Health Information National Trends Survey

hints

Health Information National Trends Survey 4 (HINTS 4)

HINTS-FDA Methodology Report

December 2015

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

Prepared by
Westat
1600 Research Boulevard
Rockville, MD 20850

Westat

Table of Contents

Appendixes

Tables

HINTS - FDA Overview 1

The Health Information National Trends Survey (HINTS) is a nationally-representative survey which has been administered every few years by the National Cancer Institute since 2003. The HINTS target population is adults aged 18 or older in the civilian non-institutionalized population of the United States. HINTS-FDA was a special round of HINTS data collection conducted by the National Cancer Institute (NCI) in partnership with the Food and Drug Administration (FDA) to combine the traditional HINTS topics of health communication, cancer knowledge, and cancer risk behaviors with an assessment of the public's knowledge of medical devices, communications related to product recalls, diet supplement labeling, risk perceptions about new tobacco products, perceptions of tobacco product harm, and tobacco product claims. This report summarizes the methodology, sampling, and procedures of HINTS-FDA. Data cleaning and weighting procedures as well as response rates are also discussed.

HINTS-FDA was conducted from May 29 through September 8, 2015. HINTS-FDA was conducted by mail using a protocol similar to that used in HINTS 4 with a goal of obtaining 3,500 completed questionnaires.

1.1 Oversampling for Smokers

Because of the unique nature of the HINTS-FDA instrument and the specific goals of FDA, the regular sampling strategy of HINTS was altered in an effort to include more current and former smokers in the study. Using data from the Behavioral Risk Factor Surveillance System (BRFSS), county-level smoking rates were used to group addresses into sampling strata of high, medium-high, medium-low, and low smoking rates. The high and the medium-high strata were then oversampled to increase the yield of current smokers. A full description of the sampling strategy can be found in Chapter 2 and the results of the sampling can be found in Chapter 6.

1.2 Identifying Potentially Spanish-speaking Households

As in HINTS 4, households in HINTS-FDA were flagged as potentially Spanish-speaking in one of three ways:

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

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

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

Sample Selection 2

The HINTS-FDA sample used a similar sample design as other rounds of data collection in HINTS 4: a two-stage design where a sample of addresses was selected from a file of residential addresses in the first stage, and one adult was selected within each sampled household in the second stage. The main difference between the two designs is in the stratification of the address frames. The HINTS 4 sample was designed to produce reliable estimates for minority subpopulations and therefore stratified the address frame by minority population. The HINTS-FDA sample, on the other hand, was designed to ensure a large number of current and former smokers in the sample. It used smoking rates to stratify the address frame.

2.1 Sampling Frame

The sampling frame consisted of a database of addresses used by Marketing Systems Group (MSG) to provide random samples of addresses. All non-vacant residential addresses in the United States

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

Rarely are surveys conducted with a sampling frame that perfectly represents the target population. The sampling frame is one of the many sources of error in the survey process. The sampling frame used for the address sample contained duplicate units because some households receive mail in more than one way. To permit adjustment for this duplication of households in the sampling frame, a question about how many different ways respondents receive mail was included on the survey instrument (see question I23).

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

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 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. As shown in table 2-1 below, addresses in census counties that had smoking rates equaling or exceeding 25.1 percent were assigned to the high-smoking stratum. Addresses in census counties that had smoking rates between 21.2 percent and 25.0 percent were assigned to the medium-high smoking stratum. Addresses in census counties that had smoking rates between 15.0 percent and 21.1 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
25.1 or more	High
21.2 – 25.0	Medium High
15.0 – 21.1	Medium Low
14.9% or less	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-FDA was 13,001: 3,566 from the high smoking stratum, 3,831 from medium-high smoking stratum, 4,693 from the medium-low smoking stratum, and 911 from the low smoking stratum. Relative to a proportional design, the high and medium-high smoking strata were oversampled by 60 percent and 20 percent respectively, while the medium-low and low smoking rates strata were under-sampled by 20 percent and 47 percent, respectively.

Table 2-2 below summarizes the address sample for HINTS-FDA, 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-FDA 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	3,566	17.1	27.4	+60.2
Medium-high	3,831	24.6	29.5	+19.9
Medium-low	4,693	45.1	36.1	-20.0
Low	911	13.2	7.0	-47.0
Total	13,001			

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, data collection for HINTS-FDA 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.

3 Data Collection

Data collection for HINTS-FDA started on May 29, 2015 and concluded on September 8, 2015. 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. 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. Most households received one survey per mailing (in English), while households that were flagged as potentially Spanish-speaking received two surveys per mailing (one English and one Spanish).

The second survey mailing was sent via USPS Priority Mail, while all other mailings were sent First Class. Just as in most cycles of HINTS 4, mailings targeted potentially Spanish-speaking households by sending Spanish language materials in the first mailing. The contents of all mailings are further described in Table 3-1 below. Cover letters in English can be found in **Appendix 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	Dates mailed	Mailing method	Materials	Type of recipients
Mailing 1	May 29, 2015	1st Class Mail	English cover letter with FAQs English questionnaire Return envelope $2 bill	All sampled households that were not identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope $2 bill	All sampled households that were identified as possibly Spanish-speaking
Postcard	June 5, 2015	1st Class Mail	Reminder/thank you postcard	All sampled households
Mailing 2	June 29, 2015	USPS Priority Mail	English cover letter with FAQs English questionnaire Return envelope	Non-responding households that were not identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope	Non-responding households that were identified as possibly Spanish-speaking
Mailing 3	July 29, 2015	1st Class Mail	English cover letter with FAQs English questionnaire Return envelope	Non-responding households that were not identified as possibly Spanish-speaking
			English cover letter with FAQs Spanish cover letter with FAQs English questionnaire Spanish questionnaire Return envelope	Non-responding households that were identified as possibly Spanish-speaking

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

Table 3-2. Number of packets per mailing

Mailing	English only	English and Spanish	Spanish only*	Total
Mailing 1	12,151	850	N/A	13,001
Mailing 2	8,609	725	3	9,337
Mailing 3	7,353	660	8	8,021
Upon Request Mailing	0	0	3	3
Total	28,113	2,235	14	30,362

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

3.2 Spanish Language Households

As in HINTS 4, extra efforts were made to elicit responses from Spanish-speaking households. Mailings that included both English and Spanish materials were sent to households who:

- Were flagged as living in a linguistically-isolated area; or
- Were flagged as having a potentially Hispanic surname.

Spanish materials were sent to these households starting with the first mailing. The outer envelope for these households included a message stating 'Please return within two weeks' in both English and Spanish so these households would know without opening the package that the contents were bi-lingual. Households that received bi-lingual materials were sent those materials in the order established in HINTS 4: English cover letter, Spanish cover letter, English questionnaire, Spanish questionnaire. Any household that had received only English materials but called to request Spanish materials then received materials in only Spanish for all subsequent mailings.

3.3 In-bound Telephone Calls

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

The two toll-free lines together received 34 calls throughout the HINTS-FDA field period (see Table 3-3 below). A majority of the in-bound calls were to request Spanish materials. The rest were refusals or respondents calling in with some form of a comment or a question. Sixteen calls were not resolved because the study staff were never able to reach the respondent.

Table 3-3. Telephone calls received

Reason for call	Number of calls received
Request for a Spanish questionnaire	11
Refusal	4
Inquiring about the importance of filling out the questionnaire	1
Calling to let us know the respondent is visually impaired and unable to complete	1
Inquiring about whether the study was still going on	1
Calls that were never resolved	16
Total	34

3.4 Incoming Questionnaires

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

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

Table 3-4. Final household status

Household Status	English	Spanish	TOTAL	
			N	%
Complete	3,697	41	3,738	28.7
Refusal			93	0.7
Undeliverable			1,775	13.7
Nonresponse			7,395	56.9
Total			13,001	100.0

The number of questionnaires returned by date during the field period can be found in Table 3-5 below. The majority of the returns are early in the field period, with 65 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 10 percent were in response to the final mailing.

Table 3-5. Survey response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: May 29	May 30 - June 8	980
Postcard: June 5	June 9 - July 1	1,443
Mailing 2: June 29	July 2 - July 31	920
Mailing 3: July 29	August 1 - September 8	395
	Total	3,738

4 Data Management

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

4.1 Scanning

All completed questionnaires were electronically scanned to capture the survey data and images. Staff reviewed each form as it was prepared for scanning. The review included:

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

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

The reviewed surveys were then sent through the high-speed TeleForm scanner to capture the responses. TeleForm read the form image files and extracted data according to HINTS-FDA 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 about data issues were recorded in a data decision log. The decision log contained the respondent ID, the value triggering the edit, the updated value, and the reason for the update. A total of 6 entries were made into the data decision log during the course of data processing. The majority of these were attributed to illogical responses on a numeric question.

A 10 percent quality control check was then conducted on the scanned data and the electronic images of the data forms. Quality Assurance (QA) staff compared the hard copy questionnaire to the data captured in the database item-for-item and the images stored in the repository page-for-page to ensure that all items were correctly captured. If needed, updates were made. 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. **Appendix D** (Variable Values and Data Editing Procedures) provides a list of the values and their definitions.
- Missing values were recoded for some responses to questions that featured a forced-choice response format and for filter questions where responses to later questions

suggested a particular response was appropriate. **Appendix D** (Variable Values and Data Editing Procedures) provides details about when and how these recodings took place.

- Variables were designed to summarize the responses for the electronic cigarette research, tobacco products research, tobacco product awareness, tobacco product use, tobacco regulation knowledge, dietary supplement reaction symptoms, reaction related dietary supplement, Hispanic ethnicity, and race questions. These variables, ECig_Cat, Tobacco_Cat, RecentTobacco_Cat, TobaccoHeard_Cat, TobaccoTried_Cat, RegulateTobacco_Cat, SuppSymptom_Cat, WhichSupp_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, rules, as described in **Appendix D** (Variable Values and Data Editing Procedures), 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 HINTS 4 surveys. Responses where a missing value of -5 (multiple responses selected) was applied were imputed. This occurred for 365 respondents for WhereSeekHealthInfo, 61 respondents forECigInfoSeek, 1 respondent for-SizeCigarsSmoked, and 3 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-FDA 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 3,814 questionnaires received, 63 were incompletely-answered, and 11 were determined to be ineligible since they were duplicates from the same household. Two were ineligible because the address ended up being a business. A total of 3,738 questionnaires were coded as eligible questionnaires. Details about the eligibility rules are listed below.

Definition of a Complete and Partial Complete Questionnaire

In HINTS-FDA, 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 143 partially complete

questionnaires. Both partially-completed and completely-answered questionnaires were retained. The 63 questionnaires which met the criteria described above for being flagged as incomplete were discarded.

Table 4-1. Completed and partially completed questionnaires

	Definition	Number of Cases	Disposition
Complete	80% of Sections A and C completed	3,595	Included in dataset
Partial Complete	50-79% of Sections A and C completed	143	Included in dataset
Incomplete	49% or less of Sections A and C completed	63	Discarded

Eligibility of Multiple Questionnaires from a Household

Twelve households returned more than one completed questionnaire. The procedures to deal with this issue followed the same guidelines that were used for households using the Next Birthday method in HINTS 4:

- If the same respondent returned multiple questionnaires, the first questionnaire received was retained unless it was a reduced version and a subsequent return was a full version.
- If the same respondent returned multiple questionnaires on the same day, the first questionnaire to complete the editing process was retained.
- If a return date was unavailable for questionnaires from the same respondent that were the same version, 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.[1]
- If, in the above situation, compliance for one or both questionnaires from a household was unclear, the first questionnaire returned was retained.
- If different respondents returned a questionnaire and the ages of household members listed in the roster question were not substantively in agreement, the earliest questionnaire received that complied with the next birthday rule was retained.

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

4.6 Codebook Development

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

5 Weighting and Variance Estimation

Every sampled adult who completed a questionnaire in HINTS-FDA received a full-sample weight and a set of 50 replicate weights. The full-sample weight is the weight which is used to calculate population and subpopulation estimates from the data collected in HINTS-FDA. 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 weighting process encompasses the procedures used to create the final full-sample and replicate weights for the survey respondents. The use of sampling weights is done to ensure valid inferences from the responding sample to the population, correcting for nonresponse and noncoverage biases to the extent possible. The computation of the full-sample weights for HINTS-FDA 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).

Each of these steps is described below.

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

5.1 Household Base Weights

The initial step in the weighting process was calculating the household-level base weight for each household in the sample. The household base weight is the reciprocal of the probability of selecting the household for the survey, which depends on the stratum the household was selected from. Generally, base weights for units in oversampled strata are smaller than those in strata that were not oversampled. In HINTS-FDA, the base weights for households in the high smoking stratum were roughly 1/5, 2/5, and 1/3 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 I23).

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

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

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

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 1.91 to a high of 4.89, and averaged 3.01 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-FDA 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 2014 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 2015 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 2015 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-FDA, a set of 50 replicate weights was assigned to each responding adult. To illustrate how the replicate weights are computed, suppose P is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A nationally representative estimator p can be computed by aggregating the adult sampling weights of all responding adults with this characteristic (e.g., all responding adults in the survey answering the survey question in a particular way). A JK1 jackknife variance estimator of the sampling variance of p can be computed in two steps:

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

Step 2. Compute the JK1 jackknife variance estimator

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

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

subset, and it should mirror the full sample design, as if it were a reduced version of the original sample.

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

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

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

Analysis of Integrated HINTS-FDA and other HINTS Data

The sampling and weighting procedures for HINTS-FDA were designed to facilitate combining HINTS-FDA data with other cycles of HINTS for analysis. Procedures previously documented (Rizzo et al., 2008) describe how to integrate HINTS data from the 2003 and 2005 HINTS surveys. These same procedures can be used to combine the HINTS-FDA data with other cycles of HINTS.

6 Response Rates

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 rate calculation. 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 33.04 percent and this did not vary significantly by strata. The percent of undeliverable households ranged from 15.8 percent to 9.3 percent across strata, with the high smoking and medium-high smoking strata having the highest undeliverable rates (15.8 and 14.9 percent respectively).

Table 6-1. Response rate calculations by Strata

Response class	High	Medium-High	Medium-Low	Low	Overall
Total sample	15,411,756	44,831,517	63,449,213	17,930,276	141,622,762
Respondents	4,486,092	12,392,737	18,481,797	5,392,860	40,753,486
Nonrespondents	8,483,813	25,745,063	37,490,874	10,864,448	82,584,198
Undeliverable/Ineligible	2,441,851	6,693,716	7,476,543	1,672,968	18,285,078
Total Households	12,969,905	38,137,800	55,972,671	16,257,308	123,337,684
Percent Undeliverable	15.84%	14.93%	11.78%	9.33%	12.91%
Household response rate	**34.59%**	**32.49%**	**33.02%**	**33.17%**	**33.04%**

6.2 Impact of Stratification by Smoking

Table 6-2 presents the weighted smoking rates and unweighted counts of smokers across the four strata of HINTS-FDA as well as four cycles of HINTS 4. 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 smokers[1] in the sample divided by the number of respondents. Nonrespondents and respondents with reporting errors were excluded from the calculation. The table shows variability in weighted smoking rates across the strata. For HINTS-FDA, the overall weighted smoking rate is 14.9 percent, with the highest rate of 24.8 percent for the high smoking stratum and lowest rate of 7.1 percent for the low smoking stratum.

The absolute number of smokers who responded was similar to those found for cycle 4 of HINTS. This was somewhat surprising, given the stratification seemed to work as planned. For example, the strata that were oversampled had significantly higher rates of smokers when compared to strata that were undersampled. However, it is the case that the smoking rates for the 'medium low' and 'low' strata were considerably lower than expected. These two strata make up more than half of the overall population (see Table 2-1). More research is needed to fully assess the impact of the stratification when compared to that used for other cycles of HINTS.

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

	% Current smokers (weighted)	n, current smokers	n, % Missing	Total
HINTS-FDA overall	14.9%	496	66 (2%)	3738
High smoking stratum	24.8%	181	15 (1%)	1038
Medium high smoking stratum	21.4%	159	19 (2%)	1059
Medium low smoking stratum	10.9%	137	21 (2%)	1367
Low smoking stratum	7.1%	19	5 (2%)	274
HINTS 4 Cycle 4	15.2%	498	49 (1%)	3677
HINTS 4 Cycle 3	19.4%	486	45 (1%)	3185
HINTS 4 Cycle 2	18.7%	586	52 (1%)	3630
HINTS 4 Cycle 1	17.8%	615	82 (2%)	3959

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

References

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

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

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

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

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

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.

This page left blank intentionally.

Appendix A

Cover Letters in English

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

FIRST MAILING

Dear {City} Resident:

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

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

Your participation is voluntary and your responses will not be linked to your name. We have enclosed $2 as a token of our appreciation for your participation.
You can find out more about HINTS at hints.cancer.gov. Westat, a research firm, will conduct the survey. If you have any questions about HINTS {or if you need more questionnaires}, or if you would like to complete this survey in a language other than English or Spanish, please call Westat toll-free at 1-888-738-6805.

Thank you in advance for your cooperation.

Sincerely,

Bradford W. Hesse, 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

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

SECOND AND THIRD MAILINGS

Dear {City} Resident:

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

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

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

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

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

Sincerely,

Bradford W. Hesse

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

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

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

This page left blank intentionally.

Appendix B

Cover Letters in Spanish

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

FIRST MAILING

Estimado residente de {City}

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

El objetivo de HINTS es averiguar qué información sobre la salud les interesa saber a las personas y dónde tratan de buscarla. Complete este cuestionario para ayudar a averiguar la información sobre la salud que usted necesita y cómo ponerla a disposición suya, de su familia y de su comunidad. **Para asegurarnos de obtener respuestas que contengan un muestreo aleatorio de la población, le pedimos que el adulto en su hogar con el próximo cumpleaños, complete y devuelva este cuestionario en las próximas dos semanas.**

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

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

Atentamente,

Bradford W. Hesse, 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.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
Bethesda, Maryland 20892

SECOND MAILING

Estimado residente de {City}:

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

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

Para asegurarnos de obtener respuestas que contengan un muestreo aleatorio de la población, le pedimos que 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.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

THIRD MAILING

Estimado residente de {City}:

Recientemente lo invitamos a participar en una importante encuesta nacional patrocinada por el Departamento de Salud y Servicios Humanos de Estados Unidos, la Encuesta Nacional de Tendencias de Información sobre la Salud (HINTS, por sus siglas en inglés). El completar esta encuesta nos ayudará a mantenerlos mejor informados en asuntos de salud que usted y su familia consideran importantes.

Si usted ya envió de regreso su encuesta y se cruzó con esta carta en el correo, le agradecemos por el tiempo que se tomó para contribuir al éxito de este estudio.

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

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

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

Atentamente,

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

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

This page left blank intentionally.

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

Q: Who is Westat?

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

Preguntas Frecuentes Encuesta Nacional de Tendencias de Información sobre la Salud

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

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

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

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

P: ¿Cómo obtuvieron mi dirección?

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

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

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

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

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

P: ¿Cuánto tiempo tomará responder las preguntas?

R: Cerca de 20 a 30 minutos.

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

R: El estudio es patrocinado por el Departamento de Salud y Servicios Humanos de EE.UU. El estudio ha sido aprobado por la Oficina de Administración y Presupuesto (OMB, por sus siglas en inglés). Esta oficina revisa todas las encuestas patrocinadas federalmente. El número de aprobación asignado por la OMB para este estudio es 0925-0538.

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:

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., LotOfEffort = missing), variables with missing values were assigned the value -6.

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

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 AnyoneRegulateTobacco SupplementRelatedProblem GivenSuppToChildren BornInUSA TobaccoUserInHH	Recoding filter/skip questions	For these filter questions (questions containing a skip instruction associated with the particular response that was selected), response patterns following the question were examined if the filter question was not answered. The 'yes' value (in the majority of cases where a 'yes' response instructed a respondent to continue answering the subsequent questions) was substituted for the missing filter question when any of the subsequent questions were answered. Similarly (when a 'no' response instructed a respondent to skip subsequent questions), the 'no' value was substituted for the missing filter question when all of the subsequent questions that a 'no' response would have directed the respondent to skip were left unanswered and the respondent answered the next applicable question all respondents were supposed to answer. Please note that if neither condition was met, the missing response code values were retained.
WhereSeekHealthInfo_IMP ECigInfoSeek_IMP SizeCigarsSmoked_IMP SexualOrientation_IMP	Imputation for multiple responses	Imputation was carried out when multiple responses were selected, resulting in one unique response for these "mark only one" variables. Respondent's multiple answers were replaced with a single imputed answer that had the same distribution over the multiple answers as occurred in the single-answer responses. Imputation was not performed on missing values for this question. Flags (indicated by suffix '_IFlag') indicate which values were imputed.
IntRsn_SharedSocNet IntRsn_ExchangedSupport IntRsn_SupportGroup IntRsn_DietWebsite IntRsn_YouTube IntRsn_SelfHealthInfo IntRsn_HealthInfoSE IntRsn_InfQuitSmoking IntRsn_HCProviderSearch IntRsn_PDADownload	Recoding missing responses for items with forced-choice response formats	Respondents were asked to select 'yes' or 'no' to a series of sub-items, allowing them to select as many responses as would apply. These 'forced-choice' response formats sometimes result in respondents indicating which sub-items apply to them by selecting the 'yes' response option for some and leaving the others unanswered. To allow the data to reflect this practice, if respondents did check one or more 'yes' response options within the group, but did not check a 'no' response option for any sub-item in

Variable	Editing Decision	Description
IntRsn_TrackedPHR Supplements_MultiVites Supplements_SingleVites Supplements_Herbs ReportSupp_FDA ReportSupp_CDC ReportSupp_HealthDep ReportSupp_Manufacturer ReportSupp_Doctor		the question, the sub-items that were missing a response were set to 'no.' However, if a respondent, in addition to leaving other sub-items unanswered, did select a 'no' response option for at least one sub-item, the unanswered sub-items were not assumed to be 'no' responses and instead remained missing.
HealthInsurance_I 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.
SmokeNow SmokeNowCigars	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.
ECig_Cat Tobacco_Cat RecentTobacco_Cat TobaccoHeard_Cat TobaccoTried_Cat RegulateTobacco_Cat SuppSymptom_Cat WhichSupp_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	Derived variables for multiple	For the variable OccupationStatus, derived variables were created to indicate each response selected, showing the unique

Variable	Editing Decision	Description
Homemaker Student Retired Disabled OtherOcc	responses	response for respondents selecting one occupation, and showing each response for respondents who did not follow the 'mark only one' response instruction.
Education IncomeRanges	Edits for multiple responses	The highest order (e.g., education level or income range) was taken when multiple responses were selected.
R_HHAdults	Derived variable	Responses to questions asking about household size as well as other information about the household (e.g., number of questionnaires returned) were compiled into a derived measure that best represented the number of adults in the household.
HHAdults_Num	Imputation for zero and missing responses	Missing values were imputed for the derived count of household adults when the derived variable had values of zero or missing. A flag (indicated by suffix '_IFlag') indicates which values were imputed.
QDisp	Derived variable	A variable was created to indicate the proportion of items respondents answered in the first two sections. This was used to determine incompletely-filled out questionnaires.
WatchTV ListenRadio PersonalInternet WeekendWatchTV WeekendRadio WeekendInternet ReadNewsPaper SelfMOB HHAdultMOB2 HHAdultMOB3 HHAdultMOB4 HHAdultMOB5	Recoding out of range responses	**WatchTV:** Any responses greater than 24 hours were recoded to "-4", which is the code for non-conforming responses. **ListenRadio:** Any responses greater than 24 hours were recoded to "-4", which is the code for non-conforming responses. **PersonalInternet:** Any responses greater than 24 hours were recoded to "-4", which is the code for non-conforming responses. **WeekendWatchTV:** Any responses greater than 48 hours were recoded to "-4", which is the code for non-conforming responses. **WeekendRadio:** Any responses greater than 48 hours were recoded to "-4", which is the code for non-conforming responses. **WeekendInternet:** Any responses greater than 48 hours were recoded to "-4", which is the code for non-conforming responses.

Variable	Editing Decision	Description
		ReadNewspaper: Any responses greater than 7 days were recoded to "-4", which is the code for non-conforming responses. **SelfMOB (Respondent's Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses. **HHAdultMOB2 (Second Adult in Household Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses. **HHAdultMOB3 (Third Adult in Household Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses. **HHAdultMOB4 (Fourth Adult in Household Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses. **HHAdultMOB5 (Fifth Adult in Household Month of Birth):** Any responses less than 1 or greater than 12 months were recoded to "-4", which is the code for non-conforming responses.
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 for 6 questionnaires.
SelfAge Age Age_I Age_IFlag	Reviewed for out-of-range responses	Responses of ages < 18 (no cases) and > 105 years old were reviewed. For 1 case that was >105, SelfAge and Age were updated to -4. Imputed age value was set to 78 and the imputation flag for age was set to 1 for "imputation performed".
MailHHAdults	Reviewed for out-of-range responses	Responses of >20 were reviewed and 4 questionnaires were recoded to -4.
MailSurveyTime_Hrs MailSurveyTime_Min	Reviewed for blanks	Responses of 0 or blank were reviwed and recoded for 2 cases. In one case, the respondent had written 0 in both fields and in one case, the respondent had written 0 in hours and left minutes blank.

Variable	Editing Decision	Description
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. For ages <18: HHAdultAge2 needed 29 updates; HHAdultAge3 needed 168 updates; HHAdultAge4 needed 129 updates; and HHAdultAge5 needed 43 updates. There were no ages > 105 listed in any of these variables.
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. For 4 questionnaires, NumberTobaccoUsers was recoded to -4.
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". 8 questionnaires were updated.
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. YearcametoUSA was converted to -4 for one respondent.

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.