



Health Information National Trends Survey (HINTS) 2007

FINAL REPORT

Authors:

David Cantor, PhD
Kisha Coa, MPH
Susan Crystal-Mansour, PhD
Terisa Davis, MPH
Sarah Dipko, MS
Richard Sigman, MS

February 2009

Prepared for:

National Cancer Institute
6120 Executive Boulevard
Bethesda, MD 20892-7195

Prepared by:

Westat
1650 Research Boulevard
Rockville, MD 20850



Table of Contents

<u>Chapter</u>		<u>Page</u>
1	Introduction	1-1
	1.1 Background.....	1-1
	1.2 Mode of HINTS 2007.....	1-1
2	Pretesting Methods and Results	2-1
	2.1 Testing of Advance Materials.....	2-1
	2.2 Pilot Studies	2-2
	2.2.1 RDD Pilot Study.....	2-3
	2.2.2 Mail Pilot Study.....	2-5
3	Instrument Development.....	3-1
	3.1 Questionnaire Development.....	3-1
	3.1.1 Working Groups.....	3-1
	3.1.2 Question Tracking System	3-2
	3.2 CATI Instrument Cognitive Testing.....	3-3
	3.3 Mail Questionnaire Development	3-3
	3.3.1 Mail Cognitive Testing: Round 1.....	3-4
	3.3.2 Mail Cognitive Testing: Round 2.....	3-5
	3.3.3 Mail Cognitive Testing: Round 3.....	3-5
	3.4 Final Instruments.....	3-6
4	RDD Study Design and Operations	4-1
	4.1 Sample Selection.....	4-1
	4.1.1 Size of RDD Sample	4-1
	4.1.2 Stratification by Mailable Status	4-2
	4.1.3 Subsampling of Screener Refusals.....	4-2

<u>Chapter</u>		<u>Page</u>
4.2	Summary of RDD Operations.....	4-3
	4.2.1 Staffing and Training.....	4-3
	4.2.2 Advance Materials	4-4
	4.2.3 Calling Protocol	4-4
4.3	Findings From the CATI Operations.....	4-6
	4.3.1 Weekly Reports	4-7
	4.3.2 Administration Times	4-8
	4.3.3 Average Calls per Case.....	4-9
	4.3.4 Cooperation Rates and Refusal Conversion.....	4-10
	4.3.5 Results of Hispanic Surname Coding.....	4-11
	4.3.6 Data Retrieval.....	4-12
	4.3.7 Imputation	4-14
	4.3.8 Interview Data Processing.....	4-14
5	Mail Study Design and Operations	5-1
	5.1 Sample Selection.....	5-1
	5.1.1 Sampling Frame for Address Sample	5-1
	5.1.2 Selection of Main-Survey Address Sample	5-3
	5.2 Mail Survey Operations.....	5-4
	5.2.1 Questionnaire Mailing Protocol	5-4
	5.2.2 Interactive Voice Response (IVR) Experiment	5-5
	5.3 Findings From the Mail Operations.....	5-6
	5.3.1 Weekly Reports	5-8
	5.3.2 Telephone Contacts	5-9
	5.3.3 IVR Experiment Results.....	5-10
	5.3.4 Survey Processing	5-12
	5.3.5 Imputation	5-12

<u>Chapter</u>		<u>Page</u>
6	Combined Data Set and Accompanying Metadata.....	6-1
	6.1 Combining Data Sets.....	6-1
	6.2 Codebooks	6-1
	6.3 Metadata Development.....	6-2
7	Sample Weights and Variance Estimation Overview.....	7-1
	7.1 Overview of Sample Weights.....	7-1
	7.2 Variance Estimation Methodology for HINTS 2007	7-2
	7.3 Base Weights.....	7-4
	7.4 Nonresponse Adjustment.....	7-5
	7.4.1 RDD Screener Nonresponse Adjustment	7-6
	7.4.2 RDD Extended Interview Nonresponse Adjustment	7-6
	7.4.3 Address-Sample Nonresponse Adjustment.....	7-8
	7.4.4 Replicate Nonresponse Adjustment	7-9
	7.5 Calculation of Composite Weights.....	7-9
	7.6 Calibration Adjustments	7-9
	7.6.1 Control Totals	7-10
8	Response Rates	8-1
	8.1 RDD Sample.....	8-1
	8.1.1 RDD Screener Response Rate.....	8-2
	8.1.2 RDD Extended Interview Response Rate.....	8-4
	8.1.3 RDD Overall Response Rate.....	8-4
	8.2 Address-Sample Response Rate.....	8-5
	8.2.1 Address-Sample Household Response Rate.....	8-5
	8.2.2 Within Household Response Rate	8-6
	8.2.3 Overall Response Rate.....	8-6
	References.....	R-1

<u>Appendixes</u>		<u>Page</u>
A	RDD Pilot Study Letters and Introductions	A-1
B	RDD Main Study Advance Letter.....	B-1
C	RDD Information Request Letter	C-1
D	RDD Screener Refusal Conversion Letter.....	D-1
E	RDD Extended Refusal Conversion Letter.....	E-1
F	Sample of Production Report by Release Group	F-1
G	Sample Weekly TRC Report From NCI.....	G-1
H	Mail Advance Letters, Cover Letters, and Postcards	H-1
I	Decisions for Combining CATI and Mail Data.....	I-1
<u>Tables</u>		<u>Page</u>
2-1	RDD pilot test sample size.....	2-3
2-2	Incentive/mail mode treatment combinations.....	2-5
2-3	Mail pilot field period schedule	2-6
2-4	Household-level response rates by incentive and mail method.....	2-7
2-5	Average proportion of questionnaires returned per household	2-7
4-1	Unweighted RDD sample by mailable status	4-2
4-2	Unweighted RDD sample results by mailable status.....	4-6
4-3	Weekly TRC production: Completed cases by week.....	4-8
4-4	Total screener level of effort: Number of call attempts by result	4-9

<u>Tables</u>		<u>Page</u>
4-5	Total extended (CATI) level of effort: Number of call attempts by result	4-10
4-6	Residential, cooperation, refusal conversion, and response rates and yield by mailable stratum, for screener and extended interviews	4-11
4-7	Data retrieval calls	4-13
5-1	Mail survey schedule and protocol.....	5-5
5-2	Household cooperation in the mail survey.....	5-6
5-3	Household response by week	5-7
5-4	Household response by mailing and strata	5-7
5-5	IVR calls.....	5-10
5-6	Live interviewer prompt calls.....	5-11
5-7	Household response by treatment in IVR experiment	5-11
8-1	Weighted estimates of percentages of residential telephone numbers that are residential in the HINTS 2007 RDD sample	8-3
8-2	Screener response rate calculations for the HINTS 2007 RDD sample.....	8-3
8-3	Extended interview response rate calculations for HINTS 2007 RDD sample.....	8-4
8-4	Overall response rate calculations for HINTS 2007 RDD sample.....	8-4
8-5	Household response rate calculations for the HINTS 2007 address sample	8-5
8-6	Weighted within-household response rate calculations for HINTS 2007 address sample	8-6

8-7 Overall response rate calculations for HINTS 2007 address
sample..... 8-6

The National Cancer Institute's (NCI's) Health Information National Trends Survey (HINTS) collects nationally representative data about the U.S. public's use of cancer-related information. This study, increasingly referenced as a leading source of data on cancer communication issues, was developed by the Health Communication and Informatics Research Branch (HCIRB) of the Division of Cancer Control and Population Sciences (DCCPS) as an outcome of NCI's Extraordinary Opportunity in Cancer Communications. HINTS strives to: provide updates on changing patterns, needs, and information opportunities in health; identify changing health communications trends and practices; assess cancer information access and usage; provide information about how cancer risks are perceived; and offer a test-bed to researchers to investigate new theories in health communication. HINTS data collection is conducted every 2-3 years in order to provide trends in the above areas of interest. This report presents a summary of the third round of HINTS data collection known as HINTS 2007.

1.1 Background

The first round of HINTS, administered in 2003, used a probability-based sample, drawing on random digit dialing (RDD) telephone numbers as the sample frame of highest penetration at that time. Due to an overall decline in RDD rates, the second cycle of HINTS, HINTS 2005, included embedded methodological experiments to compare data collected by telephone with data collected through the Internet. In addition, the field study explored the impact of various levels of incentives on response rates. Unfortunately, providing respondents with an Internet alternative, a monetary incentive for nonresponse conversion, and having an operations priority on nonresponse conversion were not successful in reducing the impact of falling response, and the overall response rate for HINTS 2005 was lower than expected.

1.2 Mode of HINTS 2007

In an effort to address dropping RDD response rates, NCI turned to work done at the Centers for Disease Control and Prevention (CDC) on the Behavioral Risk Factor Surveillance System (BRFSS).

BRFSS data collection has recently included experiments with mail surveys and mixed mode data collection (mail and telephone). Recent research by Link and colleagues (2008) suggests that use of a mail survey, with appropriate followup, can achieve a higher response rate than RDD alone. One experiment (Link & Mokdad, 2004) found that a mail survey led to significantly more responses than a web survey (43% vs. 15%), and that a mail survey with a telephone followup produced a significantly higher response rate than a RDD telephone survey (60% vs. 40%).

Following the model provided by BRFSS, HINTS 2007 used a dual-frame design that mixed modes in a complementary way. One frame was RDD, using state-of-the-art procedures to maximize the response rate. The second frame was a national listing of addresses available from the United States Postal Service (USPS). This list is relatively comprehensive (Iannacchione et al., 2003) and includes both telephone and nontelephone households. These households were administered a mail survey. The study was designed to complete 3,500 interviews with the RDD and 3,500 from the USPS frame. National estimates were developed by combining the two frames using a composite estimator.

There are a number of advantages of this dual-frame design. One is that using two modes offers the potential for improving coverage over a design that exclusively relies on RDD. In addition to landline telephone users, the use of the USPS frame also allows for the coverage of mobile-only telephone users and those without a telephone. This directly addresses the increasing difficulty RDD surveys have with reaching those who do not regularly use a landline telephone. There is also the possibility of improved measurement for a number of characteristics (e.g., those subject to social desirability bias). Moving to a dual frame leaves open the opportunity to implement other modes in the future if they are found to be appropriate.

Link and Mokdad (2004) report that unit response rates between the two modes for their experiment with the BBRFSS were generally equivalent. An important issue discussed was the tendency for mail respondents to have characteristics associated with higher socioeconomic status, such as higher income, majority race, and higher education. This finding is consistent with other studies that have examined characteristics of nonrespondents to mail surveys (e.g., Hauser, 2005). The design of the HINTS mail survey was developed to maximize response rate while minimizing the potential for nonresponse bias. In addition, experiments with incentives and delivery methods were conducted in an attempt to decrease the different nonresponse bias patterns that emerge for mail surveys (i.e., lower response rates by levels of education and minority status).

Before fielding HINTS 2007, advance materials were tested and pilot tests were conducted to refine the methodology in an effort to achieve the best possible response rates and data quality. These tests guided the finalization of the study design used for the data collection effort. This chapter describes the objectives of the focus groups and the pilot tests that were conducted, the results of these tests, and the approach that resulted from the tests.

2.1 Testing of Advance Materials

Notification letters received by potential respondents prior to telephone contact have been shown to improve response rates (e.g., Hembroff et al., 2005). Although respondents to HINTS 2005 were sent advance letters and materials, the format and content of these materials were not examined to determine whether they were optimal for encouraging study participation. Therefore, a primary goal of HINTS 2007 pretesting was to develop notification letters that focus group participants found meaningful and motivating.

A Westat-led brainstorming session with NCI investigators, held in August 2006, created the groundwork for the materials that would be reviewed by the focus groups. Investigators reviewed the advance materials used in previous HINTS data collection efforts and other similar studies directed by Westat from which they then generated ideas for HINTS 2007 materials.

Materials developed as a result of the brainstorming meeting were tested in four focus groups conducted in the fall of 2006. A total of 38 individuals living in the Rockville, Maryland, area participated. The participants were recruited from Westat's database of study volunteers. Each focus group was made up of 9 to 10 members and each individual was paid \$75 as an incentive for participating in a session lasting 90 to 120 minutes.

Each group was moderated by a Westat staff member using a semi-structured discussion guide. Participants were asked to react to multiple versions of advance letters as well as various introductions that could be used by HINTS telephone interviewers. Two groups focused on materials designed for the mail sample and two groups focused on materials designed for the RDD

telephone sample. Reactions to potential follow up mailings, designed for people who had not cooperated with prior requests for survey participation (e.g., refusal conversion letters for the telephone sample), were also obtained from two groups.

Observations from the focus groups suggested a number of ways to maximize response rates for HINTS 2007. Changes were made to many of the materials in response to the focus group comments. In addition, some materials and scripts were selected for further testing in the pilot test. Decisions resulting from the focus groups include the following:

- **Advance Letter.** Two versions of an advance letter were presented to the focus groups. One letter included factoids (brief findings from a previous survey administration) and the other version did not. Letters that included factoids appeared to be better received than those without. Further testing of the impact of both letter versions on participant response were conducted during the pilot study.
- **Frequently Asked Questions (FAQs).** Notification letters that included FAQs on the reverse side were better received by focus group participants than those without. Therefore, notification letters used in HINTS 2007 included the FAQs.
- **Refusal Conversion Letter.** The focus groups suggested that the refusal conversion letter could easily be interpreted as harsh or scolding in tone if not carefully worded. Accordingly, refusal conversion letters used in HINTS 2007 were shortened and softened.
- **Study Sponsorship.** The focus groups strongly indicated that identifying the U.S. Department of Health and Human Services (DHHS) as the sponsor rather than NCI would be a better approach from the standpoint of maximizing response rates. All participants recognized DHHS as being a Federal Government agency, while few recognized NCI as such. Furthermore, participants suggested that for people not particularly concerned about cancer, a reference to NCI may result in less interest in participating in the survey. For HINTS 2007, DHHS was identified as the study sponsor on all printed materials and in the telephone introduction.
- **Telephone Introduction.** The focus groups indicated that the introduction for telephone surveys must be short and immediately get to the purpose of the call. Two possible telephone introductions were identified. The impact of these introductions on cooperation rates were tested during the pilot study.

2.2 Pilot Studies

Before the full field study, Westat conducted pilot studies of both the RDD and mail methodologies. The pilot studies used the procedures intended for the full field effort to test the operations and

systems. The pilots also tested the impact of study material on respondent understanding and cooperation rates. A summary of the pilot studies and resulting changes to the study design are provided in the following sections.

2.2.1 RDD Pilot Study

One purpose of the RDD pilot study was to test the operations and systems to be used for the main study. The RDD pilot was designed to:

- Identify problems with the computer-assisted telephone interview (CATI) programming of either the screener or extended instrument;
- Determine the average amount of time needed to complete the CATI instrument; and
- Identify any problems with specific questionnaire items that needed revision for the field study or required additional training of interviewers.

The RDD pilot also included an embedded experiment to test the impact of advance letters and introductions on cooperation rates. Respondents were randomized to one of four conditions in which they received one of two versions of the pre-notification letter and one of two versions of the CATI screener introduction. Letters differed by either providing a summary of aspects of the study or a set of bullets highlighting previous results of the study. Introductions differed in that one characterized the study as a “national study on people’s needs for health information” while the other characterized it as a “national health study.” These letters and introductions can be found in Appendix A.

The RDD pilot was conducted from September 24 through October 15, 2007. The sample size of the RDD pilot test was 1,000 households, with 250 cases in each of the four experimental treatments (see Table 2-1).

Table 2-1. RDD pilot test sample size

	Letter A	Letter B
Introduction A	250	250
Introduction B	250	250

Because the advance letter was being tested in the pilot, only people who had addresses tied to their telephone numbers were included in the initial sample file. Refusal conversion was not conducted and no incentive was included with the advance letters.

Following the RDD pilot study field period, a 1-hour debriefing was held with interviewers. The purpose of the debriefing was to gain interviewer feedback on the following:

- Problems with individual items or sections (either respondents having difficulty answering questions or interviewers having difficulty reading questions);
- Reactions to the introductions and the screener as a whole; and
- Items requiring additional training, such as more help text or guidance on how to deal with certain responses.

Both project staff and NCI investigators attended the debriefing.

RDD Pilot Results

No CATI programming problems were identified during the pilot study. There were issues with specific questionnaire items identified from both the actual data collection activities and the interviewer debriefing. These are discussed in Section 3.4 along with a broader discussion of the RDD instrument.

The average time needed to complete the CATI instrument during the pilot test was 40.12 minutes. This was approximately 10 minutes longer than the 30-minute target time. As a result, 30 items were deleted to shorten the instrument for the main study. These changes are discussed in more detail in Section 3.4.

In the embedded experiments of the advance letter and introductory text, neither yielded statistically significant results. For the letter, the response rates were 29.0 percent (Letter A) and 25.4 percent (Letter B). For the introductions, the response rates were 27.9 percent (Introduction A) and 26.5 percent (Introduction B). Based on the reaction of the focus groups, letters containing bulleted facts were employed for the main data collection effort. Both introductions to the CATI screener were made available to the interviewers on the CATI introduction screen, allowing interviewers to select whichever they felt would be the most appropriate for a particular respondent.

2.2.2 Mail Pilot Study

One purpose of the mail pilot study was to test the operations and systems required to accomplish the postal portion of the main study. The mail pilot was designed to:

- Identify problems with the paper version of the HINTS 2007 instrument;
- Test the tracking system to ensure that both households and individual questionnaires were appropriately monitored throughout the field period; and
- Test the scanning of the instruments being done through a scanning subcontractor to ensure that systems were adequate and that the data returned to Westat were appropriate.

In addition to the focus described above, the mail pilot study contained three embedded experiments. The first two experiments were designed to determine the impact of incentives and mailing vehicle on response rates. The sample was randomized to either receive a \$2 incentive or no incentive with the initial mailing of the instrument and randomized to receive the second mailing of the instrument either via USPS or Federal Express (FedEx). These experiments consisted of 640 cases with four treatment combinations (see Table 2-2).

Table 2-2. Incentive/mail mode treatment combinations

Mail mode	Incentive	
	\$0	\$2
USPS	160	160
FedEx	160	160

The third experiment evaluated the impact of mail questionnaire length on response rates and data quality. Half of the households received a questionnaire that was 20 pages long (the long questionnaire), and the other half received a questionnaire that was 15 pages long (the short questionnaire).

The timeline for the mail component of the pilot was shorter than the timeline planned for the full fielding of the study in order to complete the pilot within the limited time available. The specific schedule for the mail pilot can be found in Table 2-3. Selected households were sent a letter introducing the study and explaining the questionnaire mailing they would receive. Two days following the mailing of the introductory letter, a package with three questionnaires was mailed to

households with instructions for each adult in the household to complete a questionnaire. One week following the initial mail out, a reminder postcard was sent to households from which no questionnaires had been received. One week after postcards were sent, a second mailing of three questionnaires was sent to all households from which no questionnaires had been received. One week after the second questionnaire mailing (4 weeks after the initial mailing), a sample of nonresponding households for which telephone numbers were available were contacted by telephone interviewers to complete the telephone version of the instrument. In comparison to the main study, this schedule considerably shortened the time between mailings.

At the close of the field period for the pilot study, all completed questionnaires were sent to the scanning subcontractor in order to test the accuracy and speed of the scanning process.

Table 2-3. Mail pilot field period schedule

Date	Activity
August 23, 2007	Advance letters sent to all households in the mail survey
August 27, 2007	First set of questionnaire packets mailed to all households
September 3, 2007	Reminder postcards sent to nonresponding households
September 10, 2007	Second set of questionnaire packets mailed to nonresponding households
September 24, 2007	Nonresponding households sent to TRC for CATI interview
October 15, 2007	All mail cases finalized and no additional questionnaires accepted.

Results of the Mail Component Pilot Test

Some issues with the paper instrument were identified during the pilot testing. These problems and resulting changes were primarily related to skip patterns embedded in the instrument and are outlined in greater detail in Section 3.4.

The tracking and scanning systems were also tested during the pilot test. Both worked well and required only minor changes in preparation for the main study.

Both the incentive and mailing method treatments significantly increased the return of the mail survey. As noted in Table 2-4, each of these treatments increased the household-level response rate by approximately 10 percentage points. The two treatments seemed to complement each other. When each was applied separately, the household-level response rate increased from 22 percent to

31 percent. When both were used together, the response rate increased an additional 10 points to 41 percent.

Table 2-4. Household-level response rates by incentive and mail method

	\$2 incentive—%	No incentive—%	Total—%
FedEx	41.1	30.9	36.1
USPS	31.0	21.8	26.3
Total	35.8	25.9	

The experiment indicated that the FedEx treatment was also more effective at increasing the within-household response rate. This is illustrated in Table 2-5, which shows the mean percentage of questionnaires returned for households. The first column provides the data for all households, including one-person households. The second column is restricted to households with at least two adults. There is no difference for either the incentive or FedEx when looking at all households. Similarly for households with at least two adults, the incentive does not affect response rates (74.4 vs. 74.9). However, in households with two or more adults, FedEx did seem to make a difference (77.6 vs. 70.0). This difference is not statistically significant ($p < .13$; two-tailed test), but the sample sizes for this test were relatively small.

Table 2-5. Average proportion of questionnaires returned per household

		All households	Households with at least two adults
Incentive	None	82.6	74.9
	\$2	84.5	74.4
Mail mode	FedEx	84.3	77.6
	USPS	83.0	70.2

As a result of the experiment, the use of both the incentive and FedEx treatments were adopted for the full sample in the main study.

There was no difference in response rates for the two different questionnaires that were sent (short vs. long). Both had a response rate of 30.8 percent. NCI opted to shorten the longer version of the mail questionnaire to keep it in line with the shortened version of the CATI questionnaire discussed earlier.

During the pilot study, telephone interviewers attempted to contact a sample of nonresponding households for which telephone numbers were available to complete the telephone version of the instrument. The response rate from the telephone followup was low (3.85%). As a result, it was decided that telephone followup to the mail questionnaire would be eliminated from the design for the main data collection effort. As an alternative, Westat proposed an embedded experiment using IVR (interactive voice recording) telephone reminders to complete the mail questionnaire 2 weeks after the second questionnaire mailing to all nonresponders. This experiment is described in Section 5.2.2.

One of the primary goals for HINTS 2007 was to preserve the methodological integrity of the survey. To this end, Westat worked closely with NCI and the HINTS stakeholders to develop the content of the HINTS instrument, ensuring that key concepts were appropriately represented in both modes of the survey.

3.1 Questionnaire Development

The development of the HINTS 2007 instrument began with NCI investigators and HINTS stakeholders completing a survey to identify important constructs to be assessed in the HINTS 2007 instrument. Constructs fell into the following categories:

- Health communication;
- Cancer communication;
- Cancer knowledge, cognitions, and affect;
- Cancer screening/cancer-specific knowledge and cognitions; and
- Cancer-related lifestyle behaviors/cancer contexts.

Stakeholders rated the priority of each construct based on a standard set of criteria. They also had an opportunity to recommend additional constructs that they felt should be captured in HINTS 2007.

3.1.1 Working Groups

Based on the results of this survey, NCI established working groups to develop and identify survey questions for the HINTS 2007 priority constructs. The following workgroups were formed:

- Health communication;
- Health services;

- Cancer screening;
- Cancer cognition;
- Energy balance (physical activity and diet);
- Tobacco use;
- Complementary and alternative treatments;
- Sun safety; and
- Health status and demographic characteristics.

Westat provided NCI with a matrix of the HINTS 2003 and 2005 items to assist in the selection of questions for HINTS 2007. The matrix included question wording, response options, and year(s) that the question was asked, so that the working groups could identify questions from previous iterations of HINTS that should be asked.

Each working group submitted a pool of possible survey items for their sections. NCI's HINTS management team developed the framework for the questionnaire, sorting the questions into five main sections:

1. Health communication;
2. Health services;
3. Behaviors and risk factors;
4. Cancer; and
5. Health status and demographics.

3.1.2 Question Tracking System

Westat staff compiled the items into an Access database question tracking system, a repository where the following information about questions was stored: question wording, response options, section, variable name, whether they were included in HINTS 2003 and/or HINTS 2005, mode, whether they underwent cognitive testing, and a description of any changes made to questions during the instrument development process. The question tracking system was maintained and updated throughout HINTS 2007 to document decisions about item deletions, additions, and

revisions. The question tracking system also provided reports that served as the basis for the development of the metadata tables discussed in Section 6.3.

3.2 CATI Instrument Cognitive Testing

Westat conducted three rounds of cognitive interviews as part of the development of the CATI instrument. The interviews were conducted in the focus group facility at Westat by project staff. Interviewers adhered to a semistructured protocol for conducting the interviews. Staff asked selected sections of the instrument and frequently probed respondents' comprehension of questions as well as any observed difficulties. The interviews were audiotaped and then closely reviewed by staff conducting the interviews. Nine Rockville, Maryland, area volunteers participated in each round of cognitive interviews. Each respondent received \$30 for their participation in a 1-hour interview. Westat staff summarized the results of each round of cognitive testing and provided recommendations to NCI about specific items and sections of the instrument. As a result of the first round of cognitive testing, 2 questions were deleted, 45 questions were altered, and 7 questions were added. As a result of the second round of cognitive testing 1 question was deleted, 6 questions were altered, and 1 question was added. As a result of the final round of cognitive testing, 9 questions were altered.

After revisions were made to the instrument based on the cognitive interview findings, Westat project staff conducted several rounds of the revised interview with volunteer family and friends to obtain preliminary timings for the administration of the instrument. This timing data, although not exact, provided insight into which sections of the instrument could be anticipated to take longer to administer than others.

Based on the cognitive testing, timed interviews, and discussions during internal NCI meetings and retreats, changes to the instrument were finalized to create the version of the CATI instrument used in the RDD pilot study described in Section 2.2.1.

3.3 Mail Questionnaire Development

Once items to be incorporated into the CATI HINTS 2007 instrument were finalized for the pilot test, development of the mail questionnaire began. Items included in the mail questionnaire were

similar to those included in the CATI, but reworded, as necessary, to reflect self-administration. In some cases, different questions to measure similar constructs were used for the mail and CATI instruments. The Dillman double-column approach was employed for the formatting of the mail instrument (Dillman, 2000). Selected sections from the mail instrument underwent three rounds of cognitive testing. The first two rounds focused on the format of the survey, while the last round focused on selecting an appropriate survey cover. Nine Rockville, Maryland, area volunteers participated in each round of testing and each volunteer was paid a \$30 incentive for participating in a 1-hour interview.

3.3.1 Mail Cognitive Testing: Round 1

The major goals of the first round of cognitive testing were to ensure that: (1) respondents could easily follow the skip pattern instructions; and (2) question wording and format were appropriate for self-administration. Reactions to the anticipated mail package as a whole were also assessed.

The participants filled out most sections of an 18-page, booklet-style questionnaire with double-sided pages, very similar to the format anticipated for the mail survey. In selecting sections for the cognitive interviews, those presenting skip instructions and items with somewhat unusual formatting or response requirements (e.g., requiring numeric entries along with indicating units such as minutes or hours) were prioritized.

Participants were asked to read and fill out the instrument on their own. They were also asked to read aloud as they completed the instrument to help assess the items that they were attending, the items that they overlooked, the difficulty of instructions, etc. Westat staff conducting the interview did very little probing—instead they focused on closely observing the participants while noting any difficulties or problems with responding.

Based on the findings from the first round of cognitive testing for the mail instrument, the following revisions were made to the formatting of the mail instrument:

- Skip instructions were changed from italics to bold;
- Indentation of items was eliminated;
- Introductions to items presented in grids were reworded to better communicate that the respondent should answer each item in the series;

- The format for questions where unit was an issue was altered (e.g., separate entry spaces for minutes and hour); and
- Font size was increased, which increased the number of pages from 18 to 20.

3.3.2 Mail Cognitive Testing: Round 2

The objectives of the second round of cognitive testing for the mail instrument were to: (1) assess the ease/accuracy of following skips and handling various item formats; (2) obtain the time required to complete the instrument (participants filled out almost all of the instrument and were asked to read to themselves, rather than aloud); and (3) obtain further reactions to the mail package and a draft cover with photos.

The format was greatly improved between the first and second rounds of cognitive testing. Skips were overlooked less frequently, and there was almost no missing data. The time to complete the survey varied from 21 minutes to 40 minutes; however, it should be noted that not all sections of the instrument were completed, so the instrument was longer than anticipated.

Since the length of the mail instrument was a concern, the effect of instrument length on response rate was tested during the mail pilot. Working group leaders were asked to identify questions that they would consider cutting to develop the short version of the instrument to be used in the pilot as described in Section 2.2.2.

The impact of the cover of the instrument was another factor explored during the second round of cognitive testing. The connection between health and the photos was not apparent to all respondents. Therefore, Dillman's general suggestion of not including photos on mail instrument covers was followed (Dillman, 2000).

3.3.3 Mail Cognitive Testing: Round 3

The third round of cognitive testing explored participants' responses to three different versions of the cover. Participants were asked to rate which cover best represented each of a series of attributes, such as most government looking, most commercial, most trivial, etc. Using the findings of this round of cognitive testing, a cover was developed that capitalized on the "government looking"

cover, since official looking covers have been found to result in higher response rates (Dillman, 2000), while softening some of the criticisms of that cover.

Following the third round of cognitive testing, the long and short versions of the mail instrument for the pilot were finalized.

3.4 Final Instruments

Following the pilot study, Westat worked closely with NCI to identify final cuts and edits to the instrument without taking out high-priority items in an attempt to reduce the length of the instruments and maintain the consistency across both modes.

Although results from the mail pilot indicated that there was no difference in response rates for the short and long mail questionnaires, NCI opted to shorten both the mail and CATI questionnaire for the main fielding to reduce the length of each to approximately 30 minutes. The basis for the revised instruments was the short version of the mail instrument, since working group leaders had previously agreed that items not included in the short instrument were possible candidates for deletion.

To assist NCI in making the final revisions to the instruments, Westat delivered question-by-question timings and frequencies. NCI also participated in a debriefing with interviewers who conducted the pilot test to obtain feedback on the administration of the instrument. Interviewers indicated items that seemed to be problematic for respondents and items that were difficult for them to code. Comments from the interviewers influenced the alteration of 9 items.

Although the goal was to maintain consistency across both modes as much as possible, some mode-specific cuts were made to the mail instrument based on an analysis of skip patterns that showed either erroneous skipping or erroneous marking of responses during the pilot study. This analysis highlighted both questions and formats for which this was especially problematic, and 5 additional questions were cut from the mail instrument.

The instruments were finalized approximately 2 months before the main fielding. The final CATI instrument contained a total of 201 items and the final mail instrument contained a total of 189 items. No single respondent was asked all questions.

This chapter summarizes the approach for the RDD component of HINTS 2007, including the sample design and the data collection protocol procedures. The chapter concludes with a description of cooperation to the RDD survey, contacts made by respondents, and other details about the RDD operations conducted.

4.1 Sample Selection

CATI data collection for HINTS 2007 used a list-assisted RDD sample. A list-assisted RDD sample is a random sample of telephone numbers from all ‘working banks’ in U.S. telephone exchanges (see, for example, Tucker, Casady, & Lepkowski, 1993). A working bank is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits) with at least one listed residential number.¹

4.1.1 Size of RDD Sample

A total of 88,530 telephone numbers were sampled. Tritone and business purging was then used to remove unproductive numbers (i.e., business and nonworking numbers). The procedure, called Comprehensive Screening Service (CSS), was performed by Market Systems Group (MSG), the vendor that provided the sampling frame. In CSS, telephone numbers are first matched to numbers in the White and Yellow Pages to identify business numbers. A second procedure, a tritone-test, identifies the nonworking numbers. A telephone number is classified as a nonresidential number if a tritone (the distinctive three-bell sound heard when dialing a nonworking number) is encountered in two separate tests. Following the CSS processing, the numbers that were not identified as nonworking or nonresidential were sent for address matching. Of those telephone numbers, 25,655 had addresses and the remaining 62,875 did not. Subsampling selected 54,576 numbers (86.8%) of the no address cases.

¹ Note that all numbers, whether listed as residential or not, are part of the sampling frame, as long as they are in working banks.

Table 4-1. Unweighted RDD main sample by mailable status

	Mailable		Nonmailable*		Total
	Total	Percent of total	Total	Percent of total	
Original numbers	17,101	32.2	36,017	67.8	53,118
Residential numbers (estimated)	13,986	87.6	1,986	12.4	15,972
Unweighted residency rate	81.8%		5.5%		30.1%

* Includes nonworking and nonresidential telephone numbers.

The resulting 80,231 telephone numbers were partitioned into a main sample and a reserve sample. The main sample consisted of approximately two-thirds of these telephone numbers (53,118), while the reserve consisted of the remainder (27,113). The reserve sample was set aside to be used in case our expectations for 3,500 completes were not met in working the main sample. Table 4-1 presents the sample sizes of the mailable and nonmailable strata for the RDD main sample. The stratification by mailable status is discussed in Section 4.1.2.

4.1.2 Stratification by Mailable Status

Table 4-1 above shows that in HINTS 2007, 32.2 percent of the main RDD sample was mailable and that 67.8 percent was nonmailable. This table also shows that although the mailable stratum is smaller in size, it contains the majority of the total estimated residences.

4.1.3 Subsampling of Screener Refusals

After the selection of a sample of telephone numbers, the remaining working residential numbers were released in batches for calling by Westat's Telephone Research Center (TRC). Telephone numbers were assigned at random to the batches so that each batch was representative of the universe of working residential telephone numbers. The subsampling of screener second refusals was implemented by excluding from the second refusal conversion cases the nonhostile screener refusals in the last two batches of the main telephone sample. This resulted in 65.4 percent of the screener second refusals being assigned to a second refusal conversion attempt. This subsampling excluded 11,804 main sample telephone numbers from the second refusal conversion process, resulting in the remaining telephone numbers receiving full (first and second) refusal conversion.

4.2 Summary of RDD Operations

The RDD component of the main data collection effort was conducted from January 7 through April 27, 2008. The following sections summarize the staffing and training and the procedures used for the RDD study including the calling protocol, related mailings, refusal conversion activities, and processing interview data. Additional detail about these procedures can be found in the *HINTS 2007 Operations Manual* dated January 2008.

4.2.1 Staffing and Training

The HINTS 2007 data collection was staffed with data collectors hired and trained by the Westat TRC. The study was staffed mainly with experienced RDD interviewers, complemented by a smaller number of newly hired staff. Approximately three-fourths of interviewing and supervisory staff for this data collection effort were home-based.

Project-specific training was developed by study staff and consisted of interviewer and trainer reference materials available online through a learning management system and a specific training agenda that included lectures, interactive sessions, and dyad role plays. Specific attention was paid to contact procedures, and the training program emphasized gaining the cooperation of respondents in the first few moments of the telephone attempt. All training was completed online, including 3.5 hours of self-paced material covering the study purpose, sponsors, and questionnaire, followed by a 2-hour WebEx session hosted “live” by a trainer, covering contact procedures and the questionnaire. Project training concluded with 2.5 hours of role plays, in which interviewers were paired up and alternated serving as respondent and interviewer, using scripted example interviews.

A total of 52 interviewers completed training. Most of the interviewers participated in one of the first three trainings conducted between January 9-15, 2008. A small training to account for attrition was held 2 weeks later, yielding five additional interviewers. The first 26 to complete training were available to start interviewing on January 14, the first day of data collection. An additional 21 trainees were available to start by January 16. There were 22 active interviewers during the first week of data collection, 39 during the second week, and by the third week 48 interviewers were actively working.

Instruction of bilingual interviewers in Spanish was completed during the initial training session by pairing up bilinguals for role play practice in the Spanish instrument. Spanish-language FAQs were also provided to these interviewers. It was important to begin Spanish language interviewing immediately, as the Hispanic surname coding procedure described in Section 4.2.3 had isolated a group of cases for initial release specifically to bilingual interviewers.

During the course of the data collection effort, telephone interviewer supervisors and other project staff continued to monitor individual interviewers. Ten percent of each interviewer's work was routinely observed to ensure the continued quality and accuracy of their work.

4.2.2 Advance Materials

Sampled households with address matches were sent a letter approximately 1 week prior to being called by an interviewer to do the screening interview. The letter alerted the household that an interviewer would call them and provided information about the study, including FAQs on the reverse side of the letter (see Appendix B). A \$2 incentive was included with the advance letter.

4.2.3 Calling Protocol

Interviewers were assigned to work on the study during TRC operating hours; that is, 9 a.m. to midnight on weekdays, 10 a.m. to 6 p.m. on Saturdays, and 2 p.m. to 10 p.m. on Sundays. Respondents were called only between 9 a.m. and 9 p.m. within their own time zones, unless they specifically requested an appointment at another time.

Interviews were conducted in either English or Spanish. If a respondent either requested to complete the interview in Spanish or if the interviewer determined that the respondent spoke only Spanish, the case was transferred to a bilingual interviewer. The bilingual interviewers conducted interviews in Spanish or went back and forth between English and Spanish as necessary.

Hispanic Surname Coding

In an effort to increase participation by Hispanic respondents in general, and specifically with those who are Spanish-speaking, a new procedure was employed for HINTS 2007. Sampled telephone

numbers that were matched to mailing addresses with surnames were compared to a Census list of surnames. Sampled telephone numbers corresponding to surnames that were Hispanic more than 75 percent of the time in the 2000 Census were flagged and loaded directly into a “Priority Hispanic” work class staffed by bilingual interviewers. This allowed the first contact with these sampled households to be made by someone who could easily transition to Spanish if needed. Results of this coding procedure are described in Section 4.3.5.

Information Requests

During the TRC calling process, some respondents were hesitant to participate until they received written information about the study. Since Westat was not able to obtain a matching address for all telephone numbers, some households did not receive an advance letter prior to the telephone call. When a respondent requested written information, he or she was sent a letter (see Appendix C) and a HINTS brochure.

Screener

The household screener was administered over the telephone using CATI. The purpose of the screening interview was to select an eligible person from the household for the extended interview. The screener involved asking the respondent how many adults live in their household and discerning the number of telephones in the household. One adult in the household was sampled for the extended interview using an algorithm designed to minimize intrusiveness.

As noted in Section 4.1.3, a subsample of households that refused to participate in the screener was selected for refusal conversion. Prior to refusal conversion contact by telephone, Westat sent a refusal conversion letter to the households for which there were address matches to request participation. The letter explained the purpose of the study as well as the importance of their participation (see Appendix D). If the case was not matched to a valid address, Westat attempted to contact the household again without sending a letter.

Extended Interview

If the screener contact was selected for the extended interview, the interviewer began the interview at this point. If someone else in the household was selected, the interviewer asked to speak to that person to conduct the extended interview. If the extended respondent was unavailable, the TRC tried to conduct the extended interview at a different time.

All extended refusals except for hostile refusals were contacted 2 weeks after their refusal to attempt refusal conversion. Prior to the refusal conversion call, all extended refusals linked to addresses were sent a refusal conversion letter intended to arrive a couple of days prior to being called (see Appendix E). If a completed interview was not obtained at the first refusal conversion attempt, a second followup call was made to elicit participation in the survey.

4.3 Findings from the CATI Operations

The field period for the RDD study was January 7 through April 27, 2008, with a total of 3,767 complete CATI interviews collected and an additional 325 partially complete CATI interviews collected, bringing the total number to 4,092² (see Table 4-2). Partial completes were defined as cases where the respondent completed the first section (Health Communications) of the interview, but that did not reach the end of the survey instrument. Respondents that did not complete at least the Health Communications section were coded as incompletes.

Table 4-2. Unweighted RDD sample results by mailable status

	Mailable		Nonmailable*		
	Total	Percent of total	Total	Percent of total	Total
Screener completes	5,988	88.6	772	11.4	6,760
Completes with initial cooperation	4,537	89.3	544	10.7	5,081
Completes with initial refusal	1,451	86.4	228	13.6	1,679
Extended Interviews	3,666	89.8	415	10.2	4,081

* Includes nonworking and nonresidential telephone numbers.

² Eleven of these completes were Spanish-speaking Hispanic respondents from the mail sample, who were transferred to telephone interviewers to complete the Spanish CATI interview and are therefore not included in Table 4-2.

4.3.1 Weekly Reports

To measure progress in meeting project goals, a series of production and management reports were generated on a regular basis during the field period. These reports provided information on response rates, cooperation rates, production to date in terms of total interviews, and cost as expressed by interviewer hours per completed interview. Reports monitoring HINTS 2007 data collection included the following:

- **Weekly Sample Performance Report.** This weekly report provided summary statistics on screener and extended interview sample status and yield including eligibility and response rates.
- **Weekly Cooperation and Conversion Rates.** This weekly report provided screener and extended interview initial cooperation and refusal conversion rates for the prior 7 days and for the study to date.
- **Weekly Summary of Interviewer Hours.** This weekly report provided information on total hours worked by the interviewing staff for the past 7 days and the study to date. The report also contained “air hours,” which reflect time spent actively dialing and interviewing sample cases. This report was used to track interviewer hours per completed interview throughout the study with a final estimate of 2.34 hours per complete.
- **Daily Interviewer Cooperation and Conversion Rates.** This daily report was used to track performance at the interviewer level. The report included screener and extended interview initial cooperation and refusal conversion rates for the past 7 days and for the study to date for every interviewer that worked on the study. This report was instrumental in identifying exceptional interviewers who might be candidates for refusal conversion work, and also those in need of refusal avoidance training due to low cooperation rates.
- **Production Report by Release Group.** This report showed the status of cases released to the TRC broken down by release group (i.e., the order of release within the TRC). This report estimated initial cooperation, refusal conversion, and response rates for both screener and extended interviews. This report was created on an ad hoc basis at several points during data collection, to inform possible changes to the protocol based on sample performance. See Appendix F for a sample.
- **Weekly TRC Production Report.** This report showed overall screener and extended interview production for the current week and cumulatively for the entire study. The report tracked screener and extended interview completes and cooperation/conversion rates, interviewer hours, hours per completed interview, and size of interviewing staff throughout the life of the study. A summary of this report is provided in Table 4-3.

- **Weekly TRC Report for NCI.** This report was sent on a weekly basis to NCI with summary information on sample status and performance for both screener and extended interviews. Please see Appendix G for a sample of this report.

Table 4-3. Weekly TRC production: Completed cases by week

Week beginning	Screener		Extended (CATI)*	
	Actual	Cumulative	Actual	Cumulative
1/14/2008	317	317	128	128
1/21/2008	637	954	293	421
1/28/2008	661	1,615	329	750
2/4/2008	682	2,297	309	1,059
2/11/2008	593	2,890	297	1,356
2/18/2008	568	3,458	323	1,679
2/25/2008	597	4,055	311	1,990
3/3/2008	556	4,611	326	2,316
3/10/2008	564	5,175	320	2,636
3/17/2008	340	5,515	238	2,874
3/24/2008	346	5,861	224	3,098
3/31/2008	287	6,148	203	3,301
4/7/2008	242	6,390	157	3,458
4/14/2008	214	6,604	149	3,607
4/21/2008	156	6,760	150	3,757

* Partial completes, 324 of which were coded following the completion of data collection, are not included in this weekly production count of extended completes.

4.3.2 Administration Times

The mean administration time for the extended telephone interview was 33.6 minutes, ranging from 16.0 to 126.8 minutes. The median length was 31.6 minutes.

4.3.3 Average Calls per Case

Before the start of calling, the CATI scheduler was configured with some standard call limits and study options. This allowed the project both the opportunity to standardize the flow of work and the flexibility to change the configuration to meet specific needs should that be necessary during the course of data collection.

Cases that never had any contact with the respondent were placed in each of seven non-contact time slices. These cases received at least one call attempt per time slice before being finalized. As resources allowed, these cases were “rested” and released additional times over several weeks for another round of seven calls in an effort to complete the case. Consequently, some cases received 14 call attempts over several weeks. Similarly, cases that were unresolved after nine calls were also released for additional calls, as resources allowed.

Queue priorities were set within the scheduler. Extended interview appointments had a higher priority than screener questionnaires. Table 4-4 details the level of effort for the screener by result code, while Table 4-5 details the level of effort for the CATI extended interview.

Table 4-4. Total screener level of effort: Number of call attempts by result

Call attempts	Completes and ineligible		Nonresponse		Nonworking and nonresidential		Noncontact	
	N	%	N	%	N	%	N	%
0	-	0.0	7	0.1	27,755	81.0	-	0.0
1-5	5,523	81.5	2,266	38.2	5,249	15.3	3,189	51.9
6-10	875	12.9	1,352	22.8	837	2.4	662	10.8
11-15	288	4.2	975	16.4	326	1.0	2,290	37.3
16-20	74	1.1	953	16.1	77	0.2	-	0.0
21-25	15	0.2	302	5.1	20	0.1	-	0.0
26-30	2	0.0	75	1.3	7	0.0	-	0.0
Total	6,777	100.0	5,929	100.0	34,271	100.0	6,141	100.0

Table 4-5. Total extended (CATI) level of effort: Number of call attempts by result

Call attempts	Completes and ineligible		Nonresponse	
	N	%	N	%
1-5	3,260	86.7	1,499	50.0
6-10	351	9.3	688	22.9
11-15	111	3.0	294	9.8
16-20	31	0.8	432	14.4
21-25	3	0.1	63	2.1
26-30	3	0.1	27	0.9
Total	3,759	100.0	3,001	100.0

4.3.4 Cooperation Rates and Refusal Conversion

Once the predictor sample had been in the field for several weeks, the initial screener cooperation rate was higher than expected—several percentage points higher than for HINTS 2005 and at the same level as HINTS 2003. Refusal conversion efforts were productive at both the first and second conversion stages, resulting in a combined conversion rate of well over 25 percent. At the extended interview stage, initial cooperation and refusal conversion rates were on par with the prior HINTS studies. Therefore, it was unnecessary to release the reserve sample.

Table 4-6 shows the percentage of residential numbers, the screener cooperation rate, and the extended-interview cooperation rates for the mailable and nonmailable strata. As was seen in HINTS 2005, both the percentage of residential numbers and the screener cooperation rates were higher among the mailable numbers than among nonmailable numbers. One reason for the higher screener cooperation rate in the mailable stratum is the \$2 incentive sent to the mailable cases. Another possible explanation is that even without the \$2 incentive, individuals in the mailable stratum may have a higher propensity to respond to the screener than those in the nonmailable stratum. On the other hand, the extended-interview cooperation rates for the mailable and nonmailable strata were approximately equal, which was also observed in HINTS 2005.

Table 4-6. Residential, cooperation, refusal conversion, and response rates and yield by mailable stratum, for screener and extended interviews

	Mailable	Mailable percent of total	Nonmailable	Nonmailable percent of total	Total
Sample used for CATI study	17,101		36,017		53,118
Residential and undetermined numbers ¹	14,326	76.0	4,521	24.0	18,847
Residential numbers (estimated) ²	13,986	87.6	1,986	12.4	16,972
Residency rate ²	81.8		5.5		30.1
Screener cooperation					
Initial cooperation rate	42.1		31.3		40.6
Refusal conversion rate	25.4		22.6		24.9
Second-refusal subsampling rate	65.4		65.6		65.4
Final cooperation rate	58.3		49.4		57.1
Screener completes					
Completes with initial cooperation	4,537	89.3	544	10.7	5,088
Completes with initial refusal	1,451	86.4	228	13.6	1,679
Total screener completes	5,988	88.6	772	11.4	6,760
Unweighted screener response rate	42.8		38.9		39.8
Extended interview cooperation					
Initial cooperation rate	62.8		52.4		61.6
Refusal conversion rate	28.8		25.5		28.4
Final cooperation rate	77.5		70.3		76.7
Extended interview completes					
Completes with initial cooperation	2,964	90.8	302	9.2	3,266
Completes with initial refusal	431	87.8	60	12.2	491
Partial completes	271	83.6	53	16.4	324
Total extended completes	3,666	89.8	415	10.2	4,081
Unweighted extended interview response rate	61.3		53.8		60.4

¹ Includes all the undetermined numbers due to answering machines or ring no answer.

² Includes only the portion of the undetermined numbers that are estimated to be residential.

4.3.5 Results of Hispanic Surname Coding

As described in Section 4.2.3, the surname coding procedure allowed for our first contact with these sample cases to be made by an interviewer who could easily transition to Spanish if necessary. Only 1,086 (4.3%) of the 25,363 numbers dialed for the telephone survey (excluding those purged prior to data collection) were coded with the Hispanic work class flag. This small part of the sample yielded 63 percent of the Spanish language completed screeners, and 56 percent of the Spanish language

extended interviews. Given the small size of the bilingual work force, with only four bilingual interviewers (8% of the staff), the surname coding was a very useful tool for streamlining the delivery of cases in need of bilingual attention to those with bilingual skills.

4.3.6 Data Retrieval

During the fourth week of data collection, three programming errors in the CATI instrument were discovered by data preparation staff reviewing preliminary frequencies. These errors were discovered on February 15, 2008, and were corrected immediately. The errors and their impact on the data are described below.

Tobacco Section: Respondents who reported hearing of telephone quit lines such as a toll-free number to call for help in quitting smoking (BR-46) were asked if they have ever called a telephone quit line (BR-51). Respondents who reported calling a quit line and who are current smokers or quit less than a year ago were supposed to be asked BR-52 (“In the past 12 months, did any doctor, dentist, nurse, or other health professional suggest that you call or use a telephone helpline or quit line to help you quit smoking?”). There was a problem with the routing, which resulted in only current smokers being asked BR-52. Respondents who quit smoking less than a year ago were not asked BR-52, resulting in missing data for 22 respondents.

Respondents who were never smokers or who had quit smoking over a year prior to the interview were supposed to go to question BR-53 (“How likely would you be to call a smoking cessation telephone quit line in the future, for any reason?”). This programming error resulted in these respondents going to BR-53a, the question after BR-53, instead. This problem resulted in missing data for 609 respondents.

Cancer Section: Respondents who reported being diagnosed as having cancer were asked at what age or in what year they were first told that they had cancer (CS-19). They could respond to this question by providing either an age or a year. All respondents who were asked CS-19 were supposed to be asked if they ever received any treatment for their cancer (CS-20). There was a problem with the routing and only respondents who answered CS-19 with a year were asked CS-20. Respondents who answered with an age were skipped to the following question (CS-21: “How long ago did you finish your most recent treatment?”). This resulted in missing data for CS-20 for the 102 respondents who answered CS-19 with an age.

A total of 673 respondents were identified as having missing data for one or more of the affected items. Due to the size of this missing data problem, it was determined that the data retrieval effort would be best conducted using a computerized scripted program, which could be customized for each case, rather than as a paper-based effort typically performed for data retrieval. Westat designed and conducted the data retrieval effort using Voxco, a survey program that allows quick and easy programming, and supports predictive dialing. A short introductory script and contact screens were programmed in both English and Spanish languages, with customized fills and displays (e.g., his/her, he/she, subject's name could all be displayed as appropriate to each case).

Data retrieval was conducted over the course of 16 days, from March 26 through April 10, 2008. Interviewers attempted these cases during the daytime, evening, and weekend shifts throughout this time period. Up to five attempts per case were made, and Voxco permitted the re-releasing of cases for additional calls (e.g., for cases resulting in “ring no answer” results or “answering machine” results across all five calls). If a respondent refused, no further call attempts were made. If a respondent had moved, we attempted to obtain a new telephone number from the original household and contact the respondent at the new number.

The data retrieval effort was very successful, with missing data obtained from a total of 515 of the 673 respondents. Our response rate for this effort was 77 percent, with an initial cooperation rate of 95 percent. Most of the nonresponse was not caused by respondent refusals, but from an inability to locate respondents who had moved and from noncontacts. Table 4-7 describes the final case outcomes and call results for this data retrieval effort.

Table 4-7. Data retrieval calls

Code	Final case results		Call results	
	Number of cases	Percent of cases	Number of calls	Percent of calls
Unable to reach respondent (nonworking or disconnected number, subject moved)	17	2.5	31	1.5
No contact (reached ring no answer or answering machine on repeated attempts)	97	14.4	1,069	52.9
Interim/unresolved (appointments, busy signals, dropped calls)	15	2.2	378	18.7
Complete: successfully obtained missing data from respondent	515	76.5	515	25.5
Refusal	29	4.3	29	1.4
Total	673	100.0	2,022	100.0

4.3.7 Imputation

For the 158 cases for which data retrieval was not successful, hot-deck imputation was used to replace missing responses with imputed data that had the same distribution as the reported data. Hot-deck imputation is a data processing procedure in which cases with missing values for specific variables have the “holes” in their records filled in with values from other cases, referred to as “donors.” Variables not containing missing data are used to create groups of similar cases. Donors are then randomly selected within each group to be the sources of imputed data for variables of cases within the group that contain missing data. For question BR-52 (“In the past 12 months, did any doctor, dentist, nurse, or other professional suggest that you call or use a telephone helpline or quit line to help you quit smoking?”), there were five imputed responses. For question CS-20 (“Did you ever receive any treatment for your cancer?”), there were 23 imputed responses. For question BR-53 (“How likely would you be to call a smoking cessation telephone line in the future, for any reason?”), there were 143 imputed responses.

4.3.8 Interview Data Processing

Throughout the field period, data preparation staff conducted a daily review of collected data to see if any updates were needed for the CATI data. On a regular basis, the data preparation staff ran frequencies and crosstabulations for categorical data. In addition to this review, to ensure that the interview data were as complete as possible, staff used proven quality control procedures including: (1) a review of interviewer comments for problems in response coding, or where the CATI system did not provide sufficient means to code a legitimate response; and (2) a review of open-ended responses to ensure consistency in the data and simplify the overall analysis and reporting operations. Westat consulted with NCI on open-ended response coding before collapsing responses into discrete categories. Coding decisions relating to rules used for open-ended response upcoding and for instrument consistency were collected in a Decision Log.

Mail Study Design and Operations

This chapter describes the process of conducting the mail survey for HINTS 2007, including the development of the mail survey instrument, the sample design, and the data collection protocol procedures. The chapter concludes with a description of cooperation to the mail survey, contacts made by respondents, and results of the IVR experiment.

5.1 Sample Selection

The mail survey included a stratified sample selected from a list of addresses that oversampled for minorities. Sampled addresses were matched to a database of listed telephone numbers, with 50 percent of the cases successfully matched to a telephone number. Matches in which a telephone number was both appended to an address-sample address and included in the RDD sample were deleted from the address sample. The final sample size for the mail survey was 7,851.

5.1.1 Sampling Frame for Address Sample

The sampling frame for the address sample was a database used by MSG to provide random samples of addresses. The decision to use this database as a sampling frame was the result of an evaluation study conducted by Link et al. (2005). This study compared five address vendors in terms of the coverage of their lists for a six-state area. Three vendors had high levels of under-coverage in one or more of the six states. Of the remaining two vendors, only MSG could provide sampling services for a single-stage sample of addresses. The use of the other vendor would have required two stages of sampling—first the sampling of carrier routes and then the sampling of individual addresses. Compared to a single-stage design, a two-stage design for selecting addresses is more costly and provides less precision for a given sample size.

The MSG address database is updated bimonthly from the USPS's Computerized Delivery Sequence (CDS) File. Licensed by the USPS to qualified address vendors, the CDS is an electronic data product that provides and updates addresses by carrier route (USPS, 2006). Address vendors must initially qualify for the CDS information for a given 5-digit ZIP Code area by having at least 90

percent but not more than 110 percent of all the addresses in the ZIP Code area. Once a vendor has qualified for a 5-digit ZIP Code area, CDS information is made available bimonthly via electronic media.

The CDS contains current information on all mailing addresses serviced by the USPS, with the exception of general delivery. CDS information is available for the following types of addresses:

- Addresses that currently receive or have received mail delivery.
- Addresses on city routes to which carriers do not deliver because of alternative delivery arrangements, e.g. to post office boxes. (Referred to as “throwbacks”, these addresses can be included in or excluded from MSG-provided samples of addresses.)
- Addresses on city routes vacant longer than 90 days and likely to be long-term vacancies, which are not considered seasonal. (Referred to as “vacants”, these addresses can also be included in or excluded from MSG-provided samples of addresses.)
- Addresses delivered seasonally. (No CDS information is available, however, on the dates of the mailing season. Referred to as “seasonals”, these addresses can also be included in or excluded from MSG-provided samples of addresses.)

Link et al. (2005) evaluated the coverage of the MSG address list for the six states of California, Illinois, New Jersey, North Carolina, Texas, and Washington. For each of the counties in this six-state study area, they compared the number of addresses on the MSG list as of April 1, 2005, to the Census Bureau’s estimated number of households for July 1, 2003. They tabulated the number of counties in which there was a high level of undercoverage, which they defined as the number of addresses on the MSG list for the county being less than the number of households in the county by at least 10 percent. They found that in counties where less than 25 percent of the population lives in an urban area, nearly 90 percent of the counties had a high level of undercoverage; whereas in counties where 75 percent or more of the population lives in an urban area, only 4.3 percent of the counties had a high level of undercoverage.

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 we chose for the address sample contained duplicate units because some households could receive mail in more than one way. To permit adjustment for this duplication of households in the sampling frame, we included a question on the mail questionnaire that asked how many different ways respondents receive mail.

In rural areas, some of the addresses on the CDS are simplified addresses, which are addresses that do not contain street addresses or box numbers. Simplified addresses contain insufficient information for the mailing of questionnaires. Consequently, alternative sources of usable addresses were used when a carrier route contained simplified addresses. This partially ameliorated the CDS's known undercoverage of rural areas, but the coverage and undeliverable rates for the used alternative sources of addresses are not known.

5.1.2 Selection of Main-Survey Address Sample

The sampling unit for the address sample was an individual address. The sampling frame was all residential addresses in the United States on the MSG database, including post office boxes, throwbacks, vacant addresses, and seasonal addresses. The sampling frame was stratified into two strata—a high-minority stratum and a low-minority stratum—by using Claritas demographic data for census block groups matched to the address ZIP+4 Codes. Addresses matched to census block groups that had a population proportion for Hispanics or a proportion for African Americans that equaled or exceeded 24 percent were assigned to the high-minority stratum. All other addresses were assigned to the low-minority stratum. An equal-probability sample of addresses was selected from each stratum. The high-minority stratum's proportion of the sampling frame was 25.1 percent, and it was oversampled so that its proportion of the sample was 50 percent.

Unlike the RDD sample, all adults in the household at a sampled address were asked to complete a questionnaire. Hence, the mail sample was a stratified cluster sample, in which the household was the cluster. Our decision to not subsample the adults in sampled households is the result of an evaluation study conducted by Battaglia et al. (2005). This study compared three respondent-selection methods for household mail surveys: (1) any adult in the household; (2) the adult in the household having the next birthday; and (3) all adults in the household. The study found that the next birthday and all-adults methods yielded household-level completion rates that were comparable to the any-adult method, the method that the researchers assumed to have the least respondent burden. Another finding from this study was that differences in response rates by gender and age were less for the all-adults methods than for the next birthday and any-adults method.

Following the selection of the address sample, telephone numbers were obtained for 50.0 percent of the sampled addresses, and these were matched to the telephone numbers in the RDD sample. There was one address-sample telephone number that had also been selected for the RDD sample.

This case was deleted from the address sample. There were a total of 7,851 sampled addresses, with 3,926 in the high-minority stratum and 3,925 in the low-minority stratum.

5.2 Mail Survey Operations

The mail survey was conducted from January 15 to April 27, 2008. The following sections summarize the procedures for the mail survey including the protocol for sending out the mailings, the IVR experiment, and the procedures for processing questionnaires. More detail about these procedures can be found in the *HINTS 2007 Operations Manual* dated January 2008.

5.2.1 Questionnaire Mailing Protocol

Data collection on the mail survey was initiated on January 15, 2008, and continued through April 27, 2008. The specific mailing schedule and details are shown in Table 5-1. Households were sent an advance letter introducing the study and explaining the questionnaires that they would receive. This letter included a set of FAQs on the back. A week later, households were sent a package of three questionnaires with a request that each adult in the household complete and return a survey. This package included a \$2 incentive. Two weeks after the initial mailing, households that had not yet responded were sent a postcard reminding them to return the surveys. Two weeks after the reminder postcard, households that had not yet responded were sent a second package of questionnaires by FedEx. Two weeks after the mailing of the second package, nonresponding households for which telephone numbers were available were entered into the IVR experiment outlined in Section 5.2.2. The field period ended on April 27, with any questionnaires received after that date not eligible for entry into the study. Advance letters, cover letters, and the postcard used for the mail study can be found in Appendix H.

Once a questionnaire was received back from a household, that household was considered “complete” and did not receive any further mailings. Households that sent back a questionnaire on which they had written that they did not want to participate were considered refusals and did not receive any additional mailings.

Table 5-1. Mail survey schedule and protocol

Date	Activity	Package contents	Mailing method
January 15, 2008	Advance letters sent to all households	Advance letter with FAQs	USPS
January 22, 2008	First set of questionnaires sent to all households	Cover letter with FAQs 3 questionnaires 3 return envelopes \$2 incentive	USPS
February 5, 2008	Reminder postcards sent to nonresponding households	Postcard	USPS
February 19, 2008	Second set of questionnaires sent to nonresponding households	Cover letter with FAQs 3 questionnaires 3 return envelopes	FedEx
March 5, 2008	IVR experiment initiated for nonresponding households		
April 27, 2008	All mail cases finalized and no additional questionnaires accepted		

All letters (advance letter and cover letters for the first and second mailing) sent to respondents referenced two toll-free telephone numbers that respondents could call to contact Westat. One of the toll-free numbers (the English number) was used by respondents who had questions, wanted to request additional questionnaires, or wanted to refuse participation in the study. The other number was dedicated to Spanish-speaking respondents who could leave a voicemail message on the designated toll-free number asking questions, requesting to complete the survey by telephone, or refusing participation in the study. This voicemail was checked regularly by Spanish-speaking study staff who addressed the requests as appropriate.

5.2.2 Interactive Voice Response (IVR) Experiment

Households that did not respond within 2 weeks of the second mailing of the instrument and for which we had a telephone number were included in an IVR experiment. Households were randomized into one of three experimental groups: (1) IVR; (2) live prompt from interviewer; and (3) control group. The prompt encouraged respondents to complete the surveys that had been mailed to their household. For both the IVR and interviewer prompt groups, it was considered a

contact if either a person was directly contacted or a voicemail reminder was left. Results of this experiment are described in Section 5.3.3.

5.3 Findings from the Mail Operations

As noted in the previous section, data collection of the mail survey was initiated on January 15, 2008, and continued through April 27, 2008. A total of 3,473 completed and 109 partially completed surveys were received during the field period. A questionnaire was considered to be complete if at least 80% of sections A, B, and C were filled in. To be considered partially complete, a questionnaire had to have between 50% and 79% of these sections filled. Questionnaires with less than 50% of sections A, B, and C filled were coded as incomplete and discarded. These 3,582 surveys represented a total of 2,581 households. The response at the household level is shown in Table 5-2.

Table 5-2. Household cooperation in the mail survey

	Number	Percent
Survey complete or partially complete	2,581	32.9
Sent to CATI for Spanish interview	23	0.3
Occupant deceased	3	0
Occupant sick and unable to complete survey	6	0.1
Refused	93	1.2
Package undeliverable	735	9.4
Nonresponse	4,411	56.2
Total	7,851	100

Survey response tended to rise following each mailing. Table 5-3 outlines the household response by week during the field period, with the household being coded as complete once a single complete or partially complete survey was returned regardless of the number of adults in the household.

Table 5-3. Household response by week

Week ending	Household coded as complete	
	Actual	Cumulative
February 2, 2008	861	861
February 9, 2008	406	1,267
February 16, 2008	311	1,578
February 23, 2008	9	1,587
March 1, 2008	597	2,184
March 8, 2008	188	2,372
March 15, 2008	90	2,462
March 22, 2008	41	2,503
March 29, 2008	35	2,538
April 5, 2008	13	2,551
April 12, 2008	15	2,566
April 19, 2008	7	2,573
April 26, 2008	7	2,580
May 1, 2008	1	2,581

Almost half the households that sent in questionnaires responded after the first mailing. Table 5-4 shows the response following each mailing and following the IVR experiment. The response is further broken down by the high- and low-minority strata.

Table 5-4. Household response by mailing and strata

Household-level response	High-minority strata		Low-minority strata		Total	
	N	%	N	%	N	%
After the first mailing	428	42.8	839	503.0	1,267	49.1
After the reminder postcard	131	13.1	187	11.8	318	12.3
After the second mailing	326	32.6	424	26.8	750	29.1
After the IVR experiment	114	11.4	132	8.3	246	9.5
Total	999	100.0	1,582	100.0	2,581	100.0

5.3.1 Weekly Reports

To regularly track progress of the mail survey, three production reports were generated on a weekly basis during the field period. These reports provided information on cooperation rates and survey completes. Regular reports included:

- **Production report.** This report showed the status of all the households in the mail survey at the time of the report. Codes included were:

- Completed survey received;
- Spanish interview requested;
- Occupant deceased;
- Package undeliverable;
- Occupant sick and unable to complete survey;
- Refusal; and
- Not yet responded.

The report showed these codes for the last 7 days, the field period to date, and by the high- and low-minority strata.

- **Response by mailing report.** This report showed the response to the specific mailings by household response. Categories in this report included:

- Household response after the first questionnaire mailing;
- Household response after the reminder postcard;
- Household response after the second questionnaire mailing; and
- Household response after the start of the IVR experiment.

The report indicated these codes for the full sample and by the low- and high-minority strata.

- **Response by IVR Status report.** This report showed the number of households included in the IVR experiment by group and the survey response by those groups. Categories in this report included:

- Not included in the experiment;

- Received an IVR call;
- Received a live interviewer call; and
- Included in experiment, but randomized to “no treatment.”

5.3.2 Telephone Contacts

As described in Section 5.2.1, households in the mail survey were provided with two toll-free telephone numbers: (1) in English to call with questions about the survey; and (2) in Spanish to call to request a Spanish-language interview. Numerous calls were received on both of the toll-free lines, with calls clustering around mailing dates. On the English line, the majority of calls related to the following:

- **Messages stating that a survey had already been sent in, but the household had received an additional mailing.** This was a result of the completed questionnaire not being received before the next scheduled mailing. Respondents were assured that if all the adult members of the household had sent in a questionnaire, they did not need to fill out an additional survey.
- **Messages stating that they did not want to participate.** Households that called requesting to withdraw from the study were marked as “refused” and did not receive any additional mailings.
- **Messages asking what to do with extra surveys.** Respondents in single-adult households or in households that got both the questionnaire mailings wanted to know if they should send back the unused questionnaires or give them to people outside their household. Respondents were instructed to throw away the extra questionnaires and not to share them with people outside the household.

Calls to the Spanish toll-free line primarily related to requests to receive the mail survey in Spanish. Most callers did not initially respond positively to the idea of doing the interview over the telephone, even with a Spanish-language interviewer. Although some callers were convinced to allow an interviewer to call them to conduct the CATI interview, most refused. A total of 23 households were sent to the TRC for followup with the Spanish-language interview.

5.3.3 IVR Experiment Results

The IVR experiment was conducted on nonresponding households starting March 5, 2008. The IVR instrument was a simple reminder prompt recorded by a female voice. Calls were made over the course of 8 days, using five revolving “time slices” within the sampled household’s time zone: 4-5 p.m., 5-6 p.m., 6-7 p.m., 7-8 p.m., and 8-9 p.m. A maximum of five calls were made to each household, until a live answer or answering machine/voicemail was reached. Of the 713 cases randomized to receive IVR calls, 592 (about 83%) successfully received the message played either to a live person or an answering machine. A total of 1,353 IVR calls were made. Details about these calls are outlined in Table 5-5.

Table 5-5. IVR calls

Code	Number of cases	Percent of cases	Number of calls	Percent of calls
Line busy	2	.28	26	1.92
No Answer	10	1.40	71	5.25
Call disconnected	28	3.93	189	13.97
Fax machine reached	10	1.40	48	3.55
Uncallable for other reason	71	9.96	427	31.56
Successfully played to live person	289	40.53	289	21.36
Successfully played to answering machine	303	42.50	303	22.39
Total	713	100.0	1,353	100.0

The calling effort for cases assigned to the “live interviewer prompt” condition was programmed and managed in Voxco. Voxco’s predictive dialing feature was used, which allowed for a very small interviewing staff to be used for these calls. The script began by verifying that we had reached someone at the correct mailing address. If so, we inquired as to whether they remembered receiving the package of HINTS questionnaires and asked whether additional copies were needed.

Live interviewer prompt calling hours were similar to those used for the IVR effort, with calls placed to sampled households between 6 p.m. and 9 p.m. within the household’s time zone over the course of 8 days, beginning on March 5, 2008. The number was called until an answering machine message was left, a household member was contacted, a refusal was encountered, or the five-call calling algorithm was otherwise fulfilled with no resolution.

Of the 690 households assigned to this treatment, 99 (14.3%) were not reachable due to nonworking numbers or nonmatching mailing addresses. Upon making contact, 139 households (20.1%) either indicated they planned to return the questionnaire or requested additional copies in order to participate. Messages were successfully left with an additional 241 households (34.9%), and 87 households (12.6%) refused. The cooperation rate for this prompting effort was 81.4 percent. Efforts to prompt the remaining 124 households (18%) resulted in repeated noncontact or unresolved callbacks. Table 5-6 provides more details on the live interviewer prompt calls.

Table 5-6. Live interviewer prompt calls

Code	Final case results		Call results	
	Number of cases	Percent of cases	Number of calls	Percent of calls
Nonworking	77	11.2	111	9.5
Wrong address	22	3.2	22	1.9
No contact (ring no answer)	91	13.2	453	38.7
Interim/unresolved	33	4.8	117	10.0
Complete: household received questionnaires and will participate	97	14.1	97	8.3
Complete: household requested replacement questionnaire(s)	42	6.1	42	3.6
Complete: message left on answering machine or with non-household member	241	34.9	241	20.6
Refusal	87	12.6	87	7.4
Total	690	100.0	1,170	100.0

Survey completions between March 5 and the end of the field period for the cases included in the experiment are shown in Table 5-7. The households receiving an IVR call had the same response rate as those that received no treatment. Households receiving a call from a live interviewer appeared to respond at a higher rate than those contacted by IVR, although this difference is not statistically significant.

Table 5-7. Household response by treatment in IVR experiment

	Sample size	Number of households responding	Percent of households responding
Randomized to IVR call	712	53	7.4
Randomized to live interviewer call	690	75	10.9
Randomized to no treatment	725	54	7.4

5.3.4 Survey Processing

Upon receipt, each questionnaire was edited for applicable comments and entered into the Survey Management System. Completed questionnaires were photocopied and sent to the scanning company 2 weeks after each mailing (first questionnaire mailing, postcard mailing, and second questionnaire mailing) and at the end of the mail field period, for a total of four batches. Once the scanning company returned the surveys, data was cleaned and updated as needed.

After scanning was complete, edits that check allowable ranges for each variable, skip logics and additional logics were run on the data to ensure logical consistency and accuracy. The data were updated as needed. All updated data were reviewed to ensure that the updates were applied correctly. Listings of data by ID were reviewed as well as crosstabulated as part of the quality control measures. All verbatim responses were reviewed for clarity and spelling errors. Responses specified as “other” were upcoded into preexisting codes when applicable.

Cycles of edits were run until the data were clean. As the final step in the editing process, the keyed data were concatenated into one batch and frequencies and crosstabulations were produced and reviewed and updates were applied as needed until the final dataset was clean.

5.3.5 Imputation

Some respondents did not follow the instructions to mark only one response for question HC02 (“The most recent time you looked for information about health or medical topics where did you go first?”) and question HC11 (“The most recent time you looked for cancer information, where did you go first?”). Instead, 315 respondents marked multiple answers for question HC02, and 188 respondents marked multiple answers for question HC11. When this occurred, 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. For example, assume the frequencies of occurrence for the answers of “Books” and “Family” were 4 percent and 6 percent, respectively, for the single-answer responses to question HC02. Then, if a respondent incorrectly answered HCO2 by selecting both “Books” and “Family” (and selected no other answers), the imputed answer would be a random selection between “Books” 4 times out of 10 and “Family” 6 times out of 10.

6.1 Combining Data Sets

After data editing and cleaning were completed on both the RDD and mail data, the two data sets were merged into a master SAS database for further editing and cleaning, including additional edit/logic checks (range, skip patterns, and consistency) to ensure the consistency of the data across instruments.

For most questions, data from the RDD and mail modes were easily combined. However, there were some items where the data had a different denominator by mode. A meeting was held with NCI investigators to address this issue and to decide how best to handle each of these items. Items were handled in one of three ways:

1. CATI data were changed to match the denominator for the mail data;
2. Mail data were changed to match the denominator for the CATI data; or
3. A new variable was created to be used by both modes.

Details about the application of these data combination methods and the variables affected are available in Appendix I.

The SAS combined dataset was delivered to NCI on September 18, 2008 and, following some updates to specific items, again on October 17, 2008.

6.2 Codebooks

Detailed codebooks of the combined data were created using COED, Westat's proprietary system for codebook development and data validation. The codebooks defined all variables in the dataset, provided the question text, listed the allowable codes, and explained the inclusion criteria for each item. Frequencies for the combined dataset were added to the codebooks, which were delivered to NCI in both hard-copy and electronic format.

6.3 Metadata Development

As noted in Section 3.1.2, Westat developed a database to track information about each item on the HINTS instrument. One of the main purposes of this database was to provide metadata to NCI to be included on the HINTS web site.

Westat worked with NCI staff to ensure that complete information was provided for each item in the database. In particular, the original source of each item had to be provided by the working group leader responsible for the item. Reports from the tracking system were provided to NCI when needed to facilitate the gathering of information. Information provided by NCI or gathered from other sources was put into the database on an ongoing basis.

Westat attended several meetings with NCI investigators and other NCI contractors to discuss the specific requirements and format of the metadata tables to be delivered. Westat provided table shells to all parties and asked for feedback before the final tables were developed.

Based on information accumulated in the database as well as feedback about the table shells, Westat developed a series of metadata tables in Excel that provided the following information on the data:

- **HINTS History.** This table indicated whether questions had been asked in the HINTS 2003 and HINTS 2005 iterations. If the question had been asked previously, whether or not modifications were made to the question was noted.
- **Mode Status.** This table indicated whether questions were in both the CATI and mail surveys, CATI only, or mail only. For questions that were in both the CATI and mail surveys, any differences in question wording were noted.
- **Sources.** This table provided the original source for each question that was not novel to HINTS 2007.
- **Inclusion Criteria.** This table listed the criteria for determining which respondents were asked each question. This table was used to identify variables that appeared in both the RDD and mail surveys but had different denominators.

These tables were updated as necessary to reflect decisions that were made in combining the mail and RDD data sets.

Sample Weights and Variance Estimation Overview

7

7.1 Overview of Sample Weights

Once data cleaning and the merging of the datasets was complete, weighting of the data was initiated. Every sampled adult who completed a questionnaire in HINTS 2007 received three full-sample weights and three sets of replicate-sample weights. Two of the three types of weights correspond to the type of samples—the address sample and the RDD sample. The address-sample weight is missing for a case in the RDD sample and vice versa. The sample-specific weights are used to calculate estimates based on data from one of the two samples. The third type of weight is a composite weight which is used to calculate estimates based on the data from both samples.

Each type of sample-specific sampling weight consists of three major components. The first component is the respondent's base weight. This base weight is the reciprocal of the probability that the respondent had of being sampled. Section 7.3 discusses the computation of base weights.

The second part of the sampling weight is an adjustment for nonresponse. There are several points at which cooperation needs to be gained. For the address sample, the household needs to be successfully reached and then each adult in the household needs to successfully complete a questionnaire. Thus, for the address-sample both a household nonresponse adjustment and a within-household nonresponse adjustment were computed. For the RDD sample, on the other hand, the household needs to be successfully reached, the screener needs to be successfully completed, and the sampled respondent within the household needs to be successfully recruited to complete the extended interview. For the RDD sample, both a screener nonresponse adjustment and an extended interview nonresponse adjustment were computed. The computation of the screener nonresponse adjustment is complicated by the fact that many residential households are never reached even after a considerable number of calls and are never completely confirmed as residential or nonresidential. These telephone numbers with unknown residential status can be categorized as NM numbers (for which only an answering machine is reached) and NA numbers (for which no contact is made of any kind). Section 7.4 discusses nonresponse adjustments in detail.

The third part of the sampling weight is a calibration adjustment. The primary purpose of the calibration adjustment is to reduce the sampling variance of estimators through the use of reliable

auxiliary information (reliable in the sense of having less sampling and nonsampling error than the corresponding HINTS estimates). For example, the total number of male and female adults in the United States is estimable by taking the summation of all (nonresponse-adjusted) base weights of responding adults in the survey by sex. There are other estimates of these same population totals with less sampling and nonsampling error that can be used to calibrate the HINTS estimates (e.g., if HINTS population estimates for males deviate from corresponding estimates from the auxiliary information, the weights of male respondents can be altered to bring HINTS estimates “in line” with the auxiliary information). This process of calibration improves the sampling error of HINTS estimates, which are correlated in the population with characteristics represented in the auxiliary information. Calibration adjustments are discussed in Section 7.6.

Composite weights permit one to calculate estimates based on the data from both samples. Section 7.5 discusses the calculation of the composite weights.

7.2 Variance Estimation Methodology for HINTS 2007

The sampling weights for each responding adult are insufficient for the computation of statistically sound nationally representative estimators based on HINTS data. It is also necessary to produce statistically valid standard errors for these estimators.

The jackknife technique is compatible with the sample design and weighting procedures for HINTS. The jackknife variance estimation technique takes carefully selected subsets of the data for each “replicate,” and for each respondent in the replicate subset 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 Westat software package, WesVar, was used to calculate variances using jackknife procedures for a wide range of estimators. This package can be obtained from the Internet by accessing the WesVar site: http://www.westat.com/westat/statistical_software/wesvar/index.cfm.

The jackknife variance estimator requires the use of replicate weights. For the HINTS 2007 data set, a set of $R=50$ replicate weights was assigned to each responding adult. To illustrate how the replicate rates are computed, suppose \mathbf{P} is a percentage of adults in the U.S. population having a particular

characteristic (e.g., answering one of the HINTS 2007 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 jackknife variance estimator of the sampling variance of p can be computed in two steps:

- Step 1. Recompute estimators $p(r)$, $r=1,\dots,R$, by aggregating the replicate sampling weights corresponding to replicate r for all responding adults with the characteristic.
- Step 2. Compute the jackknife variance estimator

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

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 2007 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 jackknife variance estimation, each sample telephone number 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). See Section 7.3 for further details on computation of the replicate base weights.

Nonresponse and calibration adjustments were then computed for each set of replicate base weights, using the replicate base weights in the computation of nonresponse and calibration adjustments in place of the original base 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 base weights, nonresponse adjustments, and calibration adjustments.

7.3 Base Weights

Base weights for the RDD sample were assigned to sampled telephone numbers. The base weight is the reciprocal of the telephone number's probability of selection and was a constant for all of the selected telephone numbers. The sum of the base weights for the sample represents the total number of telephone numbers eligible for selection in HINTS.

The base weight calculation for the address sample is slightly different. The base weight is the reciprocal of an address's probability of selection, which depends on the stratum the address was selected from. With those cases in the high-minority stratum having a base weight of approximately three times the base weight of the addresses in the low-minority stratum—reflecting the oversampling of high-minority areas.

Standard errors were computed for HINTS 2007 estimates through the use of the jackknife technique, as discussed in Section 7.2. A total of 50 replicate base weights were computed for each sample unit.³ Suppose we write as \mathcal{A} the set of all sampled adults in the study. The base weight will be indicated below as w_i (i indicating the particular responding adult). Any given survey estimate can be written as

$$Y = \sum_{i \in \mathcal{A}} w_i y_i$$

where y_i is the value of a particular survey characteristic for responding adult i , and w_i is the full-sample base weight. The r -th replicate estimate for Y can be written as:

$$Y(r) = \sum_{i \in \mathcal{A}} w_i(r) y_i \quad \text{with} \quad w_i(r) = \begin{cases} \frac{R}{R-1} w_i & i \in A(r) \\ 0 & i \in D(r) \end{cases}$$

³ The total of 50 was chosen from among a number of acceptable alternatives. Generally a large number is necessary for stable variance estimates (e.g., greater than 10), but a number much greater than, for example, 100 generates sample files that are too large in size (because of large number of replicate weight fields).

The set $A(r)$ is the replicate set corresponding to replicate r , and the set $D(r)$ is the deleted set corresponding to replicate r . $w_i(r)$ is the r -th replicate weights for responding adult i . The union of $A(r)$ and $D(r)$ for each replicate r is the full sample set A .

7.4 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, there 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 answer any part of the questionnaire) 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).

7.4.1 RDD Screener Nonresponse Adjustment

Before nonresponse adjustments were performed, the base weights were modified to account for several design features of the RDD sample:

- The subsampling of nonmailable telephone numbers. A discussion of this can be found in Section 4.1.1.
- For HINTS 2007, a reserve sample was selected and set aside. The weights were adjusted to account for those telephone numbers that were never dialed. This is discussed in Section 4.1.1.
- Refusal subsampling: A portion of the refusals did not undergo “full” refusal conversion attempts. This process is discussed in Section 4.1.3.

In addition to the adjustments that were made based on design features, an additional adjustment was made prior to screener nonresponse adjustment for those households for which no contact was made and only an answering machine (NM) or ring no answer (NA) was observed. A discussion of how this adjustment was handled can be found in Section 7.1.

Once aforementioned adjustments to the base weights were made, screener nonresponse adjustments were performed. Screener nonresponse occurs when a household is reached, but no screener gets completed. To adjust for this type of nonresponse, each completed screener received a screener nonresponse adjustment equal to the reciprocal of the estimated response rate in its screener nonresponse cell. For this adjustment (as well as the others discussed in this section), mailable status was used to classify respondents into nonresponse adjustment cells. The final screener adjustment factor for respondents with a mailable address was 2.33. The adjustment factor for respondents with a nonmailable address was 2.56.

7.4.2 RDD Extended Interview Nonresponse Adjustment

More information is available about extended interview nonrespondents compared to screener nonrespondents. This extra information comes from the completed screener (a case was not designated as an extended interview nonrespondent unless the screener was successfully completed).

Two adjustments were made to the screener weight. The first was a multiple telephone adjustment. Those households with more than one telephone number that receives calls received an adjustment factor of 2. Next, adjustments were made for the subsampling of adults in the household in the

screeners. Here the adjustment factor is equal to the number of adults reported in the screener. The resulting weight w_i is used in the nonresponse adjustment.

Twenty-two extended interview nonresponse cells were generated using cross-classifications of the following characteristics of the sampled adult and household:

- Age and gender (11 cells);
- Mailable status (2 cells).

Weighted nonresponse adjustments were computed for each extended interview cell b as follows:

$$ENRA(b) = \frac{\sum_{i \in SA(b)} w_i HNRA(a)}{\sum_{i \in SRA(b)} w_i HNRA(a)},$$

where w_i is the base weight for sampled adult i , $SA(b)$ is the set of all sampled adults (in cooperative screeners) in interview response cell b , $SRA(b)$ is the set of all sampled adults in cell b completing an extended interview (i.e., the extended interview respondents), and $HNRA(a)$ is the screener nonresponse adjustment for the screener nonresponse cell a containing household i . The denominator of $ENRA(b)$ is an unbiased estimator (adjusted for screener nonresponse⁴) of the total number of adults in the nonresponse cell who would answer an extended interview if contacted (the “population respondents”), the numerator of $ENRA(b)$ is an unbiased estimator of the total number of adults in the nonresponse cell (also adjusted for screener nonresponse), and $ENRA(b)$ is an approximately unbiased estimator of the response rate which would be obtained in cell b if the entire U.S. population were contacted for the study.

Westat also computed a weighted extended interview response rate for reporting purposes (see Chapter 6 for discussion of these rates). Write SA as the set of all sampled adults from completed screeners and SRA as the set of all sampled adults completing an extended interview. The weighted extended interview response rate was computed as follows:

⁴ Under full response, the sum of the base weights is an unbiased estimator. With the presence of nonresponse, there will be nonresponse bias from any differences between the responding and nonresponding households. This nonresponse bias is reduced in magnitude by the screener nonresponse adjustments. It cannot be expected that these adjustments eliminate all bias, so the claim of “unbiasedness” of these totals needs to receive this caveat.

$$EXTINR = \frac{\sum_{i \in SRA} w_i HNRA(a)}{\sum_{i \in SA} w_i HNRA(a)}.$$

The average nonresponse adjustment factor was 1.75 and ranged from a high of 3.43 to a low of 1.35.

7.4.3 Address-Sample Nonresponse Adjustment

For the address-sample nonresponse adjustment, there is no screener nonresponse adjustment as screening was not done at the household level. Instead, adjustments were made at the person level for the following reasons:

- Household nonresponse;
- The number of ways a household can receive mail; and/or
- Within-household nonresponse.

Household nonresponse was examined first. Not all households that were mailed surveys returned questionnaires. The base weights of the households that did return at least one questionnaire were then adjusted to reflect nonresponse by the remaining households (minus those addresses returned by the post office). Eight nonresponse cells were created based on cross-classifications of Census Region and Stratum (high vs. low minority). The overall adjustment factor for this adjustment was 2.50 and ranged from a low of 2.23 to a high of 4.02.

Adjustments were made for the number of ways that a household can receive mail. This adjustment is analogous to the multiple telephone adjustment for the RDD survey. The mean adjustment for the number of ways a household receives mail was 1.09.

Finally, adjustments were made for within-household nonresponse. Each household was asked to have all adults in the household fill out a questionnaire and return it. However, for numerous households the number of questionnaires returned did not match the number of adults reported. Therefore, the weights were adjusted to reflect this within-household nonresponse. The mean adjustment factor for within-household nonresponse was 1.29.

7.4.4 Replicate Nonresponse Adjustment

Nonresponse adjustments are themselves random variables and contribute a variance component to the overall sampling variance. This variance component is represented in the final jackknife estimator by replicating the computation of nonresponse adjustments (by replacing the original base weights by the replicate base weights, and repeating the computations described in Sections 7.3 and 7.4).

The nonresponse adjustments are the reciprocals of weighted response rates. Replicate screener response rates were computed for each screener response cell a and each replicate r by removing the deleted set corresponding to each replicate r and recomputing the response rate. In other words, Westat recomputed response rates for each replicate set as if it were the original RDD sample or original address sample.

7.5 Calculation of Composite Weights

Composite weights were created for the combined data set that includes both the RDD and address samples. In creating the composite weights, those cases from the address sample that do not have a landline telephone were given a compositing factor of 1.0000—as the only way that they can participate in HINTS was through the mail questionnaire. For those cases that had a landline telephone number, a compositing factor was used. For the RDD sample respondents a composite factor of .5629 was used, while for the address sample respondents a composite factor of .4371 was used.

7.6 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.

7.6.1 Control Totals

The American Community Survey (ACS) of the U.S. Census Bureau has much larger sample sizes than those of HINTS. The ACS estimates of any U.S. population totals have lower sampling error than the corresponding HINTS estimates, making calibration of the survey weights to ACS control totals beneficial. The ACS estimates are available via the Internet. Westat used the 2006 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 CPS characteristics correlate well with HINTS questionnaire items:

- Age,
- Gender,
- Educational Attainment,
- Marital Status,
- Race,
- Ethnicity, and
- Census Region.

In addition to characteristics from the ACS, two health-related variables were used. These variables came from the 2006 National Health Information Survey (NHIS) and correspond to questions asked in the HINTS survey. They were:

- Percent With Health Insurance, and
- Percent Ever Had Cancer.

Raking to the control totals for these variables (either alone or cross-classified with each other) was then performed. The same control totals were used for each of the three (RDD, address, and composite) weights. As a result of the raking HINTS weights to the control totals, estimates calculated from HINTS data for the control-total variables agree with those calculated from the source data for the control totals. For example, the national-level estimate of Percent Ever Had Cancer calculated from HINTS 2007 data agrees with the estimate calculated from NHIS 2006 data.

8.1 RDD Sample

Nonresponse is a continually worsening problem in RDD telephone household surveys (see, for example, Atrostic et al., 2001). In the presence of nonresponse, the RDD sample can be seen only as a representative sample of the responding portion of the population (viewing all individuals in the population as belonging to a “responding” and a “nonresponding” population, supposing that the entire population is contacted). Any difference between this responding portion of the population and the nonresponding portion of the population can lead to a bias in the survey estimates as estimators of the full population. The magnitude of this bias is the product of the nonresponse rate and the difference in means between the responding and nonresponding populations (see, for example, Groves & Couper, 1998, Section 3.2).

Under this paradigm, the potential for nonresponse bias grows linearly with the nonresponse rate. The larger the nonresponse, the larger the bias is for a given difference in means between respondents and nonrespondents. For example, a relative difference of 10 percent in the means will lead to almost no bias when nonresponse is 10 percent, but would lead to a 5-percent relative bias with a response rate of 50 percent (Biemer & Lyberg, 2003, 84). Unfortunately, without conducting a special study of nonrespondents, it is not possible to know the differences between respondents and nonrespondents. In the absence of this information, the response rate is used as an indirect measure of data quality.

The American Association for Public Opinion Research (AAPOR) provides guidelines for computing response rates, which aims to standardize the computation of response rates across surveys (AAPOR, 2006). This will allow for legitimate comparisons of survey response rates as a measure of relative survey quality. However, even this standardization allows for considerable latitude in computation of these response rates, so that it is important to report the method of computing the response rate as well as the response rate itself.

The overall response rate is computed as a product of the screener response rate and the extended interview response rate.⁵

8.1.1 RDD Screener Response Rate

Following AAPOR standards, the screener response rate is equal to the sum of weights of cooperating households (eligible or not) divided by the sum of weights of residential numbers in the sample. The latter value is not completely known. In some cases, it is only possible to get an answering machine or voicemail (hereafter designated as NM). In other cases there are numbers for which there is no answer at all, even though the number rings as if the telephone number exists (hereafter designated as NA). AAPOR standards allow for considerable latitude in estimating the number of residential numbers among these two groups, requiring only a good faith effort to do this accurately. Westat's procedure is to estimate the residential rates (denoted as ' e ') for the NM and NA telephone numbers. The method used to estimate these values was the CASRO (1982) method. In this method, e is estimated as the proportion of the resolved telephone numbers that are observed to be residential.

Let C , I , REF , O , NM and NA , respectively, denote the summation of weights of completed screeners, ineligible, refusals, other residentials, answering machine, and ring no answers. Defining EM and EA as the percentage of residential numbers among the known working numbers and all numbers, respectively, the screener response rate $SCRNR$ is computed as

$$SCRNR = \frac{C + I}{C + I + REF + O + NM \times e_{NM} + NA \times e_{NA}}.$$

Note that this screener response rate is algebraically equivalent to

$$SCRNR = \frac{C}{C + e_R \times [I + REF + O + NM \times e_{NM} + NA \times e_{NA}]}$$

$$\text{with } e_R = C / (C + I).$$

⁵ Screener in this section refers to the portion of the questionnaire that identifies a sampled adult. Extended interview refers to the remaining substantive portions of the questionnaire.

The second form of *SCRNR* though algebraically more complicated is conceptually more transparent. The response rate is completes divided by the completes plus the estimated eligible numbers among the remaining residential number (refusals and NAs). Westat estimates the eligible among the estimated residential number $REF+O+NM*e_{NM}+NA*e_{NA}$ by estimating the eligibility rate from the “known eligibility status” numbers: the completes and ineligibles. In HINTS, the eligibility rate is quite high because all adults are eligible (only households with no adults are ineligible).

Table 8-1 presents the weighted estimates of e_{NM} and e_{NA} for the RDD screener. Four estimates were calculated. These estimates are based on whether the telephone number was an NA or an NM and by whether or not an address was found for the telephone number. As can be seen these estimates range from 5.1 percent for the no address NMs to 94.6 percent for the NMs with a mailing address, with the telephone numbers with a mailable address having higher residential rates than the nonmailable cases.

Table 8-1. Weighted estimates of percentages of residential telephone numbers that are residential in the HINTS 2007 RDD sample

Response class	Mailable	Nonmailable
Answering machine (NM)	94.6	24.3
Ring no answer (NA)	79.8	5.1

Table 8-2 presents the calculation of the response rate. The numerator includes the weights of complete and ineligible numbers. The denominator includes all known residential numbers, as well as estimated residential numbers from the NMs and the NAs.

Table 8-2. Screener response rate calculations for the HINTS 2007 RDD sample

Response class	Unweighted total	Weighted total	Estimated residential
Total sample	53,118	289,352,200	79,903,711
Respondents (C+I)	6,777	33,851,392	33,851,392
Nonrespondents (REF+O)	5,929	29,819,193	29,819,193
Nonresidential	34,271	193,315,044	-
Unknown (NM, NA)	6,141	32,366,571	16,233,127
RDD screener response rate			42.37%

8.1.2 RDD Extended Interview Response Rate

Table 8-3 shows the extended interview response rate calculations for the HINTS 2007 RDD sample. The extended interview response rate is designed to be an estimator of the percentage of people who would complete an extended interview, given that the household completed the screener, if the entire population was contacted. This is estimated by taking a summation of weights for completed extended interviews, divided by a corresponding summation of weights for all sampled adults within households with completed screeners. The appropriate weights are nonresponse-adjusted screener weights (screener base weights that include a multiple-telephone number adjustment, further adjusted for nonresponding screeners), multiplied by the probability of selection of the sampled adult within the household. Partial completes are included as completes in this calculation.

Table 8-3. Extended interview response rate calculations for HINTS 2007 RDD sample

Response class	Unweighted total	Weighted total
Total sample	6,760	151,310,930
Respondents	4,081	86,541,144
Nonrespondents	2,679	64,769,786
RDD extended response rate		57.19%

8.1.3 RDD Overall Response Rate

The overall response for the RDD sample is computed by taking the product of the screener and the extended interview response rates. The overall response rate is an estimator of the percentage of the overall population in which a completed interview would be obtained, if all households were canvassed. This is a product of the overall percentage responding to the extended interview conditional on responding to the screener (estimated by the extended interview response rate from Section 8.1.2). Table 8-4 presents the calculation of the overall response rate.

Table 8-4. Overall response rate calculations for HINTS 2007 RDD sample

Response type	Rate
RDD screener	42.37%
RDD extended	57.19%
Overall response rate	24.23%

8.2 Address-Sample Response Rate

Nonresponse for the address sample occurs in one of two ways. First there is household nonresponse. The household response rate is based on those cases where at least one completed survey was returned from a given household, adjusting for those addresses that were returned to us as undeliverable. The household response rate is described in Section 8.2.1.

Unlike the RDD sample, we could not sample from the adults in the household for the address sample. So for the address sample we asked that each adult fill out a questionnaire. Three questionnaires were provided to each household. In addition, as a part of the survey we asked that the respondent record the number of adults in the household. Comparing the number of household returns to the number of adults in the household (as recorded in the survey) gives the within-household response rate. This is discussed in Section 8.2.2.

8.2.1 Address-Sample Household Response Rate

Table 8-5 shows the household response rate calculation for the address sample. This data has been weighted to account for the oversampling of addresses in high-minority areas. It can be seen in Table 8-5 that not only does the high-minority stratum have the lower deliverable rate (89.38% vs. 91.77%), it also has a significantly lower response rate (28.41% vs. 43.78%) when compared to the low-minority stratum.

Table 8-5. Household response rate calculations for the HINTS 2007 address sample

Response class	High minority	Low minority	Overall
Total sample	31,845,829	96,108,551	127,954,380
Respondents	8,087,186	38,614,824	46,702,010
Nonrespondents	20,376,139	49,584,666	69,960,806
Undeliverable	3,382,504	7,909,060	11,291,564
Households	28,463,325	88,199,490	116,662,816
Percent deliverable	89.38%	91.77%	91.18%
Household response rate	28.41%	43.78%	40.03%

8.2.2 Within-Household Response Rate

Within a household all adults were asked to respond. Unlike Table 8-5, which was a household-level estimate, Table 8-6 shows the person-level rates. The sum of the weights of those who completed a questionnaire are compared to the sum of the weights of the self-reported number of adults in a household to arrive at the within-household response rate. Unlike the household-level response rate, the within-household response rate is quite consistent across stratum. In fact, the within-household response rate for the high-minority stratum is slightly higher than for the low-minority stratum.

Table 8-6. Weighted within-household response rate calculations for HINTS 2007 address sample

Response class	High minority	Low minority	Overall
Respondents	37,249,540	118,979,026	156,228,566
Adults	47,953,263	153,877,641	201,830,904
Within-household response rate	77.68%	77.32%	77.41%

8.2.3 Overall Response Rate

The overall response for the address sample is computed by taking the product of the household and the within-household (person-level) response rates. The overall response rate is an estimator of the percentage in the overall population in which a completed interview would be obtained if all households were canvassed. Table 8-7 presents this calculation.

Table 8-7. Overall response rate calculations for HINTS 2007 address sample

Response type	Rate
Household	40.03%
Within-household	77.41%
Overall response rate	30.99%

References

- American Association for Public Opinion Research [AAPOR]. (2006). *Standard definitions: Final dispositions of case codes and outcome rates for surveys. Fourth edition*. Lenexa, Kansas: AAPOR.
- Atrostic, B.K., Bates, N., Burt, G., and Silberstein, A. (2001). Nonresponse in U.S. government household surveys: consistent measures, recent trends, and new insights. *Journal of Official Statistics*, 17, 209-226.
- Battaglia, M.P., Link, M.W., Frankel, M.R., and Mokhad, A.H. (2005). An evaluation of respondent selection methods for household mail surveys. *Proceedings of the Section on Survey Research Methods* (pp. 2727-2731), *American Statistical Association*.
- Biemer, P.P., and Lyberg, L.E. (2003). *Introduction to survey quality*. New York: John Wiley.
- Council of American Survey Research Organizations. (1982). *Special report: On the definition of response rates*. Port Jefferson, NY: CASRO.
- 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. (2000). *Mail and Internet surveys: The Tailored Design Method. 2nd Edition*. New York: John Wiley Co.
- Groves, R.M., and Couper, M.P. (1998). *Nonresponse in household surveys*. New York: John Wiley.
- Hauser, R.M. (2005). Survey response in the long run: The Wisconsin Longitudinal Study. *Field Methods*, 17(1), 3-29.
- Hembroff, L.A., Ruzs, D., Rafferty, A., McGee, H., and Ehrlich, N. (2005). The cost-effectiveness of alternative advance mailings in a telephone survey. *Public Opinion Quarterly*, 69, 232-245.
- Iannacchione, V., Staab, J.M., and Redden, D.T. (2003). Evaluating the use of residential mailing addresses in a metropolitan household survey. *Public Opinion Quarterly*, 67, 202-210.
- Link, M.W., Battaglia, M.P., Frankel, M.R., Osborn, L., and Mokdad, A.H. (2008). A comparison of address-based sampling (ABS) versus random-digit-dialing (RDD) for general population surveys. *Public Opinion Quarterly*, 72, 6-27.
- Link, M.W., Battaglia, M.P., Giambo, P., Frankel, M.R., Mokhad, A.H., and Rao, S.R. (2005, May). *Assessment of address frame replacements for RDD sampling frames*. Paper presented at the 2005 American Association for Public Opinion Research Annual Conference, Miami, FL. Available at www.m-s-g.com/Reference_Library/Address_Frame_Replacements.htm.

- Link, M. and Mokdad, A. (2004). *Are Web and mail modes feasible options for the Behavioral Risk Factor Surveillance System?* Paper presented at the Health Survey Research Methods Conference.
- 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.
- Salant, P., and Dillman, D.A. (1994). *How to conduct your own survey*. New York: John Wiley Co. (Alternate selection, Newbridge Book Club; The Executive Program.)
- Staab, J.M. and Innacchione, V.G. (2003). Evaluating the use of residential mailing addresses in a national household survey. *Proceedings of the Section on Survey Research Methods (pp. 4028-4033)*, American Statistical Association.
- Tucker, C., Casady, R., and Lepkowski, J. (1993, May). *A hierarchy of list-assisted stratified telephone sample design options*. Paper presented at the Annual Conference of the American Association for Public Opinion Research, St. Charles, IL.
- United States Postal Service, (2006). *CDS User Guide*. Retrieved April 20, 2006, from <http://ribbs.usps.gov/files/CDS/CDSug.pdf>.

Appendix A

RDD Pilot Study Letters and Introductions



ADVANCE LETTER A

Dear Sir or Madam:

I'm writing to ask 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 has interviewed thousands of people in the last few years. From it we've learned that:

- About 4 out of 5 adults believe that there are so many recommendations about nutrition that it is hard to know which ones to follow.
- About one in four adults read the health section of a newspaper or magazine every week.
- Almost half of all adults don't know the age at which to begin screening for certain types of cancer.

With information like this, the survey can help the government and companies better communicate health information to everyone.

Your household was chosen at random for this survey and cannot be replaced. An interviewer from Westat, a research firm under contract with the U.S. Department of Health and Human Services, will call within the next few weeks. It will take perhaps 20 to 30 minutes to answer our questions. What you have to say will help us find out how we can best provide the health information people need. Your answers will be kept confidential to the extent provided by law. More information about the study is provided on the other side of this letter.

Thank you in advance for your cooperation. If you have any questions about the study or to schedule your interview, call Westat toll-free at 1-888-314-1133.

Sincerely,

A handwritten signature in blue ink that reads "Bradford W. Hesse".

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-314-1133

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

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

Q: How did you get my (unlisted) telephone number?

A: Your number was randomly selected from among all of the possible telephone numbers in the nation. It was selected using scientific sampling methods. If your number was unlisted, it still is.

Q: How did you get my address?

A: An independent organization matched a list of published addresses to the randomly selected telephone numbers included in the sample for this survey. This letter was sent to every address that was matched with a telephone number in the sample. Address information is kept confidential and will be destroyed as soon as the survey is completed.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours, and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

Q: How long will the survey interview last?

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.



ADVANCE LETTER B

Dear Sir or Madam:

Americans today have access to more health information than ever. However, we don't always know how to find what we really need to know. I hope you'll take part in an important survey sponsored by the U.S. Department of Health and Human Services. By taking part, you can help us do a better job of meeting the health information needs of the nation.

Your household was chosen at random for the Health Information National Trends Survey and cannot be replaced. An interviewer from Westat, a research firm under contract with the U.S. Department of Health and Human Services, will call within the next few weeks. It will take perhaps 20 to 30 minutes to answer our questions. What you have to say will help us find out how we can best provide the health information people need. Your answers will be kept confidential to the extent provided by law. More information about the study is provided on the other side of this letter.

Thank you in advance for your cooperation. If you have any questions about the study or to schedule your interview, call Westat toll-free at 1-888-314-1133.

Sincerely,

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-314-1133

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

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: What is the study about? What kind of questions will you be asking?

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 people.

Q: How did you get my (unlisted) telephone number?

A: Your number was randomly selected from among all of the possible telephone numbers in the nation. It was selected using scientific sampling methods. If your number was unlisted, it still is.

Q: How did you get my address?

A: An independent organization matched a list of published addresses to the randomly selected telephone numbers included in the sample for this survey. This letter was sent to every address that was matched with a telephone number in the sample. Address information is kept confidential and will be destroyed as soon as the survey is completed.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours, and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

Q: How long will the survey interview last?

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.

TELEPHONE INTRODUCTIONS

Introduction A: Hello, this is {INTERVIEWER NAME} and I'm calling for the U.S. Department of Health and Human Services for a national study on people's needs for health information.

Introduction B: Hello, this is {INTERVIEWER NAME} and I'm calling for the U.S. Department of Health and Human Services for a national health study.

Appendix B

RDD Main Study Advance Letter



Dear Sir or Madam:

I'm writing to ask 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 has interviewed thousands of people in the last few years. From it we've learned that:

- About 4 out of 5 adults believe that there are so many recommendations about nutrition that it is hard to know which ones to follow.
- About one in four adults read the health section of a newspaper or magazine every week.
- Almost half of all adults don't know the age at which to begin screening for certain types of cancer.

With information like this, the survey can help the government and companies better communicate health information to everyone.

Your household was chosen at random for this survey and cannot be replaced. An interviewer from Westat, a research firm under contract with the U.S. Department of Health and Human Services, will call within the next few weeks. It will take perhaps 20 to 30 minutes to answer our questions. What you have to say will help us find out how we can best provide the health information people need. Your answers will be kept confidential to the extent provided by law. More information about the study is provided on the other side of this letter.

We know that your time is valuable. We've enclosed \$2 as a token of our appreciation for your participation. Of course, your participation is voluntary. You can keep the money even if you decide not to take part in the survey.

Thank you in advance for your cooperation. If you have any questions about the study or to schedule your interview, call Westat toll-free at 1-888-314-1133.

Sincerely,

A handwritten signature in blue ink that reads "Bradford W. Hesse".

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-314-1133

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: What is the study about? What kind of questions will you be asking?

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 people.

Q: How did you get my (unlisted) telephone number?

A: Your number was randomly selected from among all of the possible telephone numbers in the nation. It was selected using scientific sampling methods. If your number was unlisted, it still is.

Q: How did you get my address?

A: An independent organization matched a list of published addresses to the randomly selected telephone numbers included in the sample for this survey. This letter was sent to every address that was matched with a telephone number in the sample. Address information is kept confidential and will be destroyed as soon as the survey is completed.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours, and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

Q: How long will the survey interview last?

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.

Appendix C

RDD Information Request Letter



Dear Sir or Madam:

Thank you for your interest in the Health Information National Trends Survey (HINTS). As requested, I am enclosing an informational brochure about the study. For more detailed information about HINTS, I encourage you to visit the HINTS website at: <http://hints.cancer.gov>

We appreciate your participation in HINTS. Your response to the survey plays a critical role in helping us find out how we can best provide the health information people need. If you have any additional questions or concerns about the study, please feel free to call Westat toll-free at 1-888-314-1133.

Sincerely

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

Appendix D

RDD Screener Refusal Conversion Letter



Dear Sir or Madam:

We ask for your household's participation in this year's Health Information National Trends Survey. This telephone survey is an important opportunity to have your voice heard on health issues.

This research helps America learn more about the health of its people and the problems they have getting accurate health information. The results will be used to help keep Americans better informed on important health issues.

We recently called your home, but it was not a good time for anyone to speak with us. We know your time is valuable. However, your household is part of a scientific sample representing many other households like yours and your household cannot be replaced. Please take a moment to take our call. If we happen to call at an inconvenient time, you can suggest a time that is better for you.

If you have any questions, please call Westat toll free at 1-888-314-1133.

Sincerely,

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-314-1133

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

Q: How did you get my (unlisted) telephone number?

A: Your number was randomly selected from among all of the possible telephone numbers in the nation. It was selected using scientific sampling methods. If your number was unlisted, it still is.

Q: How did you get my address?

A: An independent organization matched a list of published addresses to the randomly selected telephone numbers included in the sample for this survey. This letter was sent to every address that was matched with a telephone number in the sample. Address information is kept confidential and will be destroyed as soon as the survey is completed.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours, and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

Q: How long will the survey interview last?

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.

Appendix E

RDD Extended Refusal Conversion Letter



Dear Sir or Madam:

We ask for your household's participation in this year's Health Information National Trends Survey. This telephone survey is an important opportunity to have your voice heard on health issues.

This research helps America learn more about the health of its people and the problems they have getting accurate health information. The results will be used to help keep Americans better informed on important health issues.

We recently called your home, but it was not a good time for anyone to speak with us. We know your time is valuable. However, your household is part of a scientific sample representing many other households like yours, and your household cannot be replaced. Please take a moment to take our call. If we happen to call at an inconvenient time, you can suggest a time that is better for you.

If you have any questions, please call Westat toll free at 1-888-314-1133.

Sincerely,

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-314-1133

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

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

Q: How did you get my (unlisted) telephone number?

A: Your number was randomly selected from among all of the possible telephone numbers in the nation. It was selected using scientific sampling methods. If your number was unlisted, it still is.

Q: How did you get my address?

A: An independent organization matched a list of published addresses to the randomly selected telephone numbers included in the sample for this survey. This letter was sent to every address that was matched with a telephone number in the sample. Address information is kept confidential and will be destroyed as soon as the survey is completed.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours, and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

Q: How long will the survey interview last?

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.

Appendix F

Sample of Production Report by Release Group

HINTS III Cooperation, Conversion, and Response Rates by Release Group
Final as of 4-28-08

SMPTYP	Rel'd	Called	Screener									Extended Interview					Total											
			# Weeks Ago			Refusal Conversion			Response Rate			Refusal Conversion			Response Rate													
			InitCoop	1st	2nd	n	# NA	% nr/nw	X2ndC	Final	NA Adj	InitCoop	1st	2nd	n	X2ndC	Final	Final RR										
10	15	13	40.0%	20.2%	11.8%	5,572	617	27.17%	37.83%	40.54%	42.28%	59.7%	20.0%	13.0%	1,638	55.49%	56.84%	24.03%										
20	13	12	40.2%	19.8%	14.7%	2,822	362	26.54%	35.22%	38.59%	40.47%	62.1%	23.9%	16.4%	798	55.76%	57.02%	23.07%										
21	12	11	42.0%	20.8%	12.9%	2,842	361	25.37%	34.52%	36.96%	38.63%	60.9%	23.7%	11.3%	783	55.17%	56.32%	21.76%										
22	11	10	41.6%	20.2%	13.3%	2,832	383	24.75%	32.71%	34.91%	36.54%	60.6%	22.7%	21.2%	744	53.23%	54.70%	19.99%										
23	10	9	40.8%	18.1%	14.8%	2,879	354	27.06%	34.25%	37.03%	38.70%	60.7%	22.1%	12.5%	801	51.94%	52.93%	20.49%										
24	9	8	41.1%	19.2%	12.9%	2,740	352	26.20%	32.22%	34.47%	36.12%	64.0%	19.3%	20.5%	695	54.96%	56.12%	20.27%										
25	8	7	41.6%	16.3%	<i>n/a</i>	2,864	378	24.62%	<i>n/a</i>	31.50%	32.91%	65.9%	29.5%	12.5%	679	54.79%	55.38%	18.23%										
26	7	6	38.2%	19.7%	<i>n/a</i>	2,812	413	24.61%	<i>n/a</i>	29.43%	30.92%	62.8%	21.7%	2.0%	622	53.70%	53.86%	16.65%										
											25363									6760								

Screener n represents sample size - NB, NT results; Extended n represents # of extended records generated by completed screeners.

Table created after 4/28/08 when data collection was completed.

For SMPTYP = 25 and 26, no screener 2nd refusal conversion was attempted (subsampling).

Final weighted response rate will use 2nd conversion results from SMPTYP 10 through 24 in place of SMPTYP 25 and 26.

Shaded cells represent response rates which will be increased by final weights.

Refusal conversion rates reflecting fewer than 60 cases in the denominator are shown in red italics.

Appendix G

Sample Weekly TRC Report From NCI

HINTS 2007 TRC Report

As of: 4/25/2008 (7:00am)

	RDD Sample	
	Number	Percent
SCREENER		
Total Screener Sample for TRC	25,363	
Screeners Called to date	25,363	100.0%
Final Screener Results*		
Completes & Ineligibles (CS/IE)	6,720	26.5%
Final Refusals (RB/RM/R2/R3/RX)	4,280	16.9%
Non-contact (NA/NM)	1,949	7.7%
Max Calls (LM/MC/ML)	210	0.8%
Other Nonresponse (LP/LH/NO)	124	0.5%
Nonworking/Nonresidential (NR/NW)	6,441	25.4%
Total screeners called to date	25,363	100.0%
Response Rate (C+I/(sample-NR,NW))		35.5%
Cooperation and Conversion		
<i>Initial Screener Cooperation Rate</i>		
Initial Refusals (2's)	7,371	
Initial Completes (pre conversion)	5,118	
Initial Cooperation Rate ((C+I)/(C+I+2))		41.0%
<i>Screener Refusal Conversion Rate (total)</i>		
Final Refusals (RB/RM/R3)	4,557	
Completes from conversion	1,650	
Screener Conversion Rate ((C+I)/(C+I+R))		26.6%

* screeners with interim (numeric) result codes are not reflected in these categories

HINTS 2007 TRC Report

As of: 4/25/2008 (7:00am)

		RDD Sample	
		Number	Percent
EXTENDED INTERVIEW			
Sample Size		6,702	
Final Extended Interview Results*			
Completes		3,703	55.3%
Final Refusals (RB/RM/R3/RX)		1,042	15.5%
Max Calls (LM/MC/ML/MR)		221	3.3%
Other Nonresp (LP/LH/NS/NP/NF/NL/NO/NR/NW/OE)		368	5.5%
Total Extended Sample		6,702	100.0%
Response Rate (C/extended sample)			55.3%
Cooperation and Conversion			
<i>Initial Extended Cooperation Rate</i>			
Initial Refusals (2's)		2003	
Initial Completes (pre conversion)		3243	
Initial Cooperation Rate (C/(C+2))			61.8%
<i>Extended Refusal Conversion Rate (total)</i>			
Final Refusals (RB/RM/R3)		1010	
Completes from conversion		471	
Extended Conversion Rate (C/(C+R))			31.8%

* extendeds with interim (numeric) result codes are not reflected in these categories

Appendix H

Mail Advance Letters, Cover Letters, and Postcard



ADVANCE LETTER

Dear Sir or Madam:

I'm writing to ask 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 has interviewed thousands of people in the last few years. From it we've learned that:

- About 4 out of 5 adults believe that there are so many recommendations about nutrition that it is hard to know which ones to follow.
- About one in four adults read the health section of a newspaper or magazine every week.
- Almost half of all adults don't know the age at which to begin screening for certain types of cancer.

With information like this, the survey can help the government and companies better communicate health information to everyone.

Your household was chosen at random for this survey and cannot be replaced. You will receive study questionnaires in the mail within the next few days. It will take perhaps 20 to 30 minutes to answer our questions. What you have to say will help us find out how we can best provide the health information people need.

Westat, a research firm under contract with the U.S. Department of Health and Human Services, is administering the survey. Your answers will be kept confidential to the extent provided by law. More information about the study is provided on the other side of this letter.

Thank you in advance for your cooperation. If you have any questions about the study, call Westat toll-free at 1-888-636-6540.

A handwritten signature in blue ink that reads "Bradford W. Hesse".

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-636-6536.

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

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.



COVER LETTER – FIRST MAILING

Dear Sir or Madam:

I'm writing to ask 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 has interviewed thousands of people in the last few years. From it we've learned that:

- About 4 out of 5 adults believe that there are so many recommendations about nutrition that it is hard to know which ones to follow.
- About one in four adults read the health section of a newspaper or magazine every week.
- Almost half of all adults don't know the age at which to begin screening for certain types of cancer.

With information like this, the survey can help the government and companies get valuable information on health to you and your family.

Your household was chosen at random for this survey and cannot be replaced. **We ask that each adult in this household complete a questionnaire and return it to us in the postage-paid envelope within the next two weeks.** What you have to say will help us find out how we can best provide the health information people need. We know that your time is valuable. We've enclosed \$2 as a token of our appreciation for your participation. Of course, your participation is voluntary. You can keep the money even if you decide not to take part in the survey.

Westat, a research firm under contract with the U.S. Department of Health and Human Services, is administering the survey. Your answers will be kept confidential to the extent provided by law. More information about the study is provided on the other side of this letter.

Thank you in advance for your cooperation. If you have any questions about the study or you would like to request more questionnaires, please call Westat toll-free at 1-888-636-6540.

Sincerely,

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-636-6536.

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

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

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.

REMINDER POSTCARD TEXT:

A few days ago you should have received a packet of questionnaires for your household's participation in the Health Information National Trends Survey. It was sent to your address as part of our effort to determine the best ways of communicating important health information to all Americans.

We requested that each adult in the household complete a questionnaire. If you have already returned the questionnaire(s) to us, please accept my sincere thanks. If any adult here has not yet completed and returned a questionnaire, please do so as soon as possible.

Your participation is important to the study's success.

Sincerely,

A handwritten signature in black ink that reads "Bradford W. Hesse". The signature is written in a cursive style with a large initial 'B'.

Bradford W. Hesse, Ph.D.

HINTS Project Officer

Chief, Health Communication and Informatics Research Branch

National Institutes of Health

U.S. Department of Health and Human Services



COVER LETTER – SECOND MAILING

Dear Sir or Madam:

We ask for your household's participation in this year's Health Information National Trends Survey. This survey is an important opportunity to have your voice heard on health issues.

This research helps America learn more about the health of its people and the problems they have getting accurate health information. The results will be used to help keep Americans better informed on important health issues.

We recently mailed questionnaires to your home. As of today we have not received any completed questionnaires from this home. We know your time is valuable. However, your household is part of a scientific sample representing many other households like yours and it cannot be replaced. We need to hear from you.

In the event that your questionnaires were misplaced, replacements are enclosed. **We ask that each adult in this household complete a questionnaire and return it to us in the postage-paid envelope within the next two weeks.**

If you have any questions or would like to request additional questionnaires, please call Westat toll free at 1-888-636-6540.

Sincerely,

A handwritten signature in blue ink that reads "Bradford W. Hesse".

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

Si prefiere recibir la encuesta en Español, por favor llame 1-888-636-6536.

**Some Frequently Asked Questions
about the
Health Information National Trends Survey**

Q: What is the study about? What kind of questions will you be asking?

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 people.

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. Your household was selected randomly using scientific sampling methods, in order to reach a sample that reflects the entire population of the United States. You represent thousands of other households like yours and you cannot be replaced. Your answers and opinions are very important to the success of this study, as you represent others who share your knowledge and beliefs.

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

A: Yes. Your answers will not be revealed to anyone but the researchers in a way that identifies you or your household, to the extent provided by law.

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.

Appendix I

Decisions for Combining CATI and Mail Data

DECISIONS FOR COMBINING CATI AND MAIL DATA

Item	Issue	Resolution
Unit/number items such as BR06, BR08 and CS21	Mail and CATI instruments collected the number part of this question differently.	CATI data were update to match mail data: The format for the mail data was used, with one unit and one number. CATI data was collapsed to fit into this single number variable.
Income questions HD15 and HD15a-j	Income was collected differently on the mail and CATI instruments.	The open-ended data on the CATI were maintained (question HD15) while the data in items HD15a-j were collapsed into the categories provided on the mail instrument.
Sun exposure questions BR25a-d	In the CATI, a respondent that answered '95' to item BR25a was skipped to BR26. No such skip existed on the mail instrument.	CATI data were updated to match mail data: if BR25a = 95, the remaining BR25 items were recoded to '95.'
Helpline question BR52	In the CATI, length of time as a non-smoker (BR37) was used as a skip. There was no such skip in the mail instrument.	Mail data were updated to conform with the CATI skip: if a respondent was a nonsmoker for a year or more, the answer to item BR52 was recoded to blank.
Internet question HC17	In the CATI, respondents who answered "home" in HC16 were skipped out of HC17. There was no such skip in the mail instrument.	Mail data were updated to match CATI data: item HC17 was updated to blank if the respondent selected "home" in item HC16.
CAM question HS06	On the mail instrument, questions were presented in a different order than on the CATI such that mail respondents that answered "none" for HS03 still answered HS06.	Mail data for item HS06 was updated to match the CATI data: if HS03=0, then HS06 was coded to blank.
Doctor/Internet question HS09	Unlike the CATI, on the mail instrument, there was no skip pattern depending on Internet use (Item HC15).	Mail data was updated to match CATI data: if HC15 was "no" or blank, then HS09 was recoded to blank.

Item	Issue	Resolution
Doctor/Internet question HS10	Unlike the CATI, on the mail instrument, there was no skip pattern depending on Internet use (Item HC15).	Mail data was updated to match CATI data: if HC15 was “no” or blank, then HS10 was recoded to blank.
Genetic test question HS23	On the CATI, HS23 was skipped if the respondent answered “no” to HS22. No such skip existed on the mail instrument.	Mail data was updated to match the CATI data: if HS22 was “no” or blank, HS23 was recoded to blank.
Smoking questions BR30 and BR32	The mail instrument routes sometimes smokers and everyday smokers to the same question, while the CATI instrument separates them.	Mail data was updated to match CATI data with the 2 types of smokers separated.
Exercise question BR08	On the CATI instrument, respondents that answer “0” to item BR07 and skipped from item BR08.	Mail data was updated to match CATA data: if a respondent answered “0” on item BR07, then item BR08 was updated to blank.
HPV question BR64	On the CATI instrument, respondents who are male or have not heard of HPV are not asked BR64. Neither group is excluded on the mail instrument.	The mail and CATI variables were retained and a new additional variable was created. Mail data were recoded using the CATI exclusion criteria for the new variable.
HPV question BR70	On the CATI instrument, male respondents are not asked this question. There is no such skip on the mail instrument.	The mail and CATI variables were retained and a new additional variable was created. Mail data were recoded using the CATI criteria for the new variable.
Household question HD14	On the CATI instrument, HD14 is skipped if the respondent answers “0” to question HD13. There is no such skip on the mail instrument.	The mail data were updated to match CATI data: if HD13 was either “0” or was blank, then HD14 was recoded to blank.