NCHS Rapid Surveys System

NATIONAL CENTER FOR HEALTH STATISTICS

Survey Description

Round 7: Data collected July-August 2025

Last revised January 15, 2026

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 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 examine the resulting survey data for their appropriateness for use in public health surveillance.

The RSS is used multiple times per year to gather data about issues of importance to CDC and the U.S. Department of Health and Human Services (HHS). For each round, data are shared with the public through a variety of means, such as dashboards, tables, infographics, and data files with accompanying codebooks.

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 also provides decision makers with information on the quality of the estimates.

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 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 determination of data quality.

NCHS has been using online panels for methodological work since 2015 through the Research and Development Survey. 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. These transparent quality studies are documented in the Quality Profile, a report with accompanying tables, released each round. Although the Quality Profile provides 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 civilian, noninstitutionalized 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 to collect data either in-person or over the telephone. The sample design for the seventh round of RSS (RSS-7) 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 compare against relevant benchmarks, which determines 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 to produce an RSS data file. 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 intended to be used 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 determine its utility for calibration. Information about RSS-7 weight calibration can be found in Section 3.4.

Component 4 includes questions specifically used for benchmarking and communicating the utility of health estimates collected via RSS. Over time, the benchmarks may also be helpful in determining the quality of 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 determine the quality of the estimates. See Section 2.1 for information about RSS-7 benchmarking.

Each round of RSS data collection emphasizes different topics, and the content of each component varies. Rounds 1-3 used the same set of calibration variables, and each round since has adjusted the set of calibration variables according to subsequent examination of the performance of past calibration variables. 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-7 components can be found in Section 2.

This document provides guidance and information for users of RSS-7 data and estimates. Section 2 provides an overview of RSS-7 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-7 data files and guidance on variance estimation and data quality.

2.   Round 7 overview

This section summarizes the content and methodology studies for RSS-7. Additional information about RSS-7, including codebooks, survey questionnaire, and the quality profile is available.

2.1   Round 7 content

As described in Section 1.1, RSS-7 includes four components: demographic characteristics, new emerging or supplemental topics, calibration variables, and benchmarking variables, which are described below. For complete information about RSS-7 content, please see the RSS-7 codebook and questionnaire.

Demographic Questions. Demographic data are collected from the panel profile variables and from the questionnaire. Most demographic variables, such as age, race,

Hispanic origin, education, income, and employment, come from the panel profile variables; however, demographic questions may be included on a questionnaire to standardize the question wording and information collected between the AmeriSpeak Panel and KnowledgePanel, or if the panel profile data contain many missing values. For RSS-7, this includes questions on marital status, veteran status, state of residence, employment, race and Hispanic origin.

New, Emerging, or Supplemental Topic Questions (from contributing agencies). In RSS-7 there were four new emerging or supplemental topics:

1. Cigarette smoking and barriers to access for lung cancer screening [National Center for Chronic Diseases Prevention and Health Promotion (NCCDPHP), Division of Cancer Prevention and Control]
2. Glucagon-like peptide-1 (GLP-1) medications for obesity and weight-related health conditions [NCCDPHP, Division of Nutrition, Physical Activity, and Obesity]
3. Complementary and integrative health [National Institutes of Health, National Center for Complementary and Integrative Health]
4. Lyme disease prevention methods [National Center for Emerging and Zoonotic Infectious Disease (NCEZID), Division of Vector-Borne Disease]

See the Office of Management and Budget Information clearance review package for additional information.

Calibration Questions. The questions included in RSS-7 as potential weighting variables came from a range of topics including sociodemographics, healthcare access and utilization, chronic conditions, social and work limitations, internet use, civic engagement, and language used at home and in other settings. The final set of variables selected for calibration of RSS-7 weights included ever diagnosed with high cholesterol, difficulty participating in social activities, used health information technology, has a working cell phone, and did not receive care because of cost in the past 12 months. Several panel profile variables were also used for calibration: age, sex, race and Hispanic origin, educational attainment, poverty status, and metropolitan statistical area (MSA) status. See Section 3.4 for details about the calibration variables used in RSS-7.

Benchmarking Questions. Questions included for benchmarking of RSS-7 estimates fall under eight health domains: fertility and reproductive health, health status, chronic conditions, mental health, body mass index, health behaviors, healthcare access, and healthcare utilization. Fertility and reproductive health estimates were benchmarked to the 2022-2023 National Survey of Family Growth while other estimates were compared to benchmarks from the 2025 Quarter 1 NHIS. Results of these comparisons are reported in the RSS-7 Quality Profile.

2.2   Round 7 methodological studies

The content on complementary and integrative health was included in RSS-7 for methodological purposes to assess whether the population prevalence is high enough to achieve sufficient precision if the questions are included on the 2027 NHIS.

RSS-7 included accompanying cognitive interviews conducted by the NCHS Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER). Cognitive interviews probed on survey content related to the sponsored questions outlined in Section 2.1. For RSS-7, cognitive interviews were conducted concurrent with survey fielding. Because of this, cognitive interviews should be understood as an examination of the RSS-7 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, is available on Q-Bank and the RSS Data Files and Documentation webpages. Data users should refer to this report to understand what information the survey questions captured and to frame their own analysis of the RSS-7 data. Information on cognitive interviewing is available on the CCQDER webpage.

3.   Methodology

This section describes the procedures used to collect and prepare RSS-7 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. The target population of RSS-7 is all U.S. adults aged 18 and older.

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 for the civilian noninstitutionalized population. 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 eligible to participate in RSS-7.

The KnowledgePanel has approximately 62,000 adults in 52,000 households. Panel members are primarily recruited using address-based sampling (ABS) methods. The panel 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 Ipsos's KnowledgePanel page.↗

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 NORC's AmeriSpeak Panel page. ↗

3.1.2 Sample selection

A sample was selected from each frame using the sample designs described below. In both cases, only one panel member per household was selected to participate.

KnowledgePanel. For RSS-7, 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 age, sex, race and Hispanic origin, education, household income, Census region, and MSA status obtained from the March 2024 supplement of the Current Population Survey (CPS), and language proficiency from the 2023 ACS.

AmeriSpeak. For RSS-7, the frame was stratified by age, race and Hispanic origin, education, and sex. Sample allocation across the strata was determined such that the expected distribution of complete surveys matched that of census population benchmarks. 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 were conducted primarily by web; however, telephone interviews were used for some AmeriSpeak panelists known to prefer completing surveys over the phone. Respondents are provided the option to complete the interview in Spanish. About 3-5% of respondents choose to do so in each round.

Table 3.1 Interviewing modes and recruitment protocols by panel

Mode or Protocol	KnowledgePanel	AmeriSpeak
Data collection modes	Web	Web and phone
Incentives	1,000 points (equivalent of $1) for most panelists, and 5,000 points (equivalent of $5) for panelists from households with hard-to-reach members.[1]	$5 for web and $10 for phone completes
Collection period	July 18-August 6, 2025	July 14–August 6, 2025

Mode or Protocol	KnowledgePanel	AmeriSpeak
Protocol for recruitment	A prenotification email 2 days prior to fielding. Reminders on days 3, 7, 11, 15, 18, and 20 for hard-to-reach groups,[1] otherwise reminders on day 3. All reminders after day 3 were customized with email text used only for this study.	An email invitation on the first day of data collection. Up to six email and two SMS reminders. Up to three call attempts per phone-preference case.

[1] The hard-to-reach group is defined as young adults who are ages 18 to 29, are panelists of racial and ethnic minority groups, or are panelists with a high school or less education.

The number of completed interviews for the combined sample exceeded the sample size target, 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 in the public-use file (PUF). Unweighted completion rates for eligible panelists are provided in Table 3-2. Final weighted response rates that reflect the survey completion rate, household panel recruitment rate, and household panel retention or profile rates are reported in the RSS-7 Quality Profile.

Table 3.2   Target and completed interviews by panel and mode

Category	KnowledgePanel	AmeriSpeak	Combined
Sample size	5,692	7,272	12,964
Target number of interviews	4,000	4,000	8,000
Number of completed interviews	4,012	4,106	8,118
Web interviews	4,012	3,689	7,701
Phone interviews	-	417	417
Unweighted completion rate (%)[1]	70.5	56.5	62.6

[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

For RSS-7, there were several open-ended questions related to detailed race/ethnicity and one related to health insurance.

The health insurance question was reviewed in an "upcoding" process to determine whether the text response could be coded into a common response option. 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 response option "*Any other type of health insurance or health coverage plan (please specify)*" (*EMP_INSH*). Verbatim responses to this response option were reviewed, and responses were upcoded if respondents mentioned health insurance in one of the seven listed insurance types (*EMP_INSA – EMP_INSG*). Additionally, two variables were created during upcoding: (1) "*Private insurance, Other/Unknown Source*" (*EMP_INSH_UNK*), to capture verbatim responses that did not provide the source of their private insurance (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 11 variables used for calibration, 6 had missing values and were imputed including race and Hispanic origin, ever diagnosed with high cholesterol, difficulty participating in social activities, used health information technology, has a working cell phone, and did not receive care because of cost in the past 12 months. 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 implemented with the R package mice (van Buuren and Groothuis-Oudshorn, 2011). The imputation models used the default discriminant functions for classification variables, and the predictor variables included all 11 weighting variables and the sample design variables (i.e., strata and cluster). Imputation rates ranged from 0.5% to 2.1% 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, sex, and race and Hispanic origin.

Imputed variables and imputation flags are included in the data files (Section 4.1).

3.4 Sample weights

RSS is a sample survey conducted by a commercial survey panel. That is, only a sample (subset) of the civilian noninstitutionalized population is invited to join the panel,

and a sample of those panelists are selected to participate in the survey. Additionally, not everyone invited to join the panel or selected to participate in the survey agrees to participate, which can affect the representativeness of the sample. To account for these 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 RSS. The value of the weight for a given respondent can be interpreted as the number of persons in the RSS target population represented by that respondent.

Each panel provider produced weights for analysis of their panel's data. These weights were recalibrated to the population totals in Table 3-4 and used to create a composite weight for analysis of the combined samples. Section 3.4.1 describes the panel-specific weights, and Section 3.4.2 describes the creation of the composite weight. Information about the weights available in the restricted-use file (RUF) and PUF is provided in Section 4.2.

3.4.1 Panel-specific analysis weights

This section describes the weights computed by the panel providers, *P1_INTWT* and *P2_INTWT*, for analysis of their panel's data.

KnowledgePanel. The base weights for KnowledgePanel are modeled design weights that account for their most common observed patterns of nonresponse. These were calibrated to population totals for subgroups defined by age, sex, race and Hispanic origin, education, household income, language proficiency, and Census region by metropolitan status. These totals were obtained from the March 2024 Supplement of the CPS, except for language proficiency, which came from the 2023 ACS.

AmeriSpeak. Base weights were created using weights developed for the full AmeriSpeak panel, adjusted for unequal sample selection probabilities and frame coverage limitations. The base weights were adjusted for survey nonresponse using a weighting class method, with weighting classes defined by age, sex, race and Hispanic origin, and education. A raking ratio adjustment is then applied to align the sample with population benchmarks for subgroups defined by age, sex, race and Hispanic origin, income, education, Census division, and language other than English spoken at home. These totals were obtained from the March 2024 Supplement of the CPS, except for language other than English spoken at home, which came from the 2019-2023 ACS 5-year Public Use Microdata Sample.

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. This weight is calibrated so that weighted estimates of the combined sample match the population totals estimated from NHIS, shown in Table 3-4.

First, each of the panel provider weights, *P1_INTWT* and *P2_INTWT* (Section 3.4.1), was calibrated to the NHIS totals, using iterative proportional fitting, or raking, to generate the final panel-specific calibrated weights *P1_CALWT* and *P2_CALWT*. This was implemented using the R package survey (Lumley, 2010). Then, the composite calibrated weight, *WEIGHT*, was calculated from the final panel-specific calibrated weights as:

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

The adjustment factor λ_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 λ_1.

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

Panel	n	$n_{e,i}$	λ_1
AmeriSpeak	4,106	1,511	0.354
KnowledgePanel	4,012	2,759	0.646

Table 3.4   Calibration variables and NHIS population totals

Variable	NHIS estimate
U.S. total number of adults 18+	263,876,810
Age	
18-34	76,145,167
35-49	65,432,981
50-64	61,333,313
65 and older	60,965,349
Sex	
Male	128,505,028
Female	135,371,782
Race and Hispanic origin	
Black, non-Hispanic	33,385,942

Variable	NHIS estimate
White, non-Hispanic	160,125,627
Other and multiple races, non-Hispanic	21,788,188
Hispanic	48,577,053
Educational attainment	
Less than high school	27,076,067
High school diploma	71,710,116
Some college or more	165,090,627
Poverty status	
<100% federal poverty level (FPL)	23,884,879
100% FPL - < 200% FPL	48,668,905
200% FPL +	191,323,026
Metropolitan status	
Metropolitan	227,403,918
Nonmetropolitan	36,472,892
Ever diagnosed with high cholesterol	
Yes	76,000,755
No	187,876,055
Difficulty participating in social activities	
No difficulty	232,625,135
Some difficulty/a lot of difficulty/cannot do	31,251,675
Used health information technology	
Yes	192,096,913
No	71,779,897
Did not receive care due to cost in past 12 months	
Yes	16,858,251
No	247,018,559
Adult has a working cell phone	
Yes	257,417,804
No	6,459,006

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 accessing NCHS restricted-use data, see CDC's Research Data Center page.

4.   How to use RSS data

This section describes the data files available from RSS-7, 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 users 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 12,964 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 8,118 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 users interested in requesting access to the RUF.

4.1.1   Organization of the data files

Each dataset and codebook presents the questionnaire variables as they appear in the questionnaire, followed by paradata, panel profile characteristics, and weighting variables. Recoded variables appear near their source variables. Codebook descriptions of recoded variables provide the source variable name(s). The universe for these variables can be found in the RSS-7 questionnaire.

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. For example, questions about height and weight are used to derive body mass index.

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_AGE5YRS_R* [categorical grouping of *AGE*]), a new variable is created with the suffix *_R*.

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.

Calibration Variables. Variables used in weight calibration (Section 3.4.2) are identified with the prefix *CAL_*.

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-7, 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.

Table 4.1   Examples of variance specification in statistical software

Software	Design specification
SUDAAN	DESIGN = WR WEIGHT WEIGHT; NEST P_STRATA_R P_PSU_R/ MISSUNIT;
Stata	svyset P_PSU_R [pweight = WEIGHT], strata (P_STRATA_R) vce(LINEARIZED) singleunit(centered)
SAS survey data analysis procedures	VARMETHOD = TAYLOR WEIGHT WEIGHT; STRATA P_STRATA_R; CLUSTER P_PSU_R;

Software	Design specification
IBM SPSS complex samples	CSPLAN ANALYSIS /PLAN FILE='myfile.csaplan' /PLANVARS ANALYSISWEIGHT= WEIGHT /DESIGN STRATA= P_STRATA_R CLUSTER= P_PSU_R /ESTIMATOR TYPE=WR
R survey package	mydesign<-svydesign(id=~ P_PSU_R, strata=~ P_STRATA_R, nest=TRUE, weights=~ WEIGHT, data=mydata) options(survey.lonely.psu="adjust")

4.2.2 User cautions

The data on complementary and integrative health collected as part of RSS-7 are intended for methodological purposes and benchmarking and should not be used to generate national prevalence estimates. Generally, 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 known limitations 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, March 2023, for more detail on NCHS data presentation standards for proportions and rates.

RSS data are not intended to be used for longitudinal analyses. Users should not combine data from multiple rounds.

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 provides 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, in RSS-4 the RSS platform was used for studying data collection methodologies and comparing responses between panelists in the sample who had and

had not recently responded to other surveys (active vs. inactive panelists). Based on this study, decisions were made not to adopt a more intensive data collection protocol (report forthcoming) and to exclude inactive panelists from the RSS samples.

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. 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 that is used to provide context on the effectiveness of the weighting adjustments and quality of estimates generated from RSS. Benchmarking results for Round 7 are available in the RSS-7 Quality Profile.

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 use 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 when the respondent completed the RSS-7 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.

Lumley, T. *Complex surveys: A guide to analysis using R*. John Wiley and Sons. 2010.

Van Buuren, S., and Groothuis-Oudshorn, K. Multivariate imputation by chained equations in R. *Journal of Statistical Software,* 45 (3), 1-67. 2011.

Acknowledgments

Suggested citation

Survey Description. NCHS Rapid Surveys System. Round 7. February 2026. National Center for Health Statistics. Available from: https://www.cdc.gov/nchs/data/rss/round7/RSS7-survey-description.pdf.