



Health Information National Trends Survey 4 (HINTS 4)

HINTS-FDA2 Methodology Report

July 2017

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

Prepared by
Westat
1600 Research Boulevard
Rockville, MD 20850



Table of Contents

<u>Chapters</u>		<u>Page</u>
1	HINTS – FDA2 Overview	1
2	Sample Selection	2
	2.1 Sampling Frame.....	2
	2.2 Stratification.....	3
	2.3 Selection of Address Sample	3
	2.4 Within-Household Sample Selection	4
3	Data Collection	5
	3.1 Mailing Protocol.....	5
	3.2 In-bound Telephone Calls.....	6
	3.3 Incoming Questionnaires.....	7
4	Data Management.....	8
	4.1 Scanning	9
	4.2 Data Cleaning and Editing.....	10
	4.3 Imputation.....	10
	4.4 Determination of the Number of Household Adults.....	11
	4.5 Survey Eligibility.....	11
	4.6 Codebook Development.....	13
5	Weighting and Variance Estimation	13
	5.1 Household Base Weights	14
	5.2 Household Nonresponse Adjustment	14
	5.3 Initial Person-Level Weights	16
	5.4 Calibration Adjustments	16
	5.5 Replicate Variance Estimation	18
6	Response Rates	20
	6.1 Overall Response Rate	20
	6.2 Impact of Stratification on Smoking Outcomes	20
	References.....	23

Appendixes

		<u>Page</u>
A	Cover Letters in English.....	A-1
B	Cover Letters in Spanish.....	B-1
C	Frequently Asked Questions (FAQs) – English and Spanish.....	C-1
D	Variable Values and Data Editing Procedures	D-1

Tables

2-1	Smoking rates and strata.....	3
2-2	HINTS-FDA2 sample summary by sampling stratum	4
3-1	Mailing protocol.....	6
3-2	Number of packets per mailing.....	6
3-3	Final household status	8
3-4	Survey response by date.....	8
4-1	Completed and partially completed questionnaires.....	12
6-1	Response rate calculations by Strata	20
6-2	Weighted smoking rates across the HINTS-FDA strata and HINTS 4.....	22

The Health Information National Trends Survey (HINTS) is a nationally-representative survey which has been administered every few years by the National Cancer Institute (NCI) since 2003. The HINTS target population is adults aged 18 or older in the civilian non-institutionalized population of the United States. In 2015, NCI launched a special round of HINTS data collection in partnership with the Food and Drug Administration (FDA), referred to as HINTS-FDA. This data collection combined the traditional HINTS topics of health communication, cancer knowledge, and cancer risk behaviors, with an assessment of the public’s risk perceptions about new tobacco products, perceptions of tobacco product harm, and tobacco product claims in addition to other FDA topics.

HINTS-FDA2 is a follow-on survey to the original HINTS-FDA survey of 2015. HINTS-FDA2 was conducted from January 26 through May 5, 2017 with a slightly shortened HINTS-FDA instrument. Data collection was conducted by mail using a protocol similar to that used for HINTS-FDA with a goal of obtaining 1,600 completed questionnaires.

Oversampling for Smokers

The first data collection for HINTS-FDA, conducted in 2015, used a sampling approach modified from the usual HINTS sampling strategy in an effort to include more current and former smokers in the study. Stratification of the sampling frame coupled with the use of differential sampling rates was designed to increase the yield of current smokers in the sample for analysis purposes. Addresses in the high and medium-high smoking strata were oversampled while those in the medium-low and low smoking strata were under-sampled. This strategy did not increase the number of smokers relative to HINTS. The absolute number of smokers who responded to HINTS-FDA was similar to those found for Cycle 4 of HINTS using the usual HINTS strategy. The smoking rates for the medium-low and low strata were considerably lower than expected.

To supplement the number of smokers and to obtain a higher rate of smokers in the survey, the HINTS-FDA2 sample used a similar sample design as the first HINTS-FDA sample with two main differences:

- 1. The source of the data used to stratify the address frame.** HINTS-FDA2 used county-level small area estimates of current smokers from the 2010-2011 Current

Population Survey (TUS-CPS) to define the sampling strata, whereas the first FDA cycle used adjusted¹ county-level estimates of current smokers from the 2003 Behavioral Risk Factor Surveillance System (BRFSS).

- 2. A change in the sample allocation across the sampling strata.** The change in the degree of differential sampling across the strata compared to the first FDA cycle was designed to yield a relatively larger number of expected smokers in terms of effective sample size (taking into account the design effect due to differential sampling).

Sample Selection

2

The sampling for HINTS-FDA2 consisted of a two-stage design. In the first stage, a sample of addresses was selected from a file of residential addresses. IN the second stage, one adult was selected within each sampled household.

2.1 Sampling Frame

The sampling frame for HINTS-FDA2 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 in the MSG database, including post office (P.O.) boxes, throwbacks (i.e., street addresses for which mail is redirected by the United States Postal Service to a specified P.O. box), and seasonal addresses were subject to sampling.

Rarely are surveys conducted with a sampling frame that perfectly represents the target population. The sampling frame is one of the many sources of error in the survey process. The sampling frame used for the address sample contained duplicate units because some households receive mail in more than one way. To permit adjustment for this duplication of households in the sampling frame, a

¹ The county-level smoking rates were based on the 2003 BRFSS small area estimates adjusted by the ratio of the 2011 to the 2003 BRFSS state smoking rates so that when county rates are aggregated to the state level they are in agreement with the 2011 BRFSS state-level smoking estimates.

question about how many different ways respondents receive mail was included on the survey instrument (see question G24).

In rural areas, some 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

As discussed above, stratification of the sampling frame coupled with the use of differential sampling was designed to increase the yield of current smokers in the sample for analysis purposes. The sampling frame of addresses was grouped into four explicit sampling strata based on county-level smoking rates: high, medium-high, medium-low, and low. The county-level smoking rates were based on the 2010-2011 TUS-CPS small area estimates. As shown in table 2-1 below, addresses in census counties that had smoking rates equaling or exceeding 19.2 percent were assigned to the high-smoking stratum. Addresses in census counties that had smoking rates between 15.9 percent and 19.1 percent were assigned to the medium-high smoking stratum. Addresses in census counties that had smoking rates between 12.7 percent and 15.8 percent were assigned to the medium-low smoking stratum. All addresses in the remaining counties were assigned to the low-smoking stratum.

Table 2-1. Smoking rates and strata

Smoking Rate	Stratum
19.2 or more	High
15.9 - 19.1	Medium High
12.7 - 15.8	Medium Low
less than 12.7	Low

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 HINTS-FDA2 was 5,741: 1,902 from the high smoking stratum, 1,583 from medium-high smoking stratum, 1,315 from the medium-low smoking stratum,

and 942 from the low smoking stratum. Relative to a proportional design, the high and medium-high smoking strata were oversampled by 27 percent and 8 percent respectively, while the medium-low and low smoking rates strata were under-sampled by 7 percent and 31 percent, respectively.

Table 2-2 below summarizes the address sample for HINTS-FDA2, 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.

Table 2-2. HINTS-FDA2 sample summary by sampling stratum

Sampling Stratum	Number of sample addresses	Percent of addresses in frame	Percent of sample addresses	Percent of sampled addresses oversampled(+) or under-sampled(-)
High	1,902	26.2	33.1	+26.6
Medium-high	1,583	25.5	27.6	+8.0
Medium-low	1,315	24.5	22.9	-6.7
Low	942	23.7	16.4	-30.9
Total	5,741			

2.4 Within-Household Sample Selection

The second-stage of sampling consisted of selecting one adult within each sampled household. In keeping with HINTS 4 methodology, data collection for HINTS-FDA2 implemented the Next Birthday Method to select the one adult in the household. Questions were included on the survey instrument to assist the household in selecting the adult in the household having the next birthday (see page 1 of the survey instrument).

Data collection for HINTS-FDA-2 started on January 26, 2017 and concluded on May 5, 2017. The survey was conducted exclusively by mail with a \$2 pre-paid monetary incentive to encourage participation. The specific mailing procedures and outcomes are described in detail below.

3.1 Mailing Protocol

A total of four mailings were sent out as part of HINTS-FDA-2. The mailing protocol followed a modified Dillman approach (Dillman, et. al., 2009) with a total of four mailings: an initial mailing, a reminder postcard, and two follow-up mailings. All households in the sample received the first mailing and reminder postcard, while only non-responding households received the subsequent survey mailings. All households received one English survey per mailing unless someone from the household contacted Westat to request Spanish, in which case the household received one Spanish survey per mailing for all subsequent mailings.

The second survey mailing was sent via USPS Priority Mail, while all other mailings were sent First Class. The contents of all mailings are further described in Table 3-1. Cover letters in English can be found in **Appendix A** and cover letters in Spanish are in **Appendix B**. All cover letters include a list of Frequently Asked Questions (FAQs) on the back. These FAQs in both English and Spanish are in **Appendix C**.

Table 3-1. Mailing protocol

Mailing	Date mailed	Mailing method	Materials
Mailing 1	January 26, 2017	1 st Class Mail	English cover letter with FAQs English Questionnaire Return envelope \$2 bill
Postcard	February 2, 2017	1 st Class Mail	Reminder/thank you postcard
Mailing 2	February 24, 2017	USPS Priority Mail	English cover letter with FAQs English Questionnaire Return envelope OR (upon request) Spanish cover letter with FAQs Spanish Questionnaire Return envelope
Mailing 3	March 24, 2017	1 st Class Mail	English cover letter with FAQs English Questionnaire Return envelope OR (upon request) Spanish cover letter with FAQs Spanish Questionnaire Return envelope

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

Table 3-2. Number of packets per mailing

Mailing	English only	Spanish only	Total
Mailing 1	5,750	N/A	5,750
Mailing 2	4,366	2	4,368
Mailing 3	3,596	1	3,597
Total	13,712	3	13,715

3.2 In-bound Telephone Calls

Two toll-free telephone numbers were provided to respondents: one was used for English calls and one was used for Spanish calls. Both numbers were provided in each mailing. Respondents were told that they could call the number if they had questions, concerns, or if they needed to request

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

Since HINTSV Cycle 1 was fielded at the same time as HINTS-FDA-2, it is impossible to determine which survey a caller was calling about if they did not provide identifying information. There were only three calls received throughout the HINTS-FDA-2 field period that were definitely about HINTS-FDA-2. Two of these calls were to request Spanish items and one was a call to let the study staff know that the caller had sent in her survey.

3.3 Incoming Questionnaires

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

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

Table 3-3. Final household status

Household status	English	Spanish	Total	
			N	%
Complete	1,761	1	1,762	30.6
Refusal			14	0.3
Undeliverable			662	11.5
Nonresponse			3,312	57.6
Total			5,750	100.0

The number of questionnaires returned by date during the field period can be found in Table 3-4. The majority of the returns are early in the field period, with 61 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 28 percent and the remaining 11 percent were in response to the final mailing.

Table 3-4. Survey response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: January 26	January 29 – February 4	44
Postcard: February 2	February 5 – 26	1,023
Mailing 2: February 24	February 27 – March 26	501
Mailing 3: March 24	March 27 – May 5	194
	Total	1,762

Data Management

4

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

4.1 Scanning

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

- Determining if the form was not scannable for any reason, such as being damaged in the mail. Some questionnaires or individual responses needed to be overwritten with a pen that was readable by the data capture software. Numeric response boxes were pre-edited to interpret and clarify non-numeric responses and responses written outside the capture area.
- Documenting potential problem questionnaires or pertinent comments made by respondents in a decision log. Comments in Spanish were reviewed by a Spanish-speaking staff member.

The reviewed surveys were then sent through the high-speed scanner to capture the responses. TeleForm read the form image files and extracted data according to HINTS-FDA2 rules established prior to the field period. Scanned data were then subject to validation according to HINTS specifications. If a data value violated validation rules, such as marking more than one choice box in a mark-only-one question, the data item was flagged for review by verifiers who looked at the images and the corresponding extracted data and resolved any discrepancies. Spanish forms were verified by a Spanish-speaking staff member.

In addition to the problem log mentioned above based on staff review, decisions made about data issues as a result of scanning were also recorded into a decision data log. The decision log contained the respondent ID, the value triggering the edit, the updated value, and the reason for the update. Only one entry was made into the data decision log during the course of data processing. This issue was attributed to an illogical response on a numeric question.

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

4.2 Data Cleaning and Editing

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

- Customized range and logical inconsistency edits, following predetermined processing rules to ensure data integrity, were developed and applied against the data.
- Edit rules were created to identify and recode nonresponse or indeterminate responses.
- Missing values were recoded for some responses to questions that featured a forced-choice response format and for filter questions where responses to later questions suggested a particular response was appropriate.
- Variables were designed to summarize the responses for the electronic cigarette research, tobacco products research, tobacco product awareness, tobacco product use, tobacco regulation knowledge, Hispanic ethnicity, and race questions. The variables ECig_Cat, Tobacco_Cat, RecentTobacco_Cat, TobaccoHeard_Cat, TobaccoTried_Cat, RegulateTobacco_Cat, Hisp_Cat, and Race_Cat2 indicated each response selected for respondents selecting only one response, and a multiple category for all of the respondents who answered multiple responses.
- Derived variables were created to reflect each response recorded for the occupation question in order to capture this information when respondents did not follow the instruction to mark only one response. For some variables for which a respondent was expected to select only one response, editing rules were used to determine which response was retained. For other variables, imputation, as described below, was carried out.
- “Other, specify” responses were examined, cleaned for spelling errors, categorized, and upcoded into preexisting response codes when applicable.

4.3 Imputation

For four variables featuring a mark-only-one response instruction (WhereSeekHealthInfo, ECigInfoSeek, SizeCigarsSmoked, and SexualOrientation), imputation was carried out for the questionnaires in which multiple responses were recorded. The imputation process used was the same as that carried out for the previous HINTS-FDA survey. Responses where a missing value of -5 (multiple responses selected) was applied were imputed. This occurred for 204 respondents for WhereSeekHealthInfo, 13 respondents for ECigInfoSeek, 2 respondent for SizeCigarsSmoked, and 2 respondents for SexualOrientation. A respondent’s multiple answers were replaced with a single

imputed answer that had the same distribution over the multiple answers as occurred in the single-answer responses.

In addition, hot-deck imputation was used to replace missing responses with imputed data for items used in the raking procedure of the HINTS-FDA2 weighting process. Hot-deck imputation is a data processing procedure in which a case with a missing value for a specific variable is assigned the corresponding value of a “similar” case in the same imputation class. The data record that supplies the imputed value is referred to as the “donor.” Under a hot deck approach, the resulting distribution preserves the distribution of values observed for respondents. Imputation classes are defined on the basis of variables that are thought to be correlated with the item with missing values. A donor is then randomly selected within an imputation class to supply the imputed value. Items imputed using the hot-deck approach were those involving the following characteristics: age, gender, educational attainment, marital status, race, ethnicity, and health insurance coverage.

4.4 Determination of the Number of Household Adults

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

4.5 Survey Eligibility

Of the 1,766 questionnaires received, 26 were incompletely-answered, and 3 were determined to be ineligible because they were duplicates from the same household. One was ineligible because the address was a business rather than a residence. A total of 1,736 questionnaires were coded as eligible questionnaires. Details about the eligibility rules are listed below.

Definition of a Complete and Partial Complete Questionnaire

In HINTS-FDA2, a complete questionnaire is defined as any questionnaire with at least 80 percent of the required questions answered in Sections A and C. A partial complete was defined as when between 50 and 79 percent of the required questions were answered in Sections A and C. Questionnaires with less than 50 percent of required items in Sections A and C answered were coded as incomplete. As shown in Table 4-1 below, there were 60 partially complete questionnaires. Both partially-completed and completely-answered questionnaires were retained. The 26 questionnaires which met the criteria described above for being flagged as incomplete were discarded.

Table 4-1. Completed and partially completed questionnaires

Status	Definition	Number of Cases	Disposition
Complete	80% of Sections A and C completed	1,676	Included in dataset
Partial Complete	50-79% of Sections A and C completed	60	Included in dataset
Incomplete	49% or less of Sections A and C completed	26	Discarded

Eligibility of Multiple Questionnaires from a Household

Three households returned more than one completed questionnaire. The procedures to deal with this issue followed the same guidelines that were used for households using the Next Birthday method in the previous HINTS-FDA:

- 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, questionnaires with fewer substantive questions omitted were 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

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.

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

4.6 Codebook Development

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

Weighting and Variance Estimation

5

Every sampled adult who completed a questionnaire in HINTS-FDA2 received a full-sample weight and a set of 50 replicate weights. 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 full-sample weight is the weight which is used to calculate population and subpopulation estimates from the data collected in HINTS-FDA2. The computation of the full-sample weights for HINTS-FDA2 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 are used to compute standard errors for these estimates. The name of the final full-sample weight variable is PERSON_FINWT0, and the name of the 50 replicate weight variables are PERSON_FINWT1 – PERSON_FINWT50. The replicate weights were calculated using the ‘delete one’ jackknife (JK1) replication method.

5.1 Household Base Weights

The initial step in the weighting process was the calculation of the household-level base weight for each household in the sample. The household base weight is the reciprocal of the probability of selecting the household for the survey, which depends on the stratum the household was selected from. Generally, base weights for units in oversampled strata are smaller than those in strata that were not oversampled. In HINTS-FDA2, the base weights for households in the high smoking stratum were roughly 3/5, 3/4, and 6/7 the size of those in the low smoking, medium-low smoking, and medium-high smoking strata, respectively.

If two different addresses led to the same household – for example, if a household receives mail via both a street address and a post office box – that household had twice the chance of selection of a household with only one address (and should therefore receive half the normal weight). An additional adjustment was made to the base weights of households that had multiple ways of receiving mail (as determined by the answer to survey question G24).

5.2 Household Nonresponse Adjustment

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

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

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

The weighting procedure for HINTS-FDA2 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 used to define nonresponse cells for HINTS-FDA2 were:

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

³ An inhouse macro WESSEARCH, which calls the Search software, a freeware product developed by the University of Michigan (<http://www.isr.umich.edu/src/smp/search/>).

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

$$HH_NRAF(b) = \frac{\sum_{S(b)} HH_BWT_i}{\sum_{C(b)} HH_BWT_i},$$

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

The household nonresponse adjustment factors ranged from a low of 1.79 to a high of 4.40, and averaged 2.94 across all nonresponse adjustment cells.

5.3 Initial Person-Level Weights

Each sampled adult in responding households was assigned an initial person-level weight. The initial person-level weight was calculated by multiplying the nonresponse-adjusted household weight by the reciprocal of the sample person's within-household probability of selection. Since in HINTS-FDA2 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 2015 ACS estimates that were available on the Census Bureau web site.

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

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

In addition to characteristics from the ACS, one health-related variable was used – *Percent with Health Insurance*. This variable came from the 2016 National Health Information Survey (NHIS) and corresponds to the question asked in the HINTS survey.

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 of 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 2016 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 HINTS-FDA2, a set of 50 replicate weights was assigned to each responding adult. To illustrate how the replicate weights are computed, suppose P is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A nationally representative estimator p can be computed by aggregating the adult sampling weights of all responding adults with this characteristic (e.g., all responding adults in the survey answering the survey question in a particular way). A JK1 jackknife variance estimator of the sampling variance of p can be computed in two steps:

- Step 1.** Recompute estimators $p(r)$, $r = 1, \dots, 50$, 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.

Analysis of Integrated HINTS-FDA2 and other HINTS Data Including HINTS-FDA

The sampling and weighting procedures for HINTS-FDA2 were designed to facilitate combining HINTS-FDA2 data with the first HINTS-FDA data (and other cycles of HINTS as well) for analysis. To combine HINTS-FDA2 with HINTS-FDA, simply concatenate the two files and create new weights using the procedures documented in section 4 of Rizzo et al., 2008. This document describes how to integrate HINTS data from independent HINTS surveys. These same procedures can be used to combine the HINTS-FDA2 data with other cycles of HINTS.

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

6.1 Overall Response Rate

Table 6-1 shows the response outcomes overall and by strata. The data have been weighted to account for the oversampling of addresses in areas with a higher likelihood of individuals who smoke. The overall response rate was 34.1 percent and this did not vary by more than three percentage points across the four strata. The percent of undeliverable households ranged from 7.3 percent to 15.4 percent across strata, with the high smoking and medium-high smoking strata having the highest undeliverable rates (15.8 and 10.5 percent respectively).

Table 6-1. Response rate calculations by Strata

Response class	High	Medium-High	Medium-Low	Low	Overall
Total sample	38,296,603	37,186,039	35,092,321	32,826,446	143,401,409
Respondents	10,916,040	11,636,767	10,968,018	9,934,776	43,455,601
Nonrespondents	21,490,324	21,631,248	20,521,669	20,529,555	84,172,797
Undeliverable/Ineligible	5,890,239	3,918,024	3,602,634	2,362,115	15,773,011
Total Households	32,406,364	33,268,015	31,489,687	30,464,331	127,628,398
Percent Undeliverable	15.38%	10.54%	10.27%	7.20%	11.00%
Household response rate	33.68%	34.98%	34.83%	32.61%	34.05%

6.2 Impact of Stratification on Smoking Outcomes

Table 6-2 presents the weighted smoking rates and unweighted counts of smokers across the four strata of HINTS-FDA2 as well as HINTS-FDA, HINTS 5 Cycle 1, and four cycles of HINTS 4. As described earlier, stratification of the sampling frame coupled with the use of differential sampling rates was designed to increase the yield of current smokers in the sample for analysis purposes. Addresses in the high and medium-high smoking strata were oversampled, while those in the

medium-low and low smoking strata were under-sampled. Smoking rates are defined as number of current smokers⁴ in the sample divided by the number of respondents. Nonrespondents and respondents with reporting errors related to the smoking items were excluded from the calculation. The table shows variability in weighted smoking rates across the strata. For HINTS-FDA2, the overall weighted smoking rate is 14.4 percent, with the highest rate of 19.2 percent for the high smoking stratum and lowest rate of 10.1 percent for the medium smoking stratum.

The weighted proportion of smokers in the HINTS-FDA2 sample was similar to HINTS-FDA and HINTS 5 Cycle 1, all of which were lower than HINTS 4 Cycles 1 through 4. The smoking rate varied in the expected direction across strata with the exception of the lowest smoking stratum which reported the second highest smoking rate. The smoking rate in the high-smoking stratum was consistent with the expected smoking rate for counties included in the stratum (19.2% or greater). The smoking rates for the two middle strata were below their expected ranges (15.9-19.2%; 12.7-15.9%) while the lowest smoking stratum yielded a higher percentage of smokers than the expected rate of below 12.7 percent.

This strategy does not appear to have been any more effective than the allocation used for the HINTS-FDA. Overall, FDA2 yielded a comparable but slightly smaller percentage of current smokers (based on weighted and unweighted estimates). As noted above, the low smoking stratum in FDA2 had the second highest, although this is based on a relatively small number of cases. The rate in the low for FDA2 has an especially wide confidence interval (95% CI: 7.5 to 23.6). The relatively small samples in these strata, as well as the relatively large areas used for stratification (counties) made it difficult to precisely target smokers.

⁴ Respondent is considered a current smoker if they answer 'yes' to smoking 100 cigarettes in lifetime and currently smoking 'some days' or 'every day'.

Table 6-2. Weighted smoking rates across the HINTS-FDA strata and HINTS 4

	% Current smokers, weighted (Standard Error)	n, current smokers	n, % Missing	Total
HINTS-FDA2 overall	14.4% (1.8)	205	47 (3%)	1,736
High smoking stratum	19.2% (3.2)	84	16 (3%)	543
Medium high smoking stratum	13.3% (3.0)	55	17 (3%)	496
Medium low smoking stratum	10.1% (2.5)	41	7 (2%)	411
Low smoking stratum	15.6% (4.1)	25	7 (2%)	286
HINTS 5 Cycle 1	14.9% (1.2)	414	30 (1%)	3,285
HINTS-FDA overall	14.9% (1.0)	496	66 (2%)	3,738
High smoking stratum	24.8% (2.6)	181	15 (1%)	1,038
Medium high smoking stratum	21.4% (2.3)	159	19 (2%)	1,059
Medium low smoking stratum	10.9% (1.3)	137	21 (2%)	1,367
Low smoking stratum	7.1% (2.3)	19	5 (2%)	274
HINTS 4 Cycle 4	15.2% (1.1)	498	49 (1%)	3,677
HINTS 4 Cycle 3	19.4% (1.3)	486	45 (1%)	3,185
HINTS 4 Cycle 2	18.7% (1.4)	586	52 (1%)	3,630
HINTS 4 Cycle 1	17.8% (1.0)	615	82 (2%)	3,959

References

- Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Dillman, D.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.
- Rizzo, L., Moser, R., Waldron, W., Wang, Z., and Davis, W. (2008). *Analytic Methods to Examine Changes Across Years Using HINTS 2003 and 2005 Data*. NIH Publication No. 08-6435.

Appendix A

Cover Letters in English



FIRST MAILING

Dear {City} Resident:

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

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

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

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

Thank you in advance for your cooperation.

Sincerely,

Bradford W. Hesse, Ph.D.

HINTS Project Officer

National Institutes of Health

U.S. Dept of Health and Human Services

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

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

POSTCARD TEXT

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

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

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

Sincerely,



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



SECOND AND THIRD MAILINGS

Dear {City} Resident:

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

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

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

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

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

Sincerely,

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

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

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

Appendix B
Cover Letters in Spanish



FIRST MAILING

Estimado residente de {City}

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

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

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

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

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

Gracias de antemano por su cooperación.

Atentamente,

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

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



SECOND AND THIRD MAILINGS

Estimado residente de {City}:

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

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

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

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

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

Atentamente,

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

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

Appendix C

Frequently Asked Questions (FAQs) English and Spanish

Some Frequently Asked Questions about the Health Information National Trends Survey

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

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

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

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

Q: How did you get my address?

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

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

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

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

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

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

A: About 20 to 30 minutes.

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

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

Q: Who is Westat?

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

Preguntas Frecuentes Encuesta Nacional de Tendencias de Información sobre la Salud

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

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

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

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

P: ¿Cómo obtuvieron mi dirección?

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

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

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

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

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

P: ¿Cuánto tiempo tomará responder las preguntas?

R: Cerca de 20 a 30 minutos.

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

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

P: ¿Quién es Westat?

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

Appendix D

Variable Values and Data Editing Procedures

Missing Value Definitions

Values identifying types of nonresponse or indeterminate responses:

- -1 = Valid skips or appropriately missing data following a dependent question (correctly skipped). Example: If SeekHealthInfo=2 ‘no’ and WhereSeekHealthInfo was missing, WhereSeekHealthInfo was assigned the value -1.
- -2 = Question was answered, but respondent should not have answered the question. The question was answered in error by the respondent. Example: If SeekHealthInfo=2 ‘no’ and WhereSeekHealthInfo was not missing, WhereSeekHealthInfo was assigned the value -2.
- -4 = Question was answered, but data was removed because the entry of the number or character could not be determined (e.g. unreadable or non-conforming numeric response).
- -5 = Respondent selected more response options than appropriate for the question. Example: If TrustDoctor had values 3 ‘a little’ and 2 ‘some’, TrustDoctor 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., TrustDoctor) were missing (e.g., WhereSeekInfo = missing and LotOfEffort = missing, etc...), variables with missing values were assigned the value -6.
- -9 = Missing data. Not ascertained. Question should have been answered, but no response was recorded. Example: If LotOfEffort was missing, it was assigned the value -9.

Data Editing Procedures

Variable	Editing Decision	Description
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.

Variable	Editing Decision	Description
SeekHealthInfo LookedECig UseInternet NoticeHealthInfoInternet AwareQuitlineSite UsedECigEver AnyoneRegulateTobacco BornInUSA TobaccoUserInHH	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 ECigInfoSeek_IMP SizeCigarsSmoked_IMP SexualOrientation_IMP	Imputation for multiple responses	<p>Imputation was carried out when multiple responses were selected, resulting in one unique response for these “mark only one” variables. Respondent’s multiple answers were replaced with a single imputed answer that had the same distribution over the multiple answers as occurred in the single-answer responses. Imputation was not performed on missing values for this question. The suffix “_IMP” indicates that this variable includes imputed values. Flags (indicated by suffix ‘_IFlag’) indicate which values were imputed.</p>
IntRsn_SelfHealthInfo IntRsn_HealthInfoSE IntRsn_InfQuitSmoking IntRsn_SupportGroup IntRsn_DietWebsite IntRsn_HCProviderSearch IntRsn_PDADownload IntRsn_SharedSocNet IntRsn_ExchangedSupport IntRsn_TrackedPHR IntRsn_YouTube	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 Decision	Description
Age_I MaritalStatus_I Education_I Hisp_Cat_I Race_Cat2_I HealthInsurance_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.
SmokeNow SmokeNowCigars	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 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.</p>
ECig_Cat Tobacco_Cat RecentTobacco_Cat TobaccoHeard_Cat TobaccoTried_Cat RegulateTobacco_Cat Hisp_Cat Race_Cat2	Summarized distribution of ‘mark all that apply’ responses	A variable was created to indicate each response selection a respondent made for these ‘mark all that apply’ variables. The derived variable with the suffix ‘_cat’ summarized the response selected or indicated that multiple responses were selected.
Employed Unemployed Homemaker Student Retired Disabled OtherOcc	Derived variables for multiple responses	For the variable OccupationStatus, derived variables were created to indicate each response selected, showing the unique response for respondents selecting one occupation, and showing each response for respondents who did not follow the ‘mark only one’ response instruction.
Education IncomeRanges	Edits for multiple responses	The highest order (e.g., education level or income range) was taken when multiple responses were selected.
R_HHAdults	Derived variable	Responses to questions asking about household size as well as other information about the household (e.g., number of questionnaires returned) were compiled into a derived measure that best represented the number of adults in the household.

Variable	Editing Decision	Description
HHAdults_Num	Imputation for zero and missing responses	Missing values were imputed for the derived count of household adults when the derived variable had values of zero or missing. A flag (indicated by suffix ‘_IFlag’) indicates which values were imputed.
QDisp	Derived variable	A variable was created to indicate the proportion of items respondents answered in the first two sections. This was used to determine incompletely-filled out questionnaires.
WatchTV ListenRadio PersonalInternet WeekendWatchTV WeekendRadio WeekendInternet ReadNewspaper SelfMOB HHAdultMOB[2-5]	Recoding out of range responses	<p><u>WatchTV:</u> Any responses greater than 24 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>ListenRadio:</u> Any responses greater than 24 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>PersonalInternet:</u> Any responses greater than 24 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>WeekendWatchTV:</u> Any responses greater than 48 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>WeekendRadio:</u> Any responses greater than 48 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>WeekendInternet:</u> Any responses greater than 48 hours were recoded to “-4”, which is the code for non-conforming responses.</p> <p><u>ReadNewspaper:</u> Any responses greater than 7 days were recoded to “-4”, which is the code for non-conforming responses.</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>
YearCameToUSA	Responses not in standard 4 digit format reviewed	Responses not given in the standard 4-digit year format were reviewed for scanning accuracy and updated to -4.
Age Age_I Age_IFlag SelfAge	Reviewed for out-of-range responses	Responses were examined for out of range or unlikely ages (those listing their age as < 18 and > 105).

Variable	Editing Decision	Description
MailHHAdults	Reviewed for out-of-range responses	Responses were examined for unlikely household sizes (those listing numbers >20).
MailSurveyTime_Min MailSurveyTime_Hrs	Reviewed for blanks	Responses of 0 or blank were reviewed and recoded to -4 for cases where the respondent listed “0” in both fields, or listed “0” in one field and left the other field blank.
HHAdultAge[2-5]	Review for out-of-range responses	Responses of ages < 18 and > 105 years old were reviewed and updated to -4 as needed.
R_HHAdults ChildrenInHH NumberTobaccoUsers	Review for consistency	Responses reporting more tobacco users in the household than there are adults and children in the household were reviewed. All the values for the pertinent questions were verified in a review of the scanned forms.
SexualOrientation_OS	Review of verbatim responses	Responses of “none of your business” and other similar phraseology were reviewed for scanning accuracy and rephrased as “Refused”.
YearCameToUSA SelfAge	Review for consistency	The responses to YearCameToUSA were compared to the respondent’s reported age to see whether respondents reported coming to the USA in a year that would predate their birth.
UpdateDate	Update for missing values	Due to a receipting issue, return dates for 3 households’ questionnaires were not able to be recorded. These 3 records were assigned an approximate return date based on the mailing from which their returned questionnaire was sent.

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**. Six households had missing values of **R_HHAdults** that needed to be imputed.