Health
Information
National
Trends
Survey
hints

Health Information National Trends Survey 5 (HINTS 5)

Cycle 2 Methodology Report

July 2018

Prepared for
National Cancer Institute
9609 Medical Center Drive
Bethesda, MD 20892-9760

Prepared by
Westat
1600 Research Boulevard
Rockville, MD 20850

Westat®

Table of Contents

Tables

Appendices

Cycle 2 Overview 1

The Health Information National Trends Survey (HINTS) is a nationally-representative survey which has been administered every few years by the National Cancer Institute (NCI) since 2003. The HINTS target population is adults aged 18 or older in the civilian non-institutionalized population of the United States. The most recent version of HINTS administration (referred to as HINTS 5) will include four data collection cycles over the course of four years. The second round of data collection for HINTS 5 (Cycle 2) was conducted from January 26 through May 2, 2018 with a goal of obtaining 3,500 completed questionnaires. The protocol used for the collection of Cycle 2 is similar to those used in HINTS since 2011. This report summarizes the methodology, sampling, and procedures for Cycle 2. Data cleaning and weighting procedures as well as response rates are also discussed.

Content Focus

As with all rounds of HINTS, HINTS 5 is providing NCI with a comprehensive assessment of the American public's current access to and use of information about cancer across the cancer care continuum from cancer prevention, early detection, diagnosis, treatment, and survivorship. The content of each HINTS 5 data collection cycle focuses on understanding the degree to which members of the general population understand vital cancer prevention messages. In addition to this standard HINTS content, each round of HINTS 5 has specific topical content for trending in areas of recent developments in the communication environment. For Cycle 2, the topics of special interest included:

- **Caregiving**. A series of questions focuses on the experience of people who act as caregivers and explores how much time is spent providing care, activities done, and whether the caregiver uses/needs support services. The questions are asked of caregivers for any condition, not just cancer.

- **Palliative care**. The palliative care questions are aimed at finding out how much people know about palliative care, sources of information about palliative care, and attitudes about palliative care.

- **Family cancer history**. Questions added to the Beliefs About Cancer section explore how much people know about the history of cancer in their family and whether they track that information.

Sample Selection 2

The sampling strategy for the HINTS 5 Cycle 2 survey consisted of a two-stage design. In the first stage, a stratified sample of addresses was selected from a file of residential addresses. In the second stage, one adult was selected within each sampled household.

2.1 Sampling Frame

As with prior HINTS iterations, the sampling frame for Cycle 2 consisted of a database of addresses used by Marketing Systems Group (MSG) to provide random samples of addresses. All non-vacant residential addresses in the United States present on the MSG database, including post office (P.O.) boxes, throwbacks (i.e., street addresses for which mail is redirected by the United States Postal Service to a specified P.O. box), and seasonal addresses were subject to sampling.

Rarely are surveys conducted with a sampling frame that perfectly represents the target population. The sampling frame is one of the many sources of error in the survey process. The sampling frame used for the address sample for Cycle 2 contained duplicate units because some households receive mail in more than one way. To permit adjustment for this duplication of households in the sampling frame, a question about how many different ways respondents receive mail was included on the survey instrument (see question O19).

In rural areas, some of the addresses do not contain street addresses or box numbers. Simplified addresses contain insufficient information for mailing questionnaires. Consequently, alternative sources of usable addresses were used when a carrier route contained simplified addresses. This partially ameliorated the frame's known undercoverage of rural areas although the actual coverage and undeliverable rates for this portion of the frame is not known.

2.2 Stratification

The sampling frame of addresses was grouped into two explicit sampling strata:

1. Addresses in areas with high concentrations of minority population; and
2. Addresses in areas with low concentrations of minority population.

The high and low minority strata were formed using the census tract level characteristics from the 2012–2016 American Community Survey data file. Addresses in census tracts that had a population proportion of Hispanics or African Americans that equaled or exceeded 34 percent were assigned to the high-minority stratum. All the remaining addresses were assigned to the low-minority stratum.

The purpose of creating high- and low-minority strata and then oversampling the high-minority stratum is to increase the precision of estimates for minority subpopulations. The gains in precision stem from the increase in sample sizes for the minority subpopulations produced by the oversampling.

2.3 Selection of Address Sample

An equal-probability sample of addresses was selected from within each explicit sampling stratum. The total number of addresses selected for Cycle 2 was 14,590: 10,130 from the high minority stratum and 4,460 from the low minority stratum. The high-minority stratum's proportion of the sampling frame was 26.1 percent, and it was oversampled so that its proportion of the sample was 69.4 percent. Conversely, the low minority stratum comprised 73.9 percent of the sampling frame, but made up just 31.6 percent of the sample.

Table 2-1 below summarizes the address sample for Cycle 2, showing the number of sample addresses, the percent of addresses in the frame and sample, and the percent oversampled/under-sampled relative to a proportional design, by sampling stratum. As part of the deduplication process, Westat determined that four out of the 14,590 households selected for the Cycle 2 sample were also selected for the previous cycle. To avoid overburdening these households with another HINTS survey in a relatively short time frame, these four households were excluded from data collection, resulting in a final sample size of 14,586.

Table 2-1. Cycle 2 sample summary by sampling stratum

Stratum	Number of sample addresses	Percent of addresses in the frame	Percent of sample addresses	Percent of sampled addresses oversampled (+) or undersampled (-)
High minority areas	10,130	26.1	69.4	+165.9
Low minority areas	4,460	73.9	31.6	-57.2
Total Sampled	14,590			
Deduplication	-4			
Sample for Mailing	14,586			

2.4 Within-Household Sample Selection

The second-stage of sampling consisted of selecting one adult within each sampled household. In keeping with previous cycles of HINTS, data collection for Cycle 2 implemented the Next Birthday Method to select the one adult in the household. The within-household selection was conducted by the respondents themselves. Questions were included on the survey instrument to assist the household in selecting the adult in the household having the next birthday (see Page 1 of the survey instrument).

Data Collection 3

Data collection for Cycle 2 started on January 26, 2018 and concluded on May 2, 2018. The survey was conducted exclusively by mail with a $2 pre-paid monetary incentive to encourage participation. The specific mailing procedures and outcomes are described in detail below.

3.1 Mailing Protocol

A total of four mailings were sent out as part of Cycle 2. All households in the sample received the first mailing and reminder postcard, while only non-responding households received the subsequent survey mailings. All households received one English survey per mailing unless someone from the household contacted Westat to request a Spanish survey, in which case the household received one Spanish survey per mailing for all subsequent mailings.

The mailing protocol followed a modified Dillman approach (Dillman, et al., 2009) with a total of four mailings: an initial mailing, a reminder postcard, and two follow-up mailings. The second survey mailing was sent via USPS Priority Mail, while all other mailings were sent First Class. The contents of all mailings are further described in Table 3-1. Cover letters in English can be found in **Appendix A** and cover letters in Spanish are in **Appendix B**. All cover letters include a list of Frequently Asked Questions (FAQs) on the back. These FAQs in both English and Spanish are in **Appendix C**.

Table 3-1. Mailing protocol

Mailing	Date(s) mailed	Mailing method	Materials
Mailing 1	January 26, 2018	1st Class Mail	English cover letter with FAQs English Questionnaire Postage-paid return envelope $2 bill
Postcard	February 2, 2018	1st Class Mail	Reminder/thank you postcard
Mailing 2	February 21, 2018	USPS Priority Mail	English cover letter with FAQs English questionnaire Postage-paid return envelope OR (upon request) Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope
Mailing 3	March 20, 2018	1st Class Mail	English cover letter with FAQs English questionnaire Postage-paid return envelope OR (upon request) Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope

The number of packets sent per mailing is outlined in Table 3-2. Households who sent in completed questionnaires were removed from further mailings. In addition, households with packets that were returned by the Postal Service as "undeliverable" were removed from any further mailings.

Table 3-2. Number of packets per mailing

Mailing	English	Spanish	Total
Mailing 1	14,586	N/A	14,586
Mailing 2	11,710	18	11,728
Mailing 3	10,043	24	10,067
Total	36,339	42	36,381

3.2 In-bound Telephone Calls

Two toll-free telephone numbers were provided to respondents: one was used for English calls and one was used for Spanish calls. Both numbers were provided in each mailing. Respondents were told that they could call the number if they had questions, concerns, or if they needed to request materials in Spanish. Each number had a HINTS-specific voicemail message that instructed callers to leave their contact information and the reason for the call, and then a study staff member would return their call. The Spanish line was staffed by a native Spanish speaker. When voicemails were received, they were logged into the Study Management System (SMS) and the request was either processed (such as recording their desire for a Spanish questionnaire) or the respondent was called back to ascertain the respondent's need if it was not clear from the message. Callers stating they did not want to participate in the study were coded as "refusal" and removed from any subsequent mailings.

The two toll-free lines together received 54 calls throughout the Cycle 2 field period (see Table 3-3 below). A majority of the in-bound calls were respondents calling in with some form of comment or question or refusals. The rest were to request Spanish materials. 23 calls were not resolved because they were either hang-ups or non-informative messages and study staff were never able to reach the respondent.

Table 3-3. Telephone calls received

Reason for call	Number of calls received
Request for a Spanish questionnaire	9
Refusal	7
Respondent let the study team know that the survey had been completed	6
Respondent wanted to refuse but did not leave enough information for this request to be completed	1
Respondent asked a question. Topics included: assurance that HINTS was a legitimate study, the "scoring" of the HINTS survey (referencing the data tab on the study website), the deadline to submit the survey, confirmation that they should have received a survey, requests for the survey in languages other than Spanish, and questions about eligibility	8
Calls that were never resolved due to hang ups or non-informative messages	23
Total	54

3.3 Incoming Questionnaires

Field room staff receipted all returned questionnaires into the SMS using each questionnaire's unique barcode. The SMS tracked each received questionnaire as well as the status of each household (nonresponsive or complete). Once a household was recorded as complete, it no longer received any additional mailings. Packages that came back as undeliverable were marked as such in the SMS and those addresses did not receive any further mailings.

In addition to refusing by calling the toll-free line, some respondents also refused by sending a letter stating that they did not wish to participate or asking to be removed from the mailing list. These households were marked in the system as refusals and were removed from subsequent mailings. Respondents who sent back a blank questionnaire were not considered refusals and continued to receive mailings. The status of all households at the end of data collection (but before cleaning and editing) can be found in Table 3-4.

Table 3-4. Household status at close of data collection

Household status	English	Spanish	Total	
			N	%
Complete	3,513	14	3,527	24.2
Refusal			46	0.3
Undeliverable			1,752	12.0
Nonresponse			9,261	63.5
Total			14,586	100.0

The number of questionnaires returned by date during the field period can be found in Table 3-5. The majority of the returns were early in the field period, with 60 percent of returns coming in after the first mailing of the survey and the mailing of the reminder postcard. The second mailing resulted in an additional 27 percent and the remaining 13 percent were in response to the final mailing.

Table 3-5. Survey response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: January 26	January 29- February 4	164
Postcard: February 2	February 5- February 23	1,935
Mailing 2: February 21	February 24- March 22	967
Mailing 3: March 20	March 23- May 2	461
	Total	3,527

Data Management 4

After being processed and receipted into the SMS, each returned questionnaire was scanned, verified, cleaned, and edited. Imputation procedures were also conducted. These procedures are described below.

4.1 Scanning

All completed questionnaires were scanned using a data capture software (TeleForm) to capture the survey data and images were stored in SharePoint. Staff reviewed each form as it was prepared for scanning. The review included:

- Determining if the form was not scannable for any reason, such as being damaged in the mail. Some questionnaires or individual responses needed to be overwritten with a pen that was readable by the data capture software. Numeric response boxes were pre-edited to interpret and clarify non-numeric responses and responses written outside the capture area.

- Documenting potential problem questionnaires or pertinent comments made by respondents in a decision log. Comments in Spanish were reviewed by a Spanish-speaking staff member.

The reviewed surveys were then sent through the high-speed scanner to capture the responses. TeleForm read the form image files and extracted data according to HINTS 5 Cycle 2 rules established prior to the field period. Scanned data were then subject to validation according to HINTS specifications. If a data value violated validation rules (such as marking more than one choice box in a mark-only-one question) the data item was flagged for review by verifiers who looked at the images and the corresponding extracted data and resolved any discrepancies. Spanish forms were verified by a Spanish-speaking staff member.

In addition to the problem log mentioned above based on pre-scan staff review, decisions made about data issues as a result of scanning were also recorded in a data decision log. The decision log contains the respondent ID, the value triggering the edit, the updated value, and the reason for the update. A total of 30 entries were made into the data decision log during the course of data processing. The majority of these were attributed to multiple response options selected on a gate question.

A 10 percent quality control check was then conducted on the scanned data and the electronic images of the survey. Quality Assurance (QA) staff compared the hard copy questionnaire to the data captured in the database item-for-item and the images stored in the repository page-for-page to ensure that all items were correctly captured. If needed, updates were made. In addition, QA staff closely reviewed frequencies and cross tabulations of the HINTS raw data to identify outliers and open ended items to be verified. ID reconciliation across the database, images, and the SMS, was completed to confirm data integrity.

4.2 Data Cleaning and Editing

Once scanned, the data were cleaned and edited. General cleaning and editing activities are described briefly below, with more detailed information found in **Appendix D** (Variable Values and Data Editing Procedures).

- Customized range and logical inconsistency edits, following predetermined processing rules to ensure data integrity, were developed and applied against the data.
- Edit rules were created to identify and recode nonresponse or indeterminate responses.

- Missing values were recoded for some responses to questions that featured a forced-choice response format and for filter questions where responses to later questions suggested a particular response was appropriate.

- Derived variables were created to reflect each response recorded for certain “mark-one” type questions (A2, A8, F3, F4, and O12), in order to facilitate the imputation process implemented when respondents did not follow the instruction to mark only one response. For these variables, imputation, as described in Section 4.3, was carried out. For other “mark-one” type questions where respondents marked multiple responses, editing rules were used to determine which response was retained.

- Derived variables were created to reflect each response recorded for certain “mark-one” type questions (A2, A8, F3, F4, and O12), in order to facilitate the imputation process implemented when respondents did not follow the instruction to mark only one response. For these variables, imputation, as described in Section 4.3, was carried out. For other “mark-one” type questions where respondents marked multiple responses, editing rules were used to determine which response was retained.

- Data cleaning was carried out for the two height variables: Height_Feet and Height_Inches. The rules that were applied minimized the number of out-of-range values by accounting for response measurements in incorrect boxes, responses using metric measures, responses using only one unit of measurement and other response errors.

- Data cleaning was carried out for question H5, the estimated number of calories needed daily, to maintain current weight. In 32 cases, respondents entered a number in the response box, but also checked the “Don’t know” box. For those cases, after verifying that the responses were accurately entered as written, both responses were retained.

- “Other, specify” responses were examined, cleaned for spelling errors, categorized, and upcoded into preexisting response codes when applicable. On one of these questions (E3), some of the responses were especially difficult to categorize because they could potentially have been upcoded into multiple categories. In those instances, the response was left as entered in the “Other, specify” field.

4.3 Imputation

The questions for which respondents selected more than one response were recoded to -5 and subject to imputation. A single answer was imputed by selecting one response among those selected by the respondent. The selection of the imputed response was based on the distribution of answers among the single-answer responses. This is the same imputation process as was conducted for all cycles of HINTS 4 as well as the first cycle of HINTS 5. Imputation occurred for 385 respondents on question A2, for 189 respondents on question A8, 76 respondents on question F3, 45 respondents on question F4, and 6 respondents on question O1. An imputation flag is included on the data-set to indicate imputed values.

In addition, hot-deck imputation was used to replace missing responses for items used in the raking procedure for the weighting. Specifically, this was conducted for items C7, K1, M1, O1, O5, O6, O10 and O11. Hot-deck imputation is a data processing procedure in which a value is assigned with the corresponding value of a "similar" case in the same imputation class. The data record that supplies the imputed value is referred to as the "donor." Under a hot deck approach, the resulting distribution preserves the distribution of values observed for respondents. Imputation classes are defined on the basis of variables that are thought to be correlated with the item with missing values. A donor is then randomly selected within an imputation class to supply the imputed value. Items imputed using the hot-deck approach were those involving the following characteristics: age, gender, educational attainment, marital status, race, ethnicity, health insurance coverage, and cancer diagnosis.

4.4 Determination of the Number of Household Adults

For the purpose of applying weights, a measure of the number of adults in each household (R_HHAdults) was created using questionnaire responses. The initial measure was taken from responses to demographic section questions asking for the total number of people and the number of children in the household (see items O13-O15). Implausible or missing values that resulted from the answers to those questions were substituted with values to questions on the respondent-selection page of the questionnaire and further substituted with data from the demographic section roster. Edits were carried out to reconcile different values reported within households and correct differences with the receipted number of returned questionnaires. A detailed list of the steps carried out to identify the number of adults in each household is included in **Appendix D**.

4.5 Survey Eligibility

Returned surveys were reviewed to ensure they were eligible for inclusion in the final dataset. Of the 3,547 questionnaires received, a total of 3,504 were determined to be eligible. Of the questionnaires determined to be ineligible, 2 were returned by respondents who reported an age below 18, and 2 were filled out with suspicious responses which suggested the respondents were not truthful in their answers[1]. Surveys were also reviewed for completion and duplication (more than one questionnaire returned from the same household), with 19 determined to be incomplete and 20 identified as duplicates. The processes for these reviews are detailed below.

Definition of a Complete and Partial Complete Questionnaire

The procedures in HINTS 5 Cycle 2 for determining whether or not a returned questionnaire was complete were the same as for Cycles 1, 2, and 3 in HINTS 4[2], as well as Cycle 1 in HINTS 5. A complete questionnaire was defined as any questionnaire with at least 80 percent of the required questions answered in Sections A and B. A partial complete was defined as when between 50 percent and 79 percent of the questions were answered in Sections A and B. There were 70 partially completed questionnaires. Both partially-completed and completely-answered questionnaires were retained. The 19 questionnaires with fewer than 50 percent of the required questions answered in Sections A and B were coded as incompletely-filled out and discarded. The 19 incomplete questionnaires represented 0.5% of all returns, which was slightly lower than Cycles 1, 2, and 3 in HINTS 4 and Cycle 1 in HINTS 5. Data for 70 partially completed and 3,434 competed questionnaires was included in the final dataset for a total of 3,504 surveys.

Eligibility of Multiple Questionnaires from a Household

Twenty households returned more than one filled in questionnaire. The procedures to deal with this issue followed the same guidelines that were used for HINTS 4:

- If the same respondent returned multiple questionnaires, the first questionnaire received was retained.
- If the same respondent returned multiple questionnaires on the same day, the first questionnaire to complete the editing process was retained.
- If a return date was unavailable for questionnaires from the same respondent, the questionnaire with fewer substantive questions omitted was retained.

[1] These four households were treated as nonresponding households in the response rate calculation.

[2] HINTS 4 Cycle 4 used a modified definition due to an unusually high incompletion rate, See the *HINTS 4 Cycle 4 Methodology Report*.

- If different respondents returned a questionnaire and the ages of household members listed in the roster were in agreement (or differed by only one year), the questionnaire that complied with the next birthday rule was retained.[3]
- If, in the above situation, compliance for one or both questionnaires from a household was unclear, the first questionnaire returned was retained.
- If different respondents returned a questionnaire and the ages of household members listed in the roster question were not substantively in agreement, the earliest questionnaire received that complied with the next birthday rule was retained.

4.6 Additional Analytic Variables

Included in the delivery files are two sets of new analytical variables introduced in HINTS 5: 1) rural-urban commuting area (RUCA) codes that classify census tracts using measures of population density, urbanization, and daily commuting; and 2) National Center for Health Statistics (NCHS) urban-rural classification scheme for counties.

The two RUCA codes (primary and secondary) provide a detailed and flexible way for delineating sub-county components of rural and urban areas. They are based on the 2006-10 American Community Survey (ACS) and have been updated using data from the 2010 decennial census. The primary codes (PR_RUCA2010) delineate metropolitan and nonmetropolitan areas based on the size and direction of primary commuting flows. The secondary codes (SEC_RUCA2010) further subdivide the primary codes to identify areas where classifications overlap based on the size and direction of the secondary, or second largest, commuting flow.

The NCHS Urban–Rural Classification Scheme for Counties (NCHSURCODE2013) was developed in 2013 for use in studying associations between urbanization level of residence and health and for monitoring the health of urban and rural residents. The scheme groups counties and county-equivalent entities into six urbanization levels (four metropolitan and two nonmetropolitan), on a continuum ranging from most urban to most rural.

[3] Compliance was determined by whether the person listed in the roster who matched the respondent's age and gender had a month of birth that was the first to follow the month in which the questionnaire was returned.

4.7 Codebook Development

Following cleaning, editing, and weighting (described below), a detailed codebook including frequencies was created for both the weighted and unweighted data. The codebooks define all variables in the questionnaires, provide the question text, list the allowable codes, and explain the inclusion criteria for each item. The English and Spanish instruments were annotated with variable names and allowable codes to support the usability of the delivery data.

Weighting and Variance Estimation 5

Every sampled adult who completed a questionnaire in HINTS 5 Cycle 2 received a full-sample weight and a set of 50 replicate weights. The full-sample weight is used to calculate population and subpopulation estimates. Replicate weights are used to compute standard errors for these estimates. The use of sampling weights is done to ensure valid inferences from the responding sample to the population, correcting for nonresponse and noncoverage biases to the extent possible.

The computation of the full-sample weights for Cycle 2 consisted of the following steps:

- Calculating household-level base weights;
- Adjusting for household nonresponse;
- Calculating person-level initial weights; and
- Calibrating the person-level weights to population counts (also known as control totals).

Replicate weights were calculated using the ‘delete one’ jackknife (JK1) replication method.

5.1 Household Base Weights

The initial step in the weighting process is calculating the household-level base weight for each household in the sample. The household base weight is the reciprocal of the probability of selecting the household for the survey, which depends on the stratum the household was selected from. Generally, base weights for units in the oversampled stratum are smaller than those in the stratum that was not oversampled. In Cycle 2, the base weights for households in the high minority stratum were roughly 1/6 the size of those in the low minority stratum.

If two different addresses led to the same household – for example, if a household receives mail via both a street address and a post office box – that household had twice the chance of selection of a household with only one address (and should therefore receive half the normal weight). An additional adjustment was made to the base weights of households that had multiple ways of receiving mail (as determined by the answer to survey question O19).

5.2 Household Nonresponse Adjustment

Nonresponse is generally encountered to some degree in every survey. The first and most obvious effect of nonresponse is to reduce the effective sample size, which increases the sampling variance. In addition, if there are systematic differences between the respondents and the nonrespondents, that also will be a bias of unknown size and direction. This bias is generally adjusted for in the case of unit nonrespondents (nonrespondents who refuse to participate in the survey at all) with the use of a weighting adjustment term multiplied to the base weights of sample respondents. Item nonresponse (nonresponse to specific questions only) is generally adjusted for through the use of imputation. This section discusses weighting adjustments for unit nonresponse.

The most widely accepted paradigm for unit nonresponse weighting adjustment is the quasi-randomization approach (Oh & Scheuren, 1983). In this approach, nonresponse cells are defined based on those measured characteristics of the sample members that are known to be related to response propensity. For example, if it is known that males respond at a lower rate than females, then sex should be one characteristic used in generating nonresponse cells. Under this approach, sample units are assigned to a response cell, based on a set of defined characteristics. The weighting adjustment for the sample unit is the reciprocal of the estimated response rate for the cell. Any set of response cells must be based on characteristics that are known for all sample units, responding and nonresponding. Thus questionnaire items on the survey cannot be used in the development of response cells, because these characteristics are only known for the responding sample units.

Under the quasi-randomization paradigm, Westat models nonresponse as a “sample” from the population of adults in that cell. If this model is in fact valid, then the use of the quasi-randomization weighting adjustment eliminates any nonresponse bias (see, for example, Little & Rubin (1987), Chapter 4).

The weighting procedure for Cycle 2 used a household-level nonresponse adjustment procedure based on this approach. The base weights of the households that did return the questionnaire were adjusted to reflect nonresponse by the remaining eligible households. A search algorithm[4] was used to identify variables highly correlated with household-level response, and these variables were used to create the nonresponse adjustment cells. The variables used to define nonresponse cells for Cycle 2 were:

- Sampling stratum (High Minority; Low Minority)
- Census region (Northeast; South; Midwest; West)
- Route type (Street address; other addresses such as PO Box, Rural Route, etc.)
- Metropolitan Status (county in Metro areas; county in Non-Metro areas)
- High Spanish linguistically isolated area (Yes; No)

Nonresponse adjustment factors were computed for each nonresponse cell *b* as follows:

$$HH\_NRAF(b) = \frac{\sum_{S(b)} HH\_BWT_i}{\sum_{C(b)} HH\_BWT_i},$$

where *HH_BWT_i* is the base weight for sampled household *i*, *S(b)* is the set of all eligible sampled households) in nonresponse cell b, *C(b)* is the set of all cooperating sampled households in cell *b*, and *HH_NRAF(b)* is the household nonresponse adjustment factor for nonresponse cell *b*.

The household nonresponse adjustment factors ranged from a low of 1.88 to a high of 5.57, and averaged 3.04 across all nonresponse adjustment cells.

[4] An inhouse macro WESSEARCH, which calls the Search software, a freeware product developed by the University of Michigan (http://www.isr.umich.edu/src/smp/search/).

5.3 Initial Person-Level Weights

Each sampled adult in responding households was assigned an initial person-level weight. The initial person-level weight was calculated by multiplying the nonresponse-adjusted household weight by the reciprocal of the sample person's within-household probability of selection. Because only one adult per household was selected to participate in the survey, the reciprocal of the sample person's within-household probability of selection is identical to the number of adults in the household. So, for example, if a household contained three adults and one adult was selected, the initial weight for the selected adult is equal to the nonresponse-adjusted household weight times three.

5.4 Calibration Adjustments

The purpose of calibration is to reduce the sampling variance of estimators through the use of reliable auxiliary information (see, for example, Deville & Sarndal, 1992). In the ideal case, this auxiliary information usually takes the form of known population totals for particular characteristics (called *control totals*). However, calibration also reduces the sampling variance of estimators if the auxiliary information has sampling errors, as long as these sampling errors are significantly smaller than those of the survey itself.

Calibration reduces sampling errors particularly for estimators of characteristics that are highly correlated to the calibration variables in the population. The extreme case of this would be the calibration variables themselves. The survey estimates of the control totals would have considerably higher sampling errors than the "calibrated" estimates of the control totals, which would be the control totals themselves. The estimator of any characteristic that is correlated to any calibration variable will share partially in this reduction of sampling variance, though not fully. Only estimators of characteristics that are completely uncorrelated to the calibration variables will show no improvement in sampling error. Deville and Sarndal (1992) provide a rigorous discussion of these results.

Control Totals

The American Community Survey (ACS) of the U.S. Census Bureau has much larger sample sizes than those of HINTS. The ACS estimates of any U.S. population totals have lower sampling error than the corresponding HINTS estimates, making calibration of the survey weights to ACS control totals beneficial. Westat used the 2016 ACS estimates that are publically available on the Census Bureau web site.

Calibration variables were selected among those that were on the ACS public-use file and were found to be well correlated to important HINTS questionnaire item outcomes (i.e., Westat wanted ACS-available characteristics that tend to have differing mean values for HINTS questionnaire item outcomes). The following ACS characteristics correlate well with HINTS questionnaire items:

- Age
- Gender
- Educational Attainment
- Marital Status
- Race
- Ethnicity
- Census Region

In addition to characteristics from the ACS, two health-related variables were used: *Percent with health insurance* and p*ercent of adults who have ever been diagnosed with cancer*. The *health insurance* variable came from the 2017 National Health Information Survey (NHIS) (Cohen, et al., 2017) and corresponds to the question asked in the HINTS survey (C7). The p*ercent of adults who have ever been diagnosed with cancer* variable came from the 2016 National Center for Health Statistics (U.S. Department of Health and Human Services, 2016) and corresponds to the question asked in the HINTS survey (M1).

Raking to the control totals for these variables (either alone or cross-classified with each other) was then performed. As a result of the raking HINTS weights to the control totals, estimates calculated from HINTS data for the control-total variables agree with those calculated from the source data for the control totals. For example, the national-level estimate of *Percent with health insurance* calculated from HINTS data agrees with the estimate calculated from NHIS 2017 data.

5.5 Replicate Variance Estimation

In addition to the full-sample weight, a set of 50 replicate weights were provided for each adult. These replicate weights are used to calculate standard error of estimates obtained from the HINTS data, using the delete one jackknife (JK1) replication method.

The JK1 jackknife technique is compatible with the sample design and weighting procedures for HINTS. This jackknife variance estimation technique takes carefully selected subsets of the data for each "replicate," and for each respondent in the replicate subset and determines a sampling weight, as if the replicate subset were in fact the responding sample. (This replicate subset is usually almost the entire sample, except for a group of respondents that are "deleted" for that replicate.) The resulting weights are called replicate weights.

The jackknife variance estimator requires the use of replicate weights. For the Cycle 2 data set, a set of 50 replicate weights was assigned to each responding adult. To illustrate how the replicate ~~weights~~ variance estimates are computed, suppose *P* is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A nationally representative estimator *p* can be computed by aggregating the adult sampling weights of all responding adults with this characteristic (e.g., all responding adults in the survey answering the survey question in a particular way). A JK1 jackknife variance estimator of the sampling variance of *p* can be computed in two steps:

Step 1. Recompute estimators *p(r)*, *r* =1,...,50, by aggregating the replicate sampling weights corresponding to replicate *r* for all responding adults with the characteristic.

Step 2. Compute the JK1 jackknife variance estimator

$$v(p) = \frac{R-1}{R}\sum_{r=1}^{50}(p(r)-p)^2$$

The replicate weights are computed by systematically deleting a portion of the original sample, and recomputing the sampling weights as if the remaining sample (without the deleted portion) were the actual sample. These deleted sample units should be first-stage sampling units, which in HINTS are households. The remainder of the sample with the deleted portion removed is called the replicate subset, and it should mirror the full sample design, as if it were a reduced version of the original sample.

For the purposes of JK1 jackknife variance estimation, each household was assigned to one of 50 replicate "deletion" groups *D(r), r* =1,..., 50. Each replicate sample is the full sample minus the deletion group (i.e., it is roughly 49/50 of the original sample).

The replicate sampling weights were generated in a series of steps that parallel the steps computing the full-sample sampling weights. The replicate base weight for each sampled household or adult and each replicate is either equal to *R*/*(R-1)* times the full sample base weight (if the household is contained in the replicate subset) or equal to 0 (if the household is not contained in the replicate subset, but instead is contained in the "deleted" set for that replicate).

Nonresponse and calibration adjustments were then computed for each set of replicate weights, using the replicate weights in the computation of nonresponse and calibration adjustments in place of the original weights. These calculations generated a set of replicate nonresponse and poststratification adjustments for each responding adult. The final replicate weights were products of the replicate weights, nonresponse adjustments, and calibration adjustments.

5.6 Taylor Series Variance Estimation

Even though replication is the recommended method for variance estimation for HINTS, not all software packages have a replication option to produce variance estimates. For example, SPSS has built-in options for estimating variance using Taylor's Series methods but not replication methods. To accommodate SPSS users or any end user who wants to produce variances using Taylor Series methods, Westat provided the appropriate variables on the HINTS data files to do so.

The full-sample weight (as described in the introduction of Section 5) is used as the weight to compute Taylor's Series variance estimates. The variable VarStratum indicates the variance-estimation stratum and the variable VarCluster indicates the primary sampling unit (PSU) or cluster within the variance-estimation stratum. These variables allow the analyst to produce variance estimates using Taylor's Series.

Response Rates 6

Response rates were calculated using the RR2 formula of the American Association of Public Opinion Research (AAPOR).

Table 6-1 shows the response rate outcomes overall and by strata. These data have been weighted to account for the oversampling of addresses in high-minority areas. The overall response rate was 32.9 percent; however this differed significantly by strata. The high-minority strata had the lowest response rate (23.2 percent) and the low-minority had the highest (36.6 percent). The percent of undeliverable households was similar across strata, with the low-minority strata exhibiting a slightly higher undeliverable rate (12.1 vs 12.0 percent).

Table 6-1. Response rate calculations by strata

Response class	High minority	Low minority	Overall
Total sample*	40,428,351	104,136,938	144,565,288
Respondents	8,257,281	33,505,943	41,763,224
Nonrespondents	27,330,044	58,045,836	85,375,879
Undeliverable	4,841,026	12,585,159	17,426,185
Total Households	35,587,325	91,551,779	127,139,103
Percent Undeliverable	11.97%	12.09%	12.05%
Household response rate	**23.20%**	**36.60%**	**32.85%**

*values may not sum to total sample due to rounding of weighted values to nearest single digit

References

Cohen, R.A., Martinez, M.E., and Zammitti, E.P. (2017). *Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January – March 2017*. Retrieved from https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur201708.pdf

Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method.* Hoboken, NJ: John Wiley & Sons.

Little, R., and Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Oh, H., and Scheuren, F. (1983). Weighting adjustments for unit response. In W.G. Madow, I. Olkin, and D. B. Rubin (Eds.), *Incomplete data in sampling surveys, Vol. II: Theory and annotated bibliography*. New York: Academic Press.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics (2016). *Summary Health Statistics: National Health Interview Survey, 2016*. Retrieved from https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2016_SHS_Table_A-3.pdf

Appendix A

Cover Letters in English

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

FIRST MAILING

Dear {City} Resident:

We are writing to invite you to take part in an important national survey sponsored by the U.S. Department of Health and Human Services, the Health Information National Trends Survey (HINTS). The goal of HINTS is to learn what health information people want to know and where they try to find it. By completing this questionnaire, you will help us learn what health information you need and how to make that information available to you, your family and your community.

In order to make sure we get responses from a random sample of people, we ask that the adult in your household with the next birthday complete and return this questionnaire in the next two weeks.

Your participation is voluntary and your responses will not be linked to your name. We have enclosed $2 as a token of our appreciation for your participation.

You can find out more about HINTS at hints.cancer.gov. Westat, a research firm, will conduct the survey. If you have any questions about HINTS {or if you need more questionnaires}, or if you would like to complete this survey in a language other than English or Spanish, please call Westat toll-free at 1-888-738-6805.

Thank you in advance for your cooperation.

Sincerely,

Bradford W. Hesse

Bradford W. Hesse, Ph.D.
HINTS Project Officer
National Institutes of Health
U.S. Dept. of Health and Human Services

Si prefiere recibir la encuesta en español, por favor llame al 1-888-738-6812.

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

POSTCARD TEXT

A few days ago you should have received a questionnaire packet asking for your household’s participation in the Health Information National Trends Survey. By completing the questionnaire, you can help the U.S. Department of Health and Human Services determine the best ways of communicating important health information to members of your community.

We are inviting the adult in the household with the next birthday to complete the questionnaire. If that adult has already completed the questionnaire and returned it to us, please accept my sincere thanks. If that adult has not yet completed and returned the questionnaire, we ask that he or she please do so as soon as possible.

Your household’s participation is important to the study’s success.

Sincerely,

Bradford W. Hesse

Bradford W. Hesse, Ph.D.
HINTS Project Officer
National Institutes of Health
U.S. Dept. of Health and Human Services

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

SECOND AND THIRD MAILINGS

Dear {City} Resident:

We recently invited you to participate in an important national survey sponsored by the U.S. Department of Health and Human Services (HHS). The goal of the Health Information National Trends Survey (HINTS) is to learn what health information people want to know and where they go to find it. Your responses will help us keep you, your family and members of your community better informed on the health issues that matter to you.

We have not yet received your completed questionnaire. To make sure HINTS provides accurate information, we need all the households invited to participate in this year's HINTS to complete the survey. If you did send back your survey and it crossed in the mail with this letter, thank you for the time you took to help make this study a success. In the event that your questionnaire was misplaced, an additional copy is enclosed.

In order to make sure we get responses from a random sample of people, we ask that the adult in your household with the next birthday complete and return this questionnaire in the next two weeks.

Additional information about HINTS is available at: hints.cancer.gov. If you have any questions, or would like to complete this survey in a language other than English or Spanish, please call Westat toll free at 1-888-738-6805.

Thank you in advance for contributing to this important national study.

Sincerely,

Bradford W. Hesse

Bradford W. Hesse, Ph.D.
HINTS Project Officer
National Institutes of Health
U.S. Dept. of Health and Human Services

Si prefiere recibir la encuesta en español, por favor llame al 1-888-738-6812.

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Appendix B

Cover Letters in Spanish

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

FIRST MAILING

Estimado residente de {City}

Le escribimos para invitarlo a participar en una importante encuesta nacional: Encuesta Nacional de Tendencias de Información sobre la Salud (HINTS, por sus siglas en inglés). Esta encuesta está patrocinada por el Departamento de Salud y Servicios Humanos de Estados Unidos.

El objetivo de HINTS es averiguar qué información sobre la salud les interesa saber a las personas y dónde tratan de buscarla. Complete este cuestionario para ayudar a averiguar la información sobre la salud que usted necesita y cómo ponerla a disposición suya, de su familia y de su comunidad.

Para asegurarnos de obtener respuestas que contengan un muestreo aleatorio de la población, le pedimos que el adulto en su hogar con el próximo cumpleaños, complete y devuelva este cuestionario en las próximas dos semanas.

Su participación es voluntaria y sus respuestas no se asociarán con su nombre. Hemos incluido $2 dólares como símbolo de nuestro agradecimiento por su participación.

Usted podrá encontrar más información sobre HINTS en el sitio web hints.cancer.gov. La compañía de estudios de investigación Westat está realizando esta encuesta. Si tiene alguna pregunta sobre HINTS o le gustaría completar esta encuesta en otro idioma distinto al inglés o español, llame a Westat al siguiente número de teléfono libre de cargo, 1-888-738-6812.

Gracias de antemano por su cooperación.

Atentamente,

Bradford W. Hesse

Bradford W. Hesse, Ph. D.
Oficial del Proyecto HINTS
Institutos Nacionales de la Salud
Departamento de Salud y Servicios Humanos de EE.UU.

La Encuesta Nacional de Tendencias de Información sobre la Salud está autorizada bajo la Sección 285A del USC 42.

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

SECOND AND THIRD MAILINGS

Estimado residente de {City}:

Recientemente lo invitamos a participar en una importante encuesta nacional patrocinada por el Departamento de Salud y Servicios Humanos de Estados Unidos. El objetivo de la Encuesta Nacional de Tendencias de Información sobre la Salud (HINTS, por sus siglas en inglés) es averiguar cuál es la información sobre la salud que las personas quieren saber y dónde van a buscarla. Sus respuestas nos ayudarán a mantenerlo mejor informado a usted, a sus familiares y a los miembros de la comunidad sobre los temas de salud que les interesan.

Aún no hemos recibido su cuestionario completado. Para poder estar seguros de que HINTS provea información acertada, necesitamos que todos los hogares invitados a participar en la encuesta este año, la completen. Si usted ya nos envió de regreso su encuesta y se cruzó con esta carta en el correo, le agradecemos por el tiempo que se tomó para contribuir al éxito de este estudio. En caso que su cuestionario se haya extraviado, adjuntamos una copia adicional.

Para asegurarnos de obtener respuestas que contengan un muestreo aleatorio de la población, le pedimos que el adulto en su hogar con el próximo cumpleaños, complete y devuelva este cuestionario en las próximas dos semanas.

Usted podrá encontrar más información sobre HINTS en el sitio web hints.cancer.gov. Si usted tiene preguntas o le gustaría completar esta encuesta en otro idioma distinto al inglés o español, llame a Westat al número libre de cargo, 1-888-738-6812.

Gracias de antemano por contribuir al éxito de este importante estudio nacional.

Atentamente,

Bradford W. Hesse

Bradford W. Hesse, Ph. D.
Oficial del Proyecto HINTS
Institutos Nacionales de la Salud
Departamento de Salud y Servicios Humanos de
EE.UU.

La Encuesta Nacional de Tendencias de Información sobre la Salud está autorizada bajo la Sección 285A del USC 42.

Appendix C

Frequently Asked Questions (FAQs)
English and Spanish

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: What is the study about? What kind of questions do you ask?

A: The study concerns health and how people receive health information. For example, we will ask how you usually get information about how to stay healthy, the sources of information you most trust, and how you might like to get such information in the future. We will also ask about your beliefs on what contributes to good health, how best to prevent cancer, your participation in various health-related activities, and related topics.

Q: How will the study results be used? What will be done with my information?

A: Findings will help the U.S. Department of Health and Human Services promote good health and prevent disease by determining ways of better communicating accurate health information to Americans.

Q: How did you get my address?

A: Your address was randomly selected from among all of the known home addresses in the nation. It was selected using scientific sampling methods.

Q: Why should I take part in this study? Do I have to do this?

A: Your participation is voluntary, and you may refuse to answer any questions or withdraw from the study at any time. However, your answers are very important to the success of this study and will represent thousands of others. Getting an answer from all the households chosen for the study is the best way to make sure the study results reflect the thoughts and opinions of all Americans.

Q: Will my answers to the survey be kept private?

A: Yes. Your answers will be kept private under the Privacy Act. Your answers cannot be connected to your name or any other information that could identify you or your household, to the extent provided by law. The completed questionnaires will be stored in a separate file with restricted access. Both the paper and electronic versions of the information will be destroyed shortly after the research is finalized.

Q: How long will it take to answer the questions?

A: About 20 to 30 minutes.

Q: Who is sponsoring the study? Is this study approved by the Federal Government?

A: The study is sponsored by the U.S. Department of Health and Human Services.

Q: Who is Westat?

A: Westat is a research company located in Rockville, Maryland. Westat is conducting this survey under contract to the U.S. Department of Health and Human Services.

Preguntas Frecuentes Encuesta Nacional de Tendencias de Información sobre la Salud

P: ¿De qué se trata el estudio? ¿Qué tipo de preguntas contiene?

R: El estudio trata sobre la salud y la manera en que las personas reciben información sobre la salud. Por ejemplo, le preguntaremos cómo obtiene normalmente información sobre cómo mantenerse saludable, el tipo de información en la que más confía y cómo le gustaría obtener dicha información en el futuro. También le preguntaremos sobre lo que cree que contribuye a la buena salud, cómo prevenir mejor el cáncer y su participación en varias actividades afines.

P: ¿Cómo se utilizarán los resultados del estudio? ¿Qué se hará con mi información?

R. Los hallazgos ayudarán al Departamento de Salud y Servicios Humanos de EE.UU. a fomentar la buena salud y prevenir las enfermedades mediante la determinación de formas de comunicar mejor la información sobre la salud a los estadounidenses.

P: ¿Cómo obtuvieron mi dirección?

R: Su dirección fue seleccionada al azar entre todas las direcciones conocidas en la nación usando métodos científicos de muestreo.

P: ¿Por qué debo participar en este estudio? ¿Es obligatorio hacerlo?

R: Su participación es voluntaria y usted puede rehusarse a contestar cualquiera de las preguntas o retirarse del estudio en cualquier momento. Sin embargo, sus respuestas son muy importantes para el éxito de este estudio y representan a miles de personas. El obtener respuesta de todos los hogares escogidos para este estudio es la mejor manera de asegurar que éste refleje los pensamientos y opiniones de todos los estadounidenses.

P: ¿Se mantendrá la privacidad de mis respuestas a la encuesta?

R. Sí. Se mantendrá la privacidad de sus respuestas en virtud de la Ley de Privacidad. Sus respuestas no pueden asociarse a su nombre ni a ninguna otra información que podría identificarlo a usted o a su hogar en la medida de lo permisible por ley. Los cuestionarios completos se almacenarán en un archivo separado con acceso restringido. Las versiones impresas y electrónicas de la información se destruirán poco después de la finalización de la encuesta.

P: ¿Cuánto tiempo tomará responder las preguntas?

R: Cerca de 20 a 30 minutos.

P: ¿Quién patrocina el estudio? ¿Está este estudio aprobado por el Gobierno Federal?

R: El estudio es patrocinado por el Departamento de Salud y Servicios Humanos de EE.UU.

P: ¿Quién es Westat?

R. Westat es una compañía de estudios de investigación ubicada en Rockville, Maryland. Westat realiza esta encuesta en virtud de un contrato con el Departamento de Salud y Servicios Humanos de EE.UU.

Appendix D

Variable Values and Data Editing Procedures

Missing Value Definitions

Values identifying types of nonresponse or indeterminate responses:

- -1 = Valid skips or appropriately missing data following a dependent question (correctly skipped). Example: If SeekHealthInfo=2 ‘no’ and WhereSeekHealthInfo was missing, WhereSeekHealthInfo was assigned the value -1.

- -2 = Question was answered, but respondent should not have answered the question. The question was answered in error by the respondent. Example: If SeekHealthInfo=2 ‘no’ and WhereSeekHealthInfo was not missing, WhereSeekHealthInfo was assigned the value -2.

- -4 = Question was answered, but data was removed because the entry of the number or character could not be determined (e.g. unreadable or non-conforming numeric response).

- -5 = Respondent selected more response options than appropriate for the question. Example: If CancerTrustRadio had values 3 ‘a little’ and 2 ‘some’, CancerTrustRadio was assigned the value -5. In cases where both -2 and -5 values could be assigned, the -2 value was assigned.

- -6 = Missing data in variables following a missing filter question. Example: If filter question (e.g., SeekHealthInfo) was missing and variables up until the next applicable question (e.g. CancerConfidentGetHealthInf) were missing (e.g., WhereSeekHealthInfo = missing and WhoLookingFor = missing), variables with missing values were assigned the value -6.

- -9 = Missing data. Not ascertained. Question should have been answered, but no response was recorded. Example: If CancerLotOfEffort was missing, it was assigned the value -9.

Data Editing Procedures

Variable	Editing Rule	Description of Rule
AdultsInHH	Recoding initial filter/skip question	The value of the following response, MailHHAdults, determined how missing responses to AdultsInHH were re-assigned. As an example, if AdultsInHH was missing and MailHHAdults initially had value 1 (adult in household) then AdultsInHH was assigned the value 2 'no' (indicating not more than 1 adult in the household) and MailHHAdults was assigned the 'missing value' -2 (answered inappropriately). If AdultsInHH was missing and MailHHAdults had value 2 (or greater) then AdultsInHH was assigned the value 1 'yes' (indicating more than 1 adult in the household) and the value for MailHHAdults was retained.
SeekHealthInfo SeekCancerInfo UseInternet NoticeCalorieInfoOnMenu Smoke100 UsedECigEver SeenFederalCourtTobaccoMessages HeardHPV FamBetween9and27 EverHadCancer BornInUSA	Recoding filter/skip questions	For these filter questions (questions containing a skip instruction associated with the particular response that was selected), response patterns following the question were examined if the filter question was not answered. The 'yes' value (in the majority of cases where a 'yes' response instructed a respondent to continue answering the subsequent questions) was substituted for the missing filter question when any of the subsequent questions were answered. Similarly (when a 'no' response instructed a respondent to skip subsequent questions), the 'no' value was substituted for the missing filter question when all of the subsequent questions that a 'no' response would have directed the respondent to skip were left unanswered and the respondent answered the next applicable question all respondents were supposed to answer. Please note that if neither condition was met, the missing response code values were retained.
WhereSeekHealthInfo_IMP StrongNeedCancerInfo_IMP PCStrongNeedInfo_IMP PCTrustInfo_IMP SexualOrientation_I	Imputation for multiple responses	Imputation was carried out when multiple responses were selected, resulting in one unique response for these "mark only one" variables. Respondent's multiple answers were replaced with a single imputed answer that had the same distribution over the multiple answers as occurred in the single-answer responses. Imputation was not performed on missing values for this question. The suffix "_IMP" indicates that this variable includes imputed values. Flags (indicated by suffix '_IFlag') indicate which values were imputed.

Variable	Editing Rule	Description of Rule
WhoLookingFor	Edits for multiple responses	Multiple responses (e.g., 'myself,' 'someone else') were recoded to the logically applicable third response option ('both myself and someone else').
Internet_DialUp Internet_BroadBnd Internet_Cell Internet_WiFi Electronic_SelfHealthInfo Electronic_HealthInfoSE Electronic_BuyMedicine Electronic_LookedAssistance Electronic_TalkDoctor Electronic_TrackedHealthCosts Electronic_TestResults Tablet_AchieveGoal Tablet_MakeDecision Tablet_DiscussionsHCP IntRsn_VisitedSocNet IntRsn_SharedSocNet IntRsn_WroteBlog IntRsn_SupportGroup IntRsn_YouTube ProbCare_BringTest ProbCare_WaitLong ProbCare_RedoTest ProbCare_ProvideHist HealthIns_InsuranceEmp HealthIns_InsurancePriv HealthIns_Medicare HealthIns_Medicaid HealthIns_Tricare HealthIns_VA HealthIns_IHS HealthIns_Other	Recoding missing responses for items with forced-choice response formats	Respondents were asked to select 'yes' or 'no' to a series of sub-items, allowing them to select as many responses as would apply. These 'forced-choice' response formats sometimes result in respondents indicating which sub-items apply to them by selecting the 'yes' response option for some and leaving the others unanswered. To allow the data to reflect this practice, if respondents did check one or more 'yes' response options within the group, but did not check a 'no' response option for any sub-item in the question, the sub-items that were missing a response were set to 'no.' However, if a respondent, in addition to leaving other sub-items unanswered, did select a 'no' response option for at least one sub-item, the unanswered sub-items were not assumed to be 'no' responses and instead remained missing.

Variable	Editing Rule	Description of Rule
NotAccessed_SpeakDirectly NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_ConcernedPrivacy NotAccessed_NoRecord NotAccessed_Other RecordsOnline_RefillMeds RecordsOnline_Paperwork RecordsOnline_RequestCorrection RecordsOnline_MessageHCP RecordsOnline_DownloadHealth RecordsOnline_AddHealthInfo RecordsOnline_MakeDecision ESent_AnotherHCP ESent_Family ESent_HealthApp MedConditions_Diabetes MedConditions_HighBP MedConditions_HeartCondition MedConditions_LungDisease MedConditions_Arthritis MedConditions_Depression CalorieInfo_FewerCalories CalorieInfo_MoreCalories CalorieInfo_FewerItems CalorieInfo_SmallerSizes CalorieInfo_MoreItems CalorieInfo_LargerSizes		
HealthInsurance_I GenderC_I EverHadCancer_I Age_I MaritalStatus_I Education_I Hisp_Cat_I Race_Cat2_I	Imputation for missing responses	Missing values were imputed for variables that were used in the process of assigning weights. The suffix "_I" indicates that this variable includes imputed values. Flags (indicated by suffix '_IFlag') indicate which values were imputed.
FreqGoProvider EverOfferedAccessRec AccessOnlineRecord	Recoding filter/skip questions	For these filter questions (questions containing a skip instruction associated with the particular response that was selected), response patterns following the

Variable	Editing Rule	Description of Rule
KnowledgePalliativeCare TimesModerateExercise ActiveDutyArmedForces		question were examined if the filter question was not answered. The value representing the skip response was substituted for the missing filter question if all of the subsequent questions that the response directed the respondent to skip were left unanswered, and the respondent answered the next applicable question. However, missing values were not substituted with other values if the filter question was not answered but a follow-up question was answered.
Height_Feet Height_Inches	Edits for implausible values	The rules that were applied minimized the number of out-of-range values by accounting for response measurements in incorrect boxes, responses using metric, responses using only one unit of measurement and other response errors. **Rules Applied to Edit Height Variables:** If HEIGHT_Feet was 0 or missing and HEIGHT_Inches>48 and HEIGHT_Inches<=60, then the first digit was taken as the feet value and the second digit was taken as the inches value (to correct for respondents expressing both feet and inches in the inches box). If HEIGHT_Feet was 0 or missing and HEIGHT_Inches>61 and HEIGHT_Inches<=83, then the inches value was converted to its feet-and-inches equivalent (to correct for respondents expressing height in inches, resulting in heights from 5'1" to 6'11"). If HEIGHT_Feet was 1 and HEIGHT_Inches>=3 and HEIGHT_Inches<=9 (or HEIGHT_Inches>=30 and HEIGHT_Inches<=90) then this metric value was converted to feet-and-inches (to correct for respondents using meters and tenths and hundredths of a meter to express height). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 20, 30, etc. thru 90 then the trailing 0 was removed. If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 15, 25, etc. thru 95 then the trailing 5 was removed (to correct for respondents expressing values in tenths of an inch). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 12, 23, 34, 45 etc. thru 89 then the first digit was taken (to correct for respondents

Variable	Editing Rule	Description of Rule
		giving an inch value as a range, e.g., 1-2 or 8-9 inches). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = a two digit value whereby the first digit equaled the feet value the second digit was taken as the inches value (to correct for respondents expressing the height in inches as well as in feet, e.g., 5’58” resulted in value 5’8”) If HEIGHT_Feet>6 and HEIGHT_Feet<12 and HEIGHT_Inches>3 and HEIGHT_Inches<7, then the values were switched (to correct for respondents putting measurements in the wrong boxes, resulting in edited values from 4’7” to <7 feet). If none of the preceding height editing rules were applicable: **Height_Feet (Height in Feet):** Any responses greater than 7 feet were recoded to “-4”, which is the code for non-conforming responses. **Height_Inches (Height in Inches):** Any responses greater than 11 inches were recoded to “-4”, which is the code for non-conforming responses.
HaveDevice_Cat WhoOffered_Cat CaregivingWho_Cat CaregivingCond_Cat CaregivingActivities_Cat CaregivingMedAct_Cat CaregiverTraining_Cat TobaccoMessages_Cat Cancer_Cat FamilyCancer_Cat Hisp_Cat Race_Cat2	Summarized distribution of ‘mark all that apply’ responses	A variable was created to indicate each response selection a respondent made for these ‘mark all that apply’ variables. The derived variable with the suffix ‘_cat’ summarized the response selected or indicated that multiple responses were selected.
Employed Unemployed Homemaker Student Retired Disabled	Derived variables for multiple responses	For the variable OccupationStatus, derived variables were created to indicate each response selected, showing the unique response for respondents selecting one occupation, and showing each response for respondents who did not follow the ‘mark only one’ response instruction.

Variable	Editing Rule	Description of Rule
OtherOcc		
Education IncomeRanges	Edits for multiple responses	The highest order (e.g., education level or income range) was taken when multiple responses were selected.
R_HHAdults	Derived variable	Responses to questions asking about household size as well as other information about the household (e.g., number of questionnaires returned) were compiled into a derived measure that best represented the number of adults in the household.
HHAdults_Num	Imputation for zero and missing responses	Missing values were imputed for the derived count of household adults when the derived variable had values of zero or missing. A flag (indicated by suffix '_IFlag') indicates which values were imputed.
QDisp	Derived variable	A variable was created to indicate the proportion of items respondents answered in the first two sections. This was used to determine incompletely-filled out questionnaires.
Caregiving_HoursPerWeek Weight DrinkDaysPerWeek AverageTimeSitting WhenDiagnosedCancer SelfMOB HHAdultMOB[2-5] SelfAge Age HHAdultAge[2-5] YearCameToUSA SexualOrientation_OS MailSurveyTime(_Min &_Hrs)	Recoding out of range responses	**CaregivingHoursPerWeek** Any responses greater than 168 hours were recoded to "-4", which is the code for non-conforming responses. **Weight:** Any responses less than 40 pounds or greater than 500 pounds were recoded to "-4", which is the code for non-conforming responses. **DrinkDaysPerWeek** Any responses greater than 7 days per week were recoded to "-4", which is the code for non-conforming responses. **AverageTimeSitting** Any responses greater than 24 hours were recoded to "-4", which is the code for non-conforming responses. **WhenDiagnosedCancer (Age at Time of Cancer Diagnosis):** Any responses greater than the age of the respondent were recoded to "-4", which is the code for non-conforming responses. **SelfMOB (Respondent's Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses.

Variable	Editing Rule	Description of Rule
		HHAdultMOB[2-5] (Second – Fifth Adult in Household Month of Birth): Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses. **Age Variables** Responses were examined for out of range or unlikely ages (those listing their age as < 18 and > 105). **YearCameToUSA** Responses not given in the standard 4-digit year format were reviewed for scanning accuracy and updated to -4. Additionally, the responses to YearCameToUSA were compared to the respondent's reported age to see whether respondents reported coming to the USA in a year that would predate their birth. **SexualOrientation_OS** Review of verbatim responses - Responses of "none of your business" and other similar phraseology were reviewed for scanning accuracy and rephrased as "Refused". **MailSurveyTime (_Mins & _Hrs)** Responses of 0 or blank were reviewed and recoded to -4 for cases where the respondent listed "0" in both fields, or listed "0" in one field and left the other field blank.
HaveDevice_CellPh HaveDevice_None Caregiving_No	Recoding filter/skip questions	For these "mark all that apply" filter questions ("mark all that apply" type questions where one or more response option contains a skip instruction at the "No" or "None" response), when the "No" or "None" response was selected, all responses within the question group were examined. If other responses were checked, the "No" or "None" response was recoded to "Not selected", and the other responses were retained.
FamilyCancer_None	Recoding filter/skip questions	For this "mark all that apply" question, when the "I have never had discussions…" response was selected, all responses within the question group were examined. If other responses were checked, the "I have never had discussions…" response was recoded to "Not selected", and the other responses were retained.

Variable	Editing Rule	Description of Rule
AverageCaloriesPerDay AverageCaloriesPerDay_DK	Recoding filter/skip questions	For this question, in order for TeleForm to capture the "Don't now" responses, a separate "Don't Know", selected/not selected variable was created. Normally these two variables would be combined and the "DK" variable would not be deliverable. In this cycle, 32 respondents wrote numeric responses in the boxes, as well as checked the "Don't Know" box. Because this number was significant, both variables were retained in the data.

Deriving and Imputing Measure of Household Adults

A program was developed based on the following guidelines in order to develop a single derived indicator for the number of household adults. The derived value is calculated for each household based on three sources of household size information that is solicited in the questionnaire. The guidelines were adapted from the analogous procedures used in cycle 1.

1. Create a composite variable (**RS_HHAdults**) from the raw and edited versions of **MailHHAdults**, resulting in a value of household adults for all households. This will be the raw (unedited) value of **MailHHAdults** for situations when respondents indicate that there are not more than one adult in the household (**AdultsInHH**=2) but enter a value for **MailHHAdults** that is greater than 1.

2. Create a second indicator for the number of adults in the household (**Demo_HHAdults**) based on responses to questions in the demographic section. **Demo_HHAdults** = **TotalHousehold** - **ChildrenInHH**. If **Demo_HHAdults** is negative, then reset the value of **Demo_HHAdults** to be missing.

   a. If **Demo_HHAdults** value is missing, 0, or 11 or greater, then replace value with a value from **RS_HHAdults** if **RS_HHAdults** is between 1 and 10 inclusive; name this new variable **DemoRS_HHAdults**.

   b. If **Demo_HHAdults** is 0 and **RS_HHAdults** is not between 1 and 10 inclusive, retain the value of **Demo_HHAdults** for variable **DemoRS_HHAdults**.

3. Edit/correct the variable **Demo_HHAdults** when its values are implausible by substituting in plausible values of variable **RS_HHAdults**. If **Demo_HHAdults** is between 1 and 10 inclusive or **RS_HHAdults** is not between 1 and 10 inclusive, retain the value of **Demo_HHAdults** for variable **DemoRS_HHAdults**.

4. Create a household size indicator based on the number of adults in the household as listed in the household enumeration roster. This is the sum of household members listed in the table whose ages are between 18 and 115 inclusive (**Roster_HHAdults)**.

5. Edit/correct the variable **DemoRS_HHAdults** using values of variable **Roster_HHAdults** and name the final measure of household size: **R_HHAdults**.

   a. R_HHAdults = DemoRS_HHAdults;

   b. If DemoRS_HHAdults = 0 then R_HHAdults = Roster_HHAdults.

   c. If DemoRS_HHAdults is missing and Roster_HHAdults is greater than 0, R_HHAdults = Roster_HHAdults.

   d. If Roster_HHAdults > DemoRS_HHAdults then R_HHAdults = Roster_HHAdults.

Imputation for the remaining values of zero or missing for R_HHAdults involved replacing these values with the average number of adults in responding households with non-zero or non-missing values of R_HHAdults, resulting in the variable HHAdults_Num. Nine households had missing values of R_HHAdults that needed to be imputed.