is a KnowledgePanel weight that has been calibrated to population benchmarks. Benchmarks include gender, age, race and ethnicity, education, census region, household income, homeownership status, and metropolitan area obtained from the March 2023 supplement of the Current Population Survey (CPS), and Spanish language dominance from the 2022 ACS. Assuming a 70% completion rate and a 1% data cleaning rate (i.e., interviews that will not pass the data cleaning checks), 5,790 panelists were selected to achieve the target sample size of 4,000 interviews.

**AmeriSpeak.** A probability sample was selected from the AmeriSpeak panel with a target sample size of 4,000 interviews. The sampling frame was stratified by age, race and ethnicity, education, and gender. Sample allocation across the strata was determined such that the expected distribution of complete surveys matches that of census population benchmarks. To achieve the target number of 4,000 interviews, 15,322 panelists were selected. Historical completion rate by stratum was used to calculate the number of panelists to sample.

## 3.2 Data collection and completion rates

Data collection was coordinated between NCHS and the panel providers to ensure that the samples collected from each were comparable. Table 3-1 outlines interview modes and recruitment protocols by provider and sample. Interviews for the KnowledgePanel were conducted solely by web. Interviews for the AmeriSpeak sample were also primarily web-based; however, telephone interviews were used for panelists known to prefer completing surveys over the phone.

### Table 3.1 Interviewing modes and recruitment protocols by panel

| | KnowledgePanel | AmeriSpeak |
|---|---|---|
| Data collection modes | Web | Web and phone |
| Incentives | 1,000 points (equivalent of $1) | $5 for web and $10 for phone completes |
| Collection Period | January 15–February 2, 2024 | January 18–February 15, 2024 |
| Protocol for recruitment | ▪ A prenotification email 2 days prior to fielding.<br>▪ Reminder on day 3. Additional reminders on days 7 and 11 for hard-to-reach groups.[1]<br>▪ All reminders after day 3 were customized with email text used only for this study. | ▪ Seven emails: one invitation and six reminders.<br>▪ Two SMS text reminders.<br>▪ Up to three call attempts per phone-preference case. |

[1] Hard-to-reach groups include young adults who are ages 18 to 29, members of racial and ethnic minority groups, and individuals with a high school education or less.

The number of completed interviews for both samples exceeded the sample size targets, as shown in Table 3-2. Panelists who broke off and did not fully complete the survey were considered nonrespondents for response and completion rate calculations and were not included on the final datafile. Unweighted completion rates were 72.0% for KnowledgePanel panelists and 27.4% for AmeriSpeak panelists. Final weighted response rates that reflect the survey completion rate and panel recruitment and retention rates are reported in the RSS-3 Data Quality Profile (https://www.cdc.gov/nchs/data/rss/round3/quality-profile.pdf).

### Table 3.2 Target and completed interviews by panel and mode

| | KnowledgePanel | AmeriSpeak | Combined |
|---|---|---|---|
| Sample size | 5,790 | 15,322 | 21,112 |
| Target number of interviews | 4,000 | 4,000 | 8,000 |
| Number of completed interviews | 4,170 | 4,205 | 8,375 |
| Web interviews | 4,170 | 3,745 | 7,915 |
| Phone interviews | | 460 | 460 |
| Completion rate (%)[1] | 72.0 | 27.4 | 39.7 |

[1] The unweighted survey completion rate uses the AAPOR RR5 response rate definition (AAPOR, 2023).

## 3.3 Editing and Imputation

A key component of RSS is combining data from two panels into one analytic dataset. This section describes edit and imputation steps performed on the combined dataset. Both panel providers followed the same guidelines to standardize their datasets. In cases where there were differences between the datasets, such as for the panel profile variables, the data were standardized by aligning the variables' names, value labels, and value coding.

### 3.3.1 Open-ended coding

Open-ended responses were reviewed in an "upcoding" process to determine whether the text response could be coded into a common response option. For RSS-3, respondents were asked, "*Are you covered by any of the following types of health insurance or health coverage plans?*" and given seven specific health insurance types and the open-ended question "*Any other type of health insurance or health coverage plan (please specify)*"(*EMP_INSH*). Verbatim responses to this question were reviewed, and responses were upcoded if respondents mentioned health insurance in one of the seven listed insurance types (*EMP_INSHA – EMPINSG*). Additionally, two variables were created during upcoding: (1)"*Private insurance, Other/Unknown Source*" (*EMP_INSH_UNK*), to capture verbatim responses that did not provide how the private insurance they listed was paid for (e.g., self, employer, Medicare); and (2) "*Other health insurance reported*" (*EMP_INSH_OTH*), to capture responses that did not fit any other category listed.

Verbatim responses are available on the restricted-use data file. See Section 4 for a description of data products.

### 3.3.2 Missing data imputation

Variables used for calibrating the harmonized analysis weights were imputed if data were missing. Of the 12 variables used for calibration, 6 had missing values and were imputed: 2 panel variables (Race and Hispanic origin, and Metropolitan status) and 4 composite variables derived from questionnaire variables (Marital status, Ever diagnosed with high cholesterol, Difficulty participating in social activities, and Civic engagement). See Section 3.4 for additional details about the weighting variables and weighting procedures.

Missing values were singly imputed using the fully conditional specification approach with SAS PROC MI (SAS Institute, 2015). The imputation models used the default discriminant functions for classification variables and the predictor variables included all 12 weighting variables and the sample design variables (i.e., strata and cluster). Imputation rates ranged from 0.13% to 1.62% of survey respondents.

Additional missing data imputation was conducted by the panel providers prior to data harmonization. This imputation was required to calculate their respective panel-specific

analysis weights. These variables include panel variables associated with education, gender, race and ethnicity, and income.

Imputed variables are included in the data files (Section 4.1), along with imputation flags which use the naming convention <*variable name*>*_IFLG*.

## 3.4 Sample weights

Each panel provider produced weights for analysis of their panel's data. Section 3.4.1 describes these weights, and Section 3.4.2 describes the creation of a weight for analyzing data from both panel providers at once. Section 3.4.3 describes the calibration methodology. The panel providers produced additional weights that were not recalibrated but are available in the restricted-use file (RUF). Additional information about the weights available in the RUF and public-use file (PUF) is provided in Section 4.2.

### 3.4.1 Panel-specific analysis weights

Each panel provider computed calibrated weights, which were scaled to the total population of U.S. adults aged 18 or older as estimated by the third quarter of the 2023 NHIS.

**KnowledgePanel**. The base weights for KnowledgePanel are modeled design weights that account for their most common observed patterns of nonresponse. These are calibrated to population totals for subgroups defined by race and ethnicity, education, household income, Spanish language proficiency, gender by age, and Census region by metropolitan status. These totals are obtained from the March 2023 Supplement of the CPS, except for language proficiency, which comes from the 2022 ACS.

**AmeriSpeak.** Base weights are created using weights developed for the full AmeriSpeak panel, adjusted for unequal sample selection probabilities and frame coverage limitations. The base weights are adjusted for survey nonresponse using a weighting class method, with weighting classes defined by age, race and ethnicity, gender, and education. A raking ratio adjustment is then applied to align the sample with population benchmarks from the March 2023 Supplement of the CPS, including age, gender, Census division, race and ethnicity, education, age by gender, age by race and ethnicity, and gender by race and ethnicity.

### 3.4.2 Combined panel analysis weight

A composite calibrated weight was calculated for analysis of the concatenation of respondent samples from both panel providers. The composite calibrated weight, *WEIGHT*, was calculated as a composite version of the final panel-specific calibrated weights (Section 3.4.1):

$$P1\_CALWT * \lambda_1 + P2\_CALWT * (1 - \lambda_1).$$

The adjustment factor $\lambda_1$ is a ratio of the effective sample sizes and is defined as:

$$\lambda_1 = \frac{n_{e,1}}{n_{e,1} + n_{e,2}}$$

where $n_{e,i}$ is the effective sample size for the respondent sample from the respective panel *i* calculated as:

$$n_{e,i} = \frac{(\sum_s w_k)^2}{\sum_s w_k^2}$$

where *s* is the total number of respondents from the panel, and $w_k$ is the calibrated weight for the $k^{th}$ respondent. See Table 3-3 for values of $n_{e,i}$ and $\lambda_1$.

**Table 3.3 Sample size, effective sample size and composite factor by panel**

| Panel | $n$ | $n_{e,i}$ | $\lambda$ |
|---|---|---|---|
| AmeriSpeak | 4,205 | 2171.411 | 0.391 |
| KnowledgePanel | 4,170 | 3378.805 | 0.609 |

Weighted estimates (using *WEIGHT*) from the combined samples match the calibration totals from NHIS shown in Table 3-4.

**Table 3.4 Calibration variables and NHIS population totals**

| Variable | NHIS estimate |
|---|---|
| **U.S. Population Total** | 258,491,091 |
| **Age** | |
| 18-34 | 75,188,962 |
| 35-49 | 63,069,430 |
| 50-64 | 61,793,484 |
| 65+ | 58,439,215 |
| **Sex** | |
| Male | 126,152,358 |
| Female | 132,338,733 |
| **Race and Hispanic Origin** | |
| Hispanic | 45,407,259 |
| Non-Hispanic white | 159,727,880 |
| Non-Hispanic black | 32,599,072 |
| Non-Hispanic other | 20,756,880 |
| **Education** | |
| Less than high school | 27,879,613 |

| Variable | NHIS estimate |
|---|---|
| **U.S. Population Total** | 258,491,091 |
| High school degree or equivalent | 70,452,401 |
| Some college or above | 160,159,077 |
| **Income** | |
| < $50,000 | 81,721,988 |
| $50,000 - < $100,000 | 78,943,530 |
| $100,000 + | 97,825,573 |
| **Region** | |
| Northeast | 45,453,743 |
| Midwest | 53,423,807 |
| South | 98,524,550 |
| West | 61,088,991 |
| **Home Ownership** | |
| Own or buying | 179,164,068 |
| Rent/other | 79,327,023 |
| **Metropolitan Status** | |
| Metropolitan | 222,710,521 |
| Nonmetropolitan | 35,780,570 |
| **Marital Status** | |
| Married | 130,013,752 |
| Not married | 128,477,339 |
| **Ever Diagnosed with High Cholesterol** | |
| Yes | 73,749,057 |
| No | 184,742,034 |
| **Difficulty Participating in Social Activities** | |
| No difficulty/some difficulty | 247,475,871 |
| A lot of difficulty/cannot do | 11,015,220 |
| **Civic Engagement** | |
| Yes | 156,419,481 |
| No | 102,071,610 |

### 3.4.3 Weight calibration methodology

The procedure WTADJUST in SUDAAN (RTI International, 2012) was used for all weight calibration. Starting with a respondent data set $S$ containing weights scaled to the population size, WTADJUST generates adjustment factors $\{\alpha_k\}$ that convert the scaled weights $\{d_k\}$ into calibrated weights $\{w_k\}$, that is to say, $w_k = d_k\alpha_k$. Given a vector of calibration variables, $\mathbf{z}_k$ (the levels of the categorical variables in the table above are converted into dummy variables) and a vector of totals for those variables, $\mathbf{Z}_k$, the $w_k$ are chosen to solve the calibration equation,

$$\sum_{k\in S} w_k \mathbf{z}_k = \sum_{k\in S} d_k \alpha_k \mathbf{z}_k = \mathbf{Z}_k,$$

given adjustment factors with an appropriate $\beta$ having the form:

$$\alpha_k = \alpha(\mathbf{g}^T \mathbf{z}_k) = \frac{\frac{u}{u-1}\exp(\frac{u}{u-1}\boldsymbol{\beta}^T \mathbf{z}_k)}{1+\frac{u}{u-1}\frac{\exp(\frac{u}{u-1}\boldsymbol{\beta}^T \mathbf{z}_k)}{u}},$$

where $u > 1$ is the upper bound for the adjustment factors. This is a generalization of the raking algorithm. Under that algorithm $u$ is effectively set to infinity. In practice, $u$ is set by judgment with the aim of reducing the variability of the adjustment factors and thus the final weights. Here the choices for $u$ had little impact on the variability of the adjustment factors. The value of $u$ was set to 3 for respondent samples from both panels.

## 3.5 Confidentiality

NCHS is required to follow Section 308(d) of the Public Health Service Act (42 U.S.C.m(d)), which forbids disclosure of any information that may compromise the confidentiality promised to survey respondents. In addition, confidentiality protections are mandated by the Confidential Information Protection and Statistical Efficiency Act of 2018 (Title III, Public Law No.115-435).

When releasing data files to the public, NCHS takes steps, including disclosure analysis and a formal review process, to minimize the likelihood that individuals participating in RSS can be identified. As a result, some information in the PUF is suppressed or coarsened to protect the confidentiality of respondents. Users wishing to analyze more detailed data may request access to the RUF, described in Section 4.1. For information about procedures for accessing NCHS restricted-use data, see https://www.cdc.gov/rdc/.

# 4. How to use RSS data

This section describes the data files available from RSS-3, their contents, and appropriate use, including variance estimation that accounts for the sample design.

## 4.1 Available files

The following data files and codebooks are available to researchers in an NCHS Research Data Center:

- RUF in SAS and CSV format. This file combines records for all sampled panelists, from both samples, for a total of 21,112 records.
- SAS import code for RUF.
- Codebook for RUF with frequencies.

The following data files and codebooks are publicly available:

- PUF in SAS and CSV format. This file combines respondent records from each panel, for a total of 8,375 records.
- SAS import code for PUF.
- Codebook for PUF with frequencies.
- Codebook for RUF without frequencies.

Indicators are provided in each dataset to distinguish which panel each record came from. The codebooks contain information on all variables in the corresponding dataset (public or restricted use). In addition to the variable name and label, frequency tables are provided for categorical variables and summary statistics are provided for numeric variables.

The PUF contains a subset of variables from the RUF, and some variables have been modified to protect the confidentiality of respondents (Section 3.5). The RUF contains variables not in the PUF, and for some variables, finer categories. Section 4.2 discusses which weights to use for which analyses. The RUF codebook without frequencies provides additional information about the RUF contents for researchers interested in requesting access to the RUF.

### 4.1.1 Organization of the data files

Each dataset and codebook presents the panel profile characteristics, followed by questionnaire variables as they appear in the questionnaire. This is followed by paradata, recoded variables, and weighting variables.

Codebook descriptions of recoded variables provide the source variable name(s). The universe for these variables can be found in the questionnaire (www.cdc.gov/nchs/data/rss/round3/questionnaire.pdf).

### 4.1.2 Variable naming conventions

The following naming conventions are used in the data files:

**Questionnaire Variables.** Questionnaire variables are labeled with a prefix associated with the section of the questionnaire they appear in (e.g., questions about chronic health conditions have the prefix *CHR_*). Derived variables that use multiple variables or produce a conceptually separate measure from the questionnaire variables are given a new name based on underlying construct (e.g., questions about frequency of food insecurity are used to create a Food Security Scale).

**Panel Variables and Paradata.** These variables are identified with a prefix depending on their source:

- *P_* indicates the variable is common to both providers or has been harmonized across the panels.
- *P1_* and *P2_* indicate the variable is specific to one panel provider.

**Recoded Variables**. If there is recoding that transforms the variable response categories (e.g., *P_AGE* [continuous] to *P_AGE_R* [categorical grouping of *AGE*]), a new variable is created with the suffix *_R*. If multiple recodes are performed on the same variable, a brief descriptive abbreviation is added to the *_R* to make a unique variable name, for example, *P_AGE* (continuous) to *P_AGE_R5* (5 level categorical grouping of *AGE*) and to *P_AGE_R7* (seven-level categorical grouping of *AGE*).

**Imputed Variables.** If a variable has been imputed (Section 3.3) then it has the naming convention <*varname*>*_I*. Corresponding imputation flags indicate which values were imputed and have the naming convention <*varname*>*_IFLG*. All imputation flags are defined such that:

- 0 = no imputation;
- 1 = logical assignment; and
- 2 = statistical imputation.

### 4.1.3 Missing values and reserve codes

The following reserve codes are used for all questionnaire variables where they apply:

- -5 = Edited response due to invalid logic, out of range responses, or identified outlier;
- -6 = Skipped question/Implied refusal;
- -7 = Explicit refusal (telephone interview);
- -8 = Question not asked (legitimate logical skip/out of universe); and
- -9 = Don't know.

## 4.2 Analyzing RSS data

Sample weights are required for any analysis using RSS data. Survey analysis software that can compute Taylor series variance estimates should be used to account for the complex sample design (Section 4.2.1). Sample weights are provided in both the RUF and PUF, along with strata and primary sampling unit (PSU) for variance estimation. The PUF includes three weights:

- *WEIGHT* for analysis of respondent data from both panel providers combined. This is the weight that was used to generate estimates in NCHS dashboards and web tables.
- *P1_CALWT* and *P2_CALWT* for analysis of respondent data by panel provider.

In addition, the RUF contains all weights and components produced by the panel providers. Additional information about these weights is available in the RUF codebook.

### 4.2.1 Variance estimation

Users of the PUF and RUF should use the Taylor series linearization method to estimate variances. Sample design variables are provided in the data files for this purpose. This method requires the use of statistical software with this functionality. Table 4-1 provides

examples of how to specify the sample design for variance estimation in five statistical software packages when using the weight *WEIGHT*. The same strata and PSU variables should be specified when panel-specific weights are used.

Some strata have only a single PSU, making traditional variance estimation impossible because the variance contribution from these strata cannot be computed (variance is computed as the sum of the strata-level variances). There are a variety of approaches for handling this. In Table 4-1, the examples shown for SUDAAN, Stata, and R's survey package use the same approach as used for variance estimates reported for RSS-3, which may overestimate variances slightly; by contrast the approach used by SAS and SPSS may underestimate variances slightly. Some statistical software packages provide users with multiple options; refer to software documentation for additional information. In practice, the choice of approach is unlikely to be important unless estimating the variance of a total, in which case users should exercise more caution in the choice of approach.

**Table 4.1 Examples of variance specification in statistical software**

| Software | Design specification |
|---|---|
| SUDAAN | DESIGN = WR<br>WEIGHT WEIGHT;<br>NEST P_STRATA P_PSU/ MISSUNIT; |
| Stata | svyset P_PSU [pweight = WEIGHT],<br>strata (P_STRATA) vce(LINEARIZED) singleunit(centered) |
| SAS survey data analysis procedures | VARMETHOD = TAYLOR<br>WEIGHT WEIGHT;<br>STRATA P_STRATA;<br>CLUSTER P_PSU; |
| IBM SPSS complex samples | CSPLAN ANALYSIS<br>/PLAN FILE='myfile.csaplan'<br>/PLANVARS ANALYSISWEIGHT= WEIGHT<br>/DESIGN STRATA= P_STRATA CLUSTER= P_PSU<br>/ESTIMATOR TYPE=WR |
| R survey package | mydesign<-svydesign(id=~ P_PSU, strata=~ P_STRATA, weights=~ WEIGHT, data=mydata)<br>options(survey.lonely.psu="adjust") |

### 4.2.2 User cautions

RSS data are not intended to be used for longitudinal analyses. Although additional rounds are forthcoming, users should not combine data from multiple rounds.

The RSS was not designed to replace national health surveys. Benchmarking variables in RSS were obtained using questions from other surveys and were collected for benchmarking and calibration purposes only. Because of the potential for increased bias

associated with online panels, the population estimates from these other surveys are still considered gold standard prevalence estimates. However, corresponding RSS estimates are useful for methodological purposes including benchmarking and studies of online panels.

Users of the PUF are cautioned against analyzing and interpreting near-zero or low-prevalence outcomes and estimates based on small sample sizes. PUF users are encouraged to implement suppression criteria to eliminate estimates considered to have low precision. See NCHS Vital and Health Statistics, Series 2, Number 175, August 2017 and Vital and Health Statistics, Series 2, Number 200 for more detail on NCHS data presentation standards for proportions and rates.

## 4.3 Strengths and limitations

RSS has notable strengths and significant limitations, which affect its suitability for different uses.

The primary strength of the RSS is its ability to produce data of known quality about key public health issues in a timely manner. By contrast, large-scale national surveys with well-established quality measures require more time to collect and release data, making them less able to adapt to time-sensitive needs. With multiple rounds each year, the RSS will provide ample opportunities for new topics to be explored. Questions that might not be appropriate for other surveys (e.g., opinions, beliefs, stated preferences, and hypotheticals) can be asked in the RSS. The RSS can also be used for methodological work. For example, if there is a need for data on topics for which questions have not gone through rigorous testing, the RSS can be used to test and improve measurement of the underlying constructs.

Although much faster than in-person surveys, online panel surveys face different threats to accuracy and usability. Online panel surveys often have lower response rates than large-scale national surveys and may underrepresent certain subpopulations, increasing the potential for nonresponse bias. Panel survey nonresponse occurs at many stages, including panel recruitment, panel retention, and at the individual survey level. The RSS aims to compensate for nonresponse through calibration and weighting of RSS data to gold standard NCHS surveys. However, the effectiveness of these weighting adjustments for nonresponse may vary across survey estimates and will depend on the availability of appropriate gold standard survey data. The RSS also includes a benchmarking component to facilitate bias assessments for a wide array of health-related estimates. These bias assessments will provide context on the effectiveness of the weighting adjustments and quality of estimates generated from RSS. Initial results are available the RSS-3 Quality Profile (https://www.cdc.gov/nchs/data/rss/round3/quality-profile.pdf).

Another limitation is that some of the demographic covariates, or profile variables collected by the panels separately from RSS, are collected at different times and using different questions in each panel. Although the two panels may have collected their profile data using different questions, wherever possible, RSS provides harmonized profile variables. Although they are updated regularly, it is not known whether any of these characteristics have changed between the last time the panel collected the information and the respondent completed the RSS-3 questionnaire. Throughout each round of the survey, NCHS uses its best judgment to determine when the profile variable questions asked of each panel's members are comparable enough to be used and when questions with consistent wording should be added to RSS surveys.

## References

American Association for Public Opinion Research (AAPOR). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. 10th edition. AAPOR. 2023.

RTI International. *SUDAAN user's manual*. Release 11.0. RTI International. 2012.

SAS Institute Inc. *SAS/STAT® 14.1 user's guide*. SAS Institute Inc. 2015.

## Acknowledgments

This report was prepared for the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), under Contract No. GS00F354CA with RTI International. Amy Brown served as the Contracting Officer's Representative.

## Suggested citation

Survey Description. NCHS Rapid Surveys System. Round 3. August 2024. National Center for Health Statistics. Available from: www.cdc.gov/nchs/data/rss/round3/survey-description.pdf.