

Health
Information
National
Trends
Survey

hints

Health Information National Trends Survey 5 (HINTS 5)

Cycle 3 Methodology Report

August 2019

Revised: February 2021

Prepared for

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

Prepared by

Westat
1600 Research Boulevard
Rockville, MD 20850



This page is deliberately blank.

Table of Contents

<u>Chapters</u>		<u>Page</u>
1	Cycle 3 Overview.....	1
2	Sample Selection.....	2
	2.1 Sampling Frame.....	3
	2.2 Stratification.....	4
	2.3 Selection of Address Sample.....	4
	2.4 Within-Household Sample Selection.....	5
3	Data Collection.....	6
	3.1 Mailing Protocol.....	6
	3.2 In-bound Telephone Calls.....	10
	3.3 Emails from Web Pilot Respondents.....	11
	3.4 Incoming Questionnaires.....	12
4	Data Management.....	14
	4.1 Scanning.....	14
	4.2 Data Cleaning and Editing.....	15
	4.3 Imputation.....	17
	4.4 Determination of the Number of Household Adults.....	17
	4.5 Survey Eligibility.....	18
	4.6 Additional Analytic Variables.....	19
	4.7 Codebook Development.....	20
5	Weighting and Variance Estimation.....	20
	5.1 Household Base Weights.....	21
	5.2 Household Nonresponse Adjustment.....	22
	5.3 Initial Person-Level Weights.....	23
	5.4 Calibration Adjustments.....	24
	5.5 Replicate Variance Estimation.....	25
	5.6 Taylor Series Variance Estimation.....	27
6	Response Rates.....	27
	References.....	29

<u>Tables</u>		<u>Page</u>
2-1	Summary by sampling stratum.....	5
2-2	Addresses in sample by treatment group and stratum.....	5
3-1	Mailing protocol for Cycle 3.....	8
3-2	Mailing Protocol for the Web Pilot.....	9
3-3	Number of packets per Cycle 3 mailing.....	10
3-4	Number of packets per Web Pilot mailing.....	10
3-5	Telephone calls received.....	11
3-6	Household status of Cycle 3 at close of data collection.....	12
3-7	Household status of the Web Pilot at close of data collection.....	13
3-8	Cycle 3 survey response by date.....	13
3-9	Web Pilot survey response by date.....	13
5-1	Statistics for the household nonresponse adjustment factors, by treatment group.....	23
6-1	Comparison of response rates by strata for Cycle 3 and the Web Pilot.....	28

Appendices

A	Cover Letters in English (Cycle 3).....	A-1
B	Cover Letters in Spanish (Cycle 3).....	B-1
C	Cover Letters (Web Option).....	C-1
D	Cover Letters (Web Bonus).....	D-1
E	Frequently Asked Questions in English and Spanish (Cycle 3).....	E-1

<u>Appendices</u>		<u>Page</u>
F	Frequently Asked Questions (Web Option and Web Bonus)	F-1
G	Additional Insert (Web Bonus)	G-1
H	Variable Values and Data Editing Procedures	H-1
I	Web Data Errors and Corrections: Variables and Records Revised	I-1
J	Web Data Errors and Corrections: Differences in Sample Sizes and Weighted Means for 11 Open-ended Variables Before and After Corrections	J-1

This page is deliberately blank.

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 civilian, non-institutionalized adults aged 18 or older living in 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 third round of data collection for HINTS 5 (Cycle 3) was conducted from January 22 to April 30, 2019 with a goal of obtaining 3,500 completed questionnaires. A push-to-web pilot test was conducted alongside the fielding of Cycle 3 (referred to as the Web Pilot). The Web Pilot was conducted from January 29 to May 7, 2019 with a goal of obtaining 2,046 additional completed questionnaires. This report summarizes the methodology, sampling, data collection protocols, weighting procedures, and response rates for both Cycle 3 and the Web Pilot.

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 3, the topics of special interest focused specifically on health behaviors. The types of behaviors more deeply explored in Cycle 3 include:

- Sun safety and sun protection activities;
- Dietary intake and attention to calorie labels on restaurant menus;
- Alcohol consumption and the negative effects of alcohol;
- Physical activity, sedentary activity, and awareness of physical activity guidelines;
- Sleep patterns; and
- Tobacco use, e-cigarette use, and attitudes about e-cigarettes.

Web Pilot

When HINTS moved from a telephone to a paper-mail survey in 2008, access to and use of the internet was not widespread enough to make the web a feasible survey mode. However, the advantages of offering a web option are evolving as more of the general population gets access to the internet and use it on a routine basis. The Web Pilot was designed to assess whether it is now possible to push enough people to the web to realize the advantages of web-based data collection on data quality while maintaining, or even improving, response rates.

The Cycle 3 data collection was used as a control group for the two experimental conditions in the Web Pilot:

- Web Option: offering respondents a choice between responding via paper or web; and
- Web Bonus: offering respondents a choice between responding via paper or web with an additional \$10 incentive for those responding via web.

The sampling, instrument, and mailing protocol used for the Web Pilot were almost identical to Cycle 3. In addition, weights were created combining all of the samples together. If appropriate, data from the Web Pilot can be combined with the data from Cycle 3 using this weight to enhance analysis.

Although this report describes the sampling, data collection protocol, weighting procedures and the response rates for the Web Pilot groups, more details about the Web Pilot can be found in a separate report titled *HINTS 5 Web Pilot Analysis Report*, which includes the results of more thorough analyses of response rates, data quality, and costs.

Sample Selection

2

The sampling strategy for both Cycle 3 and the Web Pilot 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 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. In previous cycles, the sampling frame used for the address sample contained duplicate units because some households receive mail in more than one way, and a question about how many different ways respondents receive mail was included on the survey instrument to permit an adjustment for the duplication of households in the sampling frame. However, because this is a rare occurrence, in Cycle 3 this question was removed from the questionnaire and, consequently, no adjustment for duplication was made.

A change to the traditional HINTS sampling design was implemented for Cycle 3 and the Web Pilot. Certain types of PO Box addresses were excluded from the address frame in an attempt to improve the rate of mail that is able to be delivered. There are two types of PO Box addresses: one type pertains to those that are linked to city-style addresses, and the other type does not. Those that are linked can get mail two ways: by the PO Box address and the city-style address. Those that are not linked can only get mail by the PO Box address. The Cycle 3/Web Pilot sample was limited to PO Box addresses classified as the only-way-to-get mail. Because PO Box addresses tend to have high undeliverable rates, having a smaller number in the sample should result in a lower rate of undeliverable packets compared to past cycles.

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 2013–2017 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 3 and the Web Pilot was 23,430: 16,740 from the high minority stratum and 6,690 from the low minority stratum. The high-minority stratum's proportion of the sampling frame was 26.5 percent, and it was oversampled so that its proportion of the sample was 71.4 percent. Conversely, the low minority stratum comprised 73.5 percent of the sampling frame, but made up just 28.6 percent of the sample.

Table 2-1 below summarizes the address sample, 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 checked whether there were households that were selected from either Cycle 1 or Cycle 2 and determined that there were no such records.

Table 2-1. 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	16,740	26.5	71.4	+169.4%
Low minority areas	6,690	73.5	28.6	-61.1%
Total Sampled	23,430			
Deduplication	0			
Sample for Mailing	23,430			

To accommodate the Web Pilot, the address sample was divided up into three representative subsamples:

- A subsample for the Paper-only treatment group also acted as the control group, where the traditional data collection procedures were used (referred to as Cycle 3);
- A subsample for the Web Option treatment group of the Web Pilot, where respondents were offered the choice to participate by web or by paper; and
- A subsample for the Web Bonus treatment group of the Web Pilot, where respondents were offered a \$10 e-gift card as an incentive to participate by web.

The sample sizes for the three treatment groups are shown in Table 2-2. The sample size for the control group, which was the approximately the same as Cycle 1 and Cycle 2, was 14,730 households (with 10,520 households from the high minority stratum and 4,210 households from the low minority stratum). The sample sizes for each of the two web-based treatment groups were 4,350 households (with 3,110 from the high minority stratum and 1,240 from the low minority stratum).

Table 2-2. Addresses in sample by treatment group and stratum

Stratum	Number of Addresses in Sample			
	Cycle 3	Web Option	Web Bonus	Total
High minority	10,520	3,110	3,110	16,740
Low minority	4,210	1,240	1,240	6,690
Total	14,730	4,350	4,350	23,430

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 3 and the Web Pilot 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 3 started on January 22, 2019 and concluded on April 30, 2019. Data collection for the Web Pilot began one week later on January 29, 2019 and concluded on May 7, 2019. Cycle 3 was conducted exclusively by mail. The Web Pilot was also conducted by mail, however, respondents were offered the choice to respond via paper (in English or Spanish) or via a web survey (in English only). All groups received a \$2 pre-paid monetary incentive to encourage participation. The specific mailing procedures and outcomes for each data collection effort are described in detail below.

3.1 Mailing Protocol

The mailing protocol for Cycle 3 and both experimental conditions of the Web Pilot followed a modified Dillman approach (Dillman, et al., 2009) and each group received a total of four mailings: an initial mailing, a reminder postcard, and two follow-up mailings.

All households in each sample received the first mailing and reminder postcard, while only non-responding households received the subsequent survey mailings. All households in each sample 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. Web Pilot respondents who requested the survey in Spanish were sent the Spanish paper instrument since the web survey was not offered in Spanish.

Each group's second survey mailing was sent via USPS Priority Mail, while all other mailings were sent First Class. The contact materials for Cycle 3 (Paper-only), Web Option, and Web Bonus groups were slightly different. The language in the cover letters varied based on whether respondents were being invited to complete the survey by web and, if so, whether they were being offered a bonus incentive to do so. The contact materials for the two Web Pilot groups included a

link to the web survey and the respondent's unique PIN or access code. Contact materials for the Web Bonus group also included information about the \$10 web bonus (a \$10 Amazon e-gift card code, which would be provided to the respondent upon submission of the web survey). Reminder postcards for both Web Pilot groups were folded and sealed so that the respondent's PIN could be included in the reminder (Cycle 3 respondents received the traditional reminder postcard). All groups received the \$2 pre-incentive.

The contents of the mailings for each group are further described in Table 3-1 and Table 3-2, below. The English cover letters and reminder postcard for the Cycle 3 sample can be found in **Appendix A** and the Spanish cover letters are in **Appendix B**. The cover letters and the reminder postcard for the Web Option group of the Web Pilot can be found in **Appendix C** and the cover letters and the reminder postcard for the Web Bonus group are in **Appendix D**. Each cover letter included a list of Frequently Asked Questions (FAQs) on the back. The FAQs for Cycle 3 respondents (in English and Spanish) are in **Appendix E** and the FAQs for all Web Pilot respondents (in English only) are in **Appendix F**¹. An additional insert was included in the mailings to the Web Bonus respondents to draw their attention to the \$10 web bonus. This insert can be found in **Appendix G**.

¹ One set of FAQs were developed for both the Web Option and Web Bonus groups of the Web Pilot.

Table 3-1. Mailing protocol for Cycle 3

Mailing	Date(s) mailed	Mailing method	Cycle 3 Materials
Mailing 1	January 22, 2019	1 st Class Mail	<ul style="list-style-type: none"> • English cover letter with FAQs • English Questionnaire • Postage-paid return envelope • \$2 bill
Postcard	January 29, 2019	1 st Class Mail	Reminder/thank you postcard
Mailing 2	February 19, 2019	USPS Priority Mail	<ul style="list-style-type: none"> • English cover letter with FAQs • English questionnaire • Postage-paid return envelope <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> • Spanish cover letter with FAQs • Spanish questionnaire • Postage-paid return envelope
Mailing 3	March 19, 2019	1 st Class Mail	<ul style="list-style-type: none"> • English cover letter with FAQs • English questionnaire • Postage-paid return envelope <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> • Spanish cover letter with FAQs • Spanish questionnaire • Postage-paid return envelope

Table 3-2. Mailing Protocol for the Web Pilot

Mailing	Date(s) mailed	Mailing method	Web Option Materials	Web Bonus Materials
Mailing 1	January 29, 2019	1 st Class Mail	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope \$2 bill 	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope \$2 bill Additional insert
Postcard	February 5, 2019	1 st Class Mail	Reminder/thank you postcard with link to web survey and unique PIN (folded and sealed)	Reminder/thank you postcard with link to web survey and unique PIN (folded and sealed)
Mailing 2	February 26, 2019	USPS Priority Mail	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope 	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope Additional insert <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope
Mailing 3	March 26, 2019	1 st Class Mail	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope 	<ul style="list-style-type: none"> English cover letter with link to web survey, unique PIN, and FAQs English Questionnaire Postage-paid return envelope Additional insert <p style="text-align: center;">OR (upon request)</p> <ul style="list-style-type: none"> Spanish cover letter with FAQs Spanish questionnaire Postage-paid return envelope

The number of packets sent per mailing is outlined in Tables 3-3 and 3-4, 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 3-3. Number of packets per Cycle 3 mailing

Mailing	English	Spanish	Total
Mailing 1	14,730	N/A	14,730
Mailing 2	12,087	17	12,104
Mailing 3	10,717	38	10,755
Total	37,534	55	37,589

Table 3-4. Number of packets per Web Pilot mailing

Mailing	English	Spanish	Total
Mailing 1	8,700	N/A	8,700
Mailing 2	7,076	0	7,076
Mailing 3	6,271	1	6,272
Total	22,047	1	22,048

3.2 In-bound Telephone Calls

Two toll-free telephone numbers were provided to all 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 80 calls throughout the Cycle 3 and Web Pilot field periods (see Table 3-5 below). The majority of the in-bound calls were respondents requesting Spanish materials. One respondent from the Web Pilot called to let the study team know that they could not access their web survey and staff helped troubleshoot the issue. The rest were respondents calling in with some form of comment or question or refusals. Only one call could not be resolved because it

was either a hang-up or a non-informative message and study staff were not able to reach the respondent.

Table 3-5. Telephone calls received

Reason for call	Number of calls received
Request for a Spanish questionnaire	64
Refusal	6
Respondent let the study team know that the survey had been completed	3
Respondent could not access their web survey	1
Respondent asked a question. Topics included: <ul style="list-style-type: none"> • whether HINTS was offering health insurance • more information about who could fill out the survey • whether the survey could be completed over the phone • if the survey could be submitted after the two week period specified in the cover letters • the deadline to submit the survey 	5
Calls that were never resolved due to hang ups or non-informative messages	1
Total	80

3.3 Emails from Web Pilot Respondents

Web Pilot respondents who accessed the link to the web survey were also given the option to email study staff if they had questions or concerns. Respondents were able to email study staff using the email address provided on the survey website or they could fill out a form on the website with their name, email address, and reason for contact. Both emails and messages sent via the online form were received in the study’s email inbox and staff responded to those messages via email. Emails received were automatically logged in the SMS and requests were processed by study staff (e.g., confirming that the respondent’s access code is valid and working). The study team received three emails throughout the Web Pilot field period:

- One respondent claimed they were not able to use their Amazon e-gift card code. Study staff contacted Amazon Support and confirmed that the e-gift card code had not been used and followed up with the respondent accordingly.

- One respondent claimed their unique PIN or access code was not valid. Study staff members investigated the issue and determined that the respondent had mixed up a digit in the code for a letter. Staff followed up with the respondent and provided the correct access code.
- One respondent did not know how to redeem their e-gift card. Study staff explained how to redeem the e-gift card code on Amazon and the respondent confirmed that they were able to redeem the e-gift card.

3.4 Incoming Questionnaires

Field room staff receipted all returned questionnaires into the SMS using each questionnaire’s unique barcode. Questionnaires submitted using the web survey were logged as complete by the SMS as soon as they were submitted. The SMS tracked each received questionnaire as well as the status of each household. 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 Cycle 3 and Web Pilot households at the end of data collection (but before cleaning and editing) can be found in Tables 3-6 and 3-7.

Table 3-6. Household status of Cycle 3 at close of data collection

Household status	Total in Cycle 3	
	N	%
Complete	3,439	23.3
Refusal	43	0.3
Undeliverable	1,273	8.6
Nonresponse	9,975	67.7
Total	14,730	100.0

Table 3-7. Household status of the Web Pilot at close of data collection

Household status	Web Option		Web Bonus		Total in Web Pilot	
	N	% in Web Option Group	N	% in Web Bonus Group	N	% in Web Pilot
Complete by paper	751	17.3	471	10.8	1,222	14.0
Complete by web	223	5.1	590	13.6	813	9.3
Refusal	15	0.3	10	0.2	25	0.3
Undeliverable	410	9.4	407	9.4	817	9.4
Nonresponse	2,951	67.8	2,872	66.0	5,823	66.9
Total	4,350	100.0	4,350	100.0	8,700	100.0

The number of questionnaires returned by date during the field periods for Cycle 3 and the Web Pilot can be found in Tables 3-8 and 3-9, below. The majority of Cycle 3 returns were early in the field period, with 62 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 25 percent and the remaining 13 percent were in response to the final mailing.

Table 3-8. Cycle 3 survey response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: January 22	January 23- January 31	331
Postcard: January 29	February 1- February 21	1,797
Mailing 2: February 19	February 22- March 21	866
Mailing 3: March 19	March 22- April 30	448
	Total	3,442

The majority of Web Pilot returns were also early in the field period, with 63 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 22 percent and the remaining 15 percent were in response to the final mailing.

Table 3-9. Web Pilot survey response by date

Date of mailing	Period of returns	Number of returns by web	Number of returns by paper	Total number of returns
Mailing 1: January 29	January 30- February 7	280	119	399
Postcard: February 5	February 8- February 28	279	610	889
Mailing 2: February 26	March 1- March 28	148	301	449
Mailing 3: March 26	March 29- May 7	106	193	299
	Total	813	1,223	2,036

After being processed and receipted into the SMS, each returned paper questionnaire was scanned, and both paper questionnaires and questionnaires submitted using the web survey were verified, cleaned, and edited. Imputation procedures were also conducted. These procedures are described below.

4.1 Scanning

All completed paper 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.
- Reviewing potential problem questionnaires or pertinent comments made by respondents. Comments in Spanish were reviewed by a Spanish-speaking staff member.

The reviewed paper 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 3 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 as a result of scanning were 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 80 entries were made into the data decision log during the course of

data scanning and processing. The majority of these were attributed to multiple response options selected on a gate question. Additional entries detail the decisions made about numeric ranges and web/paper harmonization issues.

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 the paper questionnaires had been scanned, all survey data were cleaned and edited. General cleaning and editing activities are described briefly below, with more detailed information found in **Appendix H** (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.
- A new “Missing data” value was added to account for web survey break-offs to indicate that a particular item is missing because it was never seen by the respondent. The point at which the breakoff occurs is coded with the traditional “Missing data” code, but all variables thereafter are coded with the new “Missing data – Never Seen” code.
- Derived variables were created to reflect each response recorded for certain “mark-one” type questions (A2, A7, and O7), 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.
- Variables were designed to summarize the responses for the electronic device, caregiving (who and what conditions), responses to exercise recommendations, sunburn

information (activities and protections against), Federal tobacco message exposure, cancer, Hispanic ethnicity, and race questions. These variables, HaveDevice_Cat, CaregivingWho_Cat, CaregivingCond_Cat, ExRec_Cat, SunburnedAct_Cat, SunburnedProt_Cat, TobaccoMessages_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.

- 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 G3, the estimated number of calories needed daily, to maintain current weight. In 60 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 (E4), 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.

Web Data Errors and Corrections

In October 2020, Westat discovered coding errors associated with missing values for data from web respondents in the web pilot. The errors impacted 35 survey variables, 31 of which were corrected and 4 of which cannot be corrected. There are two types of errors:

1. For some of the open ended questions, invalid skips (-9) were coded to 0 instead of -9. In total, 12 open-ended questions were directly impacted by the issue and 191 responses required revision. An additional 19 variables (and 33 responses) required minor revisions to missing value assignments; this is because one of the questions that was incorrectly coded (TimesSunburned) influenced how missing values were assigned to the 19 other variables.
 - a. **Appendix I** includes a table that summarizes the 31 variables that are affected by this coding error, the survey records that required revisions for each variable, and what the revisions are. All of these revisions were applied to data delivered in NCI in December 2020.
 - b. **Appendix J** includes a table showing the sample sizes and weighted means before and after corrections were applied to 11 of the 12 open-ended questions. The sample size and mean did not change for the 12th variable, *MailHHAdults*.

2. For the four items in the N6 matrix (How much do you think that each of the following can influence whether or not a person will develop cancer?), web respondents who chose ‘Don’t know’ should have been assigned a value of 4 but were incorrectly set to missing (-9). Unfortunately we cannot distinguish those who chose ‘Don’t know’ from those who intentionally skipped the question, therefore the error remains in the dataset. It is recommended that data users remove the web respondents from any analysis involving item N6. Westat delivered updated variable formats for these variables that alert data users to the issue.

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 the first 2 cycles of HINTS 5. Imputation occurred for 459 respondents on question A2, for 200 respondents on question A7, and 3 respondents on question O7. 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, L1, M1, O1, O2, O3, O5 and O6. 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 O8-O10). 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. A detailed list of the steps carried out to identify the number of adults in each household is included in **Appendix H**.

4.5 Survey Eligibility

Returned surveys were reviewed for completion and duplication (more than one questionnaire returned from the same household) to ensure they were eligible for inclusion in the final dataset. Of the 5,590 questionnaires received, 65 were returned blank, 45 were determined to be incompletely filled out, and 42 were identified as duplicates. The remaining 5,438 were determined to be eligible. The processes for these reviews are detailed below.

Definition of a Complete and Partial Complete Questionnaire

The procedures in HINTS 5 Cycle 3 and the Web Pilot for determining whether or not a returned questionnaire was complete were slightly modified relative to Cycles 1 and 2. A complete questionnaire was defined as any questionnaire with at least 80 percent of the required questions answered in Sections A and B. For this cycle, only questions required of every respondent were factored in to the completion rate calculation. Questions that followed skip patterns were excluded from the analysis. 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 191 partially-completed questionnaires. Both partially-completed and completely-answered questionnaires were retained. The 45 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 45 incomplete questionnaires represented 0.8% of all surveys with at least one question answered, which was consistent with Cycles 1 and 2 in HINTS 5. Data for 191 partially-completed and 5,247 completed questionnaires were included in the final dataset for Cycle 3 and the Web Pilot with a total of 5,438 surveys.

Eligibility of Multiple Questionnaires from a Household

41 households returned more than one filled in questionnaire (40 returned two, and 1 returned three questionnaires). The procedures to deal with this issue followed the same guidelines that were used for previous cycles:

- 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.
- 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.²
- 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 three sets of analytical variables: 1) rural-urban commuting area (RUCA) codes that classify census tracts using measures of population density, urbanization, and daily commuting; 2) National Center for Health Statistics (NCHS) urban-rural classification scheme for counties; and 3) Delta Regional Authority service area flag.

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

² 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.

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.

The Delta Regional Authority is a regional economic development agency serving 252 counties and parishes in parts of eight states: Alabama, Arkansas, Illinois, Kentucky, Louisiana, Mississippi, Missouri, and Tennessee. Its mission is to improve the quality of life for the residents of the Mississippi River Delta Region. The Delta Regional Authority service flag (DRA) identifies the areas served by this agency.

4.7 Codebook Development

Following cleaning, editing, and weighting (described below), a detailed codebook including frequencies was created for combined HINTS 5 Cycle 3 and Web Pilot data for both the weighted and the 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

To enhance the analysis of data from both Cycle 3 and the Web Pilot, four sets of weights were created: one set for each of the three treatment group samples described in Section 2.3 above and one set for the three samples combined. Each of the four sets of weights were created using the exact same weighting procedure as used for Cycles 1 and 2. Each set of weights are contained in a separate weighting data file.

In a given data file, every sampled adult who completed a questionnaire in HINTS 5 Cycle 3 and the Web Pilot 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 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 3, the base weights for households in the high minority stratum were roughly 1/6 the size of those in the low minority stratum.

As described in Section 2.1, in the previous HINTS cycles, an additional adjustment was made to the base weights of households that had multiple ways of receiving mail. 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, so such households would have received half the normal weight. Because this is a rare occurrence, in Cycle 3, this adjustment was not implemented.

5.2 Household Nonresponse Adjustment

Nonresponse is generally encountered to some degree in every survey. The first and most obvious effect of nonresponse is the reduction in the effective sample size, which in turn increases the sampling variance. In addition, if there are systematic differences between the respondents and the nonrespondents, there 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 3 and the Web Pilot 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 algorithm³ was used to identify variables highly correlated with household-level response, and these variables were used to create the nonresponse adjustment cells. The variables identified by the search algorithm for the three different treatment groups (Paper-only; Web Option; Web Bonus) for Cycle 3 were the same; however, the cells were defined differently. The variables used to define nonresponse cells were the same as in Cycle 2.

³ An in-house macro WESSEARCH, which calls the Search software, a freeware product developed by the University of Michigan ([Search \(software for exploring data structure\) website](#)).

- 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 .

Statistics for the household nonresponse adjustment factors for each of the three treatment groups are presented in Table 5-1. The nonresponse adjustment factors ranged from a low of 2.4 to a high of 7.9 for the Paper-only group, from 2.7 to 7.9 for Web Option group, and from 2.5 to 6.2 for Web Bonus group. The adjustment factors averaged 4.3, 4.1, and 3.7 across all nonresponse adjustment cells for three respective groups.

Table 5-1. Statistics for the household nonresponse adjustment factors, by treatment group

Treatment	Nonresponse Adjustment factor		
	Min	Average	Max
Paper-only (Cycle 3)	2.4	4.3	7.9
Web Option	2.7	4.1	7.9
Web Bonus	2.5	3.7	6.2

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 2017 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 *percent of adults who have ever been diagnosed with cancer*. The *health insurance* variable came from the 2018 National Health Information Survey (NHIS) (Cohen, et al., 2018) and corresponds to the question asked in the HINTS survey (C7, “Are you currently covered by any of the following types of health insurance or health coverage plans?”). The *percent of adults who have ever been diagnosed with cancer* variable came from the 2017 National Center for Health Statistics (U.S. Department of Health and Human Services, 2017) and corresponds to the question asked in the HINTS survey (M1, “Have you ever been diagnosed as having 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 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 3 and Web Pilot data sets, a set of 50 replicate weights was assigned to each responding adult. To illustrate how the replicate 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 $p(r)$, $r = 1, \dots, 50$, 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 $D(r)$, $r = 1, \dots, 50$. 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

This section reports the response rates for both Cycle 3 and the Web Pilot. Response rates were calculated using the RR2 formula of the American Association of Public Opinion Research

(AAPOR). A full summary of the Web Pilot results can be found in the *HINTS 5 Web Pilot Analysis Report*.

Table 6-1 shows the response rates for Cycle 3 and Web Pilot samples overall and by strata. The overall response rates have been weighted to account for the oversampling of addresses in high-minority areas.

Table 6-1. Comparison of response rates by strata for Cycle 3 and the Web Pilot

Experimental Group	n	Response Rates		
		High minority	Low minority	Overall
Cycle 3 (Paper-only)	14,730	21.6%	33.4%	30.2%
Web Pilot	8,700	23.3%	33.4%	30.6%
Web Pilot Group 1: Web Option	4,350	21.9%	32.6%	29.6%
Web Pilot Group 2: Web Bonus	4,350	24.6%	34.2%	31.5%
	23,430	22.2%	33.4%	30.3%

The Cycle 3 (Paper-only) overall response rate was 30.2 percent. The Web Pilot overall response rate was 30.6 percent, which was not significantly different from Cycle 3's overall response rate. Similar to previous cycles, the high-minority strata response rates were lower than the low-minority response rates in both samples.

The Web Pilot included an experiment testing the effectiveness of offering a \$10 'bonus' Amazon e-gift card for responding over the web. This treatment was expected to help improve response, overall, as well as increase the proportion of response by web. Web Pilot respondents who were offered the bonus incentive (Group 2: Web Bonus) had a slightly higher response rate than the control group (Group 1: Web Option) and the main Paper-only group (Cycle 3). However the difference is only significant in the high minority strata based on a weighted chi square test ($p < .01$). The bonus incentive was highly effective at increasing the proportion of responses by web (59.2 percent) relative to offering no bonus incentive (25.2 percent) ($p < .0001$).

References

- Cohen, R.A., Martinez, M.E., and Zammitti, E.P. (2018). *Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January – March 2018*. Retrieved from [Health Insurance Coverage: Early Release of Estimates from the National Health Interview Survey, January-March 2018 published on CDC.gov](#)
- 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 (2017). *Summary Health Statistics: National Health Interview Survey, 2017*. Retrieved from [Summary Health Statistics: National Health Interview Survey, 2017 published on CDC.gov](#)

This page is deliberately blank.

Appendix A

Cover Letters in English (Cycle 3)

This page is deliberately blank.



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 about how people find and use health and medical information. By completing this survey, 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 the adult in your household with the next birthday to complete the survey 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 the [Health Information National Trends Survey home page](#). Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805.

Thank you in advance for your participation.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 (HINTS). By participating in HINTS, 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 survey. If that adult has already completed the survey and returned it to us, please accept my sincere thanks. If that adult has not yet completed and returned the survey, we ask that he or she please do so as soon as possible.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 about how people find and use health and medical information. 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 survey. 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 survey was misplaced, an additional copy is enclosed.

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

Additional information about HINTS is available at the [Health Information National Trends Survey home page](#). Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll free at 1-888-738-6805.

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

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 is deliberately blank.

Appendix B

Cover Letters in Spanish (Cycle 3)

This page is deliberately blank.



FIRST MAILING

Estimado residente de {City}:

Le escribimos para invitarle 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 acerca de cómo las personas encuentran y utilizan la información sobre la salud y la información médica. Complete esta encuesta para ayudarnos 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 la encuesta 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 el [Encuesta Nacional de Tendencias de Información sobre la Salud página de inicio](#). La compañía de estudios de investigación Westat está realizando esta encuesta. Si tiene alguna pregunta sobre HINTS, llame a Westat al siguiente número de teléfono libre de cargo, 1-888-738-6805.

Gracias de antemano por su participación.

Atentamente,

Kelly D. Blake, ScD
Director, 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 AND THIRD MAILINGS

Estimado residente de {City}:

Recientemente le invitamos a participar en una importante encuesta nacional patrocinada por el Departamento de Salud y Servicios Humanos de Estados Unidos (HHS, por sus siglas en inglés). El objetivo de la Encuesta Nacional de Tendencias de Información sobre la Salud (HINTS, por sus siglas en inglés) es averiguar acerca de cómo las personas encuentran y utilizan la información sobre la salud y la información médica. 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 encuesta completada. 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 de que su encuesta 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 la encuesta en las próximas dos semanas.**

Usted podrá encontrar más información sobre HINTS en el sitio web el [Encuesta Nacional de Tendencias de Información sobre la Salud página de inicio](#). La compañía de estudios de investigación Westat está realizando esta encuesta. Si tiene alguna pregunta sobre HINTS, llame a Westat al número libre de cargo, 1-888-738-6805.

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

Atentamente,

Kelly D. Blake, ScD
Director, 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

Cover Letters (Web Option)

This page is deliberately blank.



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 about how people find and use health and medical information. By completing this survey, 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 the adult in your household with the next birthday to complete the survey in the next two weeks.** To complete the survey online, please visit:

Survey Website: [Health Information National Trends Survey website](#)

Your Access Code: {1A0784B8}

You may also fill out and return the paper survey that is included in this mailing. **You do not need to do both the online and paper versions of the survey.**

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.

The study is sponsored by the U.S. Department of Health and Human Services. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805. Thank you in advance for your participation.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 survey packet asking for your household's participation in the Health Information National Trends Survey (HINTS). By participating in HINTS, 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 survey. If that adult has already completed the survey and returned it to us, please accept my sincere thanks. If that adult has not yet completed and returned the survey, we ask that he or she please do so as soon as possible.

To complete HINTS online, please visit:

Survey Website: [Health Information National Trends Survey website](#)

Access code: {1A0784B8}

If you prefer to answer a paper version of the survey, an extra copy will be mailed to your household in the next few weeks. You do not need to do both the online and paper versions of the survey.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 about how people find and use health and medical information. 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 survey. To make sure HINTS provides accurate information, we need all the households invited to participate in this year's HINTS to complete the survey. In order to make sure we get responses from a random sample of people, **we ask the adult in your household with the next birthday to complete the survey in the next two weeks.** To complete the survey online, please visit:

Survey Website: [Health Information National Trends Survey website](#)

Your Access Code: {1A0784B8}

You may also fill out and return the paper survey included in this mailing. You do not need to do both the online and paper versions of 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. If you have any questions, please call Westat toll free at 1-888-738-6805. Thank you in advance for contributing to this important national study.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 28



This page is deliberately blank.

Appendix D

Cover Letters (Web Bonus)

This page is deliberately blank.



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 about how people find and use health and medical information. By completing this survey, 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 the adult in your household with the next birthday to complete the survey in the next two weeks.** To complete the survey online, please visit:

Survey Website: [Health Information National Trends Survey website](#)
Your Access Code: {1A0784B8}

If you complete the survey online, you will receive an additional \$10 Amazon e-gift card. You may also fill out and return the paper survey that is included in this mailing. You do not need to do both the online and paper versions of the survey.

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.

The study is sponsored by the U.S. Department of Health and Human Services. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805. Thank you in advance for your participation.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 survey packet asking for your household's participation in the Health Information National Trends Survey (HINTS). By participating in HINTS, 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 survey. If that adult has already completed the survey and returned it to us, please accept my sincere thanks. If that adult has not yet completed and returned the survey, we ask that he or she please do so as soon as possible.

To complete HINTS online, please visit:

Survey Website: [Health Information National Trends Survey website](#)

Access code: {1A0784B8}

If you complete to the survey online, you will receive an additional \$10 Amazon e-gift card. If you prefer to answer a paper version of the survey, an extra copy will be mailed to your household in the next few weeks. You do not need to do both the online and paper versions of the survey.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 about how people find and use health and medical information. 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 survey. To make sure HINTS provides accurate information, we need all the households invited to participate in this year's HINTS complete the survey. In order to make sure we get responses from a random sample of people, **we ask the adult in your household with the next birthday to complete the survey in the next two weeks.** To complete the survey online, please visit:

Survey Website: [Health Information National Trends Survey website](#)

Your Access Code: {1A0784B8}

If you complete the survey online, you will receive an additional \$10 Amazon e-gift card. You may also fill out and return the paper survey that is included in this mailing. You do not need to do both the online and paper versions of 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. If you have any questions, please call Westat toll free at 1-888-738-6805. Thank you in advance for contributing to this important national study.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
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 is deliberately blank.

Appendix E

Frequently Asked Questions in English and Spanish (Cycle 3)

This page is deliberately blank.

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 F

Frequently Asked Questions (Web Option and Web Bonus)

This page is deliberately blank.

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: I completed the survey online. Do I also need to complete the paper survey?

A: No, you only need to complete the survey once. If you completed the survey online then you do not need to return the paper survey.

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

A: You can find out more about HINTS at the [Health Information National Trends Survey home page](#). The study concerns health and how people get health information. For example, we will ask how you usually get information about health and what sources of information you most trust. We will also ask about your beliefs on what contributes to good health, how best to prevent cancer, and other health related topics.

Q: How will the study results be used?

A: The results will help the U.S. Department of Health and Human Services promote good health and prevent disease by determining the best ways to communicate accurate health information.

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: Getting answers 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. Your participation is voluntary, and you may refuse to answer any questions or leave the study at any time. However, your answers are very important to the success of this study and will represent thousands of others.

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 linked to any information that could identify you or your household, to the extent provided by law. Your completed survey will be stored in a secure file with restricted access. All contact information for your household (such as mailing address) 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?

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.

This page is deliberately blank.

Appendix G

Additional Insert (Web Bonus)

This page is deliberately blank.

ADDITIONAL INSERT



Respond to **HINTS** online at
www.HINTS-survey.org
and receive a
\$10 Amazon e-gift card!

▶ See the enclosed letter for details.

This page is deliberately blank.

Appendix H

Variable Values and Data Editing Procedures

This page is deliberately blank.

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.
- -7 = Missing data (exclusive to web mode respondent variables), question never presented due to break off in interview. Example: If web mode respondent reached any point in questionnaire and exited interview without finishing, exit point would be assigned the traditional Missing data value (-9), and all variables thereafter would be assigned the value -7.
- -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 UseInternet WearableDevTrackHealth NoticeCalorieInfoOnMenu Smoke100 UsedECigEver SeenFederalCourtTobaccoMessages HeardHPV EverHadCancer	Recoding filter/skip questions	<p>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.</p> <p>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.</p> <p>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.</p> <p>Please note that if neither condition was met, the missing response code values were retained.</p>
WhereSeekHealthInfo_IMP StrongNeedHealthInfo_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 suffixes "_IMP" and "_I" indicate that these variables include 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_BuyMedicine Electronic_TalkDoctor Electronic_TrackedHealthCosts Electronic_TestResults Electronic_MadeAppts Electronic_ECigHarms Tablet_AchieveGoal Tablet_MakeDecision Tablet_DiscussionsHCP WillingShareData_HCP WillingShareData_Fam IntRsn_VisitedSocNet IntRsn_SharedSocNet IntRsn_WroteBlog IntRsn_SupportGroup IntRsn_YouTube HealthIns_InsuranceEmp HealthIns_InsurancePriv HealthIns_Medicare HealthIns_Medicaid HealthIns_Tricare HealthIns_VA HealthIns_IHS HealthIns_Other NotAccessed_SpeakDirectly NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_ConcernedPrivacy NotAccessed_NoRecord NotAccessed_LogInProb NotAccessed_Uncomfortable NotAccessed_MultipleRec RecordsOnline_RefillMeds RecordsOnline_ViewResults RecordsOnline_RequestCorrection	Recoding missing responses for items with forced-choice response formats	<p>Respondents were asked to select ‘yes’ or ‘no’ to a series of sub-items, allowing them to select as many responses as would apply.</p> <p>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.</p> <p>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.’</p> <p>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.</p>

Variable	Editing Rule	Description of Rule
RecordsOnline_MessageHCP RecordsOnline_DownloadHealth RecordsOnline_AddHealthInfo RecordsOnline_MakeDecision ESent_AnotherHCP ESent_Family ESent_HealthApp MedConditions_Diabetes MedConditions_HighBP MedConditions_HeartCondition MedConditions_LungDisease MedConditions_Depression CalorieInfo_FewerCalories CalorieInfo_MoreCalories CalorieInfo_FewerItems CalorieInfo_SmallerSizes CalorieInfo_MoreItems CalorieInfo_LargerSizes GovPARec_HCP GovPARec_Internet GovPARec_TV GovPARec_Magazine		
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 FreqGoUrgentCare AccessOnlineRecord TimesModerateExercise SmokeNow	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.

Variable	Editing Rule	Description of Rule
Height_Feet Height_Inches	Edits for implausible values	<p>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.</p> <p>Rules Applied to Edit Height Variables:</p> <p>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).</p> <p>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").</p> <p>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).</p> <p>If HEIGHT_Feet > 3 and HEIGHT_Feet < 7 and HEIGHT_Inches = 20, 30, etc. thru 90 then the trailing 0 was removed.</p> <p>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).</p> <p>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).</p> <p>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").</p>

Variable	Editing Rule	Description of Rule
		<p>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).</p> <p>If none of the preceding height editing rules were applicable:</p> <p><u>Height Feet (Height in Feet):</u> Any responses greater than 7 feet were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>Height Inches (Height in Inches):</u> Any responses greater than 11 inches were recoded to “-4”, which is the code for non-conforming responses.</p>
HaveDevice_Cat CaregivingWho_Cat CaregivingCond_Cat ExRec_Cat SunburnedAct_Cat SunburnedProt_Cat TobaccoMessages_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.
HealthInsurance	Derived variable	Responses to questions asking about different types of health insurance (C7a-h) were compiled into a derived measure of whether or not the respondent was covered by any health insurance.
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.

Variable	Editing Rule	Description of Rule
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_HoursPerWeek2 Weight DrinkDaysPerWeek AverageTimeSitting AverageSleepNight WhenDiagnosedCancer SexualOrientation_OS Age SelfAge SelfMOB HHAdultAge[2-5] HHAdultMOB[2-5]	Recoding out of range responses	<p><u>CaregivingHoursPerWeek</u> Any responses greater than 168 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>Weight:</u> Any responses less than 40 pounds or greater than 500 pounds were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>DrinkDaysPerWeek</u> Any responses greater than 7 days per week were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>AverageTimeSitting</u> Any responses greater than 24 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>AverageSleepNight</u> Any responses greater than 24 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>WhenDiagnosedCancer (Age at Time of Cancer Diagnosis):</u> Any responses greater than the age of the respondent were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>SexualOrientation OS</u> Review of verbatim responses - Responses of “none of your business” and other similar phraseology were reviewed for scanning accuracy and rephrased as “Refused”.</p> <p><u>SelfMOB (Respondent’s Month of Birth):</u> Any responses less than 1 or greater than 12 months were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>HHAdultMOB[2-5] (Second – Fifth Adult in Household Month of Birth):</u> Any responses less than 1 or greater than 12 months were recoded to “-4”, which is the code for non-conforming responses.</p>

Variable	Editing Rule	Description of Rule
		<p>Age Variables Responses were examined for out of range or unlikely ages (those listing their age as < 18 and > 105).</p>
<p>HaveDevice_CellPh HaveDevice_None Caregiving_No</p>	<p>Recoding filter/skip questions</p>	<p>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.</p> <p>If other responses were checked, the “No” or “None” response was recoded to “Not selected”, and the other responses were retained.</p>
<p>ExRec_NotHeard Sunburned_None NotHisp</p>	<p>Recoding illogical response combinations</p>	<p>For these “mark all that apply” questions (“mark all that apply” type questions where one or more response option do not contain a skip instruction at the “No” or “None” response, but keeping a “No” or “None” response in combination with other responses does not make logical sense), when the “No” or “None” response was selected, all responses within the question group were examined.</p> <p>If other responses were checked, the “No” or “None” response was recoded to “Not selected”, and the other responses were retained.</p> <p>This edit was added for this cycle, because the web pilot instrument contained hard edits which did not allow for the “No” and “None” responses to be combined with other responses on these questions.</p>
<p>AverageCaloriesPerDay AverageCaloriesPerDay_DK</p>	<p>Recoding filter/skip questions</p>	<p>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, 60 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.</p>

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.

Appendix I

Web Data Errors and Corrections: Variables and Records Revised

This page is deliberately blank.

Question Number	Name of variable	Number of Records impacted	Recodes applied	HHID for records that were updated
Q2	MailHHAdults	4	Recoded from -4 to -9	92018111 93020866 93022418 93022736
E3	Caregiving_HoursPerWeek2	12	Recoded from 0 to -9	93019884 93019649 92015236 92016831 93020503 93019332 93021749 93023085 93020442 92018942 93023002 93023293
F7_Inches	Height_inches	9	Recoded from 0 to -9	93022480 93022157 93019568 93022714 92016078 93019151 92017701 93023094 92015789
G3	AverageCaloriesPerDay	7	Recoded from 0 to -9	92015941 92018290 93019528 93019592 93020690 93021122 93021749
G3	AverageCaloriesPerDay_DK	7	Recoded from -1 to -9	92015941 92018290 93019528 93019592 93020690 93021122 93021749

Question Number	Name of variable	Number of Records impacted	Recodes applied	HHID for records that were updated
G7	DrinkDaysPerWeek	12	Recoded from 0 to -9	93019482 93021105 92014965 93021749 93019959 93020170 93023293 93019528 92017701 93019112 93023098 93021710
G8	DrinksPerDay	6	Recoded from 0 to -9	92016221 92017874 92018915 93021092 93021972 93022460
H4	AverageTimeSitting	39	Recoded from 0 to -9	93019482 93020883 93021772 93020389 93023038 93021718 92017265 93021260 93022668 93023172 93023184 93022950 93023354 93019769 93020689 92017680 92015867 93020964 93021680 93021839 92017263 93023293 92017850 93021012 93020901 92017757 93021749 93022193 93021444 93019528 93022446 92017729 92016320 93023097 93023338 93023098 93023155 93020878 92018620

Question Number	Name of variable	Number of Records impacted	Recodes applied	HHID for records that were updated	
H10	AverageSleepNight	48	Recoded from 0 to -9	92015225 92015948 92016357 92016517 92016561 92017063 92017680 92017701 92018160 92018361 92018723 92018740 92019002 93019123 93019233 93019274 93019380 93019482 93019528 93019675 93019773 93020218 93020294 93020795	93020887 93020901 93021012 93021083 93021092 93021110 93021195 93021444 93021749 93022191 93022256 93022418 93022513 93022517 93022640 93022652 93022695 93022773 93022789 93022833 93023097 93023165 93023184 93023209
J3	TimesSunburned	26	Recoded from 0 to -9	93022480 93019482 93021256 92018160 93020567 93020837 93022599 93020689 92015867 93022714 93020242 93020901 93021749	93021444 92016519 93020192 93021718 92017680 93023165 93023293 93021012 93019528 92017315 92017701 93022191 93023097

Question Number	Name of variable	Number of Records impacted	Recodes applied	HHID for records that were updated	
J4	Sunburned_HomeOutside Sunburned_JobOutside Sunburned_SportingEvent Sunburned_Swimming Sunburned_Sunbathing Sunburned_Other Sunburned_Exercise Sunburned_DK Sunburned_DayToDay Sunburned_OutdoorEvent SunburnedAct_Cat	25	Recoded from -1 to -6	93022480 93019482 93021256 92018160 93020567 93022599 93020689 92015867 93022714 93020242 93020901 93021749 93021444	92016519 93020192 93021718 92017680 93023165 93023293 93021012 93019528 92017315 92017701 93022191 93023097
J4	Sunburned_HomeOutside Sunburned_JobOutside Sunburned_SportingEvent Sunburned_Swimming Sunburned_Sunbathing Sunburned_Other Sunburned_Exercise Sunburned_DK Sunburned_DayToDay Sunburned_OutdoorEvent SunburnedAct_Cat	1	Recoded from -2 to -7	93020837	
J5	Sunburned_SPF15 Sunburned_ProtClothing Sunburned_Shade Sunburned_None Sunburned_DontRemember SunburnedProt_Cat	25	Recoded from -1 to -6	93022480 93019482 93021256 92018160 93020567 93022599 93020689 92015867 93022714 93020242 93020901 93021749 93021444	92016519 93020192 93021718 92017680 93023165 93023293 93021012 93019528 92017315 92017701 93022191 93023097
J5	Sunburned_SPF15 Sunburned_ProtClothing Sunburned_Shade Sunburned_None Sunburned_DontRemember SunburnedProt_Cat	1	Recoded from -2 to -7	93020837	

Question Number	Name of variable	Number of Records impacted	Recodes applied	HHID for records that were updated
J6	Sunburned_Alcohol	25	Recoded from -1 to -6	93022480 92016519 93019482 93020192 93021256 93021718 92018160 92017680 93020567 93023165 93022599 93023293 93020689 93021012 92015867 93019528 93022714 92017315 93020242 92017701 93020901 93022191 93021749 93023097 93021444
J6	Sunburned_Alcohol	1	Recoded from -2 to -7	93020837
M3	WhenDiagnosedCancer	7	Recoded from 0 to -9	92018111 92017270 92017804 93022335 93020031 93019769 93022887
O8	TotalHousehold	9	Recoded from 0 to -9	93019482 93019884 92015867 92015654 93021749 93021718 92018204 93021012 93020878
O10	ChildreninHousehold	10	Recoded from 0 to -9	93021718 92018204 93021012 93019482 93019884 92015867 92015654 93021749 93022446 92018690 93021718 92018204 93021012

This page is deliberately blank.

Appendix J

**Web Data Errors and Corrections:
Differences in Sample Sizes and Weighted Means
for 11 Open-ended Variables Before and After
Corrections**

This page is deliberately blank.

Variable	Label	BEFORE CORRECTION (original delivery)		AFTER CORRECTION (new delivery)		DIFFERENCE AFTER CORRECTION (new - original)		
		Valid N	Mean	Valid N	Mean	Valid N	Mean	% diff in mean
Caregiving_HoursPerWeek2	E3. Think about the individual for whom you are currently providing the most care. About how many hours per week do you spend in an average week providing care?	723	30.26	711	30.86	-12	0.60	1.97%
Height_Inches	F7. About how tall are you without shoes? Inches:	5220	5.40	5211	5.41	-9	0.02	0.33%
AverageCaloriesPerDay	G3. About how many calories do you think a man/woman of your age and physical activity needs to consume a day to maintain your current weight?	2773	1798.75	2766	1803.06	-7	4.31	0.24%
DrinkDaysPerWeek	G7. During the past 30 days, how many days per week did you have at least one drink of any alcoholic beverage?	4844	1.37	4832	1.38	-12	0.00	0.28%
DrinksPerDay	G8. During the past 30 days, on the days when you drank, about how many drinks did you drink on average?	2514	2.46	2508	2.46	-6	0.00	0.19%
AverageTimeSitting	H4. During the past 7 days, how much time did you spend sitting on a typical day at home or at work?	5015	6.88	4976	6.94	-39	0.06	0.88%
AverageSleepNight	H10. During the past 7 days, how many hours of sleep did you get on average per night?	5273	6.74	5225	6.84	-48	0.10	1.42%
TimesSunburned	J3. During the past 12 months, how many times have you had a sunburn (even a small part of your skin turns red or hurts for 12 hours or more) from too much sun exposure?	4885	1.28	4859	1.30	-26	0.02	1.25%

Variable	Label	BEFORE CORRECTION (original delivery)		AFTER CORRECTION (new delivery)		DIFFERENCE AFTER CORRECTION (new - original)		
		Valid N	Mean	Valid N	Mean	Valid N	Mean	% diff in mean
WhenDiagnosedCancer	M3. At what age were you first told that you had cancer?	833	52.40	826	52.95	-7	0.56	1.06%
TotalHousehold	O8. Including yourself, how many people live in your household?	5189	2.99	5180	3.00	-9	0.01	0.35%
ChildrenInHH	O10. How many children under the age of 18 live in your household?	5020	0.57	5010	0.57	-10	0.00	0.41%