Health Information National Trends Survey 2005 (HINTS 2005)

FINAL REPORT

November 2005

Authors:
David Cantor, Ph.D
James Covell
Terisa Davis, MPH
Inho Park, Ph.D
Louis Rizzo, Ph.D

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

Prepared by:
Westat
1650 Research Boulevard
Rockville, MD 20850

TABLE OF CONTENTS

List of Tables

List of Figures

List of Appendices

1. INTRODUCTION

The National Cancer Institute (NCI) has the vital mission of facilitating the process by which cancer research is communicated to the public. The task of collecting data relevant to cancer communication falls on the Health Communication and Informatics Research Branch (HCIRB), Division of Cancer Control and Population Science (DCCPS) at NCI. The HCIRB seeks to advance communication and information science across the cancer continuum—prevention, detection, treatment, control, survivorship, and end of life. The primary goals for the HCIRB are (1) to encourage programmatic and interdisciplinary approaches to cancer communication research and (2) to accelerate development of innovative health communication models, theories, and research strategies in cancer prevention, control, and care.

1.1 Background

To address these goals, the NCI funded the first Health Information National Trends Survey (HINTS), which was conducted in 2003. The NCI, together with its funded partner Westat, administered the first of what is an ongoing, cross-sectional survey of the U.S. civilian, noninstitutionalized, adult population. The survey used a probability-based sample, drawing on random digit dialing (RDD) telephone numbers as the sample frame of highest penetration at the time. The purpose of funding a national probability survey to assess health communication processes was to provide communication researchers with unbiased population estimates of the prevalence of cancer-relevant knowledge, attitudes, and behaviors in the U.S. adult (18+) population. HINTS 2003 was designed to minimize errors in coverage, sampling, and measurement (Dillman, 1999). Moreover, by aligning the content of the survey with emerging theories of media usage (Viswanath & Finnegan, 1996), risk information processing (Fischhoff, Bostrom, & Quadrel, 1993; Croyle & Lerman**,** 1999), behavior change (Weinstein, 1993), health communication (Glanz, Lewis, & Rimer, 1997), and the diffusion of innovations (Rogers, 1995), population scientists should not only be in a more effective position for planning population-based interventions but they should also be in a position of using data from the survey to refine the scientific knowledge base.

The second survey for HINTS, or HINTS 2005, began in early 2005. HINTS 2005 had two priorities: (1) to preserve the methodological integrity of the survey and (2) to experiment with alternative

modes of data collection. Preserving the integrity of HINTS was done through two primary mechanisms. First, a similar methodology was used to draw the sample and collect the data (i.e., a national RDD survey of the general adult population). Second, approximately 50 percent of the content from HINTS 2003 was retained for HINTS 2005. Some of the critical content included "marker" items that serve as links to other national surveys (e.g., the Behavioral Risk Factor Surveillance System and the National Health Interview Survey) in health and communication. In addition, some items were retained to allow HINTS to monitor trends in the population over time in cancer-related health communication.

The second priority for HINTS 2005 was to experiment with incentives and alternative modes of data collection. A primary reason for this priority is the decline in response rates to telephone-based surveys due to a rise in telemarketing, an increase in tools for screening calls, and a proliferation of alternative modes of communication (e.g., mobile telephones and instant messaging). In addition to the RDD portion of the sample that was administered over the telephone to achieve the methodological integrity goal of HINTS, several experiments were conducted. One was to provide respondents with a small monetary incentive to complete the interview. The rationale for this was to provide respondents with a clearer sense of the importance of the survey and to get them to focus on the possible benefits of participating.

For the second experiment, data were collected through a web-based version of the instrument. In theory, the World Wide Web offers the advantages of computer assisted interviewing (CAI) in combination with those of self-administered questionnaires (SAQs). The web offers full control over multiple aspects of the survey's design and administration. Survey designers can manipulate the presentation of material; questions can be presented simultaneously in a single scrollable view, or they can be presented sequentially following simple or complex logic patterns (Couper, 2000; Heerwegh & Loosveldt, 2002). Designers also have control over color and graphics, and, with the spreading availability of high-speed Internet connections, designers have the ability to use sound, motion, and video to illustrate difficult concepts. In addition, designers can embed quality control checks on data entry and can reduce costs of data processing. As with SAQs, the disinhibiting nature of web-based data collection has the advantage of reducing social desirability effects to promote candid reporting of controversial items.

While Internet surveys have proliferated over the past few years, their external validity has been criticized for several reasons. Although penetration rates for the Internet have increased dramatically over the past few years (Katz & Rice, 2001), overall penetration is still much lower than the telephone or

postal mail (Couper, 2000). As a result, coverage of the sampling frame for households in the United States is less complete when using the Internet to reach respondents than are sampling frames based on RDD or area probability samples. The coverage problem becomes even more of an issue when considering the effects of the so-called "digital divide," in which respondents who are better educated and who make more money are more likely to have Internet access than those who are less socioeconomically advantaged.

Regardless of coverage issues, sampling bias poses an additional problem for the external validity of findings from Internet surveys. The majority of web-based surveys collect data opportunistically from nonprobability samples. In other words, they ask for volunteers that have visited a web site to fill out a questionnaire. As with any study that uses volunteers as respondents, there are likely to be differences between the people who are motivated enough to complete the survey and those who are not. Therefore, the ability to generalize the results of these surveys to a larger population is questionable.

To partially address these problems, an alternative approach to data collection is to pursue a "mixed-mode" or hybrid methodology. In this approach, traditional RDD sampling methods are used to construct an initial pool of potential respondents; then sampled persons (SPs) are given a choice of their preferred modality (Dillman, et al., 2001).

More research is needed to understand better the differences between Internet and RDD data collection. Chang and Krosnick (2003) have offered data comparing these two methodologies. They compared the results of three surveys conducted using three different methods. The three surveys included a traditional RDD screener followed by an RDD extended interview, a mixed RDD screener followed by an Internet extended interview, and a full Internet screener followed by an Internet extended interview. The results were equivocal. In terms of response rates, the RDD-RDD method had higher response rates than the RDD-Internet condition (response rates cannot be calculated for the Internet-Internet condition because it is a nonprobability sample). However, measures of data quality indicated that data collected on the Internet were less susceptible to social desirability than those collected through the RDD survey. The HINTS provided an opportunity to examine these differences through additional research.

1.2 Organization of the Report

This report details the design and implementation of HINTS 2005. Chapter 2 discusses how the HINTS questionnaire was constructed, including development of new content areas; the cognitive and field testing of the instrument; and the final questionnaire structure. Chapter 3 reviews the sampling plan for the study. Topics in Chapter 3 include within-household sampling, oversampling minorities, sample weights and variance estimation, and nonresponse issues. Chapter 4 contains details on data collection procedures and operations, as well as a discussion of response rates. Chapter 5 discusses conclusions to be drawn from HINTS 2005.

2. DEVELOPMENT OF THE HINTS 2005 QUESTIONNAIRE

The goals of the HINTS 2005 questionnaire were similar, but not identical, to those of HINTS 2003. In HINTS 2003, the survey was primarily intended to be used as an assessment of how the general population currently accesses and uses current communication channels to obtain health information. In addition, HINTS 2003 was designed to collect baseline data on cancer knowledge, attitudes, and behaviors. The HINTS 2005 questionnaire aimed to build on this foundation and, in addition, to focus on several specific topics within the field of health communications: scanning and seeking health information; self-efficacy in information seeking; community and social support; patient-provider communication; and mental models of cancer. As with HINTS 2003, HINTS 2005 had a secondary goal of continued surveillance of cancer-related knowledge and behavior. It is anticipated that repeated iterations of HINTS will (1) enable researchers to track the success of national intervention programs designed to improve access to information, and, at the same time, (2) enable researchers to track changes in cognitive and behavioral outcomes.

This chapter explains the process used to design the data collection instrument for HINTS 2005. It includes the review of HINTS 2003 items, the development of new items, the steps taken to select and revise items for HINTS 2005, and an outline of the resulting content of the questionnaire.

2.1 Question Development

The development of the questions for inclusion in the HINTS 2005 instrument involved both a review of HINTS 2003 items and the development of new items to address evolving NCI research interests. Westat worked closely with National Cancer Institute (NCI) investigators throughout the review of established HINTS items and the development of new measures. Although Westat provided advice and guidance, all decisions about inclusion, exclusion, or alteration of HINTS items were made by NCI investigators.

2.1.1 Review of HINTS 2003 Items

The goal of the instrument development team was to ensure that 50 percent of the HINTS 2003 items remained in HINTS 2005 to provide continuity to data collection efforts. A number of analyses were conducted to inform the team which items should be considered for exclusion from HINTS 2005. These included a psychometric analysis, an analysis of "don't know" and refused items, an examination of breakoff interviews, and systematic observations.

2.1.1.1 Psychometric Analysis

HINTS 2003 data were examined for the variability in responses for various items (i.e., dichotomous, nominal, ordinal and continuous) under the premise that low variability items might be identified for potential exclusion in HINTS 2005. In addition, correlations among subitems within a lead question were examined, as well as correlations among various sets of related items. Results of these tests indicated that there were only a few items that could be nominated for exclusion based on a lack of variability, although several redundant items were identified.

2.1.1.2 Don't Know and Refusal Analysis

The examination of "don't know" and "refused" response options was conducted under the assumption that items with a high rate of "don't know" or "refused" responses might be identified for potential exclusion from HINTS 2005. A few items were identified for possible exclusion through this examination.

2.1.1.3 Interview Breakoffs

A table was provided to NCI that outlined where breakoffs occurred during HINTS 2003 data collection to identify potential problem areas in the HINTS 2003 instrument. Analysis determined that there was no obvious pattern related to the breakoffs. As one might expect, the breakoffs tended occur at the beginning of the questionnaire (e.g., in health communications).

2.1.1.4 Systematic Observations

The HINTS 2003 electronic instrument was retrieved from the archive and interviewers at the Telephone Research Center (TRC) conducted nine HINTS 2003 interviews. These interviews were recorded and then reviewed by project staff. Each interview was systematically coded and then summarized. These systematic observations resulted in a few specific recommendations for changing questions in the HINTS 2005 instrument.

2.1.2 Development of New Content Areas

In meetings with NCI, several content areas were identified for addition or increased emphasis in HINTS 2005. These topics included patient-provider interaction, scanning and seeking health information, self-efficacy in information seeking, community and social support, and mental models of cancer. The category "mental models of cancer" included four subcomponents: (1) preventability; (2) detectability; (3) treatability (importance of early detection); and (4) manageability (survivability and quality of life). Westat provided literature and existing questionnaires on a number of these topics to assist in the narrowing of NCI interests.

Experts on the above new content areas were consulted to assist in the development or selection of new items. The panel of experts from NCI that provided insight into the measurement of the new content areas included Drs. Gary Kreps, Helen Meissner, Louise Mâsse, Stephen Marcus, Ed Maibach, Brad Hesse, Wendy Nelson, Rick Moser and Neeraj Arora. In addition, NCI consultants Drs. Barbara Rimer of the University of North Carolina, Vish Viswanath of Harvard University, Robert Hornik of the University of Pennsylvania and Alex Rothman of the University of Minnesota also participated in the development and refinement of instrument items related to the new content areas of interest.

To accommodate the new content areas, NCI identified topics from HINTS 2003 that could be reduced or eliminated. These included some cancer screening behaviors and questions related to personal risk perception. NCI investigators also decided that the number of cancer types in the survey would be reduced in favor of depth in regard to health communications about particular cancers. The types of cancers will be rotated in subsequent rounds of HINTS.

2.1.3 Cognitive Testing

Once a draft instrument was developed, Westat conducted two rounds of cognitive interviews. 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 Washington, DC, area volunteers participated in the first round of cognitive testing and 12 participated in the second round. During the second round of interviews, each participant received only a portion of the instrument so that no section of the interview was administered to more than nine participants. Westat staff summarized the results of both rounds of cognitive testing and provided recommendations to NCI about specific items and sections of the instrument (See Summary Reports for Rounds 1 & 2, April and June 2004). As a result of the first round of cognitive testing, 9 items were deleted and 29 items were altered. Following the second round of cognitive testing, 36 items were deleted and 17 were altered.

2.1.4 Timing Data

Westat project staff conducted several rounds of interviews with volunteer Westat staff members to provide preliminary timings for the administration of the instrument. Timings were provided to NCI by section, gender and age. This timing data, although not exact, provided insight into which sections of the instrument could be anticipated to take longer to administer than others.

2.1.5 Item Selection

Westat program staff met with NCI staff and consultants numerous times over the course of several months to finalize the HINTS 2005 instrument. All information gathered about the HINTS items was considered, including the analysis of HINTS 2003, systematic observations, literature reviews, expert panel reviews, the examination of existing questionnaires, cognitive testing results, and timing data. Taking all this information in consideration, HINTS items underwent a series of rankings to identify those that could be eliminated, those that needed to be altered, and those that should remain the same. Various

working groups consisting of NCI experts and consultants participated in the discussions and decisions about selected sections of the instrument over the course of development. Westat worked closely with each working group and with the NCI Project Officer to finalize a HINTS 2005 instrument that met NCI's stated goals, could be easily understood by the public, and could be administered in the shortest time possible.

2.1.6 Item Tracking

Westat project staff developed an Access database to track items on the HINTS 2005 instrument, as well as decisions about item deletions and revisions. The database included the sources of each item and whether the item was included in HINTS 2003. The database served as a resource during the development phase. Reports generated from the database were able to quickly and accurately provide information about the history of each item and the number of new versus HINTS 2003 items in the instrument at any given time during the development period.

2.2 Development of the Web Instrument

As noted in Chapter 1, an embedded experiment within HINTS 2005 sought to examine the possibility of conducting a web survey and its possible influence on response rates and data quality. The development of a web-based HINTS 2005 instrument required special planning and testing, which is described below.

2.2.1 Literature Review and Expert Review

Literature searches were conducted and articles were obtained on web-based survey administration. The issues and constraints identified in the literature for conducting surveys over the Internet informed the instrument development so that the two versions of the instrument (telephone and web) would be as similar as possible. In addition, the literature was influential in determining the ideal format and functionality of the HINTS 2005 web instrument.

Westat's Usability Testing Group conducted an expert review of a prototype section of the web-based questionnaire. Issues related to usability were discussed and a list of recommended revisions was developed.

2.2.2 Usability and Systems Testing

Following input from the expert review, a test section of the HINTS 2005 instrument was programmed in Blaise IS, and a respondent home page was developed. This version was used in subsequent usability testing with three Westat staff members who were unfamiliar with the study and web-based surveys. The Usability Testing Group reviewed the results of the testing and developed a list of recommendations for revising the web-based instrument. Another round of usability tests using the revised instrument were then conducted on nine participants. Each person filled out a brief demographic questionnaire, read the advance letter, logged into the web survey, completed the survey using a "think aloud" procedure, and completed a posttest questionnaire rating their experience with the survey. The results of the second round of usability testing were summarized and recommendations for the web-version developed (See Summary of Expert Review, September, 2004; Report of Usability Testing, January, 2005).

In addition, systems testing was conducted by Westat staff to evaluate the HINTS instrument's performance within the most common computer operating systems (such as Windows, Linux, Mac) and browsers (such as Internet Explorer, Mozilla, Netscape). Westat worked with Statistics Netherlands (the developers of Blaise IS) to solve problems identified during the systems testing.

2.3 Dress Rehearsal

A full-scale field test or "dress rehearsal" was conducted before the main data collection. The interview conditions for the dress rehearsal simulated the actual survey as closely as possible. A total of 123 respondents were randomly screened and interviewed (119 English-speakers and 4 Spanish-speakers) over an 18-day period. Eight interviews were completed on the Internet. This dress rehearsal provided an important check on both computer-assisted telephone interviewing (CATI) programming and web programming and offered insight into further training issues. In light of the dress rehearsal, revisions

were made to the programming of the CATI and the web and to the training program. In addition, the dress rehearsal highlighted the need to shorten the instrument further.

A debriefing was conducted with the interviewers who participated in the dress rehearsal 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 12 items and the deletion of 12 items. A bilingual interviewer carefully reviewed the Spanish version of the instrument to ensure appropriate translation of each question.

Following the dress rehearsal and the interviewer debriefing, Westat worked closely with NCI to identify final cuts and edits to the instrument without taking out high-priority items. A timing report was delivered to NCI that indicated that during the dress rehearsal the interview was 39.4 (SD = 12.5) minutes long on average. Section-by-section timings and indications of breakoffs were also delivered to NCI to assist in final revisions (See Pretest SMS Report; Timing Report; Interviewer Debriefing Table; Pretest Questionnaire, December, 2004; Pretest Data Delivery, January, 2005).

2.4 Final Questionnaire Structure

As noted above, the HINTS 2005 instrument development team wanted to maintain 50 percent of HINTS 2003 items. After all changes, deletions and additions to the instrument, 44 percent of the HINTS 2005 items were from HINTS 2003. Of those, 55 percent were exactly the same as in HINTS 2003, while 45 percent were slightly altered from the earlier version. Appendix A provides a table identifying the relationship between the 2003 and 2005 items.

The questionnaire is divided into two primary sections including a household screener and an extended interview. The household screener, which was always conducted over the telephone, begins with an introductory statement and a set of standard screening questions to identify respondents eligible for the survey. After a sampled person (SP) is identified, the extended portion of the interview begins with the Group Assignment section to determine eligibility and willingness to conduct the interview over the Internet.

The extended interview begins with a core module on health communication. The health communication module asks respondents to report on their use of, and preference for, various types of communication media. Special "if then" logic within the module allows the interviewer to administer questions on Internet usage and cancer information-seeking for those SPs for whom the questions are relevant.

The next section of the HINTS 2005 instrument delves into the respondents' individual and family history of cancer, as well as their information-seeking behavior specifically related to cancer. Barriers to seeking information on cancer as well as source preferences and source recognition are addressed. Again, "if then" logic allowed interviewers to restrict questions to those that were relevant to the SP.

Because cancer is a broad term relating to a whole host of risk factors and diseases spread throughout an individual's life span and across body systems, the survey then divides respondents into one of several parallel paths for questioning depending on their age and gender. Using this approach, HINTS 2005 collected specialized data on different types of cancers without overburdening individual respondents. Prostate and cervical cancers were selected as the cancers to be investigated in more depth for HINTS 2005. In addition, breast, colon, and skin cancer were superficially addressed.

The content area of the Mental Model of Cancer was added to the HINTS 2005 instrument. SPs were randomly assigned to answer mental model questions about one of three cancers: colon, lung or skin. Questions in this section were designed to measure the SP's opinion of the preventability, detectability, treatability and manageability of each cancer.

The survey converges again at the end with a set of common questions, which were asked of all respondents, on cancer-related behaviors such as tobacco use, diet and exercise, and health status. A section was added to HINTS 2005 to examine SPs' social networks. Finally, the questionnaire ends with the collection of demographic information. A set of debriefing questions was added to examine the SPs' experience with their participation in the survey.

Figure 2-1 illustrates the final structure of the HINTS 2005 instrument. In addition, the full questionnaire can be found in Appendix B. It is important to note that the appendix includes all questions

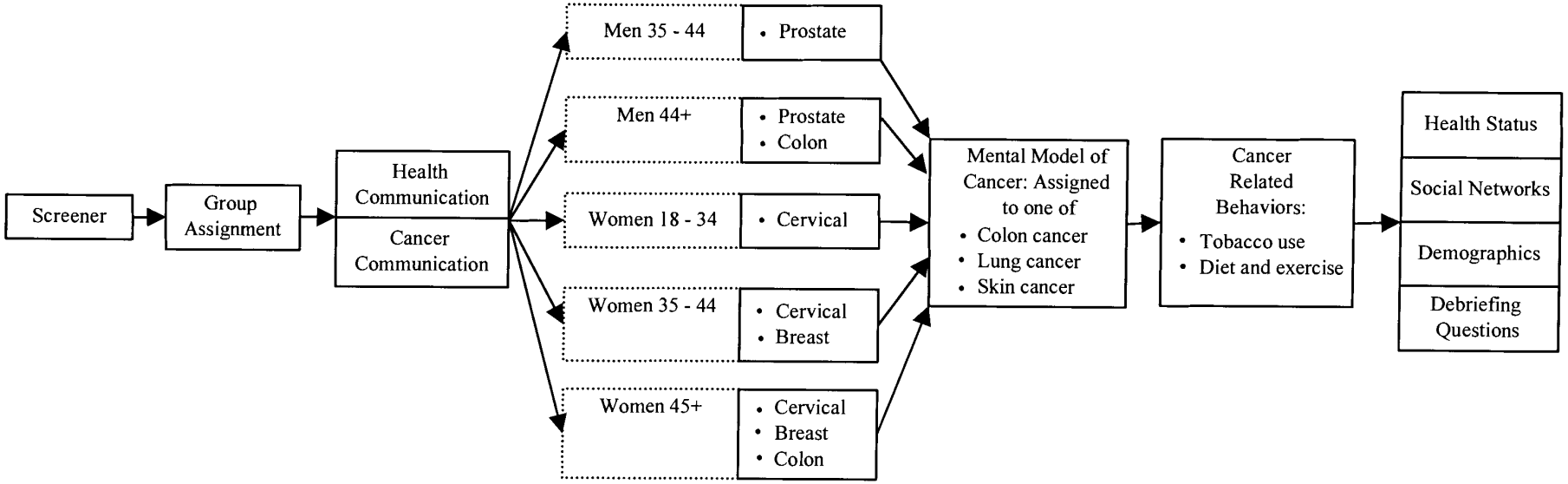


Figure 2-1. Final structure of the HINTS 2005 instrument

across all modules of the instrument. However, only a subset of those questions was asked of any single respondent based on their age, sex, and answers to previous questions. In addition, it is important to note that a few items were asked differently between the telephone and the web versions of the instrument. Appendix C contains a list of these differences.

3. HINTS 2005 SAMPLE DESIGN AND WEIGHTING

The sample design was similar to that used in HINTS 2003—both were list-assisted, random digit dialog (RDD) samples of all telephone exchanges in the United States. This approach resulted in a nationally representative sample of households.[1] During the household screener, one adult was sampled within each household and recruited for the extended interview.

The list-assisted RDD method is a random sample of telephone numbers from all "working banks" in U.S. telephone exchanges (Tucker, Casady, and Lepkowski, 1993). A working bank is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits 294-44) with at least one listed residential number.[2] The list-assisted method has been used in most RDD surveys in recent years.

3.1 Household Sampling Approach

The within-household sample involves asking the respondent how many adults are in the household, identifying the adults in a nonintrusive way (i.e., avoiding asking for names), and then sampling one adult.[3]

Westat used the same successful household sampling approach for sampling one adult (18+) per eligible household as we developed for HINTS 2003 (Rizzo, Brick, and Park, 2004) with a minor modification for three-person households. While speaking with an adult, the first step in screening is to obtain the number of adults currently living in the household. A respondent-selection algorithm is run

[1] Note that it is necessary to adjust for households with multiple telephone numbers.

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

[3] A more scientific-based method where the respondent is asked to list all the adults in the household would be much more intrusive. This could be done by asking for the full name or, to be less threatening, identifiers that do not provide personal information (e.g., made up name; age only). The computer-assisted telephone interview (CATI) then randomly selects one adult on this ordered list. The advantage of this method is that it keeps control of the sampling process in the hands of the survey administrators. The disadvantage is that it is more intrusive than the birthday method listed above and may reduce the response rate.

automatically by the computer-assisted telephone interview (CATI) system once this response is entered. The algorithm draws a random number to sample the respondent (or not) with a probability of 1/N, with N being the number of adults.

- If the respondent is sampled, sampling is complete. For one-person households, the respondent is always sampled.
- If the household contains two adults and the screener respondent is not sampled, the other adult is sampled and the process is complete.
- If the household contains three adults and the screener respondent is not sampled, the younger or older housemate is randomly sampled by the CATI and the process is complete.
- If the number of adults in the household is more than three and the screener respondent is not sampled, the so called "last birthday" method[4] will be employed. The respondent is asked which adult, other than he or she, has had the most recent birthday. This adult then becomes the sampled adult.
- If the respondent does not know the adult with the most recent birthday, the household is enumerated by initials or first names, and one adult is randomly sampled by CATI.

Use of this screening methodology minimizes the number of screener questions that are asked of a respondent and also minimizes the intrusiveness of the questions for the majority of households while still accomplishing a valid probability sample. The new method for sampling three-person households further reduces the percentage of households required for additional last-branch questions, as compared to HINTS 2003.[5]

3.2 Sample Design

Our initial sample was selected with a target of 7,000 extended completed interviews based on our sample size needs, our projections of eligibility rates, and response rates at the initiation point of the project as described in Appendix. D.

[4] This approach is described, for example, in Binson, Canchola, and Catania (2000). However, some studies have shown that the screener respondent tends to "self-select" using this method. This results in a disproportionate number of female respondents. In addition, there are some concerns that respondents may not fully implement the method because of lack of knowledge about birthdays or the general difficulty of the response task.

[5] In HINTS 2005, three-adult households were about 11.4 percent and thus this change helped to avoid additional last-branch questions for slightly less than 8 percent (2/3 of 11.4%, that is, excluding 1/3 of the chance for the screener being selected) of the households.

Table 3-1 presents the expected breakdown for the telephone sample, with the total telephone sample size required of 39,862. A slightly larger RDD sample of size 39,882 was originally considered and was divided into two waves, with 20,723 telephone numbers in Wave 1 and 19,159 numbers in Wave 2. A reserve sample of size 19,159 was also drawn for a total of 59,041 telephone numbers. In addition, a methodological experiment was planned during Wave 1 to study the effect of the mode of the extended interview (telephone interview only, Internet interview only, and respondent choice of telephone or Internet interview) and the effect of a promised incentive for completing the extended interview (no incentive, $10 promised incentive upon completion of the extended interview, and $20 promised incentive upon completion of the extended interview). The sample sizes were set so that all cross-cells defined by the two sets of three experimental groups had equal sample sizes.

Table 3-1. Overall RDD telephone sample sizes

Total telephone numbers to be drawn from the frame	39,862
Total telephone numbers after subsampling (57.0% nonmailable and 68.7% subsampling rate)	32,759
Total households in sample (47% residential rate)	15,397
Total expected completed screener (66.5% screener response rate)	10,239
Total completed interview (68.4% extended interview response rate)	7,003

In the course of conducting the pilot test, Westat found that response rates were much lower than expected, especially for the Internet interview. Given these outcomes, Westat and the National Cancer Institute (NCI) jointly decided to reorganize the sample design. Wave 1 and Wave 2 were reduced to 14,803 and 16,868 numbers, respectively, and a Wave 3 sample of size 43,791 was added. The Internet-only group was dropped. In response to comments from the Office of Management and Budget, the amount of the promised incentives was changed from $0, $10, and $20 to $0, $5, and $15. Additionally, the dropped-Internet-only numbers were reallocated into the other two mode groups, the telephone-only group and the choice group, assigning a slightly larger proportion to the telephone-only group. Table 3-2 summarizes the finalized RDD sample waves. Table 3-3 shows the size of experimental groups for the Wave 1 sample. Table 3-4 gives the allocation of the additional experimental groups for the Wave 2 and Wave 3 samples.

Table 3-2. RDD wave totals

Wave	Telephone sample size
Wave 1	14,803
Wave 2	16,868
Wave 3	43,791
Total	75,462

Table 3-3. RDD Wave 1 experimental group totals

Telephone sample size (relative size in %)	$0 promise incentive	$5 promise incentive	$15 promise incentive	Total
Telephone interview only	2,706 (18.3%)	2,705 (18.3%)	2,736 (18.5%)	8,147 (55.0%)
Telephone-Internet choice	2,222 (15.0%)	2,227 (15.0%)	2,207 (14.9%)	6,656 (45.0%)
Total	4,928 (33.3%)	4,932 (33.3%)	4,943 (33.4%)	14,803 (100.0%)

Table 3-4. Allocation of mode and promise incentive groups for Wave 2 and Wave 3 samples

Telephone sample size (Relative size in %)		Telephone only			Telephone-Internet choice			
Sample type	Batches	$5 promise incentive	$15 promise incentive	Subtotal	$5 promise incentive	$15 promise incentive	Subtotal	Grand total
Wave 2	1-2		3,365 (5.5%)	3,365 (5.5%)		2,555 (4.2%)	2,555 (4.2%)	5,920 (9.8%)
	3-6	5,474 (9.0%)	5,474 (9.0%)	10,948 (18.0%)	–	–	–	10,948 (18.0%)
Wave 3	1-3	7,065 (11.6%)	1,146 (1.9%)	8,211 (13.5%)	–	–	–	8,211 (13.5%)
	4-8	6,843 (11.3%)	6,842 (11.3%)	13,685 (22.6%)	–	–	–	13,685 (22.6%)
	9-13	6,806 (11.2%)	6,879 (11.3%)	13,685 (22.6%)	–	–	–	13,685 (22.6%)
	14-16	4,121 (6.8%)	4,090 (6.7%)	8,211 (13.5%)	–	–	–	8,211 (13.5%)
	Total	30,309 (50.0%)	27,796 (45.8%)	58,105 (95.8%)	–	2,555 (4.2%)	2,555 (4.2%)	60,660 (100.0%)

A two-phase stratification approach was applied to the entire sample, based on whether or not the telephone numbers had mailable addresses associated with them (Brick, et al. 2002). The nonmailable numbers were subselected at a rate of 68.7 percent (i.e., 31.3% of these numbers were discarded from the sample). The remaining nonmailable numbers were weighted at a rate of 1.456 to offset this subsampling. This procedure led to an increase in variance from the differential weighting, but because the nonmailable numbers were much less productive, the tradeoff led to better variance properties. Table 3-5 presents the results of the two-phase stratification.

Table 3-5. Two-phase stratification of the total sample numbers

	Total numbers in original sample	Percent	Total numbers in final sample	Percent	Weighting factor
Mailable	25,830	34.2%	25,830	46.2%	1.00
Nonmailable	49,632	65.8%	34,079	53.8%	1.46
Total	75,462	100.0%	55,909	100.0%	

3.3 Sample Weights and Variance Estimation Overview

3.3.1 Overview of Sample Weights

Every sampled adult who completed a questionnaire in HINTS 2005 received a sampling weight and a set of replicate sampling weights. These sampling weights should be used in aggregating any survey questionnaire answers for the purpose of computing nationally representative estimates.

The 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 3.4 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: the household needs to be successfully reached and the screener successfully completed and the sampled respondent within the household needs to be successfully recruited to complete the extended interview. 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 3.5 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 utilization 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. The auxiliary information used for HINTS came from the most recent Current Population Survey (March 2005), which has much larger sample sizes than HINTS. These calibration adjustments are discussed in Section 3.6.

3.3.2 Variance Estimation Methodology for HINTS 2005

The sampling weights for each responding adult are sufficient 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 all of the sample, except for a group of respondents which 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/wesvar/.

The jackknife variance estimator requires the use of replicate weights. For the HINTS 2005 data set, a set of *R=50* replicate weights was assigned to each responding adult. To illustrate how the replicate rates are computed, suppose **P** is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS 2005 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,...,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 2005 are telephone 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,...,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 3.5 for further details on computation of the replicate base weights.

Nonresponse and poststratification adjustments were then computed for each set of replicate base weights, using the replicate base weights in the computation of nonresponse and poststratification 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 poststratification adjustments. Sections 3.5 and 3.6 discuss in detail the computation of the final replicate weights.

3.4 Base Weights

Base weights were assigned to both sampled households and sampled adults within households. The base weight for the respondent is the product of four factors:

- The reciprocal of the telephone number's probability of being selected in the RDD sample (i.e., the sampling rate);
- An extra factor of 1.456 if the household was a nonmailable number (to offset the 68.7% subsampling of these numbers);
- The reciprocal of the conditional probability of the respondent being selected among the adults in the household given that the household was selected (which is equal to the number of adults in the household[6]);
- An extra factor equal to 2 if the household has more than one regular, residential telephone number, which accounts for the doubled chance of selection of the household.[7]

Standard errors were computed for HINTS 2005 estimates through the use of the jackknife technique, as discussed in Section 3.3.2. A total of 50 replicate base weights were computed for each

[6] Households with more than five adults were assigned a weighting factor of 5 to reflect their chance of selection. This is slightly biased, but the number of these households was small (about 0.15% of the HINTS 2005 screener completes). Westat wished to avoid unduly large increase of variance due to the large variation in weights.

[7] Note that cellular numbers, numbers devoted to businesses run from the household, and numbers dedicated to fax or modem usage are not considered. A small number of households had more than two regular, residential telephone numbers. We simplified the questionnaire by only asking about one or more than one and used that information in the computation of the base weight.

sample unit.[8] Suppose we write as *A* the set of all sampled adults in the study. The base weight will be indicated below as w_i (*i* indicating the particular sampled adult). Any given survey estimate can be written as

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

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

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

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* (see Section 3.3.2 for a description of the sampling procedure to select the deleted sets). The union of *A(r)* and *D(r)* for each replicate *r* is the full sample set *A*.

3.5 Nonresponse Adjustment and Response Rates

3.5.1 Introduction to 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 and Scheuren, 1983). In this approach, nonresponse cells are defined based

[8] 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).

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 quasirandomization 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 quasirandomization weighting adjustment eliminates any nonresponse bias (see, for example, Little and Rubin (1987), Chapter 4).

3.5.2 Unit Nonresponse in HINTS 2005

HINTS 2005 had two types of unit nonresponse: screener nonresponse and extended interview nonresponse. Screener nonresponse occurred when a household was reached, but the screener interview was not completed. It was also important to include in any screener nonresponse calculations any households for which a person was never reached, either because only an answering machine was reached (these are called NM numbers), or there was only a ring with no answer (these are called NA numbers), with every call made to the telephone number. Because there was no way to know if an answering machine or ring with no answer corresponds to a residential household, the number of lost residential numbers among the NA and NM numbers needed to be estimated (see Section 3.5.4).

To adjust for screener nonresponse, each completed screener received a screener nonresponse adjustment equal to the reciprocal of the estimated response rate in its screener nonresponse cell. For a discussion of the screener nonresponse cells and adjustments, see Sections 3.5.3 and 3.5.4, respectively.

Extended interview nonresponse occurred when the screener interview was completed successfully, yielding a sampled adult in the household with identifying information for this adult, and the number of adults in the household, but the sampled adult did not complete the extended interview.

To adjust for interview nonresponse, each completed extended interview received an interview nonresponse adjustment equal to the reciprocal of the weighted interview response rate in its interview nonresponse cell. (Completed extended interviews also receive a screener nonresponse adjustment.) The methodology for selecting extended interview nonresponse cells and computing extended interview nonresponse adjustments is discussed in Sections 3.5.5 and 3.5.6, respectively. Section 3.5.7 discusses the computation of replicate nonresponse adjustments.

3.5.3 Nonresponse Cells for Screener Nonresponse Adjustments

Nonresponse cells were generated using cross-classifications based on four characteristics: (1) the time of the telephone numbers' release to the field (earlier, midway, latest); (2) mailable status (whether or not an address was available for the telephone number to send them a letter); and selected sociodemographic characteristics estimated for each telephone exchange (by the vendor Genesys): (3) percentage of college graduates (exchanges with lowest percentage, next lowest percentage, and highest percentage[9]) and (4) percentage of blacks and Hispanics (three cells with roughly equal population[10]). Some of the cells were collapsed to avoid unduly large adjustment factors and/or small adjustment cell sizes. These characteristics were both related to response propensity and were correlated to unit response propensity, so that Westat believed these cells would lead to effective nonresponse adjustments.[11]

Among other characteristics for generating cells was the mailable or nonmailable status indicating whether or not a published address was available for the telephone number. These addresses were used to mail advanced letters about the study with a $2 pre-incentive and followup letters for households who had not responded. As found in HINTS 2003 and other previous surveys, response propensity tends to differ by this characteristic (telephone households with known addresses that have received mailed information respond at a higher rate than those without known addresses).

---

[9] The breakpoints were the 1/3 and 2/3 percentiles over all frame exchanges, which were calculated when the frame was constructed.

[10] The breakpoints were the 1/3 and 2/3 percentiles over all frame exchanges, which were calculated when the frame was constructed.

[11] Initially, census region of four cells and median income within exchanges of three cells were also considered for constructing the nonresponse adjustment cells, but they were dropped because of their weaker correlation to the screener response.

Cross-classifications of the above classes gave a total of 54 potential cells and further collapsing led to a total of 38 final cells as listed in Table 3-6.

Table 3-6. HINTS 2005 screener response cells, rates, and adjustment factors

Release	Mailable status	% College graduates	% Minority	Screener response rate	Household level nonresponse adjustment
Early	Mailable	Low	Low	50.6%	1.977
Early	Mailable	Low	Middle	47.7%	2.097
Early	Mailable	Low	High	42.5%	2.351
Early	Mailable	Middle	Low	45.5%	2.199
Early	Mailable	Middle	Middle	38.7%	2.586
Early	Mailable	Middle	High	36.0%	2.779
Early	Mailable	High	Low	42.5%	2.353
Early	Mailable	High	Middle	37.5%	2.665
Early	Mailable	High	High	38.2%	2.617
Early	Nonmailable	Low	Low	34.0%	2.942
Early	Nonmailable	Low	Middle	30.6%	3.272
Early	Nonmailable	Low	High	27.6%	3.625
Mixed*	Nonmailable	Middle/high	all	23.2%	4.312
Midway	Mailable	Low	Low	46.1%	2.169
Midway	Mailable	Low	Middle	41.9%	2.386
Midway	Mailable	Low	High	37.6%	2.659
Midway	Mailable	Middle	Low	38.8%	2.575
Midway	Mailable	Middle	Middle	37.1%	2.696
Midway	Mailable	Middle	High	35.6%	2.812
Midway	Mailable	High	Low	33.6%	2.978
Midway	Mailable	High	Middle	30.2%	3.311
Midway	Mailable	High	High	29.6%	3.374
Midway	Nonmailable	Low	Low	28.5%	3.510
Midway	Nonmailable	Low	Middle	28.8%	3.469
Midway	Nonmailable	Low	High	29.4%	3.405
Midway	Nonmailable	Middle	Low	22.9%	4.370
Midway	Nonmailable	Middle	Middle	20.6%	4.850
Midway	Nonmailable	Middle	High	21.4%	4.674
Latest	Mailable	Low	Low	41.9%	2.387
Latest	Mailable	Low	Middle	40.2%	2.486
Latest	Mailable	Low	High	32.1%	3.119
Latest	Mailable	Middle	Low	38.5%	2.598
Latest	Mailable	Middle	Middle	36.4%	2.746
Latest	Mailable	Middle	High	31.7%	3.153
Latest	Mailable	High	Low	36.7%	2.727
Latest	Mailable	High	Middle	33.3%	3.003
Latest	Mailable	High	High	29.8%	3.354
Latest	Nonmailable	All	All	19.8%	5.042
Total	Total	Total	Total	34.1%	

* This cell includes those of the following two combinations: (earlier, nonmailable, model/high, all) and (midway, nonmailable, high, all).

3.5.4 Screener Nonresponse Adjustments

In general, nonresponse adjustments within nonresponse cells are the reciprocals of the weighted response rates within the cell, where the respondents and nonrespondents are weighted by their (adjusted) base weight. In this case, the household base weights are unknown for screener nonrespondents because components of the base weight depend on whether the household has one or more residential telephone numbers. For this reason, Westat used telephone base weights instead. The nonresponse adjustment was set equal to the reciprocal of the weighted screener response rate for each cell. The telephone number base weights were set to the product of two factors: the reciprocal of the telephone number's probability of being selected in the RDD sample and an extra factor for the subsampled nonmailable numbers.

In principle, the weighted screener response rate (with the telephone number base weight) is equal to the weighted total number of cooperating households (eligible or not) divided by the weighted total number of residential numbers in the sample. The latter value is not completely known, because of NM and NA numbers. Write NM and NA as the counts of answering machine numbers and ring-no-answer numbers, respectively. Let *AMNA(a)* and *PNA(a)* be the counts of NM and NA numbers in cell *a*. We estimated the number of residential numbers among the NM numbers by computing the overall eligibility rate *EM* among working numbers with known eligibility status, and by computing the overall eligibility rate *EA* among all numbers with known eligibility status (working and nonworking).

With these two estimated eligibility rates applied to the NM and NA numbers, the nonresponse adjustment for cell *a* was computed as follows:

$$HNRA(a) = \frac{C(a) + I(a) + REF(a) + O(a) + (AMNA(a) * EM) + (PNA(a) * EA)}{C(a) + I(a)}$$

where *C(a)* is the weighted number of completed screeners, *I(a)* is the number of households found ineligible for the study, *REF(a)* is the number of eligible screeners who refused to participate, and *O(a)* are other residential numbers (e.g., numbers that were found to be residential, but for which a screening interview could not be completed for reasons other than refusals).

Westat also computed a study screener response rate. Writing *C*, *I*, *REF*, *O*, *AMNA*, and *PNA* as the total number of completed screeners, ineligibles, eligible screeners who refused to participate, other

residentials, answering machine NMs, and pure NAs respectively, and defining *EM* and *EA* as above, we will compute the screener response rate *SCRNR* as

$$SCRNR = \frac{C + I}{C + I + REF + O + (AMNA * EM) + (PNA * EA)}$$

Note that this screener response rate is algebraically equivalent to

$$SCRNR = \frac{C}{C + \{ER * [REF + O + (AMNA * EM) + (PNA * EA)]\}}$$

with $$ER = \frac{C}{C + I}$$

The second form of *SCRNR*, though algebraically more complicated, is conceptually more transparent. The response rate is the completes divided by the completes plus the estimated eligible numbers among the remaining residential numbers (refusals, NMs and NAs). Westat estimated the eligibles among the estimated residential numbers *REF+O+(AMNA*EM)+(PNA*EA)* by imputing the eligibility rate from the "known eligibility status" numbers: the completes and ineligibles. *SCRNR* is fully within the guidelines of American Association for Public Onion Research (AAPOR) standards regarding valid response rates.[12] Table 3-6 presents the screener response rates and the nonresponse adjustment factors by cells.

3.5.5 Nonresponse Cells for Extended Interview Nonresponse Adjustments

More information available is about extended interview nonrespondents as compared to screener nonrespondents. This extra information came from the completed screener (a case was not designated as an extended interview nonrespondent unless the screener was successfully completed). In this section, a screener was defined as completed if the key items for sampling an adult and assigning a base weight to the household were answered: the number of adults in the household and the presence of multiple telephone numbers. Note that the base weight w_i (see Section 3.4) could only be computed if the screener was completed.

[12] *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* Available on AAPOR web site: www.aapor.org.

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

- Screener respondent (yes or no);
- Size of household: number of adults in household (one or more);
- Sex of sampled person (male or female);
- Promised incentives (three cells);
- Extended interview mode groups (two cells);
- Mailable and nonmailable status (two cells).

The first three characteristics on the above list were derived directly from the screener questionnaire. The next three characteristics were determined either in designing stage or in the course of fielding. A total of 42 nonresponse cells were constructed by collapsing the 96 initial cells to avoid unduly large adjustment factors and/or small cell sizes. Table 3-7 lists those cells.

3.5.6 Extended Interview Nonresponse Adjustments

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

Table 3-7. HINTS 2005 extended interview response cells, response rates, adjustment factors

Screener	Household size	Sex of sampled person	Promised incentive	Interview mode group	Mailable status	Response rate	Adjustment factor
Yes	1	Male	$0	T	M	71.3%	1.402
Yes	1	Male	$5	C	All	57.3%	1.746
Yes	1	Male	$5	T	M	70.3%	1.422
Yes	1	Male	$15	C	All	57.9%	1.726
Yes	1	Male	$15	T	M	71.8%	1.393
Yes	1	Male	$5	T	N	71.2%	1.405
Yes	1	Male	$15	T	N	70.4%	1.419
Yes	1	Male	$0	C	All	74.1%	1.350
Yes	1	Female	$0	T	All	82.7%	1.209
Yes	1	Female	$5	C	All	74.4%	1.344
Yes	1	Female	$5	T	M	77.1%	1.298
Yes	1	Female	$15	C	All	75.8%	1.320
Yes	1	Female	$15	T	M	77.1%	1.296
Yes	1	Female	$5	T	N	68.4%	1.463
Yes	1	Female	$15	T	N	75.5%	1.325
Yes	All	Male	$0	C	All	72.3%	1.383
Yes	>1	Male	$0	T	All	72.8%	1.374
Yes	>1	Male	$5	C	All	60.5%	1.653
Yes	>1	Male	$5	T	M	68.4%	1.461
Yes	>1	Male	$15	C	All	67.7%	1.477
Yes	>1	Male	$15	T	M	70.6%	1.417
Yes	>1	Male	$5	T	N	61.1%	1.637
Yes	>1	Male	$15	T	N	67.1%	1.490
Yes	>1	Female	$0	C	All	81.9%	1.221
Yes	>1	Female	$0	T	All	79.2%	1.263
Yes	>1	Female	$5	C	All	77.8%	1.286
Yes	>1	Female	$5	T	M	78.2%	1.279
Yes	>1	Female	$15	C	All	71.1%	1.407
Yes	>1	Female	$15	T	M	80.7%	1.238
Yes	>1	Female	$5	T	N	65.8%	1.520
Yes	>1	Female	$15	T	N	77.9%	1.283
No	>1	Male	$0	T	All	43.6%	2.292
No	>1	Male	$5,$15	T	All	38.2%	2.621
No	>1	Male	$15	C	All	41.3%	2.419
No	All	Male	$15	T	M	46.8%	2.137
No	>1	All	$0,$5	C	All	32.3%	3.092
No	>1	Female	$0	T	All	49.3%	2.027
No	>1	Female	$5	T	M	48.0%	2.082
No	>1	Female	$15	C	All	43.6%	2.291
No	All	Female	$15	T	M	57.8%	1.729
No	>1	Female	$5	T	N	42.3%	2.365
No	>1	Female	$15	T	N	57.7%	1.733
Total	Total	Total	Total	Total	Total	61.2%	

NOTE: Letters T, C, M, N, respectively, represent telephone only, telephone-Internet choice, mailable, and nonmailable.

unbiased estimator (adjusted for screener nonresponse[13]) 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 4 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:

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

These weighted extended interview response rates and adjustment factors are given in Table 3-7.

3.5.7 Replicate Nonresponse Adjustments

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 3.5.3, 3.5.4, 3.5.5 and 3.5.6).

The screener nonresponse adjustments are the reciprocals of weighted screener response rates. Replicate screener response rates were computed for each screener response cell *a* and each replicate *r* by removing the deleted set (of telephone numbers) 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 of telephone numbers.

[13] 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.

Define *RS(a,r)* as the count of confirmed residential numbers in screener response cell *a* which are in replicate set *r*. (An alternative definition of *RS(a,r)* is the count of confirmed residential numbers in screener response cell *a* after the deleted set corresponding to replicate *r* has been removed from the RDD sample.) Define *C(a,r), I(a,r), REF(a,r), O(a,r), AMNA(a,r), PNA(a,r), EM(r)* and *EA(r)* similarly (see Section 4.4.3). Then a replicate nonresponse adjustment can be defined as follows:

$$HNRA(a,r) = \frac{C(a,r) + I(a,r) + REF(a,r) + O(a,r) + \left(AMNA(a,r) * EM(r)\right) + \left(PNA(a,r) * EA(r)\right)}{C(a,r) + I(a,r)}$$

The computation of interview nonresponse adjustments was also replicated. The replicate interview nonresponse adjustment for interview nonresponse cell *b* and replicate *r* was computed as follows:

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

The two nonresponse adjustments (for screener nonresponse and extended interview nonresponse) were appended to the base weight for the subject (adult):

$$SBW_i = w_i \; HNRA(a) \; ENRA(b)$$

The summation of these nonresponse-adjusted subject base weights over all responding subjects is a nonresponse-adjusted unbiased estimator of the total number of adults in the U.S. population. The corresponding replicate weights are as follows (for each replicate *r*):

$$SBW_i(r) = w_i(r) \; HNRA(a,r) \; ENRA(b,r)$$

3.6 Calibration Adjustments

3.6.1 Introduction to Calibration Adjustments

As discussed above, the purpose of calibration is to reduce the sampling variance of estimators through the use of reliable auxiliary information (e.g., see Deville and 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.

3.6.2 Control Totals from the Current Population Survey

The Current Population Survey (CPS) of the U.S. Census Bureau has much larger sample sizes than those of HINTS. The CPS estimates of any U.S. population totals have lower sampling error than the corresponding HINTS estimates, making calibration of the survey weights to CPS control totals beneficial. The CPS estimates are available via the Internet. Westat used the March 2005 CPS estimates that were available on the Census Bureau web site.

Calibration variables were selected among those that were on the CPS public-use file and were found to be well correlated to important HINTS questionnaire item outcomes (i.e., Westat wanted

CPS-available characteristics that tend to have differing mean values for HINTS questionnaire item outcomes). The following CPS characteristics will correlate well with HINTS questionnaire items:

g1. Sex

1. Male
2. Female

g2. Race/ethnicity

1. Hispanic
2. Non-Hispanic black
3. Non-Hispanic white or other.

g3. Age

1. 18 to 34 years old
2. 35 to 49 years old
3. 50 to 64 years old
4. 65 years old or older

g4. Educational level

1. Less than high school diploma (<HS)
2. High school diploma only (HS)
3. High school diploma, some college (HS+)
4. Bachelor's degree or higher (>HS+)

Westat generated a total of 96 initial cells based on cross-classifications of the above characteristics, and tabulated control totals for these cells from March 2005 CPS data. A total of 54 poststratification cells were constructed by collapsing the initial cells in a way to make the cell size more or less 30 and/or to limit the adjustment factors to no larger than 4.5. Table 3-8 lists those cells and adjustment factors. These control totals can be written as c_g, g=1,…,54. The poststratification

Table 3-8. HINTS II Poststratification Adjustment Cells, Coverage, and Factors

Gender	Age group	Race/ethnicity	Education	Coverage*	Adjustment factor
Male	18 to 34 yrs	Hispanic	<HS	92.2%	1.085
Male	18 to 34 yrs	Hispanic/black	HS	52.4%	1.910
Male	18 to 34 yrs	All	HS+	47.5%	2.107
Male	18 to 64 yrs	Hispanic	>HS+	138.3%	0.723
Male	18 to 34 yrs	Black/others	<HS	46.0%	2.174
Male	18 to 64 yrs	Black	>HS+	100.9%	0.991
Male	18 to 34 yrs	Others	HS	70.7%	1.415
Male	18 to 34 yrs	Others	>HS+	89.5%	1.117
Male	35 to 49 yrs	Hispanic/black	<HS	98.9%	1.011
Male	35 to 49 yrs	All	HS	55.5%	1.803
Male	35 to 49 yrs	Hispanic/black	HS+	42.6%	2.350
Male	35 to 49 yrs	Others	<HS	54.3%	1.840
Male	35 to 49 yrs	Others	HS+	51.0%	1.961
Male	35 to 49 yrs	Others	>HS+	89.7%	1.115
Male	50 to 64 yrs	Hispanic/black	<HS	56.1%	1.781
Male	50 to 64 yrs	All	HS	55.8%	1.793
Male	50 to 64 yrs	All	HS+	68.4%	1.463
Male	50 to 64 yrs	Others	<HS	62.2%	1.609
Male	50 to 64 yrs	Others	>HS+	118.8%	0.842
Male	65 or older	All	<HS	69.6%	1.436
Male	65 or older	All	HS	93.3%	1.072
Male	65 or older	All	>HS+	141.4%	0.707
Male	65 or older	Black/others	HS+	116.4%	0.859
Female	18 to 34 yrs	Hispanic	<HS	86.7%	1.154
Female	18 to 34 yrs	Hispanic	HS/HS+	75.1%	1.331
Female	18 to 64 yrs	Hispanic	>HS+	138.5%	0.722
Female	18 to 34 yrs	Black	<HS/HS	41.4%	2.415
Female	18 to 34 yrs	Black	HS+	48.4%	2.064
Female	18 to 49 yrs	Black	>HS+	86.1%	1.162
Female	18 to 34 yrs	Others	<HS	49.6%	2.017
Female	18 to 34 yrs	Others	HS	73.0%	1.370
Female	18 to 34 yrs	Others	HS+	65.6%	1.525
Female	18 to 34 yrs	Others	>HS+	93.9%	1.065
Female	35 to 49 yrs	Hispanic	<HS	114.8%	0.871
Female	35 to 49 yrs	Hispanic	HS	96.5%	1.036
Female	35 to 64 yrs	Hispanic	HS+	116.8%	0.856
Female	35 to 64 yrs	Black	<HS	91.8%	1.089

Female	35 to 49 yrs	Black	HS	52.5%	1.905
Female	35 to 49 yrs	Black	HS+	71.0%	1.409
Female	35 to 49 yrs	Others	<HS	106.5%	0.939
Female	35 to 49 yrs	Others	HS	81.8%	1.222
Female	35 to 49 yrs	Others	HS+	67.3%	1.486
Female	35 to 49 yrs	Others	>HS+	120.9%	0.827
Female	50 to 64 yrs	Hispanic	<HS	159.2%	0.628
Female	50 to 64 yrs	Hispanic/black	HS	81.6%	1.226
Female	50 to 64 yrs	Black	HS+	85.5%	1.170
Female	50 to 64 yrs	Black	>HS+	114.3%	0.875
Female	50 to 64 yrs	Others	<HS	83.8%	1.194
Female	50 to 64 yrs	Others	HS	90.3%	1.108
Female	50 to 64 yrs	Others	HS+	113.0%	0.885
Female	50 to 64 yrs	Others	>HS+	139.3%	0.718
Female	65 or older	Hispanic	<HS/HS/HS+	60.1%	1.665
Female	65 or older	All	>HS+	141.6%	0.706
Female	65 or older	Black	<HS	125.3%	0.798
Female	65 or older	Black	HS/HS+	70.1%	1.426
Female	65 or older	Others	<HS	78.2%	1.278
Female	65 or older	Others	HS	109.9%	0.910
Female	65 or older	Others	HS+	154.6%	0.647
Overall				84.6%	

Note: * The coverage is defined as the reciprocal of the adjustment factor, that is, $1/CA_g$.

adjustments were computed by adjusting the weights SBW_i of adult respondents by a constant factor so that the final set of weights add to the control total. That is, for each poststratification cell *PS(g)*, Westat computed CA_g equal to:

$$CA_g = \frac{c_g}{\sum_{i \in PS(g)} SBW_i} \quad .$$

Replicate versions of the *CA* adjustments were also computed for each replicate *r*. The replicate $CA_g(r)$ adjustments were computed using the same formula, but with the replicate $SBW_i(r)$ weights replacing the full sample SBW_i weights. These replicate versions of the *CA* adjustments are indicated as $CA_g(r)$.

The final subject weights W_i were assigned to each subject i with a completed interview. These weights are equal to the base weights for subject i multiplied to nonresponse adjustments and calibration adjustments, and can be written as:

$$W_i = SBW_i \, CA_g$$

The replicate weight for subject i and replicate r was as follows:

$$W_i(r) = SBW_i(r) \, CA_g(r)$$

4. OPERATIONS

4.1 Summary of Data Collection Procedures

As noted in earlier chapters, one of the priorities for HINTS 2005 was to experiment with alternative modes of data collection. A primary reason for this priority is the decline in response rates to telephone-based surveys. Therefore, the option of data collection through a web-based version of the instrument was implemented in addition to the random digit dialog (RDD) portion of the sample that was administered over the telephone to achieve the methodological integrity goal of HINTS. The following sections outline the procedures for administering both the telephone and the web instruments.

4.2 Data Collection Procedures

Telephone data collection was conducted using a Blaise computer-assisted telephone interview (CATI) instrument administered by trained Westat telephone interviewers over a period of 25 weeks. The RDD sample telephone numbers were address-matched to enable the mailing of an introductory advance letter to potential study households.

Data collection for HINTS 2005 was divided into three waves. During Wave 1, Westat conducted an experiment by testing different modes of data collection and varying the amount of the incentive. Attempts were made to screen all households over the telephone. Individuals sampled from each responding household were randomly assigned to one of five conditions: one of two modes for collecting data during the extended interview and one of three levels of incentive promised for completing the extended interview. The two levels of data collection mode were (1) a telephone administered extended interview and (2) an option for the respondent to select his or her mode of preference between the telephone and web (i.e., the “choice” mode). The three levels of promised incentives were $0, $5, and $15.

4.2.1 Advance Mailings

The release of cases to the Telephone Research Center (TRC) waves was staggered within sample waves. Before releasing cases, Westat used two vendors to obtain addresses for the sample of telephone numbers. For those households where Westat obtained a valid address, an advance letter was sent to potential respondents approximately 1 week before calling. Timing the letters to arrive at the household just before calling provided the highest likelihood that respondents would recall the advance materials.

The advance letter was printed on National Cancer Institute (NCI) letterhead. On the back of the letter, Westat included a version of the letter translated in Spanish. Mailing labels containing the sampled household addresses were produced from an electronic file and placed on the envelope. To increase the likelihood that the respondent would open the letter, Westat placed postage stamps on the envelopes, instead of using metered mail. With the letter, Westat included a HINTS brochure[14] and a $2 bill as an incentive to help the respondent remember the advance letter. This letter was addressed to Friend Residing at {ADDRESS}. For those households for which an address could not be obtained, TRC interviewer calls to these households were the first contact with potential respondents.

4.2.2 Incentive Experiment

A methodological experiment was conducted in Wave 1 to study the effect of a promised incentive. Three incentive levels were tested—(1) no incentive, (2) $5 promised incentive, and (3) $15—and were promised upon completion of the extended interview. Each case was randomly assigned one of these incentive amounts.

Based on results midway through Wave 1, it was decided to split the Wave 2 sample such that there would only be two incentive levels: $5 and $15. It was determined that paying some monetary incentive resulted in a higher response rate. Although the $15 incentive generated the highest response rates, it was unclear if there was a statistically significant difference between $5 and $15. By splitting the sample, Westat was able to continue the methodological experiment and at the same time, keep the overall project costs lower.

[14] The HINTS 2005 brochure was not available in time to include in the Wave 1 mailout. However, it will be included in the Wave 2 mailout.

Toward the end of the study, it became clear that the response rates for those with a promised $15 incentive were significantly higher than those who were promised $5. So for the remaining interim cases that were considered to be more difficult, such as initial refusals, were assigned to the $15 incentive for completing the extended interview.

4.2.3 TRC Hiring and Training

The HINTS 2005 data collection was completed from the TRC location in Rockville, Maryland. Hiring of interviewing staff for the study was done by TRC professional staff assigned to the study. The interviewers were mainly experienced RDD screeners and interviewers, complemented by a smaller number of newly hired staff. Project-specific training was developed by TRC staff and project staff and consisted of interviewer and trainer manuals, and a specific training agenda that included lectures, interactives 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.

Initial training of 16 interviewers took place in Rockville on February 19-20, 2005. Attrition trainings also took place in Rockville on April 16-17 (45 interviewers) and June 4-5 (25 interviewers) as additional interviewers were added to the calling staff. Instruction of bilingual interviewers in Spanish was completed in the week following the initial training session.

Each formal session was completed in 2 days. However, interviewers were allowed to go "live" on the telephone only when their supervisors made the assessment that they were fully prepared to start. Some interviewers required additional practice exercises during the days immediately following formal training. Those interviewers that started interviewing immediately upon conclusion of formal training were monitored using routine Westat observation and monitoring procedures to ensure that their work was of the highest quality. Any issues that supervisors observed while monitoring an interviewer's work were discussed and resolved immediately following the interview that was monitored.

During the course of the data collection effort, TRC supervisors and 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.

4.3 Findings from TRC Operations

4.3.1 Field Period

Data collection in the TRC started the week of February 21, 2005. TRC operations continued through August 14, 2005. The web-based interview option was offered to screened participants in February and completed web interviews were accepted through the end of August in an effort to capture late completions.

4.3.2 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, and production to date in terms of total interviews.

Reports monitoring HINTS 2005 data collection included the following:

- **Completed Extended Interviews by Interview Mode.** This report showed the number of completed extended interviews by actual mode (Internet and telephone) versus assigned mode (telephone only and subject choice). The report also showed the number of telephone interviews that were completed in Spanish and the number of respondents that had access to the Internet.

- **Web Status Report.** This report showed the number of extended respondents that were, at some point, provided access to the web survey (i.e., given a user ID and password). It showed the method that Westat used to provide the respondent with a user ID and password (i.e., email, letter, or over the telephone); whether the subject received a reminder email; and the status of the cases.

- **Production Report by Batch-Group.** This report showed the status of cases released to the TRC, broken down by batch-group (i.e., the order of release within the TRC). This report was available for both the screener and extended interviews. The response rate for each group was provided.

- **Production Report by Incentive Group.** This report showed the status of cases released to the TRC, broken down by incentive amount ($0, $5, $15). This report was available for both the screener and extended interviews. The response rate for each incentive group was provided.

- **Production Report by Interview Mode Group.** This report showed the status of cases released to the TRC, broken down by assigned group mode (telephone only and subject choice). This report was available for both the screener and extended interviews. The response rate for each interview mode group was provided.

- **Production Report by Mailable Status.** This report showed the status of cases released to the TRC, broken down by mailable status (i.e., mailable, nonpurged nonmailable only, and all nonmailable). This report was available for both the screener and extended interviews. The response rate for each group was provided.

- **Weekly and Monthly Production Report.** This report showed the overall production for the current week and month compared with overall production for the entire study. The report was available for both the screener and extended interviews. A summary of this report is provided in Table 4-1.

Table 4-1. Weekly TRC production: Completed cases by week

	Screener		Extended (CATI)	
Week beginning	Actual	Cumulative	Actual	Cumulative
2/21/05	212	212	87	87
2/28/05	200	412	76	163
3/7/05	177	589	81	244
3/14/05	167	756	65	309
3/21/05	123	879	54	363
3/28/05	160	1,039	70	433
4/4/05	132	1,171	70	503
4/11/05	250	1,311	57	560
4/18/05	512	1,823	273	833
4/25/05	401	2,224	250	1,083
5/2/05	275	2,499	188	1,271
5/9/05	325	2,824	220	1,491
5/16/05	431	3,255	253	1,744
5/23/05	425	3,680	253	1,997
5/30/05	476	4,156	265	2,262
6/6/05	833	4,989	461	2,723
6/13/05	664	5,653	401	3,124
6/20/05	711	6,364	426	3,550
6/27/05	560	6,924	345	3,895
7/4/05	388	7,312	273	4,168
7/11/05	449	7,761	311	4,479
7/18/05	316	8,077	258	4,737
7/25/05	340	8,417	215	4,952
8/1/05	251	8,668	213	5,165
8/8/05	117	8,785	135	5,300

4.3.3 Administration Times

The mean administration time for the extended telephone interview (5,301 respondents, not including partial completes) was 33.5 minutes.

The mean administration time for the web version of the extended interview (93 respondents, not including partial completes) was 62.7 minutes.

4.3.4 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 paced in each of seven noncontact time slices. These cases received at least one call attempt per time slice before being finalized. As with most studies, 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 more than 20 call attempts over several weeks.

Queue priorities were set within the scheduler. Extended interview appointments had a higher priority than screener questionnaires. Table 4-2 details the level of effort for the screener by result code, while Table 4-3 details the level of effort for the CATI extended interview.

Table 4-2. Total screener level of effort: Number of call attempts by result

Call attempts	Completes/ ineligibles		Non-response		Non-working		No Answer	
	N	Col %	N	Col %	N	Col %	N	Col %
0	0	0.0	16	0.3	35,140	81.6	0	0.0
1-5	7,194	81.8	2,475	49.0	6,830	15.9	0	0.0
6-10	1,304	14.8	1,294	25.6	856	2.0	352	18.0
11-15	231	2.6	372	7.4	191	0.0	1,597	81.8
16-20	55	0.6	560	11.1	42	0.0	2	0.0
21-25	9	0.0	266	5.3	8	0.0	1	0.0
26-30	1	0.0	68	1.4	1	0.0	0	0.0

Table 4-3. Total extended (CATI) level of effort: Number of call attempts by result

Call attempts	Completes and ineligibles		Nonresponse	
	N	Col %	N	Col %
1-5	4,465	84.5	921	45.2
6-10	513	9.7	372	18.2
11-15	199	3.8	178	8.7
16-20	94	1.8	499	24.5
21-25	10	0.2	64	3.1
26-30	1	0.0	6	0.3

4.4 HINTS 2005 Response Rates

4.4.1 Introduction

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 and 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 and Lyberg, 2003: 84). Unfortunately, it is difficult 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, 2004). 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.[15]

4.4.2 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 voice mail (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 impute the residential rates among the known numbers to the answering machine and pure NA numbers. This is generally considered to be "conservative"[16] (i.e., the actual residential rates among these numbers are known to be generally smaller than the residential rates among the known numbers, though it is difficult to measure exactly this rate). For answering machine NA numbers, it is known that these are working numbers, so Westat imputes to these the residential rate among working numbers with known residential status (i.e., we assume the residential rate among the answering machine numbers is equal to that of the working numbers). For pure NA numbers, we impute to these the residential rate among all numbers with known residential status. These residential rates are weighted.

Let *C, I, REF, O, AMNA* and *PNA*, respectively, denote the summation of weights of completed screeners, ineligibles, refusals, other residentials, answering machine (NMs), ring no answers

[15] 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.

[16] Conservative in this context means that it leads to an underestimation of the response rate.

(NAs). 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

$$SCRNR = \frac{C + I}{C + I + REF + O + AMNA \times EM + PNA \times EA}.$$

Note that this screener response rate is algebraically equivalent to

$$SCRNR = \frac{C}{C + ER \times [I + REF + O + AMNA \times EM + PNA \times EA]}$$

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

The second form of *SCRNR* though algebraically more complicated is conceptually more transparent. The response rate is the completes divided by the completes plus the estimated eligible numbers among the remaining residential number (refusals and NAs). Westat estimates the eligibles among the estimated residential number *REF+O+AMNA*EM+PNA*EA* by imputing 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 not eligible).

Table 4-4 presents the sum of base weights for each class of telephone numbers. These sums are estimates of the total numbers in the United States in this category. The "percent of known status" column is a percentage of each known status set as a percentage of the total known status numbers. The "percent of known working status" column is a percentage of each known working status set as a percentage of the total known working status numbers. The two important percentages are *EA* (34.55%) and *EM* (70.41%): the percentage of known residential numbers within the known status set and the known working number status set.

Table 4-5 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 imputed residential numbers from the NMs and the NAs.

Table 4-4. Weighted aggregations and percentages for the full HINTS 2005 RDD sample

Response class	Estimated totals	Percent of known-status numbers	Percent of known-status working numbers
Completed screener (C)	33,939,126	13.48%	27.52%
Ineligible (I)	29,397	0.01%	0.02%
Refusal (REF)	36,691,462	14.57%	29.75%
Other nonresponse (O)	16,615,446	6.60%	13.47%
Business (NRS)	36,068,713	14.32%	29.24%
Nonworking (NWS)	128,487,450	51.02%	
Answering machine (AMNA)	10,539,208		
Ring no answer (PNA)	14,827,024		
Total known residentials	87,275,431	34.66%	70.76%
Total known business	36,068,713	14.32%	29.24%
Total known working	123,344,145	48.98%	
Total known nonworking	128,487,450	51.02%	
Total known status	251,831,595	100.00%	
Total unknown	25,366,232		
Grand total	277,197,827		

Table 4-5. Screener response rate calculations for the HINTS 2005 RDD sample

Response class	Estimated totals	Percent residential (actual or imputed)	Actual or imputed residential
Completed screener (C)	33,939,126	100.00%	33,939,126
Ineligible (I)	29,397	100.00%	29,397
Refusal (REF)	36,691,462	100.00%	36,691,462
Other nonresponse (O)	16,615,446	100.00%	16,615,446
Business (NRS)	36,068,713	0.00%	-
Nonworking (NWS)	128,487,450	0.00%	-
Answering machine (AMNA)	10,539,208	70.76%	7,457,297
Ring no answer (PNA)	14,827,024	34.66%	5,138,493
Total residential			99,871,221
Total complete			33,968,524
Response rate			34.01%

4.4.3 Extended Interview Response Rate

Table 4-6 shows the extended interview response rate calculations for the HINTS 2005 RDD sample. The extended interview response rate is designed to be an estimator of the percentage of persons 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 4-6. Extended interview response rate calculations for HINTS 2005 RDD sample

Extended interview response category	Estimated totals	Percentage
Completes or partial	112,238,130	
Ineligible	104,383	
Refusal	71,081,101	
Out of scope	3,435,616	
Total cooperative	112,342,513	***61.25%***
Total refusal	71,081,101	38.75%
Total eligible	183,423,614	100.00%
Total out of scope	3,435,616	
Total	186,859,230	

4.4.4 Overall Response Rate

The overall response 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 in 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 4.4.3). Table 4-7 presents the calculation of the overall response rate.

Table 4-7. Overall response rate calculations for HINTS 2005 RDD sample

Response rate type	Percentage
Screener response rate	34.01%
Extended response rate	61.25%
Overall response rate	20.83%

4.5 Response Rate Comparison between Two Interview Mode Experimental Groups

Table 4-8 presents the results of the experiment on the two interview mode assignments. As discussed in Section 4.2, two interview mode groups, telephone interview only and telephone-Internet choice, were conducted for both Wave 1 sample and the first two batches of Wave 2 sample. Table 4-9

provides the comparison of weighted response rates between the two interview mode experimental groups for this portion of the sample.

Table 4-8. Sample sizes and completes for two interview mode experimental groups

Telephone numbers	Telephone interview only	Telephone-Internet choice	Overall
Total sample	11,512	9,211	20,723
Screener completes	1,601	1,286	2,887
With telephone interview by default	1,601	793	2,394
With choosing telephone interview	-	266	266
With choosing Internet interview	-	227	227
Extended interview completes	1,057	766	1,823
By telephone	1,057	611	1,668
By telephone followup	-	60	60
By Internet	-	95	95

Table 4-9. Response rates for two interview mode experimental groups

Weighted response rate	Telephone interview only	Telephone-Internet choice	Overall
Screener	39.3%	38.8%	39.1%
Extended	65.4%	57.0%	61.7%
Overall	25.7%	22.1%	24.1%

4.6 Response Rate Comparison among Three Promised Incentive Experimental Groups

Table 4-10 presents the results of the experiment on the three promised incentive assignments. As discussed in Section 3.2, all of the three promised incentive groups—$0, $5, and $15—were conducted only for the Wave 1 sample; the $15 incentive was assigned for the first two batches of Wave 2; and the $5 and $15 incentives were applied to the rest of batches of Wave 2 and Wave 3 samples. Table 4-11 provides the comparison of weighted response rates between the three promised incentive experimental groups for these breakdowns.

Table 4-10. Sample sizes and completes for three promised incentive experimental groups

Batch breakdown	Complete counts	Promised incentive			
		$0	$5	$15	Overall
Wave 1	Sample	4,928	4,932	4,943	14,803
	Screener	674	732	688	2,094
	Extended	426	458	436	1,320
Wave 2 Batches 1,2	Sample			5,920	5,920
	Screener			793	793
	Extended			503	503
Wave 2 Batches 3+	Sample		30,309	24,430	54,739
	Screener		3,244	2,654	5,898
	Extended		2,010	1,753	3,763
Total	Sample				75,462
	Screener				8,785
	Extended				5,586

Table 4-11. Response rates for three promised incentive experimental groups

Batch breakdown	Response rate	Promised incentive			
		$0	$5	$15	Total
Wave 1	Screener	38.2%	42.2%	39.3%	39.9%
	Extended	62.2%	60.2%	61.3%	61.2%
	Overall	23.8%	25.4%	24.1%	24.4%
Wave 2 Batches 1,2	Screener			37.0%	37.0%
	Extended			62.9%	62.9%
	Overall			23.2%	23.2%
Wave 2 Batches 3+	Screener		31.8%	32.3%	32.0%
	Extended		58.8%	63.8%	61.1%
	Overall		18.7%	20.6%	19.6%
Total	Screener				34.0%
	Extended				61.2%
	Overall				20.8%

The important comparisons are for the extended interview response rate for Wave 1, where equivalent samples were randomly allocated across the three conditions. There were essentially no differences between the three conditions with respect to the final response rate. Nominally, the $0 condition actually has a higher response rate. However, when examining the $5 and $15 conditions for the batches remaining in Wave 2, there is a difference between the two conditions in the expected direction.

The rationale for the decision to continue the experiment into Wave 2 was based on preliminary data during Wave 1. At the time the decision had to be made, only initial cooperation rates

were available (Table 4-12). At this point in the process, there were large differences between the $0 and $15 conditions (61.2% versus 68.4%). As can be seen from Table 4-12, while the initial cooperation rates were different, the final cooperation rates mirror the final response rates, with virtually no differences between the conditions. This pattern indicates that at least for the first wave of interviews, the $15 incentive served to get cooperation faster than the other incentive conditions. After refusal conversion was completed, however, the different incentive conditions were essentially the same. This pattern did not hold for the continuation of the experiment during the last part of the field period. The $15 condition did produce a higher response rate than the $5 condition.

Table 4-12. Final and initial cooperation rates for three promised incentive experimental groups

Batch breakdown	Cooperation rate	Promised incentive			
		$0	$5	$15	Total
Wave 1	Final	74.2%	68.6%	71.8%	70.9%
	Initial	61.2%	60.2%	68.4%	63.2%
Wave 2 Batches 1,2	Final			68.4%	68.4%
	Initial			62.2%	62.2%
Wave 2 Batches 3+	Final		67.4%	73.0%	69.9%
	Initial		61.4%	68.1%	64.5%
Total	Final	72.4%	67.6%	72.1%	34.0%
	Initial	61.2%	61.2%	67.2%	61.2%

5. CONCLUSIONS

One of the most important goals of the HINTS program is to provide an ability to track changes in how the general population accesses and uses different communication channels to obtain health information on cancer. Building on the work of HINTS 2003, this was accomplished by successfully administering the HINTS for a second time.

The work began by reviewing the HINTS 2003 instrument. A small number of HINTS 2003 interviews were administered and monitored by project staff. A set of correlations among the items were computed to summarize the psychometrics of the responses. Both of these tasks were used to assess the performance of the HINTS 2003 instrument and generate initial ideas of what may need to be changed for HINTS 2005. To develop the new instrument, a series of meetings was held to finalize the core content of each section. Once questionnaire items were finalized, they were subjected to cognitive interviews, usability tests (web survey) and a field pretest. The pretest was used to make sure the computer-assisted telephone interview (CATI) program was working properly, assess the flow and timing of the instrument.

Several measures were taken to try to increase the response rate. The telephone numbers in the sample were matched to a reverse directory and addresses were found for a subsample of the cases. For those with an address, an advance letter and $2 was sent to complete the screening interview. Extended interview respondents were provided up to $15 to complete the interview (depending on which experimental incentive group they were in). Some respondents were given a choice to fill out the survey on the web or do it by telephone.

Despite these efforts, the response rate dropped significantly relative to HINTS 2003 (approximately 13%). The incentives did increase the response rates by a few percentage points. However, the experiment giving respondents a choice of mode decreased the combined response rate by several percentage points. The decline is primarily related to a dramatic drop in the screener response rate from 54 percent to 34 percent. The overall extended interview response rate did not significantly change.

There has been a decline in response rates for random digit dialog (RDD) surveys throughout the survey industry (e.g., Curtin, et al., 2005). The general population is not as willing to participate in telephone surveys, and they are making it more difficult to contact them over the telephone (e.g., answering attendants; voice mail). Related to this is an increasing proportion of the population who

are relying on cell phones as their main telecommunication device. As of the writing of this report, this is not a significant enough coverage issue to abandon the RDD methodology. The proportion of "cell-only" households is still relatively small (approximately 6% in 2004; Tucker, et al., 2005). However, the widespread use of cell phones may be making it harder to contact households that have a landline telephone, but primarily rely on cell phones for their voice communication (e.g., younger; highly mobile). In the future, the use of cell phones may eventually lead to a coverage problem with RDD surveys. It will be necessary to develop alternative methodologies to contact and/or administer the survey.

Beyond the general design issues related to RDD, there are several areas that HINTS should consider when moving into the future. With the increasing reluctance and suspicions of respondents, it will be necessary to improve the methods used to communicate the goals of the survey. Both the written materials (e.g., prenotification letter) and the scripts interviewers use when introducing the survey are important components to obtaining high cooperation. One of the most difficult tasks for an RDD interviewer is to communicate the importance of a survey in the first few seconds of the initial household contact. A majority of the refusals to the screening interview occur at this point. This emphasizes the importance of the prenotification letter, which can prepare the respondent for the call. It also points to the importance of the message conveyed in the first moments of contacting the household.

Conveying the importance of HINTS to the general public is not a simple task. While the idea of having strategies and treatments to reduce cancer is very salient for many people, understanding how people obtain information about these strategies is not simple to communicate. Without simple communication strategies, advance letters may not be as effective as they could be. More importantly, when interviewers attempt to explain the study, they may not be armed with material that is convincing. For example, interviewers reported that there is some uncertainty around the use of "National Cancer Institute" or "Cancer" in lead introductions. Both may leave the impression that the call is asking for money because the former may be mistaken for a charitable organization fighting cancer and the latter is a common cause of many charitable organizations. The use of the word cancer may also lead to many people thinking that the call is not relevant to them if they have not had personal experience with it. On the other hand, both of these terms, if understood in the right context, should be clear strengths to convincing respondents to participate. Future administrations of the HINTS should consider experimenting with different messages because they affect respondent's willingness to participate on the survey.

The response rate is also affected by the length and complexity of the HINTS interview. Based on monitoring and interviewer reports, many of the refusals to the extended interview occurred when reading the introduction explaining how long the survey would take. Reducing the length of the interview should increase the number of people who cooperate. A related issue for HINTS is the complexity of the questionnaire. There are sections that make relatively subtle distinctions that are not immediately apparent to respondents. In addition, respondents commented on the perception that questions were redundant. At times, this complexity frustrated respondents and affected the interviewer's ability to convince them to continue. Future waves of HINTS should review complex sections and possibly redesign them to flow better from the respondent's perspective.

The experiences of HINTS 2005 point to the need to adapt the methods and instrumentation to a changing survey climate. Current thinking is that one adaptation will be to mix several different modes. The HINTS 2005 mode experiment indicated that mixing modes by offering a choice is not very effective. Perhaps using a methodology that sequenced the mode assignments would work better (e.g., Link and Mokdad, forthcoming) or use some type of dual frame design.

Regardless of how successful these efforts are, there will be a need to evaluate survey error more fully than has been done in the past. Studies examining the implications of nonresponse should be built into the next wave of HINTS. Even with low response rates, it is not clear which survey estimates are subject to significant bias. If a mixed mode design is used, then it will be important to assess differences in the measures across the modes. This may become increasingly important if the use of modes like the web increase in popularity. Access to the web is correlated with a number of important predictors of access to communication mechanisms, incidence of cancer and lifestyles (e.g., income, age, education). As HINTS moves forward with different collection methodologies, it will be critical to understand how these new methods affect the measurement of these outcomes.

REFERENCES

AAPOR. (2004). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. Third 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.

Biemer, P.P., and Lyberg, L.E. (2003). *Introduction to survey quality*. New York:, NY: John Wiley

Binson, D., Canchola, J.A., and Catania, J.A. (2000). Random selection in a national telephone survey: a comparison of the Kish, next-birthday, and last-birthday methods, *Journal of Official Statistics*, 16, 53-59.

Brick, J.M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). Two-phase list-assisted RDD sampling. *Journal of Official Statistics*, 18, 203-215.

Chang, L., and Krosnick, J.A. (2003). *National surveys via RDD telephone interviewing vs. the Internet: Comparing sample representativeness and response quality.* Ohio State University. Manuscript submitted for publication.

Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly,* 64(4), 464-494.

Croyle, R.T., and Lerman, C. (1999). Risk communication in genetic testing for cancer susceptibility. *Journal of the National Cancer Institute.* Monographs, 25, 59-66.

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

Dillman, D. (1999). *Mail and Internet surveys, the tailored design method*. John Wiley & Sons, Inc.: New York.

Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., and Berck, J. (2001, May). *Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the Internet.* Paper presented at 2001 AAPOR Annual Conference, Montreal, Quebec.

Fischhoff, B., Bostrom, A., and Quadrel, M.J. (1993). Risk perception and communication. *Annual Review of Public Health,* 14, 183-203.

Glanz, K., Lewis, F.M., and Rimer, B.K. (1997). *Health behavior and health education.* San Francisco: John Wiley & Sons, Inc.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in household surveys. New York, NY: John Wiley.

Heerwegh, D., and Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review,* 20(4), 471-484.

Katz, J.E., and Rice, R.E. (2001). The Internet, 1995-2000: Access, civic involvement, and social interaction. *American Behavioral Scientist,* 45(3), 405-419.

Kish, L. (1992). Weighting for unequal p_i, *Journal of Official Statistics 8*, 183-200.

Link, M. and Mokdad (forthcoming) Use of alternative modes for health surveillance studies: Results of a web/mail/telephone experiment. *Journal of Epidemiology*.

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

Oh, H., and Scheuren, F. (1983). Weighting adjustments for unit response. In W.G. Madow, I. Olkin, and D. B. Rubin (Eds.). *Incomplete data in sampling surveys, Vol. II: Theory and annotated bibliography*. New York: Academic Press.

Rizzo, L., Brick, J.M., and Park, I. (2004). A minimally intrusive method for sampling persons in RDD surveys, *Public Opinion Quarterly*, 68, 267-274.

Rogers, E.M. (1995). Lessons for guidelines from the diffusion of innovations. *Joint Commission Journal on Quality Improvement,* 21(7), 324-328.

Tucker, C., Brick, J.M., and Meekins, B. (2005). Household telephone service and usage patterns in the U.S. in 2004: Implications for telephone samples. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Miami, FL.

Tucker, C., Casady, R., and Lepkowski, J. (1993). 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.

Viswanath K., and Finnegan, J.R. (1996). The knowledge gap hypothesis: 25 years later. In B. Burleson (Ed.), *Communication Yearbook 19* (187-227). Thousand Oaks, CA: Sage Publications).

Weinstein, N.D. (1993). Testing four competing theories of health-protective behavior. *Health Psychology,* 12(4), 324-333.

APPENDIX A:

HINTS 2005 vs. 2003 Items

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
Group Assignment	
GA-0A	HE-12
GA-0B	HE-13
GA-0C	HE-14
GA-0D	New to HINTS 2005
GA-1	HC-20
GA-2	New to HINTS 2005
GA-3	New to HINTS 2005
GA-4A	HC-23 (altered)
GA-4B	New to HINTS 2005
GA-5	New to HINTS 2005
GA-7	New to HINTS 2005
Health Communication	
HC-01a	HC-6a
HC-01b	HC-6b
HC-01c	New to HINTS 2005
HC-02a	New to HINTS 2005
HC-02b	New to HINTS 2005
HC-02c	New to HINTS 2005
HC-03a	HC-7a
HC-03b	New to HINTS 2005
HC-03c	New to HINTS 2005
HC-04	New to HINTS 2005
HC-05	New to HINTS 2005
HC-08	New to HINTS 2005
HC-09	New to HINTS 2005
HC-10	New to HINTS 2005
HC-11	New to HINTS 2005
HC-13a	HC-18a: altered
HC-13b	HC-18b: altered
HC-13c	HC-18c: altered
HC-13d	HC-18d: altered
HC-13e	HC-18e: altered
HC-13f	HC-18f: altered
HC-13g	HC-18g: altered
HC-14a	HC-24a: altered
HC-14b	HC-24b: altered
HC-14c	HC-26a
HC-14d	HC-26b
HC-14e	HC-26c
HC-14f	New to HINTS 2005
HC-14g	New to HINTS 2005
HC-14h	New to HINTS 2005
HC-14i	New to HINTS 2005
HC-14k	HC-26g
HC-14k	HC-26g
HC-15	New to HINTS 2005
HC-16	New to HINTS 2005

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
Cancer Communication	
CA-01	CH-1: altered
CA-02	CH-2: altered
CA-03	CH-3
CA-04	New to HINTS 2005
CA-05	New to HINTS 2005
CA-06	CH-4: altered
CA-07	New to HINTS 2005
CA-08	HC-9
CA-09	HC-10: altered
CA-10	HC-11: altered
CA-11	HC-12
CA-12	HC-13: altered
CA-13	HC-14
CA-14a	HC-19b
CA-14b	HC-19e
CA-14c	HC-19g
CA14d	HC-19j
CA-15	HC-29
CA-16	HC-31: altered
CA-17	HC-15A: altered
CA-18	HC-15: altered
CA-21c	HC-34d
CA-21d	HC-34: altered
CA-21e	HC-34e
CA-21h	New to HINTS 2005
CA-21i	HC-34f: altered
CA-22	New to HINTS 2005
CA-23	New to HINTS 2005
Mental Model of Cancer	
MM-01	CC-1, TU-16, CC-1: altered
MM-02	CC-2: altered
MM-03	CC-3, CC-3: altered
MM-04	CK-4: altered
MM-05a	CK-14b: altered
MM-05b	CK-14c: altered
MM-05c	New to HINTS 2005
MM-05d	New to HINTS 2005
MM-05e	New to HINTS 2005
MM-05f	CK-14a: altered
MM-05h	CC-26b: altered
MM-05i	CC-26c: altered
MM-05l	New to HINTS 2005
MM-06	New to HINTS 2005
MM-07	New to HINTS 2005
Prostate Cancer	
PC-01	PC-3A
PC-02	PC-5

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
PC-03	PC-6: altered
PC-04	New to HINTS 2005
PC-05	New to HINTS 2005
PC-06	New to HINTS 2005
Cervical Cancer	
CV-01	CV-1
CV-02	CV-2: altered
CV-03	New to HINTS 2005
CV-04	CV-4: altered
CV-05	CV-5
CV-06	CV-6: altered
CV-07	New to HINTS 2005
CV-08	New to HINTS 2005
CV-09	New to HINTS 2005
CV-10	New to HINTS 2005
CV-11	New to HINTS 2005
CV-12	New to HINTS 2005
CV-13a	New to HINTS 2005
CV-13b	New to HINTS 2005
CV-13c	New to HINTS 2005
CV-13d	New to HINTS 2005
CV-13e	New to HINTS 2005
CV-13f	New to HINTS 2005
Breast Cancer	
BC-01	BC-5
BC-02	BC-6
Colon Cancer	
CC-01	CC-4
CC-04	New to HINTS 2005
CC-05	CC-6: altered
CC-06	CC-7
CC-07	CC-16: altered
CC-08	CC-18
Skin Protection	
SP-01a	New to HINTS 2005
SP-01b	New to HINTS 2005
SP-01c	New to HINTS 2005
SP-01d	New to HINTS 2005
SP-01e	New to HINTS 2005
SP-02	New to HINTS 2005
SP-03	New to HINTS 2005
Tobacco Use	
TU-01	TU-1
TU-02	TU-2
TU-03	TU-3
TU-04	New to HINTS 2005
TU-05	TU-4: altered
TU-06	New to HINTS 2005

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
TU-07	New to HINTS 2005
TU-08	New to HINTS 2005
TU-10	New to HINTS 2005
TU-11	TU-6: altered
TU-12	New to HINTS 2005
TU-13	TU-7: altered
TU-14	New to HINTS 2005
TU-17	TU-12
TU-18a	New to HINTS 2005
TU-18b	New to HINTS 2005
TU-18c	New to HINTS 2005
TU-18d	New to HINTS 2005
TU-18e	New to HINTS 2005
TU-18jI	New to HINTS 2005
TU-18jII	New to HINTS 2005
TU-18jII	New to HINTS 2005
TU-19	New to HINTS 2005
TU-20	New to HINTS 2005
TU-20a	New to HINTS 2005
TU-21	New to HINTS 2005
TU-22	New to HINTS 2005
TU-23	New to HINTS 2005
TU-24	New to HINTS 2005
Energy Balance	
EB-01	FV-2: altered
EB-02	FV-1
EB-03	FV-3a
EB-04	FV-3: altered
EB-05	FV-4
EB-06	New to HINTS 2005
EB-07	New to HINTS 2005
EB-08	New to HINTS 2005
EB-09	New to HINTS 2005
EB-10	EX-3: altered
EB-11	New to HINTS 2005
EB-12	New to HINTS 2005
EB-13	New to HINTS 2005
EB-15	New to HINTS 2005
EB-16	New to HINTS 2005
EB-17	New to HINTS 2005
EB-18	New to HINTS 2005
EB-19	New to HINTS 2005
EB-20	New to HINTS 2005
EB-21a	New to HINTS 2005
EB-21b	New to HINTS 2005
EB-21c	New to HINTS 2005
EB-21d	New to HINTS 2005
EB-21e	New to HINTS 2005

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
EB-22	HW-1
EB-23	HW-2
Health Status	
HS-01	HS-1
HS-02a	HS-2a
HS-02b	HS-2b
HS-02c	HS-2c
HS-02d	HS-2d
HS-02e	HS-2e
HS-02f	HS-2f
HS-03	HS-3
HS-04	HS-5
HS-05	HC-3: altered
Social Networks	
SN-01	New to HINTS 2005
SN-02	New to HINTS 2005
SN-03	New to HINTS 2005
SN-04	New to HINTS 2005
SN-05	New to HINTS 2005
SN-06	New to HINTS 2005
Demographics	
DM-01	DM-1
DM-02	DM-2: altered
DM-03	DM-6: altered
DM-04	DM-4
DM-05	DM-5
DM-06	New to HINTS 2005
DM-07	New to HINTS 2005
DM-08	New to HINTS 2005
DM-09	New to HINTS 2005
DM-11	New to HINTS 2005
DM-12	New to HINTS 2005
DM-13	DM-3: altered
DM-14	New to HINTS 2005
DM-14a	DM-7a
DM-14b	DM-7b
DM-14c	DM-7c
DM-14d	DM-7d
DM-14e	DM-7e
DM-14f	DM-7f
DM-14g	DM-7g
DM-14h	New to HINTS 2005
DM-14i	New to HINTS 2005
DM-14j	New to HINTS 2005
DM-15	New to HINTS 2005
Debriefing Questions	
DB-01	New to HINTS 2005
DB-02	New to HINTS 2005

HINTS 2005 ITEM NUMBER	HINTS 2003 ITEM NUMBER
DB-03	New to HINTS 2005
DB-04	New to HINTS 2005
DB-05	New to HINTS 2005
DB-06	New to HINTS 2005
DB-07	New to HINTS 2005
DB-08	New to HINTS 2005
DB-09	New to HINTS 2005
DB-10	New to HINTS 2005
DB-11	New to HINTS 2005
DB-13	New to HINTS 2005
DB-14	New to HINTS 2005
DB-15	New to HINTS 2005
DB-16	New to HINTS 2005

APPENDIX B:

HINTS 2005 Extended Questionnaire

GROUP ASSIGNMENT (GA)

GA-0A. [AUTOFILL. ASK ONLY IF NECESSARY: May I please have your age?]

|___|___|___| .....(GA-0C)
AGE
DK ..... (GA-0B)
RF ..... (GA-0B)

GA-0B. Are you…

less than 18 years old,	1	(END STATEMENT 2)
between 18 and 34,	2	
35 to 39,	3	
40 to 44, or	4	
45 or older?	5	
DK	9	(END STATEMENT 2)
RF	8	(END STATEMENT 2)

GA-0C. [ASK IF NOT OBVIOUS: Are you male or female?]

MALE	1
FEMALE	2

GA-0D. WHICH LANGUAGE IS THIS INTERVIEW GOING TO BE CONDUCTED IN?

ENGLISH	1
SPANISH	2

{We are interested in conducting this survey over the Internet in the future./We are giving some respondents the option to complete the questionnaire on the Internet.} So, I need to ask you a few questions about your access to the Internet.

[INTERNET REFERS TO ALL SERVICES OFFERED BY AN INTERNET SERVICE PROVIDER. IT INCLUDES THE USE OF E-MAIL, THE WORLD WIDE WEB, BULLETIN BOARDS, CHAT GROUPS, DISCUSSION GROUPS, NEWS GROUPS, ON-LINE ORDERING FACILITIES, FILE TRANSFER (FTP), WEB TV, REAL AUDIO, ETC.]

GA-1. Do you ever go on-line to use the Internet or World Wide Web, or to send and receive e-mail?

YES	1	
NO	2	(NEXT SECTION)

GA-2. Where do you go on-line from to use the Internet? [PROBE: Anywhere else?]
[CODE ALL THAT APPLY.]

HOME	11
WORK	12
SCHOOL	13
A PUBLIC LIBRARY	14
A COMMUNITY CENTER	15
SOMEONE ELSE'S HOUSE	16
SOME OTHER PLACE (SPECIFY) _____	91

BOX GA-1

IF GA-2 HAS ONLY 1 RESPONSE, GO TO BOX GA-2.
OTHERWISE, CONTINUE.

GA-3. Where do you use the Internet from most often?

HOME	1
WORK	2
SCHOOL	3
A PUBLIC LIBRARY	4
A COMMUNITY CENTER	5
SOMEONE ELSE'S HOUSE	6
[PLACE SPECIFIED IN GA-2]	7

BOX GA-2

IF GA-2 = 11 (HOME), ASK GA-4A.
OTHERWISE, ASK GA-4B.

GA-4A. When you use the Internet at home, do you mainly access it through . . .

a telephone modem,	1
a cable or satellite modem,	2
a DSL modem,	3
a wireless device such as a PDA, or	4
some other way? (SPECIFY)_____	91

GO TO BOX GA-3.

GA-4B. When you use the Internet at [FILL GA-3], do you mainly access it through . . .

a telephone modem or	1
some other way?	2

BOX GA-3

IF INTERVIEW IS IN SPANISH (GA-0D=2), GO TO NEXT SECTION.
IF RANDOMLY ASSIGNED TO TELEPHONE, GO TO NEXT SECTION.
IF RANDOMLY ASSIGNED TO CHOICE, CONTINUE.

GA-5. To make participation in this study as convenient as possible for you, you have the choice of completing the rest of the questions over the telephone or on the Internet. Which would you prefer?

TELEPHONE	1	(NEXT SECTION)
INTERNET	2	

GA-6. We would like to e-mail the information you need to access the questionnaire on the Internet. What is your e-mail address?

EMAIL: _____ (END 2)

DOESN'T HAVE AN EMAIL ADDRESS..... (GA-7)
DK ..... (GA-7)
RF ..... (GA-7)

GA-7. It is easier to give you the information for accessing the Internet questionnaire in writing, because you will need some detailed instructions. In order to mail you the information, I need your name and address.

DK ..... (GA-08)
RF ..... (GA-08)

_____
FIRST NAME

_____
LAST NAME

_____
STREET ADDRESS1

_____
STREET ADDRESS2

_____
CITY

|__|__| STATE

|__|__|__|__|__| - |__|__|__|__| ZIP CODE

GO TO END2.

GA-8. Then, let me give you the information you need to access the web site over the phone. Do you have a pencil and paper? First, let me give you the web site address. It is [INSERT URL]. On the home page, you will be asked to log in with a unique name and password. Your login name is [FILL LOGIN] and your password is [FILL PASSWORD]. The survey will be available to you starting tomorrow evening.

[END2] Thank you for your time today. Please call us at 1-888-314-1133 if you have any questions about accessing the web site to complete the rest of the questions. If we haven't received your survey in 2 weeks, we will call you back to make sure you haven't had any problems getting onto the web site.

BOX HC-1

IF RESPONDENT DOES NOT WATCH TV (HC-01a=95) THEN SKIP QUESTIONS HC-02a, HC-03b, HC-03c, HC-08, HC-09 AND HC-13g.

IF RESPONDENT DOES NOT LISTEN TO THE RADIO (HC-01b=95) THEN SKIP QUESTIONS HC-02b, AND HC-13e.

IF RESPONDENT DOES NOT USE THE INTERNET (GA-1=2 OR HC01c=95) THEN SKIP QUESTIONS HC-01c, HC-02c, HC-10, HC-11, HC-13f, AND HC-14 THROUGH HC-16.

IF RESPONDENT CANNOT READ (HC-03a=9) THEN SKIP QUESTIONS HC-04, HC-05, HC-13c AND HC-13d.

MEDIA EXPOSURE

Before the questions specifically about cancer, there are some questions about how you get information in general.

HC-01. On a typical weekday, about how many hours do you…

[IF GREATER THAN ZERO, BUT LESS THAN ONE HOUR, ENTER 1.]

[IF NOT APPLICABLE, ENTER 95.]

	HOURS
a. watch television?	|___|___|
b. listen to the radio?	|___|___|
c. use the Internet for personal reasons?	|___|___|

HC-02. During a typical weekend, including both Saturday and Sunday, about how many hours do you….

[IF GREATER THAN ZERO, BUT LESS THAN ONE HOUR, ENTER 1.]

	HOURS
a. watch television?	|___|___|
b. listen to the radio?	|___|___|
c. use the Internet for personal reasons?	|___|___|

HC-03. In the past seven days, how many days did you...

[IF NOT APPLICABLE, ENTER 9]

	DAYS
a. read a newspaper?	|___|___|
b. watch the national news on television?	|___|___|
c. watch the local news on television?	|___|___|

EXPOSURE TO HEALTH INFORMATION

The next few questions are about various ways you might get health information.

HC-04. Some newspapers or general magazines publish a special section that focuses on health. In the past 12 months, have you read health sections of the newspaper or of a general magazine?

YES..... 1
NO..... 2 (HC-08)

HC-05. About how often have you read such health sections in the past 12 months? Would you say…..

once or more per week, or..... 1
less than once per week?..... 2

HC-06 AND HC-07 DELETED.

HC-08. Some local television news programs include special segments of their newscasts that focus on health issues. In the past 12 months, have you watched health segments on the local news?

YES..... 1
NO..... 2 (HC-10)

HC-09. How often have you watched health segments on local news in the past 12 months? Would you say….

once or more per week, or..... 1
less than once per week?..... 2

HC-10. Some people notice information about health on the Internet, even when they are not trying to find out about a health concern they have or someone in the family has. Have you read such health information on the Internet in the past 12 months?

YES..... 1
NO..... 2 (HC-13)

HC-11. About how often have you read this sort of information in the past 12 months? Would you say…

once or more per month, or 1
less than once per month? 2

HC-12 DELETED.

HC-13. How much would you trust information about health or medical topics [FILL SOURCE]? Would you say a lot, some, a little or not at all? (How about from [FILL SOURCE])?

[ASK IN RANDOM ORDER.]

	A LOT	SOME	A LITTLE	NOT AT ALL
a. from a doctor or other health care professional	1	2	3	4
b. from family or friends	1	2	3	4
c. in newspapers	1	2	3	4
d. in magazines	1	2	3	4
e. on the radio	1	2	3	4
f. on the Internet	1	2	3	4
g. on television	1	2	3	4

INTERNET USAGE FOR HEALTH

HC-14. Here are some ways people use the Internet. Some people have done these things, but other people have not. In the past 12 months, have you done the following things while using the Internet?

	YES	NO
a. Looked for health or medical information for yourself?.....	1	2
b. Looked for health or medical information for someone else?.....	1	2
c. Bought medicine or vitamins on-line?.....	1	2
d. Participated in an on-line support group for people with a similar health or medical issue?.....	1	2
e. Used e-mail or the Internet to communicate with a doctor or a doctor's office?	1	2
f. Looked for information about physical activity or exercise?	1	2
g. Looked for information about diet or nutrition?.....	1	2
h. Looked for information about protecting yourself from the sun?.....	1	2
i. Looked for information about quitting smoking?.....	1	2
k. Done anything else health-related on the Internet? (SPECIFY) _____	1	2

HC-14j DELETED.

BOX HC-2

IF DID NOT LOOK FOR HEALTH INFO FOR SELF OR SOMEONE ELSE ON INTERNET (HC-14a=2 AND HC-14b=2), THEN GO TO NEXT SECTION.
OTHERWISE, CONTINUE.

HC-15. Have you ever talked to a doctor, nurse, or other health care provider about any kind of health information you have gotten from the Internet?

YES	1	
NO	2	(NEXT SECTION)

HC-16. When you talked with a health care provider, how interested were they in hearing about the information you found on-line? Were they…

very interested,	1
somewhat interested,	2
a little interested, or	3
not at all interested?	4

TOUCHED BY CANCER

The next few questions are about your personal experience with cancer.

CA-01. Have you ever been told by a doctor that you had cancer?

YES	1	
NO	2	(CA-06)
YES, BUT IT WAS A MIS-DIAGNOSIS	3	(CA-06)

CA-02. What type of cancer was it, or in what part of the body did the cancer start? [PROBE: Anything else?] [CODE ALL THAT APPLY.]

BLADDER CANCER	10
BONE CANCER	11
BREAST CANCER	12
CERVICAL CANCER (CANCER OF THE CERVIX)	13
COLON CANCER	14
ENDOMETRIAL CANCER (CANCER OF THE UTERUS)	15
HEAD AND NECK CANCER	16
HODGKIN'S LYMPHOMA	17
LEUKEMIA/BLOOD CANCER	18
LIVER CANCER	19
LUNG CANCER	20
MELANOMA	21
NON-HODGKIN'S LYMPHOMA	22
OTHER SKIN CANCER	23
ORAL CANCER	24
OVARIAN CANCER	25
PANCREATIC CANCER	26
PHARYNGEAL (THROAT) CANCER	27
PROSTATE CANCER	28
RECTAL CANCER	29
RENAL (KIDNEY) CANCER	30
STOMACH CANCER	31
OTHER (SPECIFY) _____	91

CA-03. At what age or in what year were you first told that you had cancer?
[ENTER UNIT.]

|___|
UNIT

AGE	1
YEAR	2

[ENTER {AGE/YEAR}.]

|___|___|___|___|___|
AGE/YEAR

CA-04. Did you undergo treatment for your cancer?

YES	1	
NO	2	(CA-06)

CA-05. How long ago did you finish your treatment?

[ENTER UNIT.]

|___|
UNIT

MONTHS	1	
YEARS	2	
STILL IN TREATMENT	3	(CA-06)

[ENTER NUMBER.]

|___|___|
NUMBER

CA-06. Have any of your family members ever had cancer?

[IF INDICATE "DOESN'T HAVE FAMILY," CODE AS "NO FAMILY."]

YES	1	
NO	2	(CA-08)
NO FAMILY	3	(CA-08)

CA-07. What type of cancer was it? [PROBE: Anything else?]
[CODE ALL THAT APPLY.]

BLADDER CANCER	10
BONE CANCER	11
BREAST CANCER	12
CERVICAL CANCER (CANCER OF THE CERVIX)	13
COLON CANCER	14
ENDOMETRIAL CANCER (CANCER OF THE UTERUS)	15
HEAD AND NECK CANCER	16
HODGKIN'S LYMPHOMA	17
LEUKEMIA/BLOOD CANCER	18
LIVER CANCER	19
LUNG CANCER	20
MELANOMA	21
NON-HODGKIN'S LYMPHOMA	22
OTHER SKIN CANCER	23
ORAL CANCER	24
OVARIAN CANCER	25
PANCREATIC CANCER	26
PHARYNGEAL (THROAT) CANCER	27
PROSTATE CANCER	28
RECTAL CANCER	29
RENAL (KIDNEY) CANCER	30
STOMACH CANCER	31
OTHER (SPECIFY) _____	91

INFORMATION SEEKING ABOUT CANCER

Next are some questions about looking for information on cancer. Please consider all sources of information such as the Internet, the library, friends, and health care professionals.

CA-08. Have you ever looked for information about cancer from any source?

YES	1	
NO	2	

CA-09. Not including your doctor or other health care provider, has someone else ever looked for information about cancer for you?

YES	1	
NO	2	(BOX CA-1)

CA-10. Who was that? [PROBE: Anyone else?]
[CODE ALL THAT APPLY.]

SPOUSE	10
OTHER FAMILY MEMBER	11
FRIEND	12
CO-WORKER	13
INFORMATION SPECIALIST (E.G., LIBRARIAN)	14
OTHER (SPECIFY)_____	91

BOX CA-1

IF RESPONDENT HAS NOT LOOKED FOR INFORMATION
FROM ANY SOURCE ON CANCER, NEITHER INDIVIDUALLY
NOR THROUGH SOMEONE ELSE (CA-08=2 AND CA-09=2), THEN GO TO CA-17.
OTHERWISE, CONTINUE.
IF SOMEONE ELSE LOOKED FOR INFORMATION (CA-09=1), THEN INCLUDE
BRACKETED PHRASE IN CA-11 INTRODUCTION.

Think about the most recent time you looked for cancer-related information from any source {either on your own or by someone else looking for you}.

CA-11. About how long ago was that?

[ENTER UNIT.]

|___|
UNIT

DAYS AGO..... 1
WEEKS AGO..... 2
MONTHS AGO 3
YEARS AGO..... 4

[ENTER NUMBER.]

|___|___|
NUMBER

BOX CA-2
IF SP HAS NOT LOOKED FOR INFORMATION ABOUT CANCER THEMSELF (CA-08=2), THEN GO TO CA-17.
OTHERWISE, CONTINUE.

CA-12. The most recent time you wanted information on cancer, where did you go first?

[IF SP HAS MADE MORE THAN ONE SEARCH, PROBE FOR THE MOST RECENT SEARCH.]

BOOKS..... 1
BROCHURES, PAMPHLETS, ETC. 2
CANCER ORGANIZATION 3
FAMILY 4
FRIEND/CO-WORKER 5
HEALTH CARE PROVIDER..... 6
INTERNET..... 7
LIBRARY 8
MAGAZINES..... 9
NEWSPAPERS..... 10
SOMEONE WITH CANCER 11
TELEPHONE INFORMATION NUMBER (1-800 NUMBER)..... 12
OTHER (SPECIFY)_____ 91

CA-13. What type of information were you looking for in your most recent search? [CODE ALL THAT APPLY.]

A SPECIFIC CANCER	10
CANCER ORGANIZATIONS	11
CAUSES OF CANCER / RISK FACTORS FOR CANCER	12
COPING WITH CANCER / DEALING WITH CANCER	13
DIAGNOSIS OF CANCER	14
INFORMATION ON CANCER IN GENERAL	15
PAYING FOR MEDICAL CARE / INSURANCE	16
PREVENTION OF CANCER	17
PROGNOSIS / RECOVERY FROM CANCER	18
SCREENING / TESTING / EARLY DETECTION	19
SYMPTOMS OF CANCER	20
TREATMENT / CURES FOR CANCER	21
WHERE TO GET MEDICAL CARE	22
OTHER (SPECIFY)_____	91

BARRIERS TO CANCER INFORMATION SEEKING

CA-14. Based on the results of your most recent search for information on cancer, how much do you agree or disagree with the following statements?

[IF SP HAS MADE MORE THAN ONE SEARCH, PROBE FOR THE MOST RECENT SEARCH.]

	STRONGLY AGREE	SOMEWHAT AGREE	SOMEWHAT DISAGREE	STRONGLY DISAGREE
a. It took a lot of effort to get the information you needed. Would you say you...	1	2	3	4
b. You felt frustrated during your search for the information. (Would you say you...)	1	2	3	4
c. You were concerned about the quality of the information. (Would you say you...)	1	2	3	4
d. The information you found was too hard to understand. (Would you say you...)	1	2	3	4

BOX CA-3

IF RESPONDENT DOESN'T USE INTERNET (GA-1=2), GO TO CA-17.
IF RESPONDENT'S MOST RECENT SEARCH WAS ON THE INTERNET (CA-12=7), GO TO CA-16.
OTHERWISE, CONTINUE.

CANCER INFO ON THE INTERNET

CA-15. Have you ever visited an Internet web site to learn specifically about cancer?

YES	1	
NO	2	(CA-17)

CA-16. {Thinking about all the times you've looked for cancer information on the Internet,} {How/how} useful was the cancer-related information you got from the Internet? Would you say . . .

very useful,	1
somewhat useful,	2
a little useful, or	3
not at all useful?	4

INFORMATION EFFICACY

CA-17. Overall, how confident are you that you could get advice or information about cancer if you needed it? Would you say…

completely confident,	1
very confident,	2
somewhat confident,	3
a little confident, or	4
not confident at all?	5

SOURCE PREFERENCES

CA-18. {The next time you have a strong need to get information about cancer, where will you go first?/Imagine that you had a strong need to get information about cancer. Where would you go first?}

BOOKS	1
BROCHURES, PAMPHLETS, ETC.	2
CANCER ORGANIZATION	3
FAMILY	4
FRIEND/CO-WORKER	5
HEALTH CARE PROVIDER	6
INTERNET	7
LIBRARY	8
MAGAZINES	9
NEWSPAPERS	10
TELEPHONE INFORMATION NUMBER (1-800 NUMBER)	11
SOMEONE WITH CANCER	12
OTHER (SPECIFY)_____	91

CA-19 AND CA-20 DELETED.

SOURCE RECOGNITION AND USE

CA-21. Before being contacted for this study, had you ever heard of...

	YES	NO	
c. the National Cancer Institute?.....	1	2	(CA-21e)
d. the National Cancer Institute's Cancer Information Service?.....	1	2	
e. the 1-800-4-Cancer information number?.....	1	2	
h. the 1-800-ACS-2345 cancer information number?	1	2	
i. Cancer Control of America?.....	1	2	

CA-21a, CA-21b, CA-21f AND CA-21g DELETED.

BOX CA-4

IF HAVE NEVER HEARD OF CIS OR 800-NUMBER (CA-21d AND CA-21e = 2) THEN GO TO NEXT SECTION.

OTHERWISE, ASK CA-22 FOR EACH "YES" ANSWER IN CA-21d AND CA-21e BEFORE MOVING ON TO NEXT ITEM IN CA-21.

CA-22. Have you ever contacted it for information?

YES..... 1
NO..... 2

BOX CA-5

IF RESPONDENT HAS EVER LOOKED FOR CANCER INFORMATION (CA-08=1) AND ANSWERED "NO" TO CA-22e, THEN CONTINUE.
OTHERWISE, GO TO NEXT SECTION.

CA-23. Is there a particular reason you didn't contact it?

_____

BOX PC-1

IF RESPONDENT IS MALE, IS 45 OR OLDER, AND HAS NOT HAD PROSTATE CANCER, CONTINUE.
OTHERWISE, GO TO NEXT SECTION

The next few questions are about getting tested for cancer. A Prostate-Specific Antigen test, also called a PSA test, is a blood test used to check men for prostate cancer.

PC-01. Have you ever heard of a PSA or prostate-specific antigen test?

YES	1	
NO	2	(NEXT SECTION)

PC-02. Have you ever had a PSA test?

YES	1	
NO	2	(INTRO TO PC-04)
HAD BLOOD TEST, BUT DK IF CHECKED PSA	3	(INTRO TO PC-04)

PC-03. When did you have your most recent PSA test?

A YEAR AGO OR LESS	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS AGO	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS AGO	3
OVER 5 YEARS AGO	4

The next few questions are about discussions that health care providers might have had with you about the PSA test.

PC-04. Has a health care provider such as a doctor or nurse ever talked to you about a PSA test?

YES	1	
NO	2	(NEXT SECTION)

PC-05. Thinking about the last time a health care provider talked to you about a PSA test, which of the following statements best describes your health care provider's recommendation about PSA tests?

That you should have a PSA test,	1
that you should NOT have a PSA test, or	2
your health care provider did not make a recommendation	3

PC-06. Thinking about the last time a health care provider talked to you about a PSA test, did your health care provider encourage you to ask questions or express any concerns you had about PSA testing? Would you say…..

yes, definitely,	1
yes, somewhat, or	2
no, not at all?	3
DID NOT HAVE ANY QUESTIONS OR CONCERNS ABOUT PSA	4

PC-07 DELETED.

BOX CV-1

IF RESPONDENT IS FEMALE AND HAS NOT HAD CERVICAL CANCER, CONTINUE. OTHERWISE, GO TO NEXT SECTION

The next few questions are about getting tested for cancer. A Pap smear is a test for cancer of the cervix.

CV-01. Have you ever had a Pap smear?

YES	1	
NO	2	(CV-07)

CV-02. When did you have your most recent Pap smear?

A YEAR AGO OR LESS	1
MORE THAN 1 BUT NOT MORE THAN 3 YEARS AGO	2
MORE THAN 3 BUT NOT MORE THAN 5 YEARS AGO	3
OVER 5 YEARS AGO	4

CV-03. What was the main reason that you had this Pap smear?

ROUTINE ANNUAL PAP SMEAR OR PART OF ROUTINE PHYSICAL EXAM	1
LAST PAP SMEAR WAS NOT NORMAL	2
A SPECIFIC PROBLEM	3
SOMETHING SHE HEARD / SAW / READ	4
SHE HAD NEVER HAD ONE AND THOUGHT SHE SHOULD	5
PREGNANCY / FOLLOW-UP TO BIRTH	6
OTHER (SPECIFY)_____	91

CV-04. You said your most recent Pap smear was {INSERT TIME FRAME FROM CV-02}. How long before <u>that</u> Pap smear was the previous one?

A YEAR OR LESS BEFORE	1
MORE THAN 1 BUT NOT MORE THAN 3 YEARS BEFORE	2
MORE THAN 3 BUT NOT MORE THAN 5 YEARS BEFORE	3
OVER 5 YEARS BEFORE	4
THIS WAS FIRST PAP SMEAR	5

CV-05. Have you had a hysterectomy?

[IF NEEDED: A hysterectomy is an operation to remove the uterus (womb).]

YES	1	(CV-07)
NO	2	

CV-06. When do you expect to have your next Pap smear?

A YEAR OR LESS FROM NOW	1
MORE THAN 1 BUT NOT MORE THAN 3 YEARS FROM NOW	2
MORE THAN 3 BUT NOT MORE THAN 5 YEARS FROM NOW	3
OVER 5 YEARS FROM NOW	4
AM NOT PLANNING TO HAVE ANOTHER	5
IF I HAVE SYMPTOMS	6
WHEN DOCTOR / HEALTH PROVIDER RECOMMENDS	7

CV-07. How often do you think a woman your age should have a Pap smear?

MORE THAN TWICE A YEAR	1
TWICE A YEAR / EVERY SIX MONTHS	2
ONCE A YEAR	3
EVERY TWO YEARS	4
EVERY THREE YEARS	5
MORE THAN EVERY THREE YEARS	6
NEVER	7

CV-08. Most medical organizations now recommend a Pap smear every three years for healthy adult women. Have you heard about this change in guidelines?

YES	1
NO	2

BOX CV-2

IF RESPONDENT DOES NOT PLAN TO HAVE ANOTHER PAP SMEAR (CV-06=5)
OR HAS HAD A HYSTERECTOMY (CV-05=1) , GO TO CV-10.
OTHERWISE, CONTINUE.

CV-09. Would you agree to have Pap smears every three years if your health care provider recommended it?

YES	1
NO	2

CV-10. Have you ever been treated for venereal warts or condyloma?

YES	1
NO	2

CV-11. Have you ever heard of HPV? HPV stands for Human Papillomavirus.

YES	1	
NO	2	(NEXT SECTION)

CV-12. Have you ever been told by a health care provider that you had HPV infection?

YES	1
NO	2

CV-13. Next are some questions on your opinion about HPV.

	YES	NO
a. Do you think that HPV causes cervical cancer?	1	2
b. Do you think that HPV is a sexually transmitted disease?.....	1	2
c. Do you think that HPV infection is rare?.....	1	2
d. Do you think that HPV will often go away on its own without treatment?.....	1	2
e. Do you think HPV can cause abnormal Pap smears?.....	1	2
f. Do you think that HPV can affect a woman's ability to get pregnant?	1	2

BOX BC-1

IF RESPONDENT IS FEMALE, 35 YEARS OR OLDER AND HAS NOT HAD BREAST CANCER, CONTINUE.
OTHERWISE, GO TO NEXT SECTION.

The next few questions are about breast cancer.

BC-01. A mammogram is an x-ray of each breast to look for breast cancer. Have you ever had a mammogram?

YES	1	
NO	2	(BC-03)

BC-02. When did you have your most recent mammogram to check for breast cancer?

A YEAR AGO OR LESS	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS AGO	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS AGO	3
OVER 5 YEARS AGO	4

BC-03 DELETED.

The next few questions are about colon cancer.

CC-01. Can you think of any tests that detect colon cancer? [PROBE: Anything else?]

[CODE ALL THAT APPLY. IF R DOES NOT KNOW NAME OF TEST, ENTER DESCRIPTION OF TEST IN OTHER, SPECIFY.]

[DO NOT NAME OR DEFINE TESTS FOR THE RESPONDENT.]

BARIUM ENEMA	10
BIOPSY	11
STOOL BLOOD TEST/FECAL OCCULT BLOOD TEST	12
COLONOSCOPY	13
DIGITAL RECTAL EXAM	14
ENDOSCOPY	15
PROCTOSCOPY	16
SIGMOIDOSCOPY	17
OTHER (SPECIFY)_____	91
NO/NOTHING	0

CC-02 AND CC-03 DELETED.

BOX CC-1

IF RESPONDENT IS LESS THAN 45 YEARS OLD OR HAS HAD COLON CANCER,
GO TO NEXT SECTION.
OTHERWISE, CONTINUE.

CC-04. Has a doctor, nurse or other health professional ever advised you to get a test to check for colon cancer?

YES	1	
NO	2	

CC-05. A stool blood test, also known as a Fecal Occult Blood Test, is a test done to check for colon cancer. It is done at home using a set of 3 cards to determine whether the stool contains blood. Have you ever done this test using a home kit?

YES	1	
NO	2	(CC-07)

CC-06. When did you do your most recent stool blood test using a home kit to check for colon cancer?

A YEAR AGO OR LESS	1	
MORE THAN 1 BUT NOT MORE THAN 2 YEARS AGO	2	
MORE THAN 2 BUT NOT MORE THAN 5 YEARS AGO	3	
OVER 5 YEARS AGO	4	

CC-07. A sigmoidoscopy and a colonoscopy are both tests that examine the bowel by inserting a tube in the rectum. Have you ever had either a colonoscopy or a sigmoidoscopy?

YES	1	
NO	2	(NEXT SECTION)

CC-08. When did you have your most recent sigmoidoscopy or colonoscopy to check for colon cancer?

A YEAR AGO OR LESS	1	
MORE THAN 1 BUT NOT MORE THAN 5 YEARS AGO	2	
MORE THAN 5 BUT NOT MORE THAN 10 YEARS AGO	3	
OVER 10 YEARS AGO	4	

BOX MM-1

RESPONDENTS WILL BE RANDOMLY ASSIGNED TO RECEIVE MENTAL MODEL QUESTIONS ON COLON, LUNG OR SKIN CANCER. IF RESPONDENT IS ASSIGNED TO COLON CANCER, CONTINUE.

IF RESPONDENT IS ASSIGNED TO COLON CANCER BUT HAS HAD COLON CANCER, GO TO THE NEXT SECTION.

IF RESPONDENT IS NOT ASSIGNED TO COLON CANCER, GO TO THE NEXT SECTION.

MM-01. How likely do you think it is that you will develop colon cancer in the future? Would you say your chance of getting colon cancer is . . .

very low,	1
somewhat low,	2
moderate,	3
somewhat high, or	4
very high?	5

MM-02. Compared to the average person your age, would you say that you are . . .

more likely to get colon cancer,	1
less likely, or	2
about as likely?	3

MM-03. How often do you worry about getting colon cancer? Would you say . . .

rarely or never,	1
sometimes,	2
often, or	3
all the time?	4

MM-04. What are some things that people can do to reduce their chances of getting colon cancer? [CODE ALL THAT APPLY.]

DON'T DRINK ALCOHOL	10
DON'T SMOKE	11
EAT FIBER	12
EAT FRUITS AND VEGETABLES	13
EAT HEALTHY / BETTER NUTRITION	14
EXERCISE	15
GET SCREENED FOR CANCER / GET TESTED FOR CANCER	16
HAVE REGULAR CHECK-UPS	17
OTHER (SPECIFY)	91

MM-05. Do you agree or disagree with the following statements?

	AGREE	DISAGREE
a. There's not much you can do to lower your chances of getting colon cancer. Would you say you…	1	2
b. There are so many different recommendations about preventing colon cancer that it's hard to know which ones to follow. (Would you say you)…	1	2
c. Colon cancer develops over a period of several years. (Would you say you…)	1	2
d. There are ways to slow down or disrupt the development of colon cancer. (Would you say you….)	1	2
e. Colon cancer is most often caused by a person's behavior or lifestyle. (Would you say you….)	1	2
f. It seems like almost everything causes colon cancer. (Would you say you…).	1	2
h. You are reluctant to get checked for colon cancer because you fear you may have it. (Would you say you…)	1	2
i. Getting checked regularly for colon cancer increases the chances of finding cancer when it's easy to treat. (Would you say you…)	1	2
l. People with colon cancer would have pain or other symptoms prior to being diagnosed. (Would you say you...)	1	2

MM-05g, MM-05j AND MM-05k DELETED.

MM-06. What are the common symptoms of colon cancer?
[CODE ALL THAT APPLY.]

BLOOD IN STOOL	10
CHANGE IN BOWEL HABITS	11
CONSIPATION	12
DIARRHEA	13
LOSS OF APPETITE	14
RECTAL BLEEDING	15
STOMACH PAIN	16
SWELLING	17
TIREDNESS / FATIGUE	18
VOMITING	19
OTHER (SPECIFY)_____	91

MM-07. Overall, how many people who develop colon cancer do you think survive at least 5 years? Your best guess is fine. Would you say…..

less than 25 percent,	1
about 25 percent,	2
about 50 percent,	3
about 75 percent, or	4
nearly all?	5

SKIN PROTECTION (SP)

The next few questions are about things people might do to protect themselves from the sun.

SP-01. When you go outside for more than 1 hour on a warm, sunny day, how often do you...

[IF R DOES NOT GO OUTSIDE ON A SUNNY DAY FOR MORE THAN ONE HOUR, CODE 95.]

	ALWAYS	OFTEN	SOMETIMES	RARELY	NEVER	DOES NOT GO OUT ON SUNNY DAY	
a. wear sunscreen? Would you say	1	2	3	4	5	95	(SP-02)
b. stay in the shade? (Would you say…).....	1	2	3	4	5		
c. wear a hat that shades your face, ears and neck? (Would you say…).....	1	2	3	4	5		
d. wear a long-sleeve shirt? (Would you say…).....	1	2	3	4	5		
e. wear long pants? (Would you say…).....	1	2	3	4	5		

SP-02. How many times in the past 12 months have you used indoor tanning devices such as a sun lamp, a sun bed, or a tanning booth?

0 TIMES 1
1-2 TIMES 2
3-10 TIMES 3
11-24 TIMES..... 4
25 TIMES OR MORE 5

SP-03. How many times in the past 12 months have you used sunless tanning products? These are products that you either have sprayed on or which you apply to darken the color of your skin.

0 TIMES 1
1-2 TIMES 2
3-10 TIMES 3
11-24 TIMES..... 4
25 TIMES OR MORE 5

SP-04 DELETED.

BOX MM-1

RESPONDENTS WILL BE RANDOMLY ASSIGNED TO RECEIVE MENTAL MODEL QUESTIONS ON COLON, LUNG OR SKIN CANCER.
IF RESPONDENT IS ASSIGNED TO SKIN CANCER, CONTINUE.

IF RESPONDENT IS ASSIGNED TO SKIN CANCER, BUT HAS HAD SKIN CANCER, GO TO THE NEXT SECTION.

IF RESPONDENT IS NOT ASSIGNED TO SKIN CANCER, GO TO THE NEXT SECTION.

The next few questions are about skin cancer.

MM-01. How likely do you think it is that you will develop skin cancer in the future? Would you say your chance of getting skin cancer is . . .

very low,	1
somewhat low,	2
moderate,	3
somewhat high, or	4
very high?	5

MM-02. Compared to the average person your age, would you say that you are . . .

more likely to get skin cancer,	1
less likely, or	2
about as likely?	3

MM-03. How often do you worry about getting skin cancer? Would you say . . .

rarely or never,	1
sometimes,	2
often, or	3
all the time?	4

MM-04. What are some things that people can do to reduce their chances of getting skin cancer? [CODE ALL THAT APPLY.]

DO NOT USE TANNING BEDS / TANNING SALONS	10
GET SCREENED FOR CANCER / GET TESTED	11
HAVE REGULAR CHECK UPS	12
STAY OUT OF THE SUN	13
WEAR PROTECTIVE CLOTHING / HAT	14
WEAR SUNSCREEN	15
OTHER (SPECIFY)	91

MM-05. Do you agree or disagree with the following statements?

	AGREE	DISAGREE
a. There's not much you can do to lower your chances of getting skin cancer. Would you say you….....	1	2
b. There are so many different recommendations about preventing skin cancer that it's hard to know which ones to follow. (Would you say you)….....	1	2
c. Skin cancer develops over a period of several years. (Would you say you….)	1	2
d. There are ways to slow down or disrupt the development of skin cancer. (Would you say you….).....	1	2
e. Skin cancer is most often caused by a person's behavior or lifestyle. (Would you say you….).....	1	2
f. It seems like almost everything causes skin cancer. (Would you say you….).	1	2
h. You are reluctant to get checked for skin cancer because you fear you may have it. (Would you say you…)	1	2
i. Getting checked regularly for skin cancer increases the chances of finding cancer when it's easy to treat. (Would you say you…).....	1	2
l. People with skin cancer would have pain or other symptoms prior to being diagnosed. (Would you say you...).....	1	2

MM-05g, MM-05j AND MM-05k DELETED.

MM-06. What are the common symptoms of skin cancer?
[CODE ALL THAT APPLY.]

ABNORMAL GROWTHS..... 10
BLISTERS..... 11
BLOTCHES..... 12
LESIONS..... 13
MOLE / CHANGE IN MOLE..... 14
PIGMENT DISCOLORATION..... 15
RASH..... 16
REDNESS OF SKIN..... 17
OTHER (SPECIFY)_____ 91

MM-07. Overall, how many people who develop skin cancer do you think survive at least 5 years? Your best guess is fine. Would you say…..

less than 25 percent, ... 1
about 25 percent,..... 2
about 50 percent,..... 3
about 75 percent, or..... 4
nearly all?..... 5

TOBACCO SCREENER

Next are some questions about your use of cigarettes.

TU-01. Have you smoked at least 100 cigarettes in your entire life?

[IF NEEDED: 5 Packs = 100 Cigarettes.]

YES	1	
NO	2	(TU-17)

TU-02. Do you now smoke cigarettes . . .

every day,	1	
some days, or	2	(TU-4)
not at all?	3	(TU-11)
DK	9	(TU-17)
RF	8	(TU-17)

TU-03. On the average, how many cigarettes do you now smoke a day?

[IF NEEDED: 1 Pack = 20 Cigarettes.]

[IF LESS THAN ONE A DAY, ENTER 0. IF 76 OR MORE, ENTER 76.]

|___|___|
NUMBER OF CIGARETTES

GO TO TU-7.

TU-04. On how many of the past 30 days did you smoke cigarettes?

I___I___I
DAYS

TU-05. On the average, on those {FILL WITH TIME PERIOD FROM TU-04} days, how many cigarettes did you usually smoke each day?

[IF NEEDED: 1 Pack = 20 Cigarettes.]

[IF LESS THAN ONE A DAY, ENTER 0. IF 76 OR MORE, ENTER 76.]

|___|___|
NUMBER OF CIGARETTES

QUIT ATTEMPTS

BOX TU-1

IF SMOKED 12 OR MORE OF THE PAST 30 DAYS (TU-04=>12), GO TO TU-07.
OTHERWISE, CONTINUE.

TU-06. During the past 12 months, have you tried to quit smoking completely?

YES..... 1
NO..... 2

GO TO TU-10.

TU-07. Have you ever stopped smoking for one day or longer because you were trying to quit smoking?

YES..... 1
NO..... 2 (TU-09)

TU-08. How many times during the past 12 months have you stopped smoking for one day or longer because you were trying to quit smoking?

I___I___I___I
TIMES

TU-09 DELETED.

TU-10. Are you seriously considering quitting smoking within the next 6 months?

YES..... 1
NO..... 2

BOX TU-2

IF SMOKE EVERY DAY OR SOME DAYS (TU-02=1 OR 2), GO TO TU-17.
OTHERWISE, CONTINUE.

FORMER SMOKERS

TU-11. About how long has it been since you completely quit smoking cigarettes?

I___I___|___|___|___|___| |___|
TIME UNIT

[ENTER UNIT.]

DAYS 1
WEEKS 2
MONTHS..... 3
YEARS 4

TU-12. Have you ever smoked cigarettes every day for at least 6 months?

YES..... 1
NO..... 2 (BOX TU-3)

TU-13. When you last smoked every day, how many cigarettes did you usually smoke each day?

[IF NEEDED: 1 Pack = 20 Cigarettes.]

[IF LESS THAN ONE A DAY, ENTER 0. IF 76 OR MORE, ENTER 76.]

|___|___|
NUMBER OF CIGARETTES

BOX TU-3

IF QUIT SMOKING OVER 1 YEAR AGO, GO TO TU-17.
IF QUIT SMOKING 1 YEAR AGO OR LESS, CONTINUE.

TU-14. Around this time 12 months ago, were you smoking cigarettes…..

every day,..... 1
some days, or 2
not at all?..... 3

TU-15 AND TU-16 DELETED.

ALL RESPONDENTS

TU-17. If a new cigarette were advertised as less harmful than current cigarettes, how interested would you be in trying it? Would you say . . .

very interested, ... 1
somewhat interested, or 2
not interested? 3

TU-18. Next are some questions about new types of tobacco products that have been recently introduced. Have you ever heard of a tobacco product called….(How about…)

	YES	NO
a. Eclipse?	1	2
b. Quest?	1	2
c. Marlboro UltraSmooth?	1	2
d. Ariva?	1	2
e. Revel?	1	2

TU-18f, TU-18g, TU-18h AND TU-18i DELETED.

ASK TU-19 FOR EACH "YES" ANSWER IN TU-18
BEFORE ASKING NEXT ITEM IN TU-18.

TU-18jI. Have you heard of any other types of new tobacco products? These would include products like Accord, Advance, Omni, Exalt and Stonewall.

YES..... 1
NO..... 2 (BOX TU-5)

BOX TU-5

IF TU-18a THROUGH TU-18jI = NO, THEN GO TO NEXT SECTION.

TU-18jII. What other new products have you heard of?
[CODE ALL THAT APPLY.]

ACCORD..... 10
ADVANCE..... 11
OMNI..... 12
EXALT..... 13
STONEWALL..... 14
OTHER (SPECIFY)_____91

ASK TU-19 INDIVIDUALLY FOR EACH ITEM LISTED IN TU-18jII.

TU-19. How did you hear about {FILL TU-18a-jII}? [PROBE: Anything else?]
[CODE ALL THAT APPLY.]

FROM A DOCTOR OR HEALTH CARE PROVIDER	10
FROM FAMILY OR FRIENDS	11
IN A NEWSPAPER	12
IN A MAGAZINE	13
ON THE RADIO	14
ON THE INTERNET	15
ON TELEVISION	16
SAW IT IN A STORE	17
OTHER (SPECIFY) _____	91

TU-20. You said you have heard of {FILL WITH LIST OF ITEMS MENTIONED IN TU-18a-jII}. Have you ever tried {this product/any of these products}?

YES	1	
NO	2	(NEXT SECTION)

BOX TU-5a

IF RESPONDENT HAS ONLY HEARD OF 1 PRODUCT,
GO TO TU-21.
OTHERWISE, CONTINUE.

TU-20a. Which {one/ones}?
[CODE ALL THAT APPLY.]

ACCORD	10
ADVANCE	11
ARIVA	12
ECLIPSE	13
EXALT	14
MARLBORO ULTRASMOOTH	15
OMNI	16
QUEST	17
REVEL	18
STONEWALL	19
{FILL WITH OTHER, SPECIFY IN TU-18jII}	20

BOX TU-5b

IF ONE PRODUCT NAMED IN TU-20a, ASK TU-21 THROUGH TU-24 FOR THIS PRODUCT.

IF MORE THAN ONE PRODUCT NAMED IN TU20a, SELECT ONE PRODUCT AT RANDOM FROM GROUP 1 AND ONE FROM GROUP 2. ASK TU-21 THROUGH TU-24 FOR THE RANDOMLY SELECTED PRODUCTS IN EACH GROUP.

GROUP 1: ECLIPSE, QUEST, MARLBORO ULTRASMOOTH, ACCORD, ADVANCE AND OMNI.

GROUP 2: ARIVA, REVEL, EXALT AND STONEWALL.

TU-21. Do you now use {FILL}……

every day..... 1
some days, or 2
not at all?..... 3

BOX TU-6

IF FORMER USER OF NEW PRODUCT (TU-21=3), THEN USE PAST TENSE IN TU-22 AND TU-23.

IF CURRENT USER OF NEW PRODUCT (TU-21=1 OR 2), THEN USE PRESENT TENSE IN TU-22 AND TU-23.

TU-22. {Now that you are using/When you tried or used} {FILL}, {did/do} you smoke…..

less of your usual brand of cigarettes,..... 1
about the same number of your usual brand of cigarettes, or..... 2
more of your usual brand of cigarettes? 3
I DID NOT SMOKE USUAL BRAND OF CIGARETTES AT ALL / I COMPLETELY SWITCHED TO THIS PRODUCT..... 4

TU-23. What is the main reason you {use/tried or used} {FILL}?

INSTEAD OF QUITTING, AS A WAY TO REDUCE HEALTH RISKS..... 1
AS A WAY OF CUTTING DOWN THE NUMBER OF CIGARETTES/ OTHER CIGARETTES SMOKED..... 2
TO HELP YOU QUIT SMOKING..... 3
BECAUSE OF THE TASTE..... 4
BECAUSE LOWER COSTS/CHEAPER THAN CIGARETTES/ OTHER CIGARETTES/OTHER TYPES OF TOBACCO PRODUCTS. 5
OUT OF CURIOSITY 6
BECAUSE NOT AS STRONG/LESS NICOTINE/LIGHTER/LESS TAR THAN CIGARETTES/OTHER CIGARETTES 7
BECAUSE LESS HARMFUL TO USE AROUND FAMILY/FRIENDS/CHILDREN THAN CIGARETTES/OTHER CIGARETTES 8
BECAUSE CAN USE IN PLACES WHERE CAN'T SMOKE CIGARETTES/OTHER CIGARETTES..... 9
OTHER (SPECIFY) _____ 91

TU-24. As far as you know, is {FILL}……

less harmful than conventional cigarettes,..... 1
equally harmful, or..... 2
more harmful than conventional cigarettes? 3

REPEAT ITEMS TU-21 THROUGH TU-24 FOR ITEM IN GROUP 2 IFAPPLICABLE.

BOX MM-1

RESPONDENTS WILL BE RANDOMLY ASSIGNED TO RECEIVE MENTAL MODEL QUESTIONS ON COLON, LUNG OR SKIN CANCER.

IF RESPONDENT IS ASSIGNED TO LUNG CANCER, CONTINUE.

IF RESPONDENT IS ASSIGNED TO LUNG CANCER, BUT HAS HAD LUNG CANCER, GO TO THE NEXT SECTION.

IF RESPONDENT IS NOT ASSIGNED TO LUNG CANCER, GO TO THE NEXT SECTION.

The next few questions are about lung cancer.

MM-01. How likely do you think it is that you will develop lung cancer in the future? Would you say your chance of getting lung cancer is . . .

very low,	1
somewhat low,	2
moderate,	3
somewhat high, or	4
very high?	5

MM-02. Compared to the average person your age, would you say that you are . . .

more likely to get lung cancer,	1
less likely, or	2
about as likely?	3

MM-03. How often do you worry about getting lung cancer? Would you say . . .

rarely or never,	1
sometimes,	2
often, or	3
all the time?	4

MM-04. What are some things that people can do to reduce their chances of getting lung cancer? [CODE ALL THAT APPLY.]

AVOID ASBESTOS	10
AVOID POLLUTED AIR	11
DON'T SMOKE / QUIT SMOKING	12
EAT HEALTHY	13
EXERCISE	14
GET SCREENED FOR CANCER / GET TESTED	15
HAVE REGULAR CHECK UPS	16
STAY AWAY FROM SECOND-HAND SMOKE	17
OTHER (SPECIFY)	91

MM-05. Do you agree or disagree with the following statements?

		AGREE	DISAGREE
a.	There's not much you can do to lower your chances of getting lung cancer. Would you say you….....	1	2
b.	There are so many different recommendations about preventing lung cancer that it's hard to know which ones to follow. (Would you say you)….....	1	2
c.	Lung cancer develops over a period of several years. (Would you say you…)	1	2
d.	There are ways to slow down or disrupt the development of lung cancer. (Would you say you….).....	1	2
e.	Lung cancer is most often caused by a person's behavior or lifestyle. (Would you say you….).....	1	2
f.	It seems like almost everything causes lung cancer. (Would you say you….).	1	2
h.	You are reluctant to get checked for lung cancer because you fear you may have it. (Would you say you…).....	1	2
i.	Getting checked regularly for lung cancer increases the chances of finding cancer when it's easy to treat. (Would you say you…).....	1	2
l.	People with lung cancer would have pain or other symptoms prior to being diagnosed. (Would you say you...).....	1	2

MM-05g, MM-05j AND MM-05k DELETED.

MM-06. What are the common symptoms of lung cancer?
[CODE ALL THAT APPLY.]

BRONCHITIS..... 10
CHEST PAIN..... 11
COUGHING..... 12
DEPRESSION..... 13
DIFFICULTY BREATHING / SHORTNESS OF BREATH / WHEEZING .. 14
EXCESSIVE PHLEGM / MUCUS..... 15
FATIGUE / TIREDNESS..... 16
LOSS OF APPETITE..... 17
PNEUMONIA..... 18
SPITTING UP BLOOD..... 19
SWELLING OF NECK AND/OR FACE..... 20
WEAKNESS..... 21
WEIGHT LOSS..... 22
OTHER (SPECIFY) _____ 91

MM-07. Overall, how many people who develop lung cancer do you think survive at least 5 years? Your best guess is fine. Would you say…..

less than 25 percent,	1
about 25 percent,	2
about 50 percent,	3
about 75 percent, or	4
nearly all?	5

NUTRITION

These next few questions are about how often you ate or drank different kinds of foods during the past 30 days, for example, twice a week, three times a month, and so forth. Include all foods you ate or drank, both at home and away from home.

EB-01. During the past 30 days, how often did you drink 100% fruit juice such as orange, apple, and grape juices? Do not include fruit drinks like Kool-Aid or Hi-C.

[ENTER NUMBER.]
[IF NEVER, ENTER 95.]

|___|___| TIMES |___| UNIT

[ENTER UNIT.]

PER DAY..... 1
PER WEEK..... 2
PER MONTH..... 3
PER YEAR..... 4

EB-02. During the past 30 days, how often did you eat fruit? Include fresh, canned, or frozen fruit.

[ENTER NUMBER.]

[IF NEVER, ENTER 95.]

|___|___| TIMES |___| UNIT

[ENTER UNIT.]
PER DAY..... 1
PER WEEK..... 2
PER MONTH..... 3
PER YEAR..... 4

EB-03. During the past 30 days, how often did you eat potatoes? Do not include things like fried potatoes, french fries, or rice.

[ENTER NUMBER.]
[IF NEVER, ENTER 95.]

|___|___| TIMES |___| UNIT

[ENTER UNIT.]
PER DAY..... 1
PER WEEK..... 2
PER MONTH..... 3
PER YEAR..... 4

EB-04. During the past 30 days, how often did you eat vegetables other than potatoes? Include things like salad, cooked dried beans, corn, and broccoli.

[ENTER NUMBER.]

[IF NEVER, ENTER 95.]

|___|___| |___|
TIMES UNIT

[ENTER UNIT.]

PER DAY..... 1
PER WEEK..... 2
PER MONTH..... 3
PER YEAR..... 4

EB-05. How many servings of fruits and vegetables do you think a person should eat each day for good health?

[IF R GIVES RANGE, PROBE FOR AN EXACT NUMBER. IF DON'T KNOW, DO NOT PROBE.]

|___|___|
SERVINGS

EXERCISE

The next few questions are about your exercise, recreation, and physical activity patterns.

EB-06. In a typical week, how many days do you do any moderate-intensity physical activity or exercise comparable to walking as if you were in a hurry?

I___I___I
DAYS

NONE..... 95 (EB-08)

EB-07. On the days that you do any moderate physical activity or exercise, how long are you typically doing these activities?

I___|___|___|___| |___|
NUMBER UNIT

[ENTER UNIT.]

MINUTES..... 1
HOURS..... 2

EB-08. How many days a week of physical activity or exercise are recommended for the average adult to stay healthy?

I___I___I
DAYS

NONE..... 95 (EB-10)

EB-09. On those days, how long should the average adult be physically active to stay healthy?

I___|___|___|___| |___|
NUMBER UNIT

[ENTER UNIT.]

MINUTES 1
HOURS 2

EB-10. As far as you know, does physical activity or exercise increase the chances of getting some types of cancer, decrease the chances of getting some types of cancer, or does it not make much difference?

INCREASES CHANCES OF CANCER 1
DECREASES CHANCES OF CANCER 2
MAKES NO DIFFERENCE 3

WEIGHT LOSS

EB-11. Have you tried to lose any weight in the past year?

YES 1
NO 2

EB-12. Are you aware of low carbohydrate, high protein diets such as the Atkins Diet, the Zone, Sugar Busters, or the South Beach Diet?

YES 1
NO ..... 2 (BOX EB-1)

EB-13. Have you tried a low carbohydrate, high protein diet in the past 12 months?

YES 1
NO 2

EB-14 DELETED.

EB-15. Do you think that a low carbohydrate, high protein diet is a healthy way to lose weight?

YES 1
NO 2

BOX EB-1

RESPONDENTS RANDOMIZED TO RECEIVE EITHER NUTRITION OR EXERCISE SERIES IN EB-16 THROUGH EB-21.

DIET AND EXERCISE INFORMATION

EB-16. When you hear or read a new recommendation about {nutrition/physical activity or exercise}, are you more likely to pay attention to it or to ignore it?

PAY ATTENTION TO IT	1	
IGNORE IT	2	(EB-18)

EB-17. Think about the last time you heard a new recommendation about {nutrition/physical activity or exercise}. Which of the following things did you do in response to the new recommendation?

I changed what I do,	1
I did not change what I do, or	2
I waited to get more information.	3

EB-18. How much do you agree or disagree with the following statement? There are so many different recommendations about {nutrition/physical activity or exercise} that it's hard to know which ones to follow. Would you say you ….

strongly agree,	1
somewhat agree,	2
somewhat disagree, or	3
strongly disagree?	4

BOX EB-2

IF RESPONDENT WAS RANDOMIZED TO RECEIVE NUTRITION SERIES, GO TO EB-20. OTHERWISE, CONTINUE.

EB-19. People who are overweight can lose a significant amount of weight by doing 30 minutes of moderate activities such as brisk walking on most days of the week. Would you say you. . . .

strongly agree,	1
somewhat agree,	2
somewhat disagree, or	3
strongly disagree?	4

EB-20. Now think about {nutrition/physical activity or exercise} and cancer. Within the past 12 months, have you seen, heard, or read anything about {nutrition/physical activity or exercise} and cancer?

YES	1	
NO	2	(EB-22)

BOX EB-3

IF RESPONDENT DOES NOT WATCH TV (HC-01a=95) THEN
SKIP EB-21a.

IF RESPONDENT CANNOT READ (HC-03a=9) THEN
SKIP EB-21b AND EB-21c.

IF RESPONDENT DOES NOT USE THE INTERNET (GA-1=2) THEN
SKIP EB-21d.

EB-21. Thinking about the past 12 months only, how much have you heard about {nutrition/physical activity or exercise} and cancer [FILL SOURCE]? Would you say a lot, some, a little or not at all? (How about from [FILL SOURCE])?

[ASK IN RANDOM ORDER.]

	A LOT	SOME	A LITTLE	NOT AT ALL
a. on television?	1	2	3	4
b. in newspapers	1	2	3	4
c. in magazines	1	2	3	4
d. on the Internet	1	2	3	4
e. from a doctor or other health care professional?	1	2	3	4

HEIGHT AND WEIGHT

The next questions are about your health now.

EB-22. About how tall are you without shoes?

[ENTER FEET.]

|___|___|
FEET

[ENTER INCHES. ROUND FRACTIONS OF INCHES DOWN TO WHOLE INCH.]

|___|___|
INCHES

EB-23. About how much do you weigh without shoes?
[ROUND FRACTIONS UP TO WHOLE NUMBER.]

|___|___|___|
POUNDS

HS-01. In general, would you say your health is...

excellent, ... 1
very good,..... 2
good, ... 3
fair, or..... 4
poor?..... 5

DEPRESSION

Next are some questions about feelings you may have experienced over the past 30 days.

HS-02. During the past 30 days, how often did you feel [FEELING]? Would you say all of the time, most of the time, some of the time, a little of the time, or none of the time?

	ALL OF THE TIME	MOST OF THE TIME	SOME OF THE TIME	A LITTLE OF THE TIME	NONE OF THE TIME
a. so sad that nothing could cheer you up	1	2	3	4	5
b. nervous	1	2	3	4	5
c. restless or fidgety	1	2	3	4	5
d. hopeless	1	2	3	4	5
e. that everything was an effort	1	2	3	4	5
f. worthless	1	2	3	4	5

BOX HS-1

IF RESPONDENT HAD ANY OF THE FEELINGS IN HS-02
ALL, MOST, OR SOME OF THE TIME, CONTINUE.
OTHERWISE, GO TO HS-04.

HS-03. The last few questions were about a number of feelings you had during the past 30 days. Altogether, how much did these feelings interfere with your life or activities? Would you say . . .

a lot, ... 1
some, ... 2
a little, or..... 3
not at all?..... 4

HEALTH COVERAGE

HS-04. Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?

YES	1
NO	2

HS-05. During the past 12 months, not counting times you went to an emergency room, how many times did you go to a doctor, nurse or other health care provider to get care for yourself?

1 TIME	1
2 TIMES	2
3 TIMES	3
4 TIMES	4
5-9 TIMES	5
10 OR MORE TIMES	6
NONE	95

Next are some questions about your participation in community organizations. By community organization we mean any group that you might participate in such as church, sports leagues, self-help groups, service clubs, or professional organizations.

SN-01. How many community organizations are you currently a member of?

|___|___|
NUMBER

NONE..... 95 (SN-03)

SN-02. {Does this/Do any of these} community organization{s} provide you with information on health?

YES..... 1
NO..... 2

SN-03. Do you have friends or family members that you talk to about your health?

YES..... 1
NO..... 2 (SN-05)

SN-04. How frequently do you talk to these friends or family members about health? Would you say…..

very frequently, ... 1
somewhat frequently, or 2
not very frequently?..... 3

SN-05. How many people live near you who you can rely on in case you need a ride to visit your health care provider?

[ENTER NUMBER.]

[IF NONE, ENTER 95.]

|___|___|
NUMBER

SN-06. Not including funerals and weddings, how often do you attend religious services? Would you say….

every week, ... 1
once or twice a month,. 2
a few times a year, or 3
never? 4

It's getting close to the end of the survey. There are just a few more questions.

DM-01. Are you currently . . .

employed for wages,	1
self-employed,	2
out of work for more than one year,	3
out of work for less than one year,	4
a homemaker,	5
a student,	6
retired, or	7
unable to work?	8

DM-02. Are you . . .

married,	1
divorced,	2
widowed,	3
separated,	4
never been married, or	5
living with a partner?	6

DM-03. What is the highest level of school you completed?

NEVER ATTENDED SCHOOL OR ONLY ATTENDED NUSERYSCHOOL/KINDERGARTEN	1
GRADES 1 THROUGH 5 (ELEMENTARY)	2
GRADES 6 THROUGH 8 (MIDDLE)	3
GRADES 9 THROUGH 12 (SOME HIGH SCHOOL BUT NO DIPLOMA)	4
HIGH SCHOOL GRADUATE (HIGH SCHOOL DIPLOMA OR EQUIVELENT, E.G., GED, FOREIGN EQUIVALENT)	5
VOCATIONAL OR TRADE SCHOOL GRADUATE	6
SOME COLLEGE, BUT NO DEGREE	7
ASSOCIATE DEGREE IN COLLEGE	8
BACHELOR'S DEGREE	9
MASTER'S DEGREE	10
PROFESSIONAL SCHOOL OR DOCTORATE DEGREE (MD, DDS, JD, DVM, Ph.D., EdD, etc)	11

DM-04. Are you Hispanic or Latino?

YES	1
NO	2

DM-05. Which one or more of the following would you say is your race? Are you American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, or White?

[CODE ALL THAT APPLY. IF R SAYS "HISPANIC," PROBE FOR ONE OF THE LISTED RACE CATEGORIES.]

WHITE	11
BLACK	12
ASIAN	13
AMERICAN INDIAN OR ALASKA NATIVE	14
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	15

DM-06. Where you born in the United States?

YES	1	(BOX DM-1)
NO	2	

DM-07. In what year did you come to live in the United States?

|___|___|___|___|
YEAR

BOX DM-1

IF RESPONDENT IS HISPANIC OR WAS NOT BORN IN THE UNITED STATES (DM-04=1 OR DM-06=2), CONTINUE.
OTHERWISE, GO TO DM-09.

DM-08. How comfortable do you feel speaking English? Would you say….

completely comfortable,	1
very comfortable,	2
somewhat comfortable,	3
a little comfortable, or	4
not at all comfortable?	5
ENGLISH IS NATIVE LANGUAGE	6
DOES NOT SPEAK ENGLISH	95

DM-09. Do you currently rent or own your home?

OWN	1
RENT	2
OCCUPIED WIHTOUT PAYING MONETARY RENT	3

DM-10 DELETED.

DM-11. Including yourself, how many people live in your household?

I___I___I (IF "1", GO TO DM-14)
NUMBER

DM-12. How many people in your household are related to you by blood, marriage, or adoption?

I___I___I
NUMBER [IF 0, GO TO DM-14]

DM-13. How many of these family members are under the age of 18?

I___I___I
NUMBER

DM-14. {Thinking about members of your family living in this household, what/What} is your {combined} annual income, meaning the total pre-tax income from all sources earned in the past year?

$ I___I___I___I___I___I___I___I___I

IF DK OR RF, ASK: Is your annual household income from all sources…..

		YES		NO	
a.	less than $25,000?	1		2	(GO TO DM-14e)
b.	less than $20,000?	1		2	(DM-15)
c.	less than $15,000?	1		2	(DM-15)
d.	less than $10,000?	1	(DM-15)	2	(DM-15)
e.	less than $35,000?	1	(DM-15)	2	
f.	less than $50,000 ($35,000 to less than $50,000)?	1	(DM-15)	2	
g.	less than $75,000 ($50,000 to less than $75,000)?	1	(DM-15)	2	
h.	less than $100,000? ($75,000 to less than $100,000)?	1	(DM-15)	2	
i.	less than $200,000? ($100,000 to $200,000)?	1	(DM-15)	2	
j.	$200,000 or more?	1		2	

DM-15. At the end of the month, how much money are you able to put aside? Your best estimate is fine.

NOTHING	1
$100 OR LESS	2
$101 TO $250	3
$251 TO $500	4
$501 TO $1,000	5
MORE THAN $1,000	6

DEBRIEFING QUESTIONS (DB)

There are just a few final questions for you about your participation in this study.

BOX DB-1

IF THE RESPONDENT COMPLETED THE QUESTIONNAIRE BY TELEPHONE, GO TO BOX DB-2.

IF RESPONDENT COMPLETED THE QUESTIONNAIRE BY INTERNET, CONTINUE.

INTERNET COMPLETERS

DB-01. You were given a choice of whether to do this questionnaire over the phone or over the Internet. Why did you decide to do this survey over the Internet?

_____

DB-02. Did you receive an e-mail with information about how to complete the survey over the Internet?

Yes 1 (DB-04)
No 2

DB-03. Did you receive a letter with information about how to complete the survey over the Internet?

Yes 1
No 2

DB-04. Did you have any difficulties accessing the survey?

Yes 1
No 2 (DB-06)

DB-05. What difficulties did you have?

_____

DB-06. How satisfied were you with the speed of your connection to the survey? Would you say…..

very satisfied, ... 1
somewhat satisfied, ... 2
somewhat dissatisfied, or 3
very dissatisfied? 4

DB-07. Did you complete this survey all in one sitting, or did you do it in more than one sitting?

I completed the survey all in one sitting 1
I completed the survey in more than one sitting 2

DB-08. During the time that you have been completing this survey, were you also doing other things such as talking to someone, reading a newspaper, or caring for children?

Yes 1
No 2 (DB-10)

DB-09. What other things were you doing?

_____

DB-10. Where were you when you completed this survey?

Home	11
Work	12
School	13
A public library	14
Some other place (specify)_____	91

GO TO DB-15

TELEPHONE COMPLETERS

BOX DB-2

IF RESPONDENT HAS BEEN RANDOMIZED TO SKIP THE DEBRIEFING QUESTIONS, GO TO THE NEXT SECTION.
IF RESPONDENT HAS BEEN RANDOMIZED TO ANSWER THE DEBRIEFING QUESTIONS AND IS IN THE TELEPHONE-ONLY GROUP, GO TO DB-15.

IF RESPONDENT HAS BEEN RANDOMIZED TO ANSWER THE DEBRIEFING QUESTIONS AND IS IN THE CHOICE GROUP, CONTINUE.

DB-11. You were given a choice of whether to do this questionnaire over the phone or over the Internet. Why did you decide to do this survey over the telephone?
[CODE ALL THAT APPLY.]

COMPUTER NOT WORKING	10
DIDN'T THINK I WOULD DO IT	11
DON'T HAVE TIME TO DO IT ON THE INTERNET	12
DON'T USE THE INTERNET VERY OFTEN	13
DON'T USE THE INTERNET VERY WELL	14
OTHER, SPECIFY _____	91

DB-12 DELETED.

BOX DB-3

IF RESPONDENT ORIGINALLY CHOSE TELEPHONE (GA-05=1), GO TO DB-15.

IF RESPONDENT ORIGINALLY CHOSE INTERNET (GA-05=2), THEN CONTINUE.

DB-13. Did you receive an e-mail with information about how to complete the survey over the Internet?

YES	1	(DB-15)
NO	2	

DB-14. Did you receive a letter with information about how to complete the survey on the Internet?

YES	1

NO..... 2

ALL RESPONDENTS

DB-15. What do you think about the length of this survey? Would you say. . . .

The survey was too short,..... 1
The survey was too long, or..... 2
The survey was a good length?. 3

BOX DB-4

IF RESPONDENT IS NOT RECEIVING AN INCENTIVE, GO TO THE NEXT SECTION. OTHERWISE, CONTINUE.

DB-16. How likely is it that you would have participated in this survey if you were not receiving {FILL AMOUNT}? Would you say……

very likely,..... 1
somewhat likely,..... 2
somewhat unlikely, or..... 3
very unlikely?..... 4

APPENDIX C:

Variation in Web vs. Telephone Instrument

VARIATION IN TELEPHONE VS. WEB VERSIONS OF INSTRUMENT

Item	How the web version differed from the telephone version
HEALTH COMMUNICATION	
HC-01a	• Soft check not conducted. • Respondents were not allowed to enter "95". • Range =0-24. • No skip based on "0"
HC-01b	• Soft check not conducted. • Respondents were not allowed to enter "95". • Range =0-24. • No skip based on "0"
HC-01c	• Respondents were not allowed to enter "95". • Range =0-24. • No skip based on "0". • Soft check not conducted.
HC-02a	• Soft check not conducted. • Range =0-48. • If both HC-01a=0 and HC-02a=0 then enter skip pattern.
HC-02b	• Soft check not conducted. • Range =0-48. • If both HC-01b=0 and HC-02b=0 then enter skip pattern.
HC-02c	• Soft check not conducted.
HC-03a	• Respondents were not allowed to enter "9". • Range = 0-7. • No skips based on "0".
HC-03b	• Respondents were not allowed to enter "9". • Range = 0-7.
HC-03c	• Respondents were not allowed to enter "9". • Range = 0-7.
CANCER COMMUNICATION	
CA-02	• Only the following response options showed to respondents: Breast cancer, cervical cancer, colon cancer, lung cancer, melanoma, other skin cancer, prostate cancer, other (specify).
CA-07	• Only the following response options showed to respondents: Breast cancer, cervical cancer, colon cancer, lung cancer, melanoma, other skin cancer, prostate cancer, other (specify).
CA-13	• No response categories provided.
COLON CANCER	
CC-01	• No response categories provided.
MENTAL MODEL OF CANCER (COLON, SKIN AND LUNG)	
MM-04	• No response categories provided.
MM-06	• No response categories provided.
TOBACCO USE	
TU-03	• Soft check not conducted.
TU-05	• Soft check not conducted.
TU-11	• Soft check not conducted.
TU-13	• Soft check not conducted.
ENERGY BALANCE	
EB-01	• If never, respondents entered "0" rather than "95."
EB-02	• If never, respondents entered "0" rather than "95."
EB-03	• If never, respondents entered "0" rather than "95."
EB-04	• If never, respondents entered "0" rather than "95."
EB-05	• Soft check not conducted.
EB-06	• If never, respondents entered "0" rather than "95."
EB-07	• Soft check not conducted.
EB-08	• If never, respondents entered "0" rather than "95."
EB-09	• Soft check not conducted.
EB-22	• Soft check not conducted.

Item	How the web version differed from the telephone version
EB-23	• Soft check not conducted.
SOCIAL NETWORKS	
SN-01	• Respondent entered "0" for none instead of "95." • Soft check not conducted.
SN-05	• Respondent entered "0" for none instead of "95." • Soft check not conducted.
DEMOGRAPHICS	
DM-11	• Soft check not conducted.
DM-14	• Soft check not conducted.

APPENDIX D:

Original HINTS 2005 Sampling Plan

NCI HINTS 2005 Sample Design

Lou Rizzo, Inho Park, Westat

April 5, 2004

The sample design is a list-assisted RDD sample from all telephone exchanges in the U.S. This will result in a nationally representative sample of households. During the household screener, one adult will be sampled within each household and recruited for the extended interview.

The list-assisted RDD method is a random sample of telephone numbers from all 'working banks' in U.S. telephone exchanges (see for example Tucker, Casady, and Lepkowski 1993). A working bank is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits 294-44) with at least one listed residential number[17].

1. Overall Sample Design for HINTS 2005

Our sample design for HINTS 2005 will include two waves, with a methodological experiment carried out in Wave 1. Section 2 discusses the details of that methodological experiment. We will also decide on a similar approach for Wave 2, but this will be done only after results from Wave 1 are evaluated.

Table 1 presents expected sample sizes for the full study under our design. A total of 32,759 telephone numbers are expected to be sampled, with an expected yield of 7,000 completed interviews. These are similar to HINTS I sample sizes. A reserve sample of 15,759 telephone numbers will also be sampled and set aside to be used in case expectations are not met (i.e., a total of 48,518 telephone numbers will be initially sampled, with 15,759 then set aside as the reserve). We will subsample out 31.3% of the nonmailable numbers (numbers for which we have no address information), as discussed in Section 3. Before this subsampling takes place, the sample size will be 39,862 numbers, from which the 32,759 will be drawn (see Table 6).

It should be noted that the Wave 1 expected extended interview response rate is lower than the Wave 2 expected extended interview response rate due to our plan to carry out the internet experiments, which will likely reduce response rates. Appendix A provides a detailed breakdown of expected response rates in Wave 1 (leading to the aggregate extended interview response rate of 66%).

[17] 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 1. Expected sample sizes for full study.

	Wave 1	Wave 2	Total
Sampled telephone numbers	17,000	15,759	32,759
Residency rate	47%	47%	47%
Residentials	7,990	7,407	15,397
Screener response rate	66.5%	66.5%	66.5%
Completed screeners	5,313	4,925	10,239
Extended interview response	65.9%	71.1%	68.4%
Yield of extended interviews	3,502	3,502	7,004

*Note that all figures in the table are rounded, leading to arithmetic inconsistencies (a*b equals c, but rounded a * rounded b is not equal to rounded c).

An important question in any survey design is the size of the overall sample. Is 7,000 the correct size, or should it be smaller or larger? Our goal for the overall survey is to be able to generate 95% confidence intervals for percentages of adults answering various questionnaire items in a particular way that were no wider than ±4 percentage points, for the domain of all adults and for the domains Hispanic adults and black adults. We are not oversampling any minority stratum in HINTS 2005, so that the expected sample allocation to each race/ethnicity group should be proportional to that of the population. Table 2 presents Current Population Survey (March 2003 supplement) estimates of adults within the domains of interest, with expected sample sizes proportional to these estimates. The effective sample sizes (the sample size of a simple random sample with the same precision) are smaller by a factor of 1.3: we expect a design effect[18] of 1.3, allowing for adult selection within households (generating variable weights for adults for differing size households) which generally has a design effect of 1.2, mailable-nonmailable subsampling (see Section 3), and nonresponse weighting adjustments.

Table 2. Expected percentages by race for the HINTS 2005 survey.

	Adults in U.S. population (in 1,000s)	Percentage of adults	Expected sample size	Expected effective sample size
Hispanic	24,631	11.76%	824	634
NonHispanic Black	23,328	11.14%	780	600
NonHispanic White and Other Race	161,496	77.10%	5,400	4,154
Total	209,454	100.00%	7,004	5,388

Table 3 presents standard errors for sample percentages ranging from 10% to 50%, using the standard binomial distribution standard error formula, with the effective sample sizes from Table 2 in the denominator. As can be seen, all are in the range 0.41% to 2.04%, giving 95% confidence intervals with half-widths in the range .80% to 4.0%, as desired. Any overall sample size smaller than 7,000 will give confidence intervals for some sample percentages for Blacks and Hispanics larger than ±4%,

[18] Design effect is defined as the ratio of the actual sample variance to the variance of a simple random sample with the same sample size. See for example Kish (1965), p. 162.

unless oversampling is done. We do not wish to do oversampling in this study (in contrast to HINTS I) as it will reduce precision levels for the experimental component of the study, as described in Section 2.

Table 3. Estimated standard errors for domain sample percentages.

Sample percentage	Hispanics	Non Hispanic Blacks	Non Hispanic Whites and other race	All adults
50%	1.98%	2.04%	0.78%	0.68%
40%	1.94%	2.00%	0.76%	0.67%
30%	1.82%	1.87%	0.71%	0.62%
20%	1.59%	1.63%	0.62%	0.54%
10%	1.19%	1.22%	0.47%	0.41%

2. Methodological Experiment in Wave 1

The methodological experiment in Wave 1 studies the effect of the mode of the extended interview, and the effect of a promised incentive for completing the extended interview. Three levels of mode will be tested:

- Telephone interview only;
- Internet interview only;
- Respondent choice of telephone or internet interview;

The telephone-interview-only respondents will not have any internet-based assessment. The internet-interview-only respondents will be asked whether they have access to the internet, and if so will be asked to complete the instrument via the internet. If an internet-interview-only respondent does not have access to the internet, then the instrument will be completed via telephone.

The respondent-choice group will be asked whether they have access to the internet. If not, they will be interviewed via telephone. If they do have access to the internet, they will be asked if they prefer to complete the instrument by telephone interview or over the internet. Their choice will be implemented.

Table 4 provides our expected breakdown for Wave 1. Our expected residency rate of 47% is higher than the current residency rates for national RDD samples because we will be oversampling mailable numbers (see Section 3). Our expected screener response rates based on HINTS I experience is 66.5% (note that we are planning to include a $2 incentive with the advance letter to make this response rate possible). Table 4 assumes differing extended interview response rates by experimental group. The internet-only experimental group is lower because of lower response rates expected for the internet interviews. Appendix A provides a breakdown of our expectations for the various experimental groups which aggregate to the extended interview response rates given in Table 4. The final yield of extended interviews from Wave 1 is expected to be 3,507. Note that this includes both telephone and internet interviews.

Table 4. Experiment group sample sizes for screeners and total telephone numbers.

	Telephone only	Internet only	Telephone internet choice	Total
Sampled telephone numbers	5,667	5,667	5,667	17,000
Residency rate	47%	47%	47%	47%
Residentials	2,663	2,663	2,663	7,990
Screener response rate	66.5%	66.5%	66.5%	66.5%
Completed screeners	1,771	1,771	1,771	5,313
Extended interview response	71.0%	61.1%	65.7%	65.9%
Yield of extended interviews	1,257	1,082	1,164	3,503

*Note that all figures in the table are rounded, leading to arithmetic inconsistencies (a*b equals c, but rounded a * rounded b is not equal to rounded c).

The three experimental groups telephone-only, internet-only, and telephone-internet choice will be further subdivided by three into a total of nine equal-sized groups. Three levels of incentive will be tested:

- No incentive;
- $10 promised incentive upon completion of extended interview;
- $20 promised incentive upon completion of extended interview.

Table 5 presents the expected breakdown by experimental group, under the 'null hypothesis' that the incentives have no effect and response rates will be the same across the incentive groups.

Table 5. Wave 1 expected experiment subgroup sample sizes.

Mode group	Incentive group	Telephone sample size	Expected screener sample size	Expected completed extended interviews
Telephone only	No incentive	1,889	590	419
Telephone only	$10 promised incentive	1,889	590	419
Telephone only	$20 promised incentive	1,889	590	419
Telephone only	All incentive groups	5,667	1,771	1,257
Internet only	No incentive	1,889	590	361
Internet only	$10 promised incentive	1,889	590	361
Internet only	$20 promised incentive	1,889	590	361
Internet only	All incentive groups	5,667	1,771	1,082
Respondent choice	No incentive	1,889	590	388
Respondent choice	$10 promised incentive	1,889	590	388
Respondent choice	$20 promised incentive	1,889	590	388
Respondent choice	All incentive groups	5,667	1,771	1,164
All mode groups	All incentive groups	17,000	5,313	3,503

*Note that all figures in the table are rounded, leading to arithmetic inconsistencies (a*b equals c, but rounded a * rounded b is not equal to rounded c).

Appendix B provides power calculations for two primary research questions. The first research question is whether extended interview response rates are increased or not by the promised incentives. The second research question is whether there are mode differences between the internet and the telephone as a means for completing the interview (i.e., whether persons systematically answer some items differently when they answer by internet and when they answer by telephone).

3. Stratification by Mailable Status

We will utilize stratification by mailable status (see for example Brick et al 2002). Numbers that are 'mailable' are those for which we have an address. In HINTS I, 43% of the RDD sample was mailable, with 57% nonmailable. We expect similar percentages in HINTS 2005 (though there may be some change). The mailable numbers have a much higher percentage of residential numbers, and we have also seen in HINTS I that we obtain a higher response rate both at the screener level and the extended interview level among the mailable numbers, especially when we send incentives with an advance letter to the mailable numbers, as we plan to do in HINTS 2005. The computations in Appendix C, using HINTS I results, show that each completed extended interview among the nonmailables can be expected to be more than twice as costly as a completed extended interview among the mailables, thereby justifying an explicit stratification by mailable status. The optimal rate for the nonmailable stratum is 68.7% of that of the mailable stratum, as is found in Appendix C. Table 6 presents the sample design with this subsampling rate.

Table 6. Proposed mailable stratification sample design for HINTS 2005.

	Mailable	Mailable percent of total	Non-mailable	Non-mailable percent of total	Total
Original numbers	17,141	43.0%	22,721	57.0%	39,862
Subsampling rate	100%		68.7%		
Sampled telephone numbers	17,141	52.3%	15,618	47.7%	32,759
Residency rate	79%		12%		47.0%
Residentials	13,524	87.8%	1,874	12.2%	15,398
Screener response rate	67.9%		56.4%		66.5%
Extended interview response rate	68.8%		64.8%		68.4%
Overall response rate	46.7%		36.5%		45.5%
Yield of extended interviews	6,318	90.2%	685	9.8%	7,003
Subsampling rate	100%		68.7%		
Weighted extended interviews	6,318		996		7,314
Effective sample size					6,887

*Note that all figures in the table are rounded, leading to arithmetic inconsistencies (a*b equals c, but rounded a * rounded b is not equal to rounded c).

The residency rates and mailable percentages are from HINTS I. The differential screener response rates (67.9% for the mailable stratum and 56.4% for the nonmailable stratum) reflect the difference in response rates (11.5 percentage points) that we saw in HINTS I between the nonmailable stratum and the mailable stratum with a $2 advance letter incentive. Both rates are higher however than HINTS I, reflecting the fact that we are using other incentives as well in HINTS 2005.

The differential extended interview response rates (4 percentage points higher for the mailable stratum) also reflect the difference we saw in extended interview response rates in HINTS I between the mailable stratum with the $2 advance letter incentive, and the nonmailable stratum.

Appendix A. Expected Distribution of Telephone Numbers in Wave 1

As discussed in Section 2, we are carrying out a randomized experiment in Wave 1 of HINTS 2005, randomly assigning 1/3 of the telephone numbers to a telephone-interview-only branch, 1/3 of the numbers to an internet-interview-only branch, and 1/3 of the numbers to a telephone-internet respondent choice branch. Figures A-1, A-2, and A-3 show our expectations as to the distribution of respondents and nonrespondents in the various relevant subgroups of interest.

A very important dichotomization is between households which have internet access, and households which do not have internet access. The most recent Current Population Survey Computer Use and Internet Supplement (September 2001) indicates that roughly 60% of U.S. households have internet access. The percentage is probably increasing, and will be larger in 2005, but we will use the 60% figure (this will be conservative with regard to estimating expected internet interviews). For the internet-only group, we will ask all of the internet-access households to do the interview via internet. For the telephone-internet respondent-choice group, we will give them a choice to do the interview by internet or by telephone. For the telephone-only group, we will ask all households to do the interview by telephone. Households without internet access will all do the interviews by telephone (whatever their assigned mode branch).

A second dichotomization within the households with internet access are households which give us access information (e.g., their email address). Without this access information, we will be obliged to ask them to log on to our website themselves to complete the internet interview. We expect 70% of internet-access households to give us internet access information, which will give us the chance to email to them followup prompts.

Figure A-1 below presents our expectations as to the telephone-only branch. The breakout into households with and without internet access (60% in the former group and 40% in the latter group) is not directly relevant to the fielding of the telephone-only branch, but is relevant to making comparisons between the telephone-only branch and the two internet branches. Any mode comparison of telephone interview vs. internet interviews is best restricted to households with internet access, as households without internet access may be a different population, and there can be no internet interviews in these households. We will be asking all of the households (in all branches) whether or not they have internet access, for the purpose of this necessary subgrouping for mode comparisons.

We assume an overall 71% response rate for telephone interviews in the telephone-only branch, expecting a 71% response rate for the non-internet-access households, a 75% response rate for the internet-access households which give access information, and a 61.7% response rate for the internet-access households which do not give access information. A difference in the response rates for the two internet-access household groups is likely, as households willing to give us internet access information are certainly more likely to be cooperative for the telephone interview than the complement group, though we don't know to what degree. The 75% vs 61.7% difference is our assumed difference (note that this averages to 71%: the overall assumed rate, averaging over the assumed 70% in the access-providing group and the 30% in the complement group).

Figure A-1

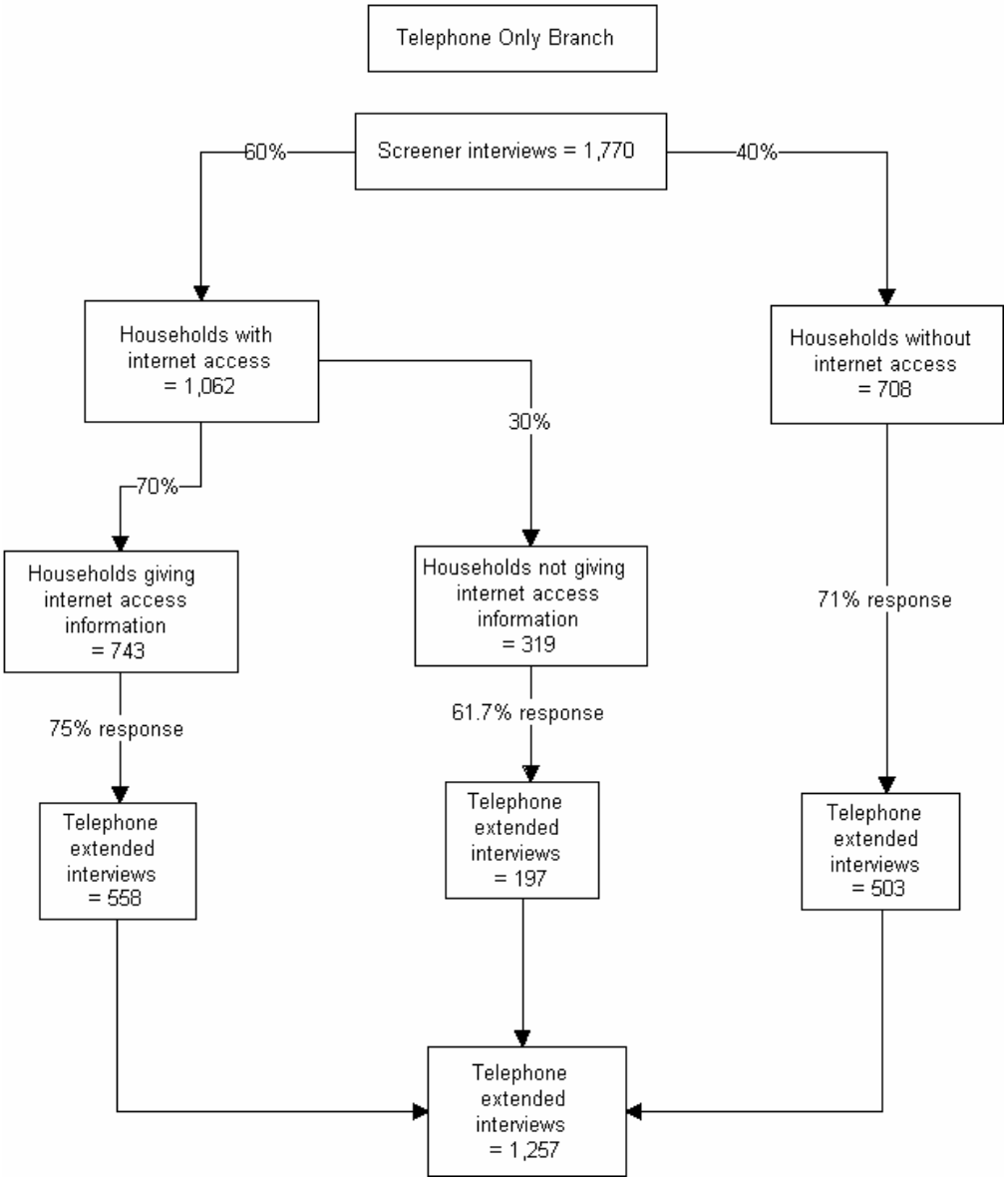

Telephone Only Branch
Screener interviews = 1,770
60%
40%
Households with internet access = 1,062
Households without internet access = 708
30%
70%
Households giving internet access information = 743
Households not giving internet access information = 319
71% response
75% response
61.7% response
Telephone extended interviews = 558
Telephone extended interviews = 197
Telephone extended interviews = 503
Telephone extended interviews = 1,257

Figure A-2 provides our expected breakdown for the internet-only branch. The households without internet access have an identical result as for the telephone-only branch (they are all interviewed by telephone, with an expected 71% response rate). The households with internet access split into those giving internet access information (assumed 70%) and those not giving internet access information (30%). For those households not giving internet access information, we will request that they log in themselves to our website and complete the interview. We expect only a small proportion (15%) to do this. Of those who do give us access information, we expect a 35% response rate for a final internet interview.

We will also follow up any of those who do not complete an internet interview by attempting to complete a telephone interview. We expect to pick up 22% of the access-providing group by follow-up telephone interview, and 35% of the non-access-providing group (for a total response rate internet or telephone of 57% for the access-providing households and 50% for the non-access-providing households).

Figure A-2

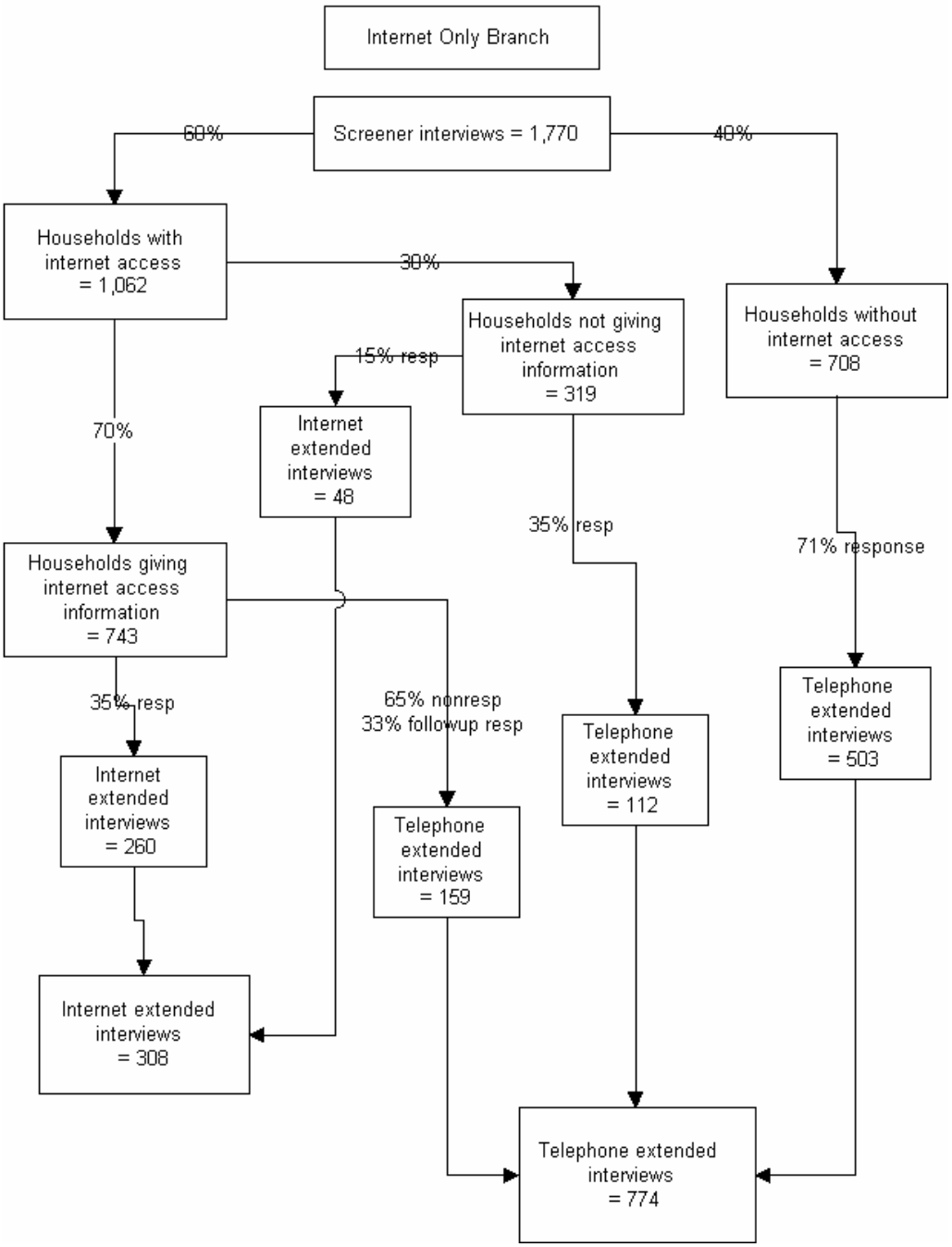
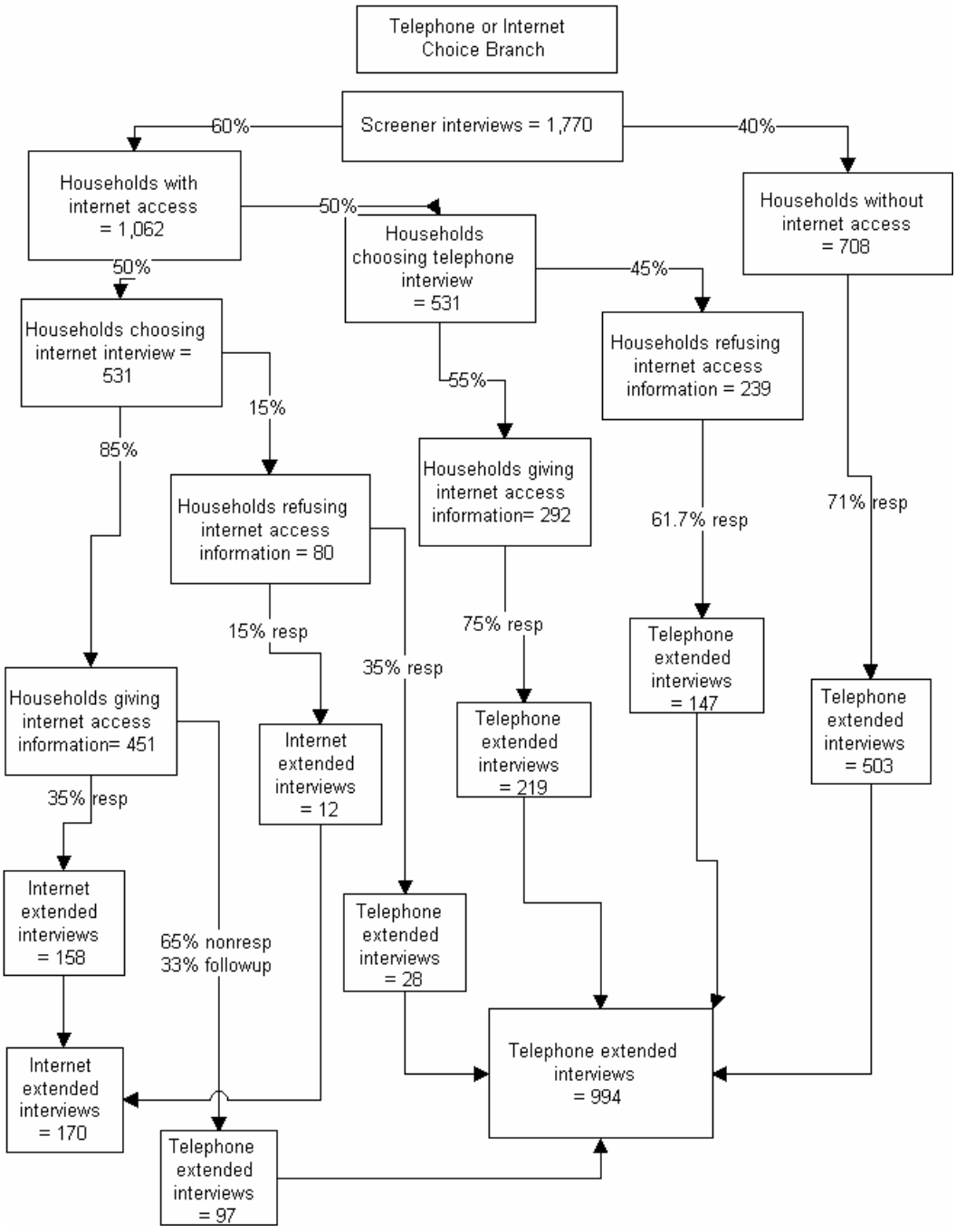

Internet Only Branch
Screener interviews = 1,770
60%
40%
Households with internet access = 1,062
30%
Households not giving internet access information = 319
Households without internet access = 708
15% resp
70%
Internet extended interviews = 48
35% resp
71% response
Households giving internet access information = 743
35% resp
65% nonresp
33% followup resp
Telephone extended interviews = 503
Telephone extended interviews = 112
Internet extended interviews = 260
Telephone extended interviews = 159
Internet extended interviews = 308
Telephone extended interviews = 774

Figure A-3 below provides the expected breakdown for the telephone-internet respondent choice branch. The households without internet access are identical to those of the telephone-only or internet-only branches: they are recruited to do telephone interviews with the same expected recruitment rates. The households with internet access are asked whether or not they wish to do the interview by telephone or by internet. We expect 50% to choose internet, and 50% telephone (this expectation is entirely speculation given our lack of experience here).

The next dichotomization is by whether the households provides internet access information (the 'access-provider' group and the 'non access-provider' group). We expect 70% of internet access households to be access providers (as assumed in Figures A-1 and A-2). In this case, though, we expect a much larger percentage of those who choose the internet as their interview mode to also be access providers (85%). We expect only 55% of the telephone preference group to be access providers. The remaining branches and percentages in Figure A-3 are self-explanatory, mirroring similar calculations for the telephone-only and internet-only groups.

Figure A-3


Telephone or Internet Choice Branch
Screener interviews = 1,770
60%
40%
Households with internet access = 1,062
Households without internet access = 708
50%
Households choosing telephone interview = 531
50%
45%
Households choosing internet interview = 531
Households refusing internet access information = 239
55%
15%
85%
Households giving internet access information= 292
Households refusing internet access information = 80
61.7% resp
71% resp
15% resp
75% resp
35% resp
Telephone extended interviews = 147
Telephone extended interviews = 503
Households giving internet access information= 451
Internet extended interviews = 12
Telephone extended interviews = 219
35% resp
Internet extended interviews = 158
65% nonresp
33% followup
Telephone extended interviews = 28
Telephone extended interviews = 994
Internet extended interviews = 170
Telephone extended interviews = 97

Appendix B. Power Calculations for Experimental Comparisons

One of the hypotheses we will be studying in the Wave 1 experiment is whether the $10 ($20) promised incentive makes a difference by increasing the extended interview response rate. Note that the $10 ($20) promised incentive should have no effect on screener response rate, as it will be offered after screener cooperation has occurred. We will check the null hypothesis that the response rate is the same for no incentive as for a $10 incentive, and as for a $20 incentive (i.e., that the three response rates are the same). The p-value computed will be based on a one-sided test: the alternative hypothesis is that the $10 incentive group response rate is strictly higher than the no-incentive group, and the $20 incentive group response rate is strictly higher than the $10 incentive group response rate.

We will do this test for the no-incentive, $10 incentive, and $20 incentive groups combined over the three mode branches (telephone only, internet only, respondent choice). This test is comparing mean response rates over the mode branches (averaging together telephone interviews and internet interviews from the various subgroups). We will also do a test of the no-incentive, $10 incentive, and $20 incentive groups within the individual modes, checking for interaction between the incentive and the mode.

Table B-1 below presents power calculations for the null hypothesis that the response rate of the $10 promised incentive group $p_1(0)$ and the no-incentive group $p_0(0)$ are both equal to 66%. These null-hypothesis response rates are averaged over the mode branches, so that the base sample sizes are $n_0=n_1=1{,}770$: the overall expected completed screeners in the no-incentive groups and $10 promised incentive groups (see Table 5). The critical region is a one-sided critical region for the difference of the two response rates $\hat{d} = \hat{p}_1 - \hat{p}_0$:

$$V_{NULL}(\hat{d}) = 1.3 * \left(\frac{p_0(0)(1-p_0(0))}{n_0} + \frac{p_1(0)(1-p_1(0))}{n_1} \right)$$

The quantity in parentheses is the variance of the difference of two independent binomial sample percentages. The 1.3 factor is our assumed design effect. Note that the actual variance estimates that will be generated for the Wave 1 experiment will use the weighted response rates and the HINTS replicate weights: our calculations here are an anticipation of the replicate sample variances we will see for this experiment (i.e., we are expecting in effect that the replicate variance of the difference in weighted response rates will be 1.3 times as much as the variance we would compute assuming independent binomial sample percentages, with no weights).

The critical region computed for the tables is the normal approximation one-sided critical region:

$$CR = \left\{ \hat{d} : \hat{d} > z_{.05} * \sqrt{V_{null}(\hat{d})} \right\}$$

The power is computed as the probability that $\hat{d}$ is in the critical region under the alternative (that $p_1>p_0$). Table B-1 gives the smallest value of p_1 that has 80% power with these sample sizes and design effect. Note that under the null we assume that all groups are equal, and equal to 65.95% (our

average over the three branches: see Table 1). The minimum response rate for the $10 incentive level group which is detectably different at 80% power from 65.95% is 70.45%. The minimum response rate for the $20 incentive level group which is detectably different at 80% power from the $10 level of 70.45% is 74.89%.

Table B-1. Null-alternative pairs of overall response rates with 80% power for testing differences in response rates using the one-sided critical region with significance level 5%.

	Under Alternative			Under Null		
Incentive level	Response rate (%)	Overall yield	Screener completes	Response rate (%)	Overall yield	Screener completes
$0	65.95	1,167	1,770	65.95	1,167	1,770
$10	70.44	1,247	1,770	65.95	1,167	1,770
$20	74.89	1,326	1,770	65.95	1,167	1,770
Total	70.43	3,740	5,310	65.95	3,503	5,310

Table B-1 tells us that we expect to be able to detect roughly a 4.5% point difference in response rates between the no-incentive and $10 promised incentive group with 80% power, with a similar power level for determining a difference in response rates between the $10 promised incentive group and the $20 promised incentive group. Note that this puts all of the mode branches together. The power for determining interactions between incentive effects and mode effects will be smaller.

Power Calculations for Response Rate Differences Between Telephone and Internet Interviews

We are interested in testing the hypothesis that there are response rate differences between those who are interviewed by telephone, and those who are interviewed by internet. We suspect from our limited experience with internet interviews that the response rate by internet will be lower than that by telephone. We wish to research this question in HINTS 2005 by comparing the response rates of the three experimental mode branches: telephone-only, internet-only, and respondent choice.

For this comparison to be as powerful as possible, it is better to set aside the non internet access group (households with no access to the internet), as they will all be interviewed by telephone. Our comparisons, then, will be between response rates for the internet-access groups within the three experimental groups. Table B-2 presents power calculations with expected sample sizes for the internet-access groups in each experimental mode branch (see Appendix A for our expected breakdowns within each experimental mode branch). In this case, we present power calculations using a two-sided test of the null hypothesis of no difference, rather than a one-sided test as in Table B-1 (as we don't have a completely defined direction in which a difference should occur as we do when testing incentives). The alternatives are given in terms however of the internet group having a lower response rate (as we suspect that that may be the favored direction).

Table B-2. Null-alternative pairs with 80% power for testing difference in overall interview response rates of telephone-only group and of internet-only group among households with internet access, using the two-sided critical region with significance level 5%.

Households with internet access from Telephone only group	Households with internet access from Internet only group	Design effect	Telephone interview mode and internet interview mode null percent	Internet interview mode alternative percent with 80% power
1,062	1,062	1.3	71.0%	64.6%
1,062	1,062	1.3	65.0%	58.3%
1,062	1,062	1.3	60.0%	53.2%
1,062	1,062	1.3	55.0%	48.1%
1,062	1,062	1.3	50.0%	43.1%

Our experimental plan will allow us to detect roughly a 7% point difference in the response rates between the telephone-only-branch-internet-access households and the internet-only-branch-internet-access households. We may in fact see a considerably larger difference, so we should have sufficient power to see it with these sample sizes.

Power Calculations for Mode Effect Differences Between Telephone and Internet Interviews

Another primary concern we aim to address in our experimental procedures for Wave 1 is the issue of mode differences between telephone interviews and internet interviews. This will occur if persons, all else being equal, answer certain questions differently on the telephone and on the internet, for whatever reason.

Suppose p_0 and p_1 are the percentages (among the full population who has internet access) who answer a particular HINTS question in a particular way via the telephone interview and via an internet interview respectively. The null hypothesis is that these two percentages ($p_0(0)$ and $p_1(0)$, where the argument *0* indicates under the null) are equal. Tables B-3 through B-5 below present power calculations for the null hypothesis for the full range of possible relevant values (10% through 50%). The alternative in this case is a two-sided alternative $p_0(1) \neq p_1(1)$. The critical region is a two-sided critical region for the difference of the two response rates $\hat{d} = \hat{p}_1 - \hat{p}_0$, with variance

$$V_{NULL}(\hat{d}) = 1.3 * \left(\frac{p_0(0)(1-p_0(0))}{n_0} + \frac{p_1(0)(1-p_1(0))}{n_1} \right)$$

The quantity in parentheses is the variance of the difference of two independent binomial sample percentages. The 1.3 factor is our assumed design effect: the higher degree of variance generated from the sample weights that we use in the weighted response rates.

The critical region computed for the tables is the normal-approximation two-sided critical region:

$$CR = \left\{ \hat{d} : |\hat{d}| > z_{.025} * \sqrt{V_{null}(\hat{d})} \right\}$$

The power is computed as the probability that $\hat{d}$ is in the critical region under the alternative (that $p_1 \neq p_0$). Tables B-3 through B-5 gives the smallest values of p_1 that have 80% power with these sample sizes and design effect, for a range of values of p_0.

Table B-3 below presents power calculations for a comparison of telephone and internet interview aggregated outcomes, using in this case all telephone interviews and all internet interviews (from any branch). For a given 3,025 (=1,257+774+994) overall telephone interviews and 478 overall internet interviews (see Figures A-1 through A-3), 80% power would be obtained with differences ranging from 7.9% (=57.9%-50.0%) to 5.0% (=15.0%-10.0%). This comparison using all telephone and all internet interviews gives the largest sample sizes, but both groups include heterogeneous populations (for example, the internet interview group includes no households without internet access, whereas the telephone interview group does), making it necessary to make assumptions about the subsets having equal percentages (e.g., the internet-access group and the non internet-access group having equal percentages) for the null hypothesis to make any sense.

Table B-3. Null-alternative pairs with 80% power for testing interview mode difference using the two-sided critical region with significance level 5%, with sample sizes 3025, 478, and design effect 1.3.

Overall telephone interviews	Overall internet interviews	Design effect	Telephone interview mode and internet interview mode null percent	Internet interview mode alternative percent with 80% power
3,025	478	1.3	50.0%	57.9%
3,025	478	1.3	40.0%	47.8%
3,025	478	1.3	30.0%	37.4%
3,025	478	1.3	20.0%	26.5%
3,025	478	1.3	10.0%	15.0%

Table B-4 presents a comparison putting together on one side all telephone interviews from the telephone-only group and on the other side all interviews of all types from the internet-only group. The third set that is compared is all interviews from the respondent-choice group. This is cleaner than the Table B-3 analysis in that the two compared groups are based on randomized groups (the telephone-only and internet-only experimental branches), and thus represent the same overall population, but the power of this analysis will likely be less with the telephone interviews being intermixed with internet

interviews in the internet-only group. Table B-4 provides power calculations for an experiment group comparison based on expected sample sizes of 1,257 and 1,082 (=308+774) extended interviews from the telephone-only and internet-only groups.. Since the choice experimental group is expected to yield a total of 1,164 interviews similar in size to the others, any associated comparison (telephone-only vs. choice; internet-only vs choice) should also be similar.

Table B-4. Null-alternative pairs with 80% power for testing mode experimental group difference using the two-sided critical region with significance level 5%, with sample sizes 1257, 1082, and design effect 1.3.

Overall interviews from telephone only group	Overall interviews from Internet only group	Design effect	Telephone only and internet only null percent	Internet only alternative percent with 80% power
1,257	1,082	1.3	50.0%	56.7%
1,257	1,082	1.3	40.0%	45.8%
1,257	1,082	1.3	30.0%	35.5%
1,257	1,082	1.3	20.0%	24.8%
1,257	1,082	1.3	10.0%	13.6%

The last type of comparison we work with compares telephone interviews with internet interviews, but within similar populations: households having internet access. The first group are telephone interviews from the telephone-only branch from households in this set. The second group are internet interviews from the internet-only branch from households in this set. Table B-5 gives power calculations for a mode-experiment-specific comparison restricting to these two household sets. The expected sample sizes (see Figures A-1 and A-2) are 754 and 308 respectively. Two groups were based on 754 extended interviews by telephone mode within telephone only group and 308 extended interviews by internet mode within internet only group. Such smaller base yields necessitate larger differences to attain 80% power, ranging from 10.8% (=60.8%-50.0%) to 6.9% (=16.9%-10.0%).

Table B-5. Null-alternative pairs with 80% power for testing mode-experiment specific difference using the two-sided critical region with significance level 5%, with sample sizes 754, 308, and design effect 1.3.

Telephone interviews from telephone only group HHs with internet access	Internet interviews from Internet only HHs with internet access	Design effect	Telephone interview mode and internet interview mode null percent	Internet interview mode alternative percent with 80% power
754	308	1.3	50.0%	60.8%
754	308	1.3	40.0%	50.7%
754	308	1.3	30.0%	40.1%
754	308	1.3	20.0%	28.9%
754	308	1.3	10.0%	16.9%

Appendix C. Subsampling of the Non-mailables

Brick et al (2002) recommended the subsampling of nonmailables to increase the efficiency of the sampling process, as the nonmailables are much more costly per completed extended interview, due to lower response and residency rates. This was done in the last wave of HINTS I: the nonmailable stratum was subsampled at an 80% rate (i.e., 20% of these numbers were deleted from the sample).

The theory underlying this is essentially assigning optimal sampling rates to strata based on stratum sizes, stratum variability, and cost per sample unit within the strata. Defining N_h as the number of sample units in the population in stratum h, n_h as the sample size in stratum h, S_h as the population variance in stratum h, and c_h as the relative cost of completing an interview in stratum h, Cochran (1977) Section 5.5 defines the optimal sampling rate by

$$\frac{n_h}{N_h} \propto \frac{S_h}{\sqrt{c_h}}$$

In this application, we have two strata: h=1 corresponding to the mailable numbers and h=2 corresponding to the nonmailable numbers. We will compute relative rates based on setting the S_h values as equal[19]. Following the general model of Brick et al (2002), we can define the cost within each stratum as $C_h = n_h * c_h$, where n_h is the number of completed extended interviews, and c_h is the non-overhead cost per completed extended interview, breaking this down as follows:

$$C_h = n_h \left\{ c_{ce} + \frac{c_{re} * (1 - r_{eh})}{r_{eh}} + \frac{c_s}{r_{eh}} + \frac{c_{rs} * (1 - r_{sh})}{r_{eh} * r_{sh}} + \frac{c_{nr} * (1 - R_h)}{r_{eh} * r_{sh} * R_h} \right\}$$

c_{ce}	Cost of carrying out extended interview with cooperative respondent;
c_{re}	Cost of finalizing extended interview with extended interview nonrespondent;
r_{eh}	Extended interview response rate in stratum h;
c_s	Cost of completing screener with cooperative household;
c_{rs}	Cost of finalizing screener-nonresponding households (refusal or other nonresponse);
r_{sh}	Screener response rate in stratum h;
c_{nr}	Cost of finalizing nonresidential telephone number;
R_h	Residency rate in stratum h.

Note that dividing by r_{eh} gives completed screeners, dividing by $r_{eh} * r_{sh}$ gives total residential numbers, and dividing by $r_{eh} * r_{sh} * R_h$ gives total sample numbers. The unit costs are assumed to be equal across the strata. In HINTS I, we found that the screener response rate for the nonmailable numbers was 11.5% less than that for the mailable numbers when $2 incentives were included in the original mailings (which we will do in HINTS 2005)[20]. The extended interview response rate for the

[19] This will be the case for any questionnaire items which do not differ in variability between adults in mailable and adults in nonmailable households.

[20] 60.7% response rate for the pre-incentive, no refusal incentive Wave 1 HINTS I group, and 49.2% for the Wave 1 non-mailable group.

nonmailable numbers was also 4% less under the same conditions[21]. Our overall projected rates are 65% for the screener response rate and 68% for the extended interview response rate. Using the cost formula above with the designated unit costs and designating target response rates for HINTS 2005 for the mailable and non-mailable subsets (assuming a 4% difference for the extended response rate and a 11.5% difference for the screener response rate, as in HINTS I), Table C-1 presents a computation of the unit costs C_h/n_h for each stratum.

Table C-1. Relative costs for the mailable and nonmailable stratum.

		Mailable	Non-mailable
Cost of completed extended	C_{ce}	3.3	3.3
Cost of extended refusal	C_{re}	2.05	2.05
Cost of completed screener	C_{cs}	1	1
Cost of refused or other screener	C_{rs}	2.65	2.65
Cost of nonresidential	C_{nr}	0.36	0.36
Extended response rate	r_{eh}	68.80%	64.80%
Screener response rate	r_{sh}	67.90%	56.40%
Residency rate	R_h	78.00%	12.00%
Overall response	$r_{eh}*r_{sh}$	46.72%	36.55%
Cost of complete extendeds	C_{ce}	3.30	3.30
Cost of refusing extendeds	$C_{re}*(1-r_{eh})/r_{eh}$	0.93	1.11
Cost of completed screeners	C_s/r_{eh}	1.45	1.54
Cost of refusing screeners	$C_{rs}*(1-r_{sh})/(r_{eh}*r_{sh})$	1.82	3.16
Cost of finalizing nonresidentials	$C_{nr}*(1-R_h)/(r_{eh}*r_{sh}*R_h)$	0.22	7.22
Total relative cost		7.72	16.34
Cost ratio			2.12

Assuming the validity of the unit costs and response and residency rates, the ratio of the costs of completing an extended interview in the nonmailable stratum to that of the mailable stratum is 2.12. This is due to the much higher number of nonresidential numbers, and the higher number of refusals. The ratio of sampling rate of nonmailable to mailable stratum should be $1/\sqrt{2.12}$, which is 68.7%. This is reflected in the Table 4 sample sizes.

[21] 64.0% extended interview response rate for the pre-incentive, no refusal incentive Wave 1 HINTS I group, and 59.8% for the Wave 1 non-mailable group.

References

Brick, J. M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). 'Two-phase list-assisted RDD sampling', in *Journal of Official Statistics, 18,* 203-215.

Cochran, W. G. (1977). *Sampling Techniques, 3rd edition.* New York: John Wiley & Sons.

Kish, L. (1965). *Survey Sampling.* New York: John Wiley & Sons.

Tucker, C., Casady, R., and Lepkowski, J. (1993). '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.