Health Information National Trends Survey 4 (HINTS 4)

Cycle 4 Methodology Report

February 2015

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

Prepared by
Westat
1600 Research Boulevard
Rockville, MD 20850

Westat®

Chapters			Page

Tables		

Cycle 4 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 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 4) includes four data collection cycles over the course of three years. This report summarizes the methodology, sampling, and procedures of the fourth of these cycles (Cycle 4). Data cleaning and weighting procedures as well as response rates are also discussed.

Cycle 4 was conducted from August 19 through November 17, 2014. Cycle 4 was conducted by mail using a protocol similar to that used in Cycles 1, 2, and 3 with a goal of obtaining 3,500 completed questionnaires.

1.1   Identifying Potentially Spanish-speaking Households

As in the other HINTS 4 cycles, households in Cycle 4 were flagged as potentially Spanish-speaking in one of three ways:

- **Linguistically isolated areas.** The US Census Bureau defines linguistically isolated households as those in which everyone over 14 years old speaks a language other than English and does not speak English very well. Sampled households from Census tracks with relatively high proportions of linguistically isolated Spanish households were flagged as potentially Spanish-speaking.

- **Hispanic surname match.** The surnames provided by the address vendor were compared to typical Hispanic surnames. Households identified as having a Hispanic surname were flagged as potentially Spanish-speaking.

- **Respondent request.** Respondents who called Westat to request Spanish materials were flagged as Spanish-speaking.

Westat®

1.2   Cover Design Experiment

In an effort to continue improvements to the appeal and usability of the HINTS instruments, there were two embedded methodological experiments included in Cycle 4. The first embedded experiment varied the cover of the survey instrument. The goal of the experiment was to understand whether and how cover design characteristics impact response rates. Two cover design characteristics were manipulated:

- **Overall page contrast.** What is the effect of having a predominately white background with dark text versus a predominately blue or green background with white text? There is evidence that a distinctive background may impact the likelihood that a respondent returns the questionnaire. It is hypothesized that a distinct background may make it more likely that respondents notice the questionnaire once it is out of the package (e.g., sitting on a desk or table for a few days) (Dillman, 2007; Dillman and Dillman, 1995; Nederhof, 1988).

- **Image placement.** What is the effect of having images grouped together in a collage versus spread about the cover page? It was hypothesized that if there were fewer large pictures on the cover, the respondent's attention might be more focused on reading important elements such as the government sponsorship, which adds credibility to the survey.

The cover used for previous cycles of HINTS 4 was maintained as the control condition, with four different conditions in the experimental groups. See **Appendix A** for the cover versions.

1.3   Grid Questions Experiment

The second embedded experiment was related to the grid questions on the instrument. A persistent issue in HINTS questionnaires is higher rates of missing data on grid questions compared to individual questions. Evidence from Cycles 1 and 2 shows that differences in the rates of missing data are driven largely by respondents with lower levels of education and weaker English language skills. It was hypothesized that the grids may be difficult for respondents to navigate, which increases the likelihood that a respondent will stop answering questions within the grid. One design feature that might improve navigation in the grid items is alternating shading of the rows in the table. The shading could help respondents navigate the grid more easily by allowing them to visualize which questions have been answered and where the next response is required. The shading of grid items on the Cycle 4 instrument was varied to test this hypothesis. See **Appendix B** for sample questions with shading in alternating rows in the grid questions.

1.4 Allocation of the Sample to the Experiments

The Cycle 4 study design incorporated both the cover design and grid shading conditions. The majority of the sample was allocated to the traditional HINTS 4 instrument design, which has a non-collage, low contrast cover, and does not have shaded rows in the grids. The cover design and grid experiment were fully crossed with each other for the remainder of the sample. Details of this sample allocation can be found in Chapter 2.

Sample Selection 2

The design for the HINTS Cycle 4 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

The sampling frame 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 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 N20).

Westat®

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 three explicit sampling strata:

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

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

3. Addresses located in counties comprising Central Appalachia regardless of minority population.

The high and low minority strata were formed using the block group level characteristics from the 2010 Decennial Census Summary File. Addresses in census block groups that had a population proportion of Hispanics or African Americans that equaled or exceeded 40 percent were assigned to the high-minority stratum. All the remaining addresses were assigned to the low-minority stratum. Addresses in counties comprising Central Appalachia were assigned to the Central Appalachia stratum regardless of minority status.

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. In Cycle 4, HINTS continued with stratifying Central Appalachia separately from the rest of the sample even though there were no separate precision requirements as there were in Cycles 1 and 2. Stratifying Central Appalachia separately allowed for a proportional sample from the stratum so as to maintain the gain in precision from the oversampling in Cycles 1 and 2.

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 4 was 14,000: 8,855 from the high minority stratum, 5,025 from the low minority stratum, and 120 from the Central Appalachia stratum. The high-minority stratum's proportion of the sampling frame was 25.0 percent, and it was oversampled so that its proportion of the sample was 63.3 percent. Conversely, the low minority stratum comprised 74.2 percent of the sampling frame, but made up just 35.6 percent of the sample. The Central Appalachia stratum was sampled proportionally with the stratum comprising about 0.8 percent of both the sampling frame and sample.

To carry out the data collection experiment described in Chapter 1, the address sample was divided into nine groups with about 70 percent of the sample addresses assigned to the control treatment group and the remaining 30 percent of the sample addresses assigned to eight experimental treatment groups. As shown in Table 2-1 below, households were allocated randomly so that 10,000 ended up in the control group and 500 ended up in each of the eight experimental treatment groups.

Table 2-1.   Allocation of the study sample to experimental conditions

	Low contrast cover	High contrast cover	Total
Current Design (control)	10,000	N/A	10,000
Revised Non-Collage Cover	1,000	1,000	2,000
No Shading in grids	500	500	1,000
Shading in grids	500	500	1,000
Collage Cover	1,000	1,000	2,000
No Shading in grids	500	500	1,000
Shading in grids	500	500	1,000
Total	12,000	2,000	14,000

The control group received the same cover and grid formatting design that were used in earlier cycles of HINTS. The eight treatment groups received one of four experimental covers; half were sent questionnaires with the shaded grid format and half saw the original grid format in their questionnaires. The two experimental factors (covers and grid formatting) were completely crossed within the experimental treatments which resulted in eight unique groups. All Spanish language questionnaires in the experimental treatment groups had shaded grids. This assignment rate was made uniform across all three strata.

Table 2-2 below summarizes the address sample for Cycle 4, showing the number of sample addresses by sampling stratum and treatment group.

Table 2-2.   Cycle 4 sample sizes by stratum and treatment group

Stratum	Total	Control treatment group	Experimental treatment groups
High minority areas	8,855	6,324	2,531
Low minority areas	5,025	3,590	1,435
Central Appalachia	120	86	34
Total	14,000	10,000	4,000

As part of the deduplication process, Westat determined that four out of the 14,000 households selected for the Cycle 4 sample were also selected for one of the three prior cycles. 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 13,996.

2.4   Within-Household Sample Selection

The second-stage of sampling consisted of selecting one adult within each sampled household. In keeping with Cycles 2 and 3, data collection for Cycle 4 implemented the Next Birthday Method to select the one adult in the household. Questions were included on the survey instrument to assist the household in selecting the adult in the household having the next birthday.

Data Collection 3

Data collection for Cycle 4 started on August 20, 2014 and concluded on November 17, 2014. 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 4. All households in the sample received the first mailing and reminder postcard, while only non-responding households received the subsequent survey mailings. Most households received one survey per mailing (in English), while households that were flagged as potentially Spanish-speaking received two surveys per mailing (one English and one Spanish).

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 final mailing contained the shortened version of the Spanish questionnaire. Just as in Cycles 2 and 3, mailings targeted potentially Spanish-speaking households by sending Spanish language materials in the first mailing. The contents of all mailings are further described in Table 3-1 below. Cover letters in English can be found in **Appendix C** and cover letters in Spanish are in **Appendix D**. All cover letters include a list of Frequently Asked Questions (FAQs) on the back. These FAQs in both English and Spanish are in **Appendix E**.

Table 3-1.   Mailing protocol

Mailing	Dates mailed	Mailing method	Materials	Type of recipients
Mailing 1	August 19-21, 2014	1st Class Mail	English cover letter with FAQs English questionnaire Return envelope $2 bill	All sampled households that <u>were not</u> identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope $2 bill	All sampled households that <u>were</u> identified as possibly Spanish-speaking
Postcard	August 26 & 28, 2014	1st Class Mail	Reminder/thank you postcard	All sampled households
Mailing 2	September 16-17, 2014	USPS Priority Mail	English cover letter with FAQs English questionnaire Return envelope	Non-responding households that <u>were not</u> identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope	Non-responding households that <u>were</u> identified as possibly Spanish-speaking

Table 3-1. Mailing protocol (continued)

Mailing	Dates mailed	Mailing method	Materials	Type of recipients
Mailing 3	October 14-15, 2014	1st Class Mail	English cover letter with FAQs English questionnaire Return envelope	Non-responding households that <u>were not</u> identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope	Non-responding households that <u>were</u> identified as possibly Spanish-speaking

The number of packets sent per mailing is outlined in Table 3-2 below. 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 only	English and Spanish	Spanish only*	Total
Mailing 1	11,717	2,279	N/A	13,996
Mailing 2	8,562	1,937	7	10,506
Mailing 3	7,337	1,742	11	9,090
Total	27,616	5,958	18	33,592

*Only includes households that contacted Westat to specifically request Spanish materials

3.2 Spanish Language Households

As in previous cycles, extra efforts were made to elicit responses from Spanish-speaking households. Mailings that included both English and Spanish materials were sent to households who:

- Were flagged as living in a linguistically-isolated area; or

- Were flagged as having a potentially Hispanic surname.

Spanish materials were sent to these households starting with the first mailing. The outer envelope for these households included a message stating 'Please return within two weeks' in both English and Spanish so these households would know without opening the package that the contents were bi-lingual. Households that received bi-lingual materials were sent those materials in the order established in previous HINTS cycles: English cover letter, Spanish cover letter, English

questionnaire, Spanish questionnaire. Any household that had received only English materials but called to request Spanish materials then received materials in only Spanish for all subsequent mailings.

3.3   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 comments, 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 33 calls throughout the Cycle 4 field period (see Table 3-3 below). A majority of the in-bound calls were to request Spanish materials. The rest were refusals or respondents calling in with some form of a comment or a question. Six calls were not resolved because the 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	13
Refusal	6
Asking for guidance on how to answer the health insurance question/asking why the question was being asked	3
Calling to let us know the respondent is elderly and required help to complete the survey	2
Requesting confirmation that we had received the completed survey	1
Informing us that the only household member was deceased	1
Asking about the possibility of doing the survey over the phone	1
Calls that were never resolved	6
Total	33

Westat®

3.4   Incoming Questionnaires

Field room staff receipted into the SMS all received questionnaires 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 our 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 final status of all households can be found in Table 3-4 below.

Table 3-4.   Final household status

Household Status	English	Spanish	TOTAL	
			N	%
Complete	3,521	156	3,677	26.3
Refusal			65	0.5
Undeliverable			1,791	12.8
Nonresponse			8,463	60.4
Total			13,996	100.0

The number of questionnaires returned by date during the field period can be found in Table 3-5 below.

Table 3-5.   Survey response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: August 19 - 21	August 20-26	205
Postcard: August 26 and 28	August 27-September 17	2,195
Mailing 2: September 16 - 17	September 18-October 16	923
Mailing 3: October 14 - 15	October 17-November 17	407
Total		3,730

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 electronically scanned to capture the survey data and images. 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. Response boxes were pre-edited to remove non-numeric responses and response options entered outside the capture area were corrected.

- 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 TeleForm scanner to capture the responses. TeleForm read the form image files and extracted data according to HINTS Cycle 4 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.

Decisions made about data issues were recorded in a data decision log. The decision log contained the respondent ID, the value triggering the edit, the updated value, and the reason for the update. A total of 37 entries were made into the data decision log during the course of data processing. The majority of these were attributed to illogical responses on a numeric question.

Westat®

A 10 percent quality control check was then conducted on the scanned data and the electronic images of the data forms. 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 F** (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. **Appendix F** provides a list of the values and their definitions.

- 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. **Appendix F** provides details about when and how these recodings took place.

- Variables were designed to summarize the response for the electronic device, medical information exchange, cancer, Hispanic ethnicity, and race questions. These variables, HaveDevice_Cat, MedInfo_Cat , Cancer_Cat, Hisp_Cat, and Race_Cat2 indicated each response selected for respondents selecting only one response, and a multiple category for all of the respondents who answered multiple responses.

- Derived variables were created to reflect each response recorded for certain "mark-one" type questions (A2 and N2), 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 (Imputation), was carried out. For other "mark-one" type questions where respondents marked multiple responses, rules, as described in **Appendix F** (Variable Values and Data Editing Procedures), 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, responses using only one unit of measurement and other response errors. A list of the rules applied to clean these variables is included in **Appendix F**.

- "Other, specify" responses were examined, cleaned for spelling errors, categorized, and upcoded into preexisting response codes when applicable.

4.3 Imputation

Imputation was used for two types of variables. The variable featuring a mark-only-one response instruction (WhereSeekHealthInfo) was imputed for the questionnaires in which multiple responses were recorded. The imputation process used was the same as that carried out for Cycles 1, 2, and 3. Responses where a missing value of -5 (multiple responses selected) were imputed. This occurred for 366 respondents for WhereSeekHealthInfo. A 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.

In addition, hot-deck imputation was used to replace missing responses with imputed data for items used in the raking procedure of the Cycle 4 weighting process. Hot-deck imputation is a data processing procedure in which a case with a missing value for a specific variable is assigned 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. 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 F**.

4.5   Survey Eligibility

Returned surveys were reviewed to ensure they were eligible for inclusion in the final dataset. Of the 3,742 questionnaires received, a total of 3,677 were determined to be eligible. Of the questionnaires determined to be ineligible, two were returned by respondents who reported an age below 18. Surveys were also reviewed for completion and duplication (more than one questionnaire returned from the same household), with 51 determined to be incomplete and 12 identified as duplicates. The processes for these reviews are detailed below.

Definition of a Complete and Partial Complete Questionnaire

Applying the completion rules traditionally used in HINTS, it was noted that Cycle 4 had an unusually high number of returned questionnaires flagged as incomplete. Cycle 4 had an incompletion rate of 2.4 percent compared to the rates of 1.5 percent in Cycle 2 and 1.2 percent in Cycle 3. There are two potential factors related to this incompletion rate increase:

- Cycle 4 had the fewest number of items in sections A and B compared to earlier cycles. There were 13 section A and B items in Cycle 4 compared to 16 items in Cycle 3 and 15 items in Cycle 2. The result of this reduction in items for sections A and B means that each item carries significantly more weight in determining completion in Cycle 4.

- Some Cycle 4 respondents skipped the entire first page of questions (n = 72). Although this behavior was also observed in previous cycles, it occurred more frequently in Cycle 4. The first page of the Cycle 4 instrument includes the entirety of Section A which is 7 of the 13 items used for determining completion. Using the traditional completion rules, all of these questionnaires were flagged as incomplete since more than half of their responses were missing. It should be noted that most of these respondents completed the rest of the questionnaire with a missing rate comparable to other completed questionnaires.

In the interest of retaining as much data as possible, the traditional criteria for flagging a questionnaire as incomplete was adjusted for Cycle 4 as shown in Table 4-1 below. Under the new criteria, questionnaires can be missing as much as 70 percent of the items in Sections A and B and still be considered 'partially complete' as long as the respondent did not skip 2 or more other Sections (C through N).

Table 4-1. Summary of rules used for determining questionnaire completeness in Cycles 1, 2, and 3 compared to Cycle 4

	Traditional HINTS criteria	Cycle 4 criteria
Definition of "Complete"	80% of questions in Sections A and B answered	80% of questions in Sections A and B answered
Definition of "Partial Complete"	50-79% of questions in Sections A and B answered	50-79% of questions in Sections A and B answered OR 30-49% of questions in Sections A and B answered and no more than one other section completely missing
Definition of "Incomplete"	49% or fewer questions in Sections A and B answered	29% or fewer of questions in Sections A and B answered OR 30–49% of questions in Sections A and B answered and more than 1 other section completely missing

Applying these revised completion criteria, 148 questionnaires were flagged as partially complete. Both partially-completed and complete questionnaires were retained. 51 questionnaires (1.4% of 3,677 eligible returns) were flagged as incomplete using the new rules and were excluded from the delivered dataset.

Eligibility of Multiple Questionnaires from a Household

Twelve households returned more than one completed questionnaire. The procedures to deal with this issue followed the same guidelines that were used for Cycles 1, 2, and 3:

- If the same respondent returned multiple questionnaires, the first questionnaire received was retained unless it was a reduced version and a subsequent return was a full version.

- 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 that were the same version, the questionnaire with fewer substantive questions omitted was retained.

Westat®

- 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.[1]

- 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   Codebook Development

Following cleaning and editing, 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 both versions of the 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 Cycle 4 received a full-sample weight and a set of 50 replicate weights. The full-sample weight is the weight which is used to calculate population and subpopulation estimates from the HINTS data collected in Cycle 4. Replicate weights are used to compute standard errors for these estimates.

The weighting process encompasses the procedures used to create the final full-sample and replicate weights for the survey respondents. 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 4 consisted of the following steps:

---

[1] 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.

- 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).

Each of these steps is described below.

The replicate weights were calculated using the 'delete one' jackknife (JK1) replication method. Details of replication used for variance estimation can be found in section 5.5.

5.1   Household Base Weights

The initial step in the weighting process was 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 oversampled strata are smaller than those in strata that were not oversampled. In Cycle 4, the base weights for households in the high minority stratum were roughly 1/5 and 1/2 the size of those in the low minority and Central Appalachia strata, respectively.

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 N20).

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 4 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[2] 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 4 were:

- Sampling stratum (High Minority; Low Minority; Central Appalachia)

- Treatment group (control group; experimental treatment groups)

- 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)

---

[2] 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/.)

- 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 2.16 to a high of 5.76, and averaged 3.32 across all nonresponse adjustment cells.

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. Since in Cycle 4 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.

Westat®

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. The ACS estimates are available via the Internet. Westat used the 2013 ACS estimates that were 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. These variables came from the 2014 National Health Information Survey (NHIS) and correspond to questions asked in the HINTS survey. They were:

- Percent With Health Insurance

- Percent Ever Had Cancer

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 Ever Had Cancer* calculated from HINTS data agrees with the estimate calculated from NHIS 2014 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 4 data set, a set of 50 replicate weights was assigned to each responding adult. To illustrate how the replicate weights 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.

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 calculation. These data have been weighted to account for the oversampling of addresses in high-minority areas. The overall response rate was 34.44 percent; however this differed significantly by strata. The High Minority strata had the lowest response rate (25.91 percent) and the Low Minority had the highest (37.24 percent). The percent of undeliverable households ranged from 11.5 percent to 14 percent across strata, with the High Minority and Central Appalachia strata having the highest undeliverable rates (13.54 and 13.33 percent respectively).

Table 6-1.   Response rate calculations by strata

Response class	High minority	Low minority	Central appalachia	Overall
Total sample*	34,994,266	103,796,143	1,165,355	139,955,764
Respondents	7,840,613	34,226,907	349,607	42,417,126
Nonrespondents	22,407,395	57,671,409	660,368	80,739,173
Undeliverable	4,738,354	11,897,827	155,381	16,791,561
Total Households	30,255,912	91,898,316	1,009,974	123,164,203
Percent Undeliverable	13.54%	11.46%	13.33%	12.00%
Household response rate	25.91%	37.24%	34.62%	34.44%

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

Westat®

7.1   Cover Design Experiment

In Cycle 4, two cover design factors were manipulated: picture placement (collaged or not) and contrast (white or dark background). Four types of new covers were compared in addition to the original HINTS cover to see whether the cover design might impact the likelihood that households responded to the survey. Examples of the five cover designs can be found in **Appendix A**. Table 7-1 compares the response rate outcomes by strata for the control cover (N = 10,000) (which was the cover used in other HINTS cycles) and the experimental covers (N = 1000 per cover).

Table 7-1.   Response rate for **five cover conditions** by strata sorted by overall response rate

		High minority	Low minority	Central Appalachia	Overall
Non-Collage, Low Contrast	Total Households	2,082,663	6,568,592	67,979	8,719,234
	Household response rate	27.70%	38.68%	14.29%	35.87%
Collage, High Contrast	Total Households	2,149,845	6,899,087	58,268	9,107,199
	Household response rate	24.08%	38.62%	66.67%	35.37%
Non-Collage, High Contrast	Total Households	2,161,701	6,589,248	87,402	8,838,350
	Household response rate	23.95%	38.24%	33.33%	34.70%
Control	Total Households	21,664,435	65,376,078	728,347	87,768,859
	Household response rate	26.14%	37.03%	33.33%	34.31%
Collage, Low Contrast	Total Households	2,197,268	6,465,312	67,979	8,730,559
	Household response rate	25.72%	35.46%	42.86%	33.07%

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

Although the experimental covers collectively yielded a marginally higher response rate (34.8%) than the control cover (34.3%), the difference was not statistically significant (p < .1). The 'Non-collage, white background' cover yielded the highest response rate of all the experimental covers (35.9%). This was the cover most similar in design to the control cover. The 'Collage, dark background'

yielded the lowest response rate (33.0%) and was the only experimental cover to achieve a response rate lower than the control cover. This cover was the most dissimilar in design to the control cover.

This pattern of response by cover was generally uniform across the High Minority and Low Minority strata. Response rates appear to be much more variable in the Central Appalachia stratum in the experimental treatment groups; however the response rates for this stratum are calculated based on a small number of households relative to the other two strata and therefore the difference is less reliable.

Overall there were no statistically significant differences in response by type of cover (p < .1). The results suggest that response rates were not significantly affected by the cover design aspects that were tested.

7.2   Grid Item Shading Experiment

In Cycle 4, approximately 15 percent of sampled households were sent experimental questionnaires with grid items formatted so that each alternating row of the grid was shaded grey or white. The remaining 85 percent of the sample received the standard HINTS grid item format with each row shaded the same color (white). Examples of the shaded grid format can be found in **Appendix B**. The goal of the experiment was to examine whether the experimental grid design could help reduce item missing data rates within grid items.

Cycle 4 of HINTS included 87 grid items spread across 19 distinct grids. 576 respondents returned questionnaires with shaded grids and 3,101 respondents returned questionnaires with unshaded grids. Table 7-2 shows the unweighted average item-missing data rate across the 87 grid items by questionnaire format, language, and respondent education and English-language proficiency.

The grid shading appears to have a larger impact on reducing item-missing data rates in the Spanish language questionnaires than in the English language questionnaires. Within the Spanish language questionnaires, the average missing data rate for those who saw shaded grid items was about 3 points lower than for those who saw unshaded grids. Spanish language respondents with lower education and lower English language proficiency appear to have benefited most from the presence of shaded grids as these groups exhibited the largest difference in item-missing data rates between the shaded and unshaded grid formats. Significance tests found that the differences demonstrated in

Westat®

the table are not statistically significant. It is important to note that any potential positive effects of shading are probably limited to the Spanish form where there are relatively few cases for analysis.

Table 7-2.   Average grid item missing data rate

ENGLISH

Mean proportion of missing data across 87 grid items		
	Shaded	Not shaded
Overall		
% Missing	4.4%	4.2%
n=	527	2994
Education		
High school or less	6.8%	6.3%
	123	756
More than HS	2.5%	2.8%
	376	2122
English Proficiency		
Speaks very well	3.3%	3.3%
	442	2520
Less than very well	4.8%	7.4%
	44	266

SPANISH

Mean proportion of missing data across 87 grid items		
	Shaded	Not shaded
Overall		
% Missing	9.8%	12.8%
n=	49	107
Education		
High school or less	7.8%	14.9%
	31	68
More than HS	9.6%	9.6%
	16	34
English Proficiency		
Speaks very well	3.0%	2.2%
	4	5
Less than very well	8.9%	13.7%
	43	99

Although the experiment lacked sufficient power to detect significant differences, the results imply that it might be mildly beneficial to incorporate grid shading in future rounds of HINTS data collection and that doing so does not have any known disadvantages.
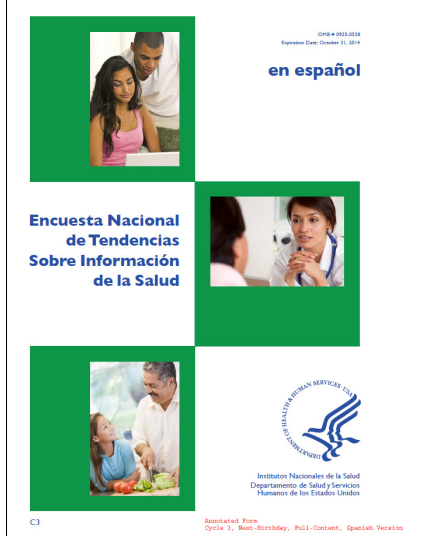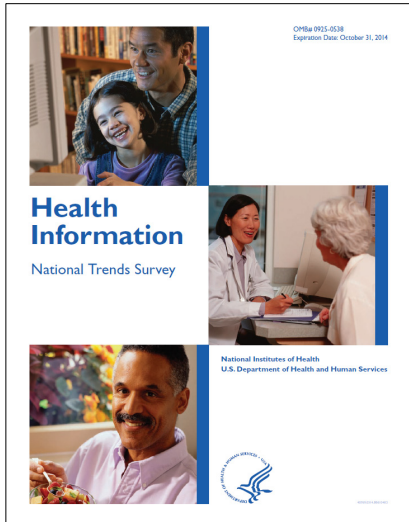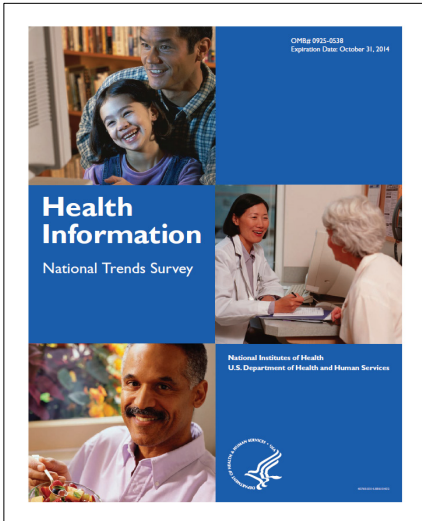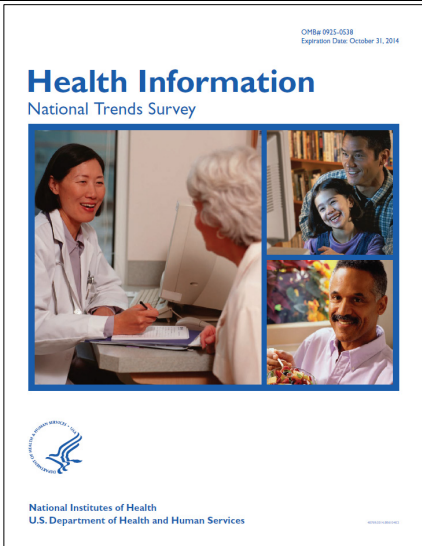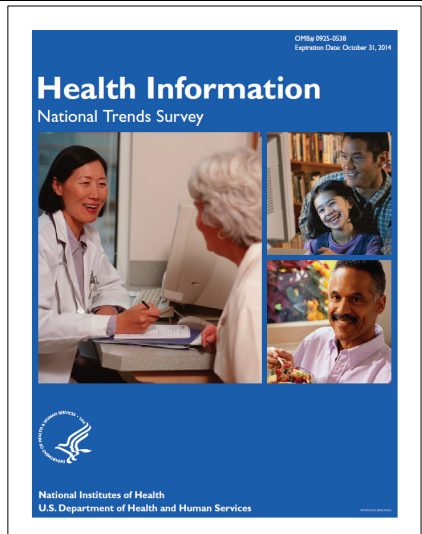
References
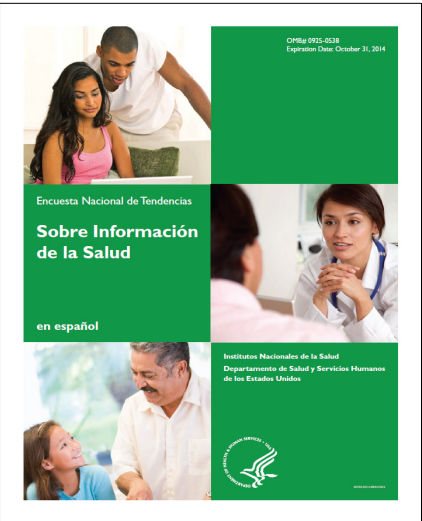
Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Dillman, D. *Mail and Internet Surveys: The Tailored Design Method.* 2nd ed. John Wiley & Sons, Inc. New Jersey, 2007. Pp. 135-139.

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.

Dillman, J., Dillman, D. (1995). *The influence of questionnaire cover design on response to mail surveys.* Paper presented at the International Conference on Measurement and Process Quality, Bristol, England.

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

Nederhof, A. (1988). Effects of a final telephone reminder and questionnaire cover design in mail surveys. *Social Science Research*, 17. 353-361.

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.

Westat

This page left blank intentionally.

Appendix A

Examples of Control and Experimental Covers Used in Cycle 4

CONDITON	English cover name and design	Spanish cover name and design
Control		
Non-collage, Low contrast		

CONDITON	English cover name and design	Spanish cover name and design
Non-collage, High contrast		
Collage, Low contrast		
Collage, High contrast		

Appendix B

Question Shading Example

A5. Based on the results of your most recent search for information about cancer, how much do you agree or disagree with <u>each</u> of the following statements?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. It took a lot of effort to get the information you needed	☐	☐	☐	☐
b. You felt frustrated during your search for the information	☐	☐	☐	☐
c. You were concerned about the quality of the information	☐	☐	☐	☐
d. The information you found was hard to understand	☐	☐	☐	☐

A7. In general, how much would you trust information about cancer from <u>each</u> of the following?

	Not at all	A little	Some	A lot
a. A doctor	☐	☐	☐	☐
b. Family or friends	☐	☐	☐	☐
c. Newspapers or magazines	☐	☐	☐	☐
d. Radio	☐	☐	☐	☐
e. Internet	☐	☐	☐	☐
f. Television	☐	☐	☐	☐
g. Government health agencies	☐	☐	☐	☐
h. Charitable organizations	☐	☐	☐	☐
i. Religious organizations and leaders	☐	☐	☐	☐

Westat

Appendix C

Cover Letters in English

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 <u>the adult in your household with the next birthday</u> 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, 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, Ph.D.
HINTS Project Officer
National Institutes of Health
U.S. Dept. of Health and Human Services

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 <u>the adult in your household with the next birthday</u> 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, 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.

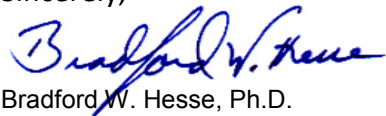The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

This page left blank intentionally.

Appendix D

Cover Letters in Spanish

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 <u>el adulto en su hogar con el próximo cumpleaños</u>, 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, 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.

SECOND MAILING

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 <u>el adulto en su hogar con el próximo cumpleaños</u>, complete y devuelva este cuestionario en las próximas dos semanas.

Usted podrá encontrar más información sobre HINTS en el sitio web <u>hints.cancer.gov</u>. 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, 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.

THIRD MAILING

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, la Encuesta Nacional de Tendencias de Información sobre la Salud (HINTS, por sus siglas en inglés). El completar esta encuesta nos ayudará a mantenerlos mejor informados en asuntos de salud que usted y su familia consideran importantes.

Si usted ya 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.

Si aún no ha tenido la oportunidad de completar la encuesta, comprendemos que a veces es difícil encontrar el tiempo para participar en un estudio como HINTS. Para reducir la cantidad de tiempo que va a tomarle, hemos incluido una versión más corta de la encuesta. Esta versión corta se concentra en asuntos que nos informan la manera como personas como usted buscan y utilizan información sobre la salud.

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 con este importante estudio nacional.

Atentamente,

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

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

This page left blank intentionally.

Appendix E

**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. The study has been approved by the Office of Management and Budget (OMB), the office that reviews all federally-sponsored surveys. The OMB approval number assigned to this study is 0925-0538.

Westat®

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. El estudio ha sido aprobado por la Oficina de Administración y Presupuesto (OMB, por sus siglas en inglés). Esta oficina revisa todas las encuestas patrocinadas federalmente. El número de aprobación asignado por la OMB para este estudio es 0925-0538.

Westat®

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 F

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., LookElsewhere= 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.

Westat®

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 Smoke100 HeardHPV FamBetween9and27 EverHadCancer UndergoCancerTreatment 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	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. Flags (indicated by suffix '_IFlag') indicate which values were imputed.
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	Recoding missing responses for items with forced-choice response	Respondents were asked to select 'yes' or 'no' to a series of sub-items, allowing them to select as many responses as would apply.

Westat®

Variable	Editing Rule	Description of Rule
Internet_WiFi HealthApps_AchieveGoal HealthApps_MakeDecision HealthApps_NewQuestions OfferedAccessHCP OfferedAccessInsurer MedConditions_Diabetes MedConditions_HighBP MedConditions_HeartCondition MedConditions_LungDisease MedConditions_Arthritis MedConditions_Depression CancerTx_Chemo CancerTx_Radiation CanncerTx_Surgery CancerTx_Other	formats	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.
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 TimesModerateExercise SmokeNow HowLongFInishTreatment_Cat ActiveDutyArmedServices	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 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.

Westat®

Variable	Editing Rule	Description of Rule
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 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).

Variable	Editing Rule	Description of Rule
HaveDevice_Cat MedInfo_Cat Cancer_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 OtherOcc	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.
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.
AverageDailyTVGames Height_Feet Height_Inches Weight WhenDiagnosedCancer SelfMOB HHAdultMOB2 HHAdultMOB3 HHAdultMOB4 HHAdultMOB5	Recoding out of range responses	**AverageDailyTVGames (TV Games):** Any responses greater than **24 hours** were recoded to "-4", which is the code for non-conforming responses. **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.

Variable	Editing Rule	Description of Rule
		Weight: Any responses less than 40 pounds or greater than 500 pounds 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. **HHAdultMOB2 (Second 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. **HHAdultMOB3 (Third 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. **HHAdultMOB4 (Fourth 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. **HHAdultMOB5 (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.

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.