HINTS

HEALTH INFORMATION NATIONAL TRENDS SURVEY

Final Report

October 2003

NATIONAL CANCER INSTITUTE (NCI)

The Privacy Act requires us to tell you that we are authorized to collect this information by Section 411.285a, 42 USC. You do not have to provide the information requested. However, the information you provide will help the National Cancer Institute's ongoing efforts to promote good health and prevent disease. There are no penalties should you choose not to participate in this study.

The information we collect in this study is in accordance with the clearance requirements of the paperwork Reduction Act of 1995. We may not conduct or sponsor, and you are not required to respond to, a collection of information unless it displays a valid control number from the Office of Management and Budget in the Federal Government. We estimate that it will take you between 20 and 30 minutes to answer our questions in this interview. This includes the time it takes to hear the instructions, gather the necessary facts, and complete the interview. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-xxxx)

OMB # 0925 - 0507
Expiration Date: 8/31/03

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

List of Appendices

List of Tables

TABLE OF CONTENTS (continued)

Table Page

List of Figures

Figure

1. INTRODUCTION

The National Cancer Institute (NCI) has the vital mission of facilitating the process by which cancer research is communicated to the public. Traditionally, communications about cancer have used mass media to reach the public. During the 1990's, the so-called "Information Revolution" began breaking down the distinction between previously restricted scientific communication channels and mass communication channels available to the public. By the year 2000, a Harris Poll revealed that of all those who had used the Internet for work or pleasure, a full 70 percent had conducted some type of search on the topics of "disease" or "medical condition." Fifteen percent of those searches were focused directly on cancer.[1]

While a shift in communication channels is unquestionably taking place, surprisingly little is known about the effect of this shift on cancer-related communication. Early theories of public health communication were anchored solidly in the broadcast milieu of traditional mass media. Clearly, new evidence is needed[2] to inform cancer communication programs, and to monitor the progress of ongoing communication efforts.

Through its Health Communication and Informatics Research Branch (HCIRB), NCI has funded the Centers of Excellence in Cancer Communication Research to refine the scientific knowledge base associated with cancer communication. The HCIRB has made a number of investments as part of a calculated strategy to improve the ways in which people become aware of, and adopt, cancer prevention messages. What has been missing, however, is a mechanism for determining how well these messages have served in the aggregate to create a better-educated citizenry, to change behavior, and to reduce the Nation's overall burden from cancer.

To address that need, the NCI funded the Health Information National Trends Survey (HINTS) in June 2001. Together with its funded partner Westat, the NCI and the Westat research team set out to create the first administration of what should be an ongoing, cross-sectional survey of the U.S. civilian, non-institutionalized, adult population. The survey used a random digit dialing (RDD) approach to select a probability sample of telephone numbers in the U.S. The purpose of funding a national

[1] Rice, R. E. (2001). The Internet and health communication: A framework of experiences. In R. Rice & J. E. Katz (Eds.) *The Internet and Health Communication: Experiences and Expectations.* Thousand Oaks, CA: Sage Publications.

[2] See T. R. Eng & D. H. Gustafson (Eds., 1999) *Wired for Health and Well-Being: The Emergence of Interactive Health Communication.* U.S. Department of Health and Human Services, Office of Public Health and Science.

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+), civilian, non-institutionalized population. Just as screening tools at the "micro-environmental" level must be accurate in order to minimize selection bias from Type I and Type II errors, so too should the NCI's hallmark communication survey be conducted with exacting performance in order to minimize errors in coverage, sampling, and measurement (Dillman, 2000). Moreover, by aligning the content of the survey with emerging theories of media usage (e.g., Viswanath & Finnegan, 1996), risk information processing (e.g., 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 able to use data from the survey to refine the scientific knowledge base.

This report describes in detail the design and implementation of the HINTS. Chapter 2 discusses how the HINTS questionnaire was constructed, including (a) the sources for items, (b) the conceptual framework for the questionnaire, (c) the testing procedures, and (d) the final questionnaire structure. Chapter 3 reviews the sampling plan for the survey. Topics discussed in Chapter 3 include within-household sampling, oversampling minorities, sample weights and variance estimation, and non-response issues. In Chapter 4, data collection procedures and operations are detailed. Chapter 5 outlines the operational findings from the HINTS and also outlines the changes Westat recommends for HINTS II.

2. DEVELOPMENT OF THE HINTS QUESTIONNAIRE

There were two primary goals in the development of the HINTS questionnaire. First, this survey was intended to be used to provide an assessment of how the general population currently accesses and utilizes current communication channels to obtain health information. At the same time, it was intended to be a vehicle to collect baseline data on cancer knowledge, attitudes, and behaviors. Future iterations of the survey should (a) enable researchers to track the success of national intervention programs designed to improve access to information, and, at the same time, (b) enable researchers to track changes in cognitive and behavioral outcomes. In order to achieve these goals, respondents were asked questions about their health, health-related risk behaviors, medical conditions, and treatments.

This chapter explains the process used to design the data collection instrument for the HINTS. It includes descriptions of other instruments that were reviewed, the steps taken to select and revise items for the HINTS, and an outline of the resulting content of the questionnaire.

2.1 Review of Items for Inclusion

The context through which cancer communication occurs has changed dramatically with the availability of new communication channels and technologies. New evidence must be gathered to develop a new generation of cancer communication programs. Some extant Federal surveys ask questions about cancer-related behavior but do not emphasize cancer knowledge or communication. Typically, the communication items that have been included in these surveys do not account for recent technological advances in communication channels. Private surveys targeting the use of new media are often focused on specific technologies and do not assess individuals' use of health information across communication channels. With the HINTS, surveys in the areas of cancer and health communication were reviewed to develop a new questionnaire designed to understand how individuals use the new array of communication options to prevent cancer, support treatment, or preserve quality of life.

The major topic areas covered by the HINTS include (a) health communication, including a heavy emphasis on the Internet and other new media, (b) cancer knowledge, (c) cancer-related health behaviors, (d) history of cancer, and (e) demographic information describing the respondent. Before developing the HINTS instrument, the research team canvassed major data collection efforts to assess the

degree to which other surveys collect and report data relevant to these areas. The following is a brief summary of the major sources reviewed.

Communication Media

- *Pew Charitable Trust.* Various instruments designed by the Pew Charitable Trust were reviewed in designing the health communication questions for the HINTS. Many of these instruments can be found at http://www.pewinternet.org/index.asp.

- *Other Sources.* A number of data collection efforts on Internet usage were reviewed, including those administered by (a) the Departments of Commerce and Education; (b) universities such as the University of California at Los Angeles, Georgia Tech, Rutgers, and Carnegie Mellon; and (c) private companies such as Harris Opinion Polling. Several members of the project were well versed in surveys regarding health communication, and at least one member had conducted survey work in the area of electronic communication since 1986. Finally, Elaine Arkin, a consultant to the project, participated in extensive discussions of potentially comparable survey collection efforts.

Cancer Behavior, Knowledge, Attitudes, and Beliefs

- *Behavioral Risk Factor Surveillance System (BRFSS).* This survey is sponsored by the Centers for Disease Control and Prevention (CDC) and was designed to monitor, through telephone interviews, health-promoting and disease-preventing behaviors in the general U.S. population. The BRFSS covers topics such as preventive health practices (e.g., exercise, healthy diet, cancer screenings, condom use), risk behaviors (e.g., tobacco use, alcohol abuse, drug abuse), access to health care, general health status, and demographic information.

- *National Health Interview Survey (NHIS).* The NHIS is a cross-sectional survey conducted annually by the National Center for Health Statistics. It is a probability sample representing the adult, civilian, non-institutionalized population of the U.S. Items from the NHIS core pertaining to cigarette smoking, alcohol intake, and leisure-time physical activity were reviewed. In addition, the 2000 NHIS cancer control module was reviewed. It covers topics such as diet and nutrition, physical activity, tobacco, cancer screening, genetic testing, and family history.

- *Current Population Survey (CPS).* The CPS is a monthly survey of approximately 50,000 households that is conducted by the Bureau of the Census for the Bureau of Labor Statistics. The sample is scientifically selected to represent the adult, civilian, non-institutionalized population of the U.S. Items on tobacco use contained in the CPS were reviewed for inclusion in this survey.

- *Five-A-Day Survey.* The National Cancer Institute conducted the National 5-A-Day for Better Health Followup Survey to measure five-year trends in fruit and vegetable intakes, as well as trends in knowledge, attitudes, and beliefs about diet and nutrition.

The study findings will allow NCI to assess the effectiveness of the National 5-A-Day for Better Health Program and other intervention efforts through a telephone survey of approximately 2,600 adults. Items from this survey on fruit and vegetable intake were reviewed.

- *Medical Expenditure Panel Survey (MEPS).* MEPS is the third (and most recent) in a series of national probability surveys conducted by the Agency for Healthcare Research and Quality on the financing and utilization of medical care in the U.S. Items on contact with healthcare providers were reviewed for inclusion in the HINTS.

- *Consumer Assessments of Health Plans (CAHPS).* The CAHPS project is a multi-year initiative funded by the Agency for Healthcare Research and Quality. CAHPS includes multiple survey instruments designed to assess the experiences of respondents with a range of health care products and services. Items on contact with healthcare providers were reviewed for inclusion in the HINTS.

- *Other Sources.* Other cancer resources such as the American Cancer Society (ACS) were contacted to assess comparability of data collection efforts on cancer. In addition, advice was sought from content consultants such as Neil Weinstein, Ph.D., of Rutgers University. Dr. Weinstein contributed significantly to the development of many of the questionnaire items related to individuals' knowledge, attitudes, and behaviors regarding cancer prevention. The development of these items was based on his "Precaution-Adoption" model, a theoretical framework addressing health behavior change.

Results of the source review indicated that no existing survey adequately covered the topic areas central to the HINTS. Items from the existing Internet surveys (e.g., UCLA, Pew Charitable Trusts, Georgia Tech, and Harris Poll) covered topics related to general Internet usage, but did not relate on-line communication directly to relevant issues regarding cancer or cancer communication. Similarly, items in the health surveys (e.g., NHIS-Cancer Supplement and BRFSS) obtained data about respondents' behaviors and contained a limited number of knowledge and attitude questions, but did not connect specific knowledge about cancer to health communication variables. When possible, items from these existing surveys that were relevant to the HINTS were adopted to allow comparability to existing data sources. However, new items were developed as needed.

2.2 Item Selection

After reviewing extant surveys as sources for items on health communication and cancer knowledge, attitudes, and behaviors, the item pool was narrowed using several techniques. Initial item selection was guided by a conceptual framework for the project and by the criteria for item inclusion developed by NCI. Next, the items were refined through cognitive interviewing. Remaining items were

then reviewed and eliminated from the HINTS questionnaire in order to reduce the length of the interview. Finally, a dress rehearsal was conducted prior to full implementation to examine the administration of the items in a field setting. The questionnaire was finalized based on the results of the dress rehearsal. Each of these steps in the process is discussed in further detail below.

2.2.1 Conceptual Framework and Inclusion Criteria

NCI and its contractor, Westat, began development of the HINTS questionnaire by constructing a conceptual framework through which the primary research questions could be answered. The framework, as developed by the team, is depicted in Figure 2-1. It draws heavily on the extensive body of social science research that exists in the areas of health communication, and then adds to it the implications of new research in the areas of medical informatics, human factors psychology, and social ecology. Updated theories from the communication literature suggest that a continuum of usage patterns exists, which ranges from mere exposure at one end to highly interactive information-seeking at the other. Communication media differ in their support of information needs along the continuum. Traditional broadcast media (television, radio, newspapers, and magazines) work best as vehicles for broad exposure. The so-called "new media" (Web sites, e-mail, and Personal Digital Assistants) support the more interactive, information-seeking behaviors at the other end of the continuum. The HINTS was designed to take into account the full range of information-consumptive behaviors along the continuum and also to relate those behaviors to the full range of media options available to the modern health information consumer. Previous communication surveys have concentrated primarily on exposure variables at one end that are found with traditional broadcast media, only. The HINTS is the first survey to provide in-depth data on the specific ways in which health information consumers utilize both traditional and new media to meet cancer information needs along the full breadth of the continuum.

The theoretical framework underlying the HINTS also incorporates data points suggested by modern "stage" theories of health behavior change. The most critical of these is the "Precaution-Adoption" model proposed by Dr. Neil Weinstein of Rutgers University. Dr. Weinstein served as a consultant on the project and was responsible for drafting many of the questionnaire items related to individuals' knowledge, attitudes, and behaviors regarding cancer prevention. These questions about cancer prevention use Dr. Weinstein's theoretical model to explore and substantiate the structural determinants of specific cancer prevention behaviors.

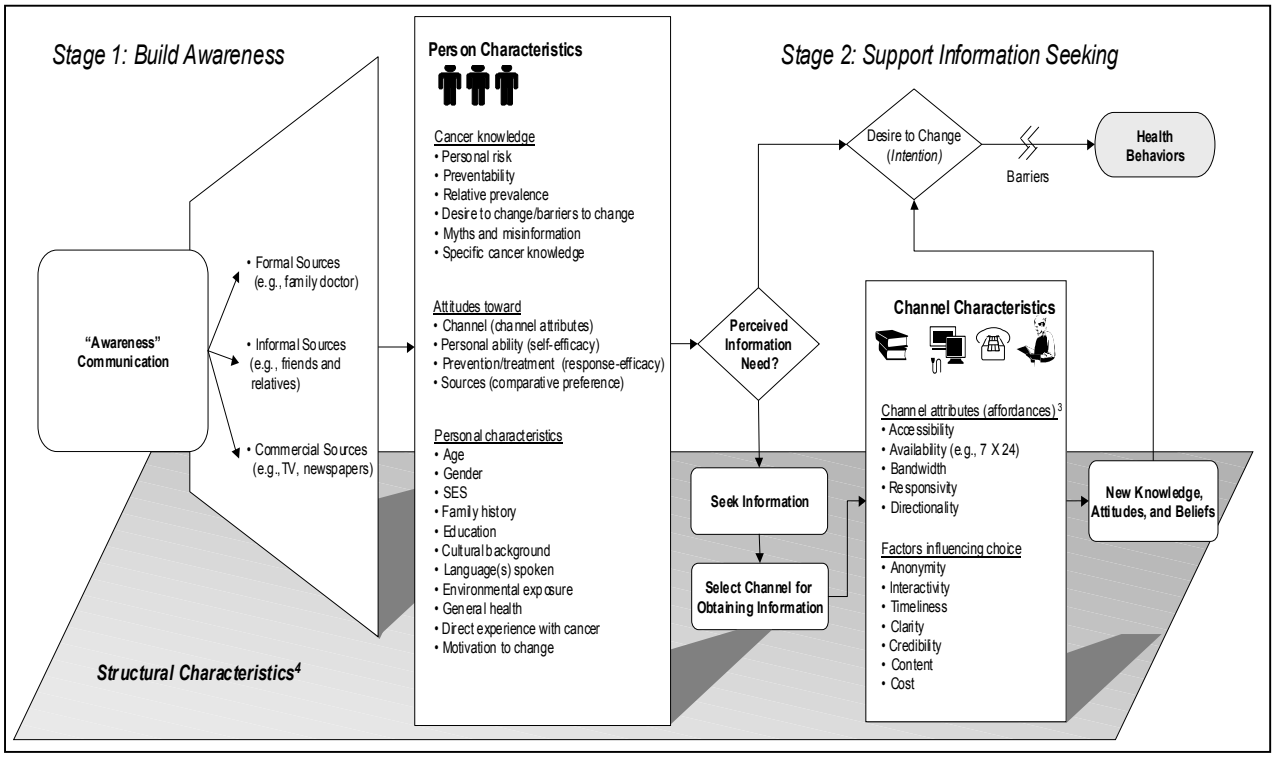

Figure 2-1. HINTS Framework. Derived from extant theory in cancer communication and the behavioral sciences, the HINTS framework served as a guide to item development.

Survey designers used the above-referenced theoretical framework to develop the structure and content of the HINTS. When possible, designers used or adapted questions from other Federal surveys to measure parallel HINTS constructs. Using such questions has the dual advantage of building on tested measurement strategies while providing a link to other Federal data. When constructs were new, designers drafted original questions following best practice in survey design, and then tested a preliminary version of those questions using cognitive interviewing within Westat's cognitive laboratory.

In addition to the conceptual model, questionnaire development was informed by specific guidelines developed by NCI in the fall of 2000. There were three primary criteria that guided item inclusion. These criteria included scientific validity, data utility, and implementation. The criterion of scientific validity involved the following (1) the questions were well-established for assessing cancer-related information, (2) the questions could be self-reported accurately by the adult population, and (3) the sample size was adequate to produce reliable estimates in analyses. Data utility encompassed priorities such as selecting and retaining items that would inform NCI's research agenda and program efforts in health communication and health promotion. Other considerations regarding this criterion were (1) the ability to monitor Healthy People 2010 goals and trends in prevalence estimates over time, and (2) the needs of the people within NCI or other agencies who had specific plans for analyzing and

disseminating information based on the data. The final criterion, implementation, referred to the administration of the survey instrument. Examples of this criterion included (1) an item was able to be administered over the telephone, (2) there was an equitable distribution of questions among topics, and (3) respondent burden was reduced as much as possible.

2.2.2 Cognitive Testing

Questionnaire pretesting consisted of cognitive interview sessions with nine English-speaking respondents. The goal of these interviews was to get enough information about respondents' comprehension, including the preparation of their responses, in order to assess whether they understood questions and responses as the researchers intended. These one-on-one sessions provided valuable insight into how individuals comprehend a question and how they generate their response. The formal report outlined the summarized findings of the interviews (largely by focusing on response problems observed) and presented recommendations for changes. These findings and recommendations brought about a number of revisions to the HINTS questionnaire. A second round of cognitive interviewing, with nine additional English-speaking individuals, was completed with a different set of potential questions to further refine and revise the questionnaire. As with the first round, the report about the second round of cognitive interviews summarized findings and recommended additional changes.

2.2.3 Timings Data

While cognitive interviewing provides valuable information about the interpretation of questions and responses, it does not provide information about the actual length of the interview because the interviewer frequently probes the respondent for additional information. Therefore, several interviews were given with Westat staff in order to obtain an estimate of the interview length. The target length of the interview was 30 minutes. Based on timings using the questionnaire that was revised from cognitive interviewing, the average length of the interview was almost 50 minutes. Therefore, an additional 20 minutes' worth of questions had to be eliminated from the interview. Several techniques were used to identify items for deletion. First, content experts on the HINTS team reviewed their respective sections of the interview for non-essential items that could be eliminated. Second, it was decided that some items did not have to be asked of every respondent (e.g., general cancer knowledge, facts, and myths). It was also decided that these items would be randomly assigned so that half of them would be given to some

respondents and the other half given to the remainder. Finally, some sections of the interview were identified as being lower priority for the initial administration of the HINTS (e.g., skin cancer). These lower-priority sections will be considered for future waves of data collection.

2.2.4 Dress Rehearsal

A full-scale telephone field test or "dress rehearsal" was conducted prior to the main data collection. The interview conditions for the dress rehearsal simulated the actual survey as closely as possible. A total of 172 respondents was randomly screened and interviewed (165 English-speakers and 7 Spanish-speakers) over an 11-day period. This dress rehearsal provided an important check on computer-assisted telephone interviewing (CATI) programming and offered insight into further training issues. The dress rehearsal also provided an initial set of data to examine for variability. In light of the dress rehearsal, revisions were made to both the programming and training program. In addition, the dress rehearsal highlighted the need to shorten the instrument further. Following the dress rehearsal, Westat worked closely with NCI to identify final cuts to the instrument without taking out high-priority items. As a result of these discussions, 33 items were eliminated, shortening the average length of the interview by an additional 7 minutes.

2.3 Final Questionnaire Structure

The questionnaire is divided into two primary sections including a household screener and an extended interview. The household screener 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 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. Figure 2-2 presents a schematic representing the structure of the HINTS instrument.

The next section of the HINTS instrument delves into the respondents' individual and family history of cancer, as well as their knowledge of recommendations regarding the prevention and treatment

of cancer. It begins with a general module aimed at (a) assessing the SP's overall sense of risk for cancer and (b) gauging the degree to which the SP is currently engaged in, or plans to engage in, cancer-prevention behaviors. Cancer, however, 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. For this reason, the survey then divides respondents into one of several parallel paths for questioning depending on their age and gender. Using this approach, the HINTS collected specialized data on different types of cancers without overburdening individual respondents. The survey converges again at the end with a set of common questions, which were asked of all respondents, on cancer-related behaviors (e.g., diet and exercise), health status, and demographic information.

The final version of the questionnaire can be found in Appendix A. It is important to note that the appendix includes all questions 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, according to the flow diagram illustrated in Figure 2-2.

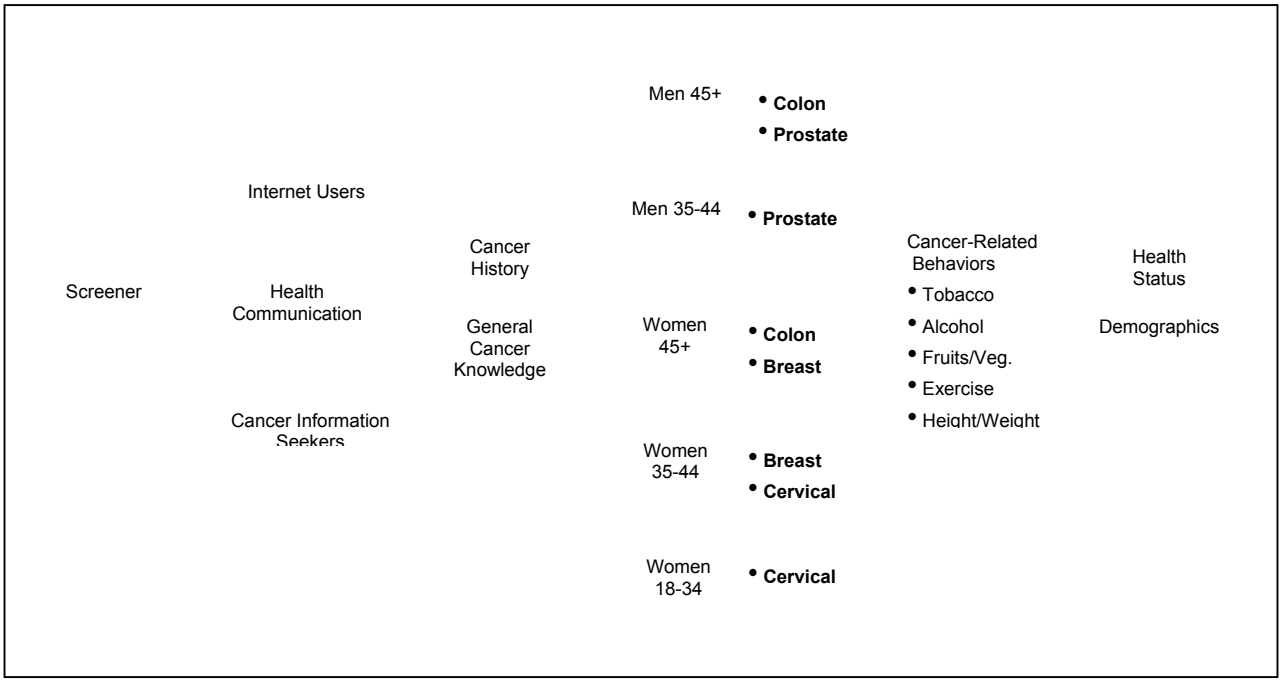

Figure 2-2. Structure of the HINTS instrument. Note that the instrument is designed so that the relevant cancer modules will be administered to selected subsamples of respondents.

3. NCI HINTS SAMPLE DESIGN AND WEIGHTING PLAN

The sample design is a list-assisted RDD sample from all telephone exchanges in the U.S., with oversampling of exchanges with high numbers of blacks and Hispanics. This oversampling resulted in a nationally representative sample of households. 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 (see Tucker, Casady, & Lepkowski 1993). A working bank is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits 294-44) with at least one listed residential number[1].

3.1 Within Household Sampling

Our plan was to sample one adult within each sampled household, with each adult having an equal chance of selection. One approach is the last birthday method (Binson, Canchola, and Catania, 2000). With this approach, we asked the respondent how many adults were in the household, and then asked which adult had had the most recent birthday. That adult became the selected adult.

Our proposed plan was designed to be as unintrusive as possible while still giving each adult an equal chance of selection. The steps were as follows:

- Confirm that the respondent is an adult. Ask the respondent how many adults are in the household. The respondent answers N=1, 2, 3, ...
- The CATI system then accesses a preselected random number RAND1 for the sampled household.
- If RAND1 is less than or equal to 1/N, then the respondent is selected. No further sampling steps are necessary (note that if N is 1, the respondent is automatically selected).
    - If RAND1 is greater than 1/N, and N=2, then the respondent is informed that his/her housemate has been selected. No further sampling steps are necessary.

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

- If RAND1 is greater than 1/N, and N>2, then adult sampling continues. The respondent is asked if he/she knows the birthdays of his/her housemates.

    1. If the respondent says yes, then the respondent is asked to identify the housemate with the most recent birthday (excluding themselves). No further sampling steps are necessary.

    2. If the respondent says no, then the respondent is asked to give the first names or first initials of his/her housemates. The CATI will then sort these names in alphabetical order. A preselected random number RAND2 will be accessed to choose one adult. (For example, if there are two housemates named Jim and Mary, then CATI sorts the names as Jim followed by Mary. If RAND2 is less than or equal to 0.5, then Jim is selected. If RAND2 is greater than 0.5, then Mary is selected.)

This sampling plan was designed to minimize non-response through limiting the number of questions asked of the respondent. If there were one or two adults in the household, then only the first question about number of adults was necessary. If there were N=3 or more adults, then there was a 1/N chance that the respondent would be selected, precluding further questions. Otherwise, the birthday method should have been sufficient in most cases to complete adult sampling. The intrusive question about first names and first initials was only a last resort which was not asked often.

The preselection of respondents allowed us to avoid the possibility of self-selection, which may have been occurring with the standard birthday method.

3.2 Oversampling Minorities

Part of the protocol for the HINTS was a requirement to achieve high precision for the subdomain of blacks and Hispanics. To fulfill this requirement, we oversampled from telephone exchanges that had a higher percentage of blacks and Hispanics. Our subcontractor, Genesys, had estimates of the percentage of minorities for each active telephone exchange in the U.S. We had studied a number of different possible stratifications for past RDD surveys, and the best approach, we believed, was to assign to a high-minority stratum all exchanges with an estimated 15 percent or more blacks and Hispanics, with the complement set of exchanges becoming the low-minority stratum. If we oversampled this high-minority stratum, we could increase the expected number of black and Hispanic households. We

wanted to oversample at a high enough rate to give us maximum percentage standard errors no higher than 0.015 for both black and Hispanic domain estimates.

In computing standard errors for this allocation and others, we needed to account for the design effect produced by the within-household sampling. Using March 1998 Current Population Survey data, we could estimate that 31.9 percent of households had one adult, 53.8 percent of households had two adults, and 14.3 percent of households had three or more adults, with a mean number of 3.9 adults in the three-or-more adult households. We assigned a household weighting factor of 1 for the one-adult households, a household weighting factor of 2 for the two-adult households (to account for sampling of one of two adults), and a larger weighting factor for the remaining households. A rough estimate for the design effect incurred for these differential weighting factors was (1+CV2), where CV was the coefficient of variation in the weighting factors (see Kish 1992). Our estimate from the CPS data was 1.22. In terms of standard errors, this was an increase of 10.5 percent from the simple random sampling result. The effective sample sizes below incorporated this Kish factor of 1.22.

We could achieve 1.6 percent standard errors for blacks and Hispanics with 8,000 extended interviews with the following sample design. We proposed to reach 14,000 households, and expected 80 percent response at the screener level from the recruited households, resulting in 11,200 completed screeners. These households and screeners were allocated to the high minority and low minority in a 66:41 proportion (oversampling the high-minority stratum, as the relative share of the two strata in the population, is 51.2:48.8). This was done by oversampling the high-minority exchanges at a 1.5428 rate (i.e., each high-minority exchange telephone number has a probability of selection 1.5428 times higher than the low-minority exchange telephone numbers).

Table 3-1 presents estimates as to how the expected households and completed screeners were allocated to black, Hispanic, and other domains within the two sampling strata, given the oversampling of the high-minority stratum at the 1.5428 rate.

Our plan was then to subsample other race/ethnicity families in the high-minority stratum at a 64.8 percent rate (the reciprocal of 1.5428). This equated the final sampling rate for other race/ethnicity families within the high-minority exchange stratum to that of other race/ethnicity families within the low-minority exchange stratum, improving efficiency for this group (as we do not need “extra” households within this domain). The expected attempted interviews in Table 3-1 reflect the expected sample sizes at this point. The final column of Table 3-1 shows the expected final interviews, which are the expected attempted interviews multiplied by 0.85, which was the expected extended interview response rate.

Table 3-1. Stratum-domain percentages and sample sizes for proposed sample design[2]

Stratum	Domain	Portion of pop'n	Portion of stratum	Expected attempted screeners	Expected completed screeners	Family subsmpg rate	Expected attempted interviews	Expected final interviews
High minrty	Black	10.47%	20.47%	1,760	1,408	100.0%	1,408	1,152
High minrty	Hispanic	10.72%	20.95%	1,801	1,441	100.0%	1,441	1,179
High minrty	Other	29.98%	58.58%	5,038	4,030	64.8%	2,612	2,137
High minrty	All	51.17%	100.0%	8,600	6,880		5,461	4,469
Low minrty	Black	1.35%	2.77%	150	120	100.0%	120	98
Low minrty	Hispanic	1.64%	3.36%	181	145	100.0%	145	119
Low minrty	Other	45.84%	93.87%	5,069	4,055	100.0%	4,055	3,319
Low minrty	All	48.83%	100.0%	5,400	4,320		4,320	3,536
All	All			14,000	11,200		9,781	8,005

Using the formulas for the variances of a stratified random sample with the Kish factor for the design effect for within-household sampling, we have the following results for this sample design:

Table 3-2. Standard errors and effective sample sizes for black and Hispanic domains

Domain	Expected completed interviews	Standard error	Effective sample size
Black domain	1,250	1.58%	1,002

[2] This table is an updated version of the table given in the January 10, 2002, version of this report, and is based on final percentages of minorities within the two exchange strata as calculated at the time of RDD sampling.

Hispanic domain	1,298	1.55%	1,037
All adults	8,005	0.62%	6,472

The standard errors are for population percentages of 0.5. The effective sample size is the sample size for a simple random sample that would achieve the same precision for population percentages of 0.5.

3.3 RDD Sample Sizes

At the initiation point of the project, based on our sample size needs and our projections of eligibility rates and response rates, we made the following projections. We assumed that roughly 43 percent of the telephone numbers were residential in the working banks. We assumed a screener response rate of 80 percent and an extended interview response rate of 85 percent. Table 3-3 presents our expected breakdown for the telephone sample, with the total telephone sample size required as 32,560.

Table 3-3. Overall RDD telephone sample sizes

Total Telephone Sample	32,560
Total Households in Sample (43 percent residential rate)	14,000
Total Expected Completed Screeners (80 percent response rate)	11,200
Total Attempted Interviews (after subsampling)	9,425
Total Completed Interviews (85 percent response rate)	8,000

This RDD sample size was divided into two waves, with 16,280 numbers in Wave 1 and 16,280 numbers in Wave 2; Wave 1 was fielded. In the course of fielding Wave 1, we found that our residency rates and response rates were much lower than expected. In response to this, we added 50 percent to the Wave 2 sample, and added a further reserve sample of 12,210 numbers as well.

Table 3-4. RDD wave and reserve totals

	Telephone sample size
Wave 1	16,280
Wave 2	24,420
Reserve	12,210

Total	52,910

We also applied a two-phase stratification approach to the reserve sample, based on whether or not the telephone numbers had mailable addresses associated with them or not. This was introduced in Brick, et al (2002). The non-mailable numbers were subselected at an 80 percent rate, i.e., 20 percent of these numbers were discarded from the sample. We weighted the remaining non-mailable numbers at a rate of 1.25 to offset this subsampling. This led to an increase in variance from the differential weighting, but the non-mailable numbers were much less productive, so that the tradeoff led to better variance properties.

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

	Total numbers in original sample	Percent	Total numbers in final sample	Percent	Weighting factor
Mailable	8,730	71.50%	8,730	75.82%	1.00
Non-mailable	3,480	28.50%	2,784	24.18%	1.25
Total reserve	12,210	100.00%	11,514	100.00%	

3.4 Sample Weights and Variance Estimation

Every sampled adult who completed a questionnaire in the HINTS 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 consisted of three major components. The first component was the respondent's base weight. This base weight is the reciprocal of the probability that the respondent had of being sampled. Section 3.6 discusses the computation of base weights. The second part of the sampling

weight was an adjustment for non-response. There were several points at which cooperation needed to be gained: (a) the household needed to be successfully reached and the screener successfully completed, and (b) the sampled respondent within the household needed to be successfully recruited to complete the extended HINTS interview. Both a screener non-response adjustment and an extended interview non-response adjustment were computed. The computation of the screener non-response adjustment was complicated by the fact that many residential households were never reached even after a considerable number of calls, and were never completely confirmed as residential or non-residential. These telephone numbers with unknown residential status can be categorized as NM numbers (for which only an answering machine was reached) and NA numbers (for which no contact of any kind was made). Section 3.7 discusses non-response adjustments in detail.

The third part of the sampling weight was a calibration adjustment. The primary purpose of the calibration adjustment was to reduce the sampling variance of estimators through the utilization of reliable auxiliary information (reliable in the sense of having less sampling and non-sampling error than the corresponding HINTS estimates). For example, the total number of male and female adults in the U.S. was estimable by taking the summation of all (non-response-adjusted) base weights of responding adults in the survey by sex. There are other estimates of these same population totals with less sampling and non-sampling error that can be used to calibrate the HINTS estimates (e.g., if the HINTS population estimates for males deviate from corresponding estimates from the auxiliary information, the weights of male respondents can be altered to bring the HINTS estimates "in line" with the auxiliary information). This process of calibration improved the sampling error of the HINTS estimates which were correlated in the population with characteristics represented in the auxiliary information. The auxiliary information used for the HINTS came from the most recent Current Population Survey (probably March 2001), which had much larger sample sizes than the HINTS. These calibration adjustments are discussed in Section 3.8 below.

3.5 Jackknife Variance Estimation

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

The jackknife technique was compatible with the sample design and weighting procedures for the 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 was usually almost all of the sample, except for a group of respondents which was "deleted" for that replicate. The resulting weights were called replicate weights.

The jackknife procedure is the standard operating procedure for variance estimation at Westat for surveys such as the HINTS. The Westat software package, WesVarPC, can be used to calculate these variance estimators. It can be obtained from the Internet by accessing the WesVarPC site: http://www.westat.com/wesvar/.

We retained on the output data files the necessary implicit and explicit stratification variables necessary to use linearization software packages such as SUDAAN.

The jackknife variance estimator was computed in the following way for the HINTS. A set of $R=50$ replicate weights was assigned to each responding adult. Suppose **P** is a percentage of adults in the U.S. population having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A nationally representative estimator p can be computed by aggregating the adult sampling weights of all responding adults with this characteristic (e.g., all responding adults in the survey answering the survey question in a particular way). A jackknife variance estimator of the sampling variance of p can be computed in two steps:

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.

2. Compute the jackknife variance estimator:

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

The replicate weights were computed by systematically deleting a portion of the original sample, and then 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 the HINTS were telephone households. The remainder of the sample with the deleted portion removed was called the "replicate subset," and it should mirror the full sample design, as if it were a reduced version of the original sample. The HINTS RDD sample was a stratified sample, so each replicate subset

had a sample from each stratum, with that sample reduced by the deleted portion (i.e., each stratum has a piece deleted from it of roughly equal size, guaranteeing that the replicate subset interpenetrates every stratum to an equal degree).

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*. This was done in such a way that *1/50* of each sampling stratum was assigned to each group *D(r)* (i.e., the deletion groups included parts of each sampling stratum). Each replicate sample was 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 paralleled the steps computing the full sample sampling weights. The replicate base weight for each sampled household or adult and each replicate was either equal to *R/(R-1)* times the full sample base weight (if the household was contained in the replicate subset) or equal to 0 (if the household was not contained in the replicate subset, but instead was contained in the "deleted" set for that replicate). See Section 3.6 for further details on computation of the replicate base weights.

Non-response and poststratification adjustments were then computed for each set of replicate base weights, using the replicate base weights in the computation of non-response and poststratification adjustments in place of the original base weights. These calculations generated a set of replicate non-response and poststratification adjustments for each responding adult. The final replicate weights were products of the replicate base weights, non-response adjustments, and poststratification adjustments. Sections 3.6 and 3.7 discuss in detail the computation of the final replicate weights.

3.6 Base Weights

Base weights were assigned to both sampled households and sampled adults within households. The base weight for the respondent was 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);
- 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);

- 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[3].

- An extra factor of 1.25 if the household was a non-mailable number in the reserve sample (to offset the 80% subsampling of these numbers).

The base weight was indicated below as w_i (i indicating the particular sampled adult).

Standard errors were computed for the HINTS estimates through the use of the jackknife technique, as discussed in Section 3.5. A total of 50 replicate base weights was computed for each sample unit[4]. Suppose we write as A the set of all sampled adults in the study. 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:

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

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 3.5 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.

3.7 Non-response Adjustment and Response Rates

Non-response is generally encountered to some degree in every survey. The first and most obvious effect of non-response is to reduce the effective sample size, which increases the sampling

[3] Note that cellular numbers, numbers devoted to businesses run from the household, and numbers dedicated to fax or modem usage were not considered. There were a small number of households with more than two regular, residential telephone numbers, but this number was small. We simplified the questionnaire by asking only about one or more than one, and used that information in the computation of the base weight.

[4] 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). However, a number much greater than, for example, 100, generates sample files that are too large in size (because of the large number of replicate weight fields).

variance. In addition, if there are systematic differences between the respondents and the non-respondents, there also will be a bias of unknown size and direction. This bias is generally adjusted for in the case of unit non-respondents (non-respondents who refuse to answer any part of the questionnaire) with the use of a weighting adjustment term applied to the base weights of sample respondents. Item non-response (non-response to specific questions only) is generally adjusted for through the use of imputation. This section discusses weighting adjustments for unit non-response, and calculations of response rates.

The most widely accepted paradigm for unit non-response weighting adjustment is the quasi-randomization approach (Oh & Scheuren, 1983). In this approach, non-response cells are defined based on those measured characteristics of the sample members that are known to be related to response propensity. For example, if it is known that males respond at a lower rate than females, then gender should be one characteristic used in generating non-response cells.

Under this approach, sample units were assigned to a response cell, based on a set of defined characteristics. The weighting adjustment for the sample unit was the reciprocal of the estimated response rate for the cell. Any set of response cells must be based on characteristics which are known for all sample units, both responding and non-responding. Thus, the questionnaire answers on the survey cannot be used in the development of response cells because these characteristics are known only for the responding sample units.

Under the quasi-randomization paradigm, we modeled non-response as a "sample" from the population of adults in that cell. If this model were in fact valid, then the use of the quasi-randomization weighting adjustment eliminated any non-response bias (Little & Rubin, 1987).

3.7.1 Unit Non-response in the HINTS

There were two types of unit non-response in the HINTS: screener non-response and extended interview non-response. Screener non-response occurred when a household was reached, but the screener interview was not completed. We also included in any screener non-response calculations any households for which we never reached a person, either because we only reached an answering machine (these are called NM numbers), or only got a ring with no answer (these are called NA numbers), with every call made to the telephone number. Since we did not know if an answering machine or ring no answer corresponded to a residential household, the number of lost residential numbers among the NA and NM numbers needed to be estimated (see Section 3.7.3 below).

To adjust for screener non-response, each completed screener received a screener non-response adjustment equal to the reciprocal of the estimated response rate in its screener non-response cell. For a discussion of the screener non-response cells and adjustments, see Sections 3.7.2 and 3.7.3, respectively.

Extended interview non-response occurs when the screener interview is completed successfully, yielding a sampled adult in the household along with identifying information for this adult (and the number of adults in the household), but the sampled adult does not complete the extended interview.

To adjust for interview non-response, each completed extended interview received an interview non-response adjustment equal to the reciprocal of the weighted interview response rate in its interview non-response cell. (Completed extended interviews also received a screener non-response adjustment.) The methodology for selecting extended interview non-response cells and computing extended interview non-response adjustments is discussed in Section 3.7.4. Section 3.7.5 discusses the computation of replicate non-response adjustments.

3.7.2 Non-response Cells for Screener Non-response Adjustments

Non-response cells were generated using cross-classifications based on selected sociodemographic characteristics estimated (by our vendor Genesys) for each telephone exchange, and address status (whether or not an address was available for the telephone number).

The estimated exchange percentages from Genesys were used to assign each exchange to cells, based on the following characteristics:

- Four cells based on geography (Census region): Northeast, South, Midwest, West;
- Three cells (with roughly equal populations)[5] by percent college graduates (exchanges with lowest percentage, next lowest percentage, and highest percentage);
- Three cells (with roughly equal populations) by median income;

[5] The breakpoints will be the 1/3 and 2/3 percentiles over all frame exchanges, which will be calculated when the frame is constructed.

- Three cells by percent blacks and Hispanics (two within the high-minority stratum, and the undivided low-minority stratum).

We judged that these characteristics may be both related to response propensity and correlated to item response to the HINTS questionnaire items, so that these cells would lead to effective non-response adjustments.

The other characteristic for generating cells was whether or not a published address was available for the telephone number. These addresses were used first to mail advance letters about the study and then to mail follow-up letters for households who had not responded. We had found in previous surveys that response propensity may differ by this characteristic (telephone households with known addresses that have received mailed information respond at a higher rate than those without known addresses).

Cross-classifications of these sociodemographic classes and the address status gave a potential total of 216 cells (though some of the cells were empty). We collapsed these cells to attain a minimum cell size of 10 sample units and a maximum cell adjustment of 3.0, using our in-house COLL_ADJ software.

3.7.3 Screener Non-response Adjustments

In general, non-response adjustments within non-response cells are the reciprocals of the weighted response rates within the cell, where the respondents and non-respondents were weighted by their (adjusted) base weights. In this case, the household base weights were unknown for screener non-respondents, since components of the base weight depended on whether the household had one or more residential telephone numbers. For this reason, the non-response adjustment was set equal to the reciprocal of the unweighted screener response rate for each cell.

In principle, the unweighted screener response rate was equal to the total number of cooperating households (eligible or not) divided by the total number of residential numbers in the sample. The latter value was not completely known because of NM and NA numbers. 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 also by computing the overall eligibility rate *EA* among all numbers with known eligibility status (working and non-working).

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

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

where *C(a)* is the 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)* is the number of other residential numbers (e.g., numbers which were found to be residential, but for which a screening interview could not be completed for reasons other than refusals).

We also computed a study screener response rate. Using *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 NA's, and pure NA's respectively, and defining *EM* and *EA* as above, we computed 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:

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

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

The second form of *SCRNR* is both algebraically more complicated and conceptually transparent. The response rate was the completes divided by the completes plus the estimated eligible numbers among the remaining residential numbers (refusals and NA's). We 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* adheres to the guidelines of AAPOR standards regarding valid response rates[6].

[6] Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Available on AAPOR (American Association for Public Opinion Research) website www.aapor.org.

3.7.4 Extended Interview Response Cells

There was more information available about extended interview non-respondents as compared to screener non-respondents. This extra information came from the completed screener (a case was not designated as an extended interview non-respondent 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 only if the screener was complete were we able to compute the base weight w_i (see Section 3.5).

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

1. Sex of sampled adult;
2. Size of household: number of adults in household (1, 2, or more than 2);
3. Census region (4 cells);
4. Telephone number in high, medium, or low-minority exchange (3 cells);
5. Telephone number in high, medium, or low-college-educated exchange (3 cells);
6. Telephone number in high, medium, or low-median-income exchange (3 cells).

The first two characteristics on the list were derived directly from the screener questionnaire. The remaining four characteristics were derived from the telephone exchange, and were identical to those utilized in constructing screener non-response cells (see Section 3.7.2). Non-response cells were constructed by collapsing the initial cells to meet the criteria that the cell sample size should be no smaller than 10 and that the non-response adjustment should be no bigger than 3.0. This was done using Westat's in-house software COLL_ADJ.

Weighted non-response 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 is the screener non-response adjustment for the screener non-response cell a containing household i. The denominator of *ENRA(b)* is an unbiased estimator (adjusted for screener non-response[7]) of the total number of adults in the non-response 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 non-response cell (also adjusted for screener non-response). *ENRA(b)* is an approximately unbiased estimator of the response rate that would be obtained in cell b if the entire U.S. population were contacted for the study.

We computed a weighted extended interview response rate, for reporting purposes. 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:

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

3.7.5 Replicate Non-response Adjustments

Non-response adjustments were themselves random variables, and contributed a variance component to the overall sampling variance. This variance component was represented in the final jackknife estimator by replicating the computation of non-response adjustments (by replacing the original base weights by the replicate base weights, and by repeating the computations described in Sections 3.7.2, 3.7.3, and 3.7.4).

[7] Under full response, the sum of the base weights is an unbiased estimator. With the presence of non-response, there will be non-response bias from any differences between the responding and non-responding households. This non-response bias is reduced in magnitude by the screener non-response adjustments. We can't expect these adjustments to eliminate all bias, so the claim of "unbiasedness" of these totals needs to receive this caveat.

The screener non-response adjustments were the reciprocals of unweighted 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, we recomputed response rates for each replicate set as if it were the original RDD sample of telephone numbers.

RS(a,r) was defined as the count of confirmed residential numbers in screener response cell *a* that 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.) *AMNA(a,r)*, *PNA(a,r)*, *EM(r)*, *EA(r)*, *C(a,r)*, *I(a,r)*, *REF(a,r)*, and *O(a,r)* were defined similarly (see Section 3.7.3). Then we could define a replicate non-response adjustment 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 non-response adjustments was also replicated. The replicate interview non-response adjustment for interview non-response 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 non-response adjustments (for screener non-response and extended interview non-response) were appended to the base weight for the subject (adult):

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

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

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

3.8 Calibration Adjustments

The purpose of calibration is to reduce the sampling variance of estimators through the use of reliable auxiliary information. One recent source for this theory is 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 discussion of these results.

3.8.1 Control Totals from the Current Population Survey

The Current Population Survey (CPS) of the U.S. Bureau of the Census has much larger sample sizes than those of the 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: we utilized the most current estimates available on the Census website.

Any potential calibration variable needs to be on the CPS public use file, and to be well-correlated to important HINTS questionnaire item outcomes (i.e., we want CPS-available characteristics which tend to have differing mean values for the HINTS questionnaire item outcomes). We believe the following CPS characteristics correlated well with the 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
2) High school diploma only
3) High school diploma, some college
4) Bachelor's degree or higher

Our plan was to generate 96 potential poststratification cells based on cross-classifications of these characteristics, and then to tabulate control totals for these cells from the most current CPS data. We can write these control totals as c_g, g=1,…,96. The poststratification 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, i.e., for each poststratification cell $PS(g)$, we computed CA_g equal to:

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

We checked each cell to make sure that the sample size was at least 20 and that the adjustment was not too much larger than the other adjustments. If any cell had too small a sample size or had too large an adjustment, we collapsed it with other cells. Westat's in-house software COLL_ADJ does this adjustment process with input from a statistician, if it is necessary.

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 non-response 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

Data collection for the HINTS was conducted using a Blaise CATI instrument administered by trained Westat telephone interviewers over a period of five and one-half months. The RDD sample phone numbers were address-matched to enable the mailing of an introductory advance letter to potential study households. The sample was released in two waves, followed by two smaller reserve samples.

RDD screening and interviewing were administered by experienced Westat staff from the Telephone Research Centers (TRCs) in Rockville, MD and Chambersburg, PA. Over 40 experienced and project-trained interviewers worked on the project. The data collection was carefully supervised and monitored to ensure data quality and completeness.

Study response rates were negatively impacted by the winter holiday season, when respondents were less likely to either answer the phone or, once screened, to continue with a 30-minute interview. At the conclusion of data collection, the screener response rate was reasonably good, but the response to the extended interview was lower than desired.

4.2 Data Collection Procedures

4.2.1 Implementation of the CATI Instrument

The CATI instrument for the HINTS was developed in Blaise. The decision was made to use Blaise since this interviewing system could be used by both the contractor and any subsequent contractor(s). Also, Blaise is easily ported to any Windows-based hardware platform. Its specialized database and data handling conventions are explicitly designed for the demands of interviewing.

Once the content of sections of the instrument was finalized, programming specifications were written. The Blaise CATI instrument was then programmed and tested based on consultation with NCI. Demonstrated assurances of proper functionality of the CATI instrument were incorporated into the design and implementation schedule.

Westat used its Call Scheduling System to support management of the HINTS CATI operations. The CATI Call Scheduling System managed the flow and assignment of cases to interviewers in two phone centers—one in Rockville, MD and one in Chambersburg, PA. The scheduler analyzes a number of factors when determining what case to assign to an interviewer, including contact attempt (call) history, interviewer skills required, time zone, day of the week, time of the day, and project-specific priorities for types of cases. Similarly, the scheduling system analyzes a number of factors when prioritizing and scheduling a case for subsequent calls. During the course of data collection, the scheduler (a) spread calls out to improve the likelihood of reaching respondents; (b) re-queued cases automatically when encountering a busy signal; (c) scheduled appointments for either exact times or general time-windows; and (d) queued refusals for subsequent processing.

During the course of data collection, the TRCs generated reports for managing caseloads and case progress. These reports were also used for generating statistics on outcome status that were used by TRC supervisors and project management staff in order to monitor production.

Westat produced a Spanish translation of the English version of the instrument. The project employed a number of bilingual interviewers proportional to the expected number of Spanish speakers in the sample, resulting in 336 extended interviews having been completed in Spanish. The Blaise instrument allowed interviewers to change languages with a keystroke.

4.2.2 Field Test of Instrument

A field test of the HINTS instrument was undertaken in August, 2002. The purpose of the field test was to evaluate the instrument with trained and experienced Westat interviewers and randomly selected respondents prior to its general administration.

The field test sought (a) to test the ease of administration, question flow and skip patterns; (b) to provide a further assessment of the cognitive demands of individual questions; (c) to test the Spanish translation of the instrument; and (d) to provide "real world" timings of individual component sections of the questionnaire.

Seven experienced Westat interviewers were provided with a brief but thorough training on the instrument. The field test was completed within a week's time, ending on September 1, 2002. Interviewers completed interviews with 172 respondents.

Many calls during the pretest were monitored by Westat project staff in order to gain additional insight into how the CATI instrument performed and to ensure that questions were properly administered and understood by respondents as had been intended.

Following the pretest, interviewing staff attended a debriefing session with project staff. Interviewers offered their thoughts as to which questions were difficult to administer, and which ones they felt were difficult for the respondents to understand or answer. This session was valuable in that interviewers provided a fresh perspective on the instrument and provided the project with a substantive set of recommendations.

Observations and information gleaned from the pretest were incorporated into the training sessions and materials for the main study.

4.2.3 Advance Mailings

Telephone numbers drawn for the RDD sample were matched to mailing addresses in order to allow the mailing of an introductory letter prior to screening. This matching was done through the commercial firms Telematch and Acxiom. Approximately 42 percent of sample phone numbers were matched to addresses, which was at the low end of the range expected. Household advance letters introducing the study were prepared on NCI letterhead under the name of the NCI project officer, Dr. David Nelson. The envelopes featured a return address with the National Institutes of Health, DHHS, and a Westat mailing address.

Advance letters were prepared for distribution to respondent households during Wave I of data collection during the week of October 14, 2002. The Wave II mailing took place in late November, 2002, and reserve sample letters were mailed in early February, 2003. A total of 22,286 advance letters was mailed to potential respondent households.

4.2.4 The OMB-mandated Incentive Experiment

During September, 2002, the project team finalized methods for implementing the incentive experiment requested by OMB in their approval of the research. (See Appendix B for a more detailed discussion of this experiment).

The experiment sought to further explore the benefits of providing a pre-incentive to potential study respondents and a refusal conversion incentive to those who initially refused to participate. Initially, the amount of each incentive treatment was to be $5, but was reduced to $2 (with the use of a $2 bill), to reduce overall costs of the experiment.

The experiment was embedded in Wave I, using sample telephone numbers with address matches. The sample was randomly assigned to one of four equally-sized groups that received different incentive treatment as follows:

- Group YY: Households were sent the pre-incentive and a refusal conversion incentive;
- Group YN: Households were sent the pre-incentive but no refusal conversion incentive;
- Group NY: Households were not sent the pre-incentive but did receive a refusal conversion incentive;
- Group NN: Households were not sent either incentive.

4.2.5 Implementing the Results of the OMB Experiment

As stated above, a $2 incentive was sent to 3,453 of the Wave I advance letters as part of the incentive experiment. No advance incentives were included in subsequent mailings, with the exception of a few remails.

The results of the incentive experiment generally showed an improvement in refusal conversion rates among those receiving the incentive. Therefore, in an effort to boost the refusal conversion effort, all Wave I, II and reserve sample refusal letters included a $2 incentive. The impact of this effort, however, was limited due to the overall 42 percent address match. Refusals were then contacted again by telephone interviewers specializing in refusal conversion.

4.2.6 Telephone Research Center Hiring and Training

The HINTS data collection was completed in two locations due to workload and availability of interviewing staff. The 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, and a few 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, interactive sessions, and dyad role plays. Specific attention was focused on the RDD screener contact procedures and emphasized gaining cooperation of respondents in the first moments of the phone attempt.

Training sessions took place in Chambersburg, PA on October 26-27, 2002, and in Rockville, MD on November 16–17 of the same year. Attrition training took place January 18-20, 2003, to train additional interviewers new to the study.

Each formal training session was completed in 2 days. However, interviewers were allowed to go "live" on the phones only when, based upon the assessment of their supervisors and project staff, they were fully prepared to start. Some interviewers required additional practice exercises over the course of the first week. The interviewers that started interviewing directly following the formal training were monitored using routine Westat observation and monitoring procedures to ensure that their work was of the highest quality. Any problems or issues that the supervisors observed while monitoring an interviewer's work were discussed and resolved immediately following the phone interview that was monitored.

Instruction of bilingual interviewers in Spanish was completed during the November training session in Rockville.

During the course of the data collection, TRC supervisors and project staff continued to monitor individual interviewers. Routinely, 10 percent of each interviewer's work is observed to ensure the continued quality and accuracy of their work.

4.3 Findings from TRC Operations

4.3.1 Weekly Report Generation

Project and TRC staff monitored study progress on a daily basis through the use of a series of reports generated specifically for the study. These included daily, weekly, and summary reports.

The HINTS Screener Report is a daily and weekly report that shows the status of the RDD sample and the extended interview. This report details the RDD sample status of both the finalized and

interim households, as well as non-households, in the sample. It further details completes at the extended level, and provides a series of calculated response rates for both the screener and extended interviews.

Another tool used extensively in the management of the study is an internal planning spreadsheet that documents the progress of the study with the actual production. This report summarizes progress by data collection week, showing (a) the amount of new work drawn during the week; (b) the planned weekly and cumulative screener completes and ineligibles, along with the actual progress; (c) the plan for extended completes and the actual progress; (d) the interviewer hours per the plan and in actual production; and (e) the hours per case for the extended interview.

4.3.2 Issues of Survey Administration

Due to the work done prior to the main data collection, such as cognitive testing, testing of the CATI instrument during the course of programming, and the completion of the pretest, administration problems during the main data collection were few.

An issue that interviewers and supervisors raised was the content of the introduction to the screener—the first words the interviewers use to try to gain cooperation. In the standard Westat introduction, the interviewer mentions their name and that they are calling from Westat, a research company, on behalf of the National Cancer Institute.

In early January, that introduction was changed to more prominently feature NCI. Anecdotally, interviewers felt they were getting more initial cooperation using the new introduction.

Another administration issue was one that was known at the start of data collection. It involved the slight delay between the screener and the extended interview. This delay was caused by the running of the sampling algorithm to choose the household respondent to the extended interview. As the algorithm ran, the interviewer asked the screener respondent to wait a moment while the "computer" chose the respondent. Many screener respondents found this to be an opportunity to break off the call.

To alleviate this problem, interviewers were provided with the hard copy of the first few screens of the extended interview. Interviewers were trained to recognize those situations in which the screener respondent would be the extended respondent as well. In such cases, the interviewer bridged the time gap caused by the algorithm by asking the next few questions (which were essentially confirming

demographic information provided in the screener). Again, anecdotally, the interviewers felt they were getting fewer break-offs once they employed this strategy.

A related issue during this pause between the screener and extended interview was the introduction to the extended interview. This introduction necessarily mentioned the average time that the extended interview was expected to require. Upon mention of the expected 30-minute interview duration, many respondents used this as an opportunity to either break off or ask for a call back (the latter of which is often a soft refusal). While indicating the length of the interview is necessary, this issue underscores the need to keep telephone interviews to a maximum of 20 minutes in order to avoid adversely impacting response rates.

4.3.3 Weekly Completion Rates

The following table represents the number of completed interviews during the field period of the study.

Table 4-1. Number of completed interviews during the field period of the study

	Screener Completes		Extended Completes	
Week beginning	Actual	Cumulative	Actual	Cumulative
10/28/02	435	435	143	143
11/4/02	488	923	154	297
11/11/02	513	1,436	145	442
11/18/02	581	1,731	252	696
11/25/02	524	2,544	208	904
12/2/02	427	2,971	226	1,130
12/9/02	308	3,279	226	1,356
12/16/02	265	3,544	199	1,555
12/23/02	659	4,203	185	1,740
12/30/02	598	4,801	231	1,971
1/6/03	674	5,474	393	2,364
1/13/03	586	6,060	394	2,758
1/20/03	698	6,757	429	3,187
1/27/03	568	7,325	351	3,538
2/3/02	587	7,912	282	3,820
2/10/03	355	8,267	299	4,119
2/17/03	410	8,677	230	4,349
2/24/03	624	9,299	373	4,722
3/3/03	600	9,895	333	5,054
3/10/03	386	10,281	316	5,320
3/17/03	378	10,659	225	5,595
3/24/03	425	11,082	218	5,813
3/31/03	326	11,408	171	5,984
4/7/03	264	11,672	168	6,152

4.3.4 Administration Times

The mean administration time for the HINTS extended interview (6,034 respondents, not including partial completes) was 31.89 minutes. The range of response times was a minimum of 13 minutes to a maximum of 388 minutes. The standard deviation for the duration of the interview was 12.6.

4.3.5 Average Calls Per Case

Prior to the start of calling, the CATI scheduler was configured with some standard call limits and some 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 had never had any contact with the respondent were placed in each of seven non-contact time slices. These cases received at least one call per time slice before being finalized. In the HINTS, initially it took nine calls to reach a maximum number of attempts. As with most studies, these cases were "rested" and released one or two more times over several weeks for another round of seven calls in an effort to complete the interview. Consequently, some cases ultimately received 20 to 30 call attempts over several months.

Within the scheduler, queue priorities were set. Extended questionnaires had a higher priority than screener questionnaires.

The table below details the number of call attempts for both the screener and the extended interview. The table shows that the sample was worked efficiently and effectively, with non-response cases worked very thoroughly.

The screener table shows that most screeners (73%) were completed in the first five attempts; 90 percent were completed in the first 10 attempts. Further, the table shows that non-response cases were worked intensely, with 45 percent of these cases receiving over 10 call attempts. Likewise, answering machine or no answer cases were worked thoroughly, with 99 percent receiving at least 11 call attempts.

The extended table (Table 4-2) shows that the number of interviews completed in the first five attempts was also quite high at 73 percent (46% in the first attempt). Table 4-3 shows the total extended level of effort: number of call attempts by result.

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

Call Attempts	Completes & Ineligibles		Non-Response		Non-Working/Non-Residential		No Answer / Answering Machine	
	N	Col %	N	Col %	N	Col %	N	Col %
0	0	0.0%	2	0.0%	19,688	**68.0%**	0	0.0%
1-5	8,540	**73.0%**	2,297	**29.3%**	7,481	25.8%	29	0.8%
6-10	1,965	16.8%	1,979	**25.8%**	1,084	3.8%	9	0.2%
11-15	678	5.8%	1,042	13.3%	435	1.5%	3,085	**82.6%**
16-20	288	2.5%	725	9.3%	159	0.6%	227	6.1%
21-25	150	1.3%	852	10.9%	98	0.3%	47	1.3%
26-30	51	0.4%	940	12.0%	24	0.7%	337	9.0%

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

Call Attempts	Completes & Ineligibles		Non-Response	
	N	Col %	N	Col %
1-5	4,702	76.4%	1,276	32.4%
6-10	813	13.2%	831	21.1%
11-15	379	6.2%	556	11.9%
16-20	166	2.7%	441	5.7%
21-25	84	1.4%	559	14.2%
26-30	13	0.2%	370	9.4%

4.4 HINTS I 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 1.2).

Under this paradigm, the nonresponse bias grows linearly with the nonresponse rate, though the constant of proportionality (the difference in means) will differ between characteristics, and is in general unknown. As nonresponse bias is one of the most problematic components of total survey error since its magnitude is very difficult to measure, an important quality measure for a survey is its response rate.

Recently, the American Association for Public Opinion Research has published guidelines for computing response rates, which aims to standardize the computation of response rates across surveys (AAPOR 2000). 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. This memorandum follows this requirement. The method for computing response rates is within the acceptable bounds prescribed by the AAPOR guidelines.

The overall response rate is computed as a product of the screener response rate and the extended interview response rate[1]. Both response rates are 'weighted': they are ratios of aggregations of base weights. The base weights used for the screener response rate are the original probabilities of selection of the numbers (excluding the multiple telephone number adjustment, which cannot be computed for the nonresponding screeners). The base weights for the extended interview response rate are the screener-nonresponse-adjusted household weights multiplied by a factor for adult selection within the household. Using the base weights makes the response rates in effect estimated response rates 'for the population': the response rates one would get if one contacted every household on the frame (rather than a sample of households).

[1] Screener in this section refers to the portion of the questionnaire which identifies a sampled adult. Extended interview refers to the remaining substantive portions of the questionnaire.

4.4.2 Screener Response Rate

In principle (following AAPOR standards), the screener response rate is equal to the sum of the 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, because of answering machine NA numbers (numbers for which only an answering machine has been reached) and pure NA numbers (numbers for which only a ring-no-answer has occurred). 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. Our procedure is to impute the residential rates among the known numbers as the assumed residential rates among the answering machine and pure NA numbers. This is generally considered to be 'conservative[2]' (i.e., the actual residential rates among these numbers is known to be generally smaller than among the known numbers, though it is difficult to measure exactly this rate). For answering machine NA numbers, we know these are working numbers, so we impute the residential rate among working numbers with known residential status. For pure NA numbers, we impute the residential rate among all numbers with known residential status. These residential rates are weighted.

Writing *C, I, REF, O, AMNA*, and *PNA* as the summation of weights of completed screeners, ineligibles, screeners who refused to participate, other residentials, answering machine NA's, and pure NA's respectively, and defining *EM* and *EA* as the percentage of residential numbers among the known working numbers and all numbers respectively, we 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

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

numbers among the remaining residential numbers (refusals and NA's). We estimate 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. In HINTS the eligibility rate is quite high, as all adults are eligible (only households with no adult are not eligible).

Table 4-4 below presents the sum of base weights for each class of telephone number. These sums are estimates of the total numbers in the U.S. 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 (40.09%) and EM (72.41%): the percentage of known residential numbers within the known status set and the known working number status set.

Table 4-4. Weight aggregations and percentages for the full HINTS I telephone sample

Response Class	Sum of weights	Percent of known-status numbers	Percent of known-status working numbers
C	58,600,253	23.93%	43.23%
I	101,672	0.04%	0.08%
REF	32,179,589	13.14%	23.74%
O	7,274,068	2.97%	5.37%
AMNA	3,651,769		
PNA	14,753,088		
NRS	37,399,239	15.27%	27.59%
NWS	109,289,456	44.64%	
Total known resid	98,155,582	40.09%	***72.41%***
Total known business	37,399,239	15.27%	27.59%
Total known working	135,554,822	55.36%	100.00%
Total known nonworking	109,289,456	44.64%	
Total known status	244,844,277	100.00%	
Total unknown	18,404,857		
Grand total	263,249,134		

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 NM's and the NA's.

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

Response Class	Sum of weights	Percent residential (actual or imputed)	Actual or imputed residential
C	58,600,253	100.00%	58,600,253
I	101,672	100.00%	101,672
REF	32,179,589	100.00%	32,179,589
O	7,274,068	100.00%	7,274,068
AMNA	3,651,769	72.41%	2,644,255
PNA	14,753,088	40.09%	5,914,363
NRS	37,399,239	0.00%	0
NWS	109,289,456	0.00%	0
Total residential			106,714,200
Total complete			58,701,925
Response rate			***55.01%***

4.4.3 Extended Interview Response Rate

Table 4-6 shows the extended interview response rate calculations for the HINTS I 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. We estimate this 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 including a multiple-telephone number adjustment, further adjusted for nonresponding screeners), multiplied to 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 the HINTS I RDD sample

Extended interview category	Sum of weights	Percentage
Completes or partial completes	119,183,586	***60.07%***
All others	79,212,602	39.93%
Total	198,396,189	100.00%

4.4.4 Overall Response Rate

The overall response rate is computed by taking the product of the screener and the extended interview response rate. We view the overall response rate as 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 screener (estimated by the screener response rate from Section 4.4.2) and 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). This calculation is given in Table 4-7.

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

Response rate type	Response rate
Screener response rate	55.01%
Extended interview response rate	60.07%
Overall response rate	33.05%

5. CONCLUSIONS

The first administration of the HINTS accomplished the goal set forth at the inception of the study, which was to provide a vehicle for collecting data to assess the way in which the general population accesses and utilizes communication channels to obtain health information on cancer.

This goal was accomplished through a collaborative survey instrument design process that yielded an instrument that was subjected to extensive cognitive testing and field testing prior to the commencement of actual data collection. This pretesting also served to test the functionality of the CATI instrument.

Despite the successful result of the first administration of the HINTS, the survey had disappointing response rates for both the screening and extended interviews.

Achieving a high response rate on the survey proved to be a challenge. While the low survey response rate was not unexpected, given the general experience of the survey community with RDD surveys, it is a concern. On the HINTS, Westat attempted to improve the rates by using various mailings where an address could be found and by providing a small pre- and post-survey monetary incentive. Mailings included FedEx letters, refusal conversion letters, introductory letters, and brochures. Traditionally, these types of methods have been used to improve response rates. In general, with each successive attempt, a slight gain is made in the response rate. However, because the initial rate on the HINTS was low, improvement did not get it up to a level that is generally thought to be acceptable. Short of a large monetary incentive (about $15-20), which methods research has shown to be effective with certain population groups, it is unlikely that these kinds of refusal conversion approaches will solve the type of problem that was encountered in the HINTS.

A solution likely to yield a considerably higher response rate is to field an in-person survey. However, the cost of a survey with a large screening effort (12,000-15,000 households) and a large extended interview sample (6,000-8,000 subjects) would be several times the current cost of the RDD vehicle used for the HINTS.

Statisticians and other survey methodologists at Westat have been considering other options that might be less costly than an in-person survey. These include exploring ways to assess any bias that may exist due to the low response rate and also exploring methods for adjusting the sample/data to

account for bias that may be introduced. One possible option might be to conduct both an in-person interview in a small area-based probability sample and an RDD survey in the same compact geographic area. One could then compare the characteristics of respondents in the two samples. It might also be possible to determine a method for using the in-person sample to adjust for bias in the RDD sample.

A further consideration for attaining higher response rates for future HINTS is shortening the length of the interview. A 30-minute interview is a disincentive to potential respondents, regardless of the content of the interview. Interviewing and supervisory staff anecdotally reported that they felt the instrument was long and that the length complicated gaining cooperation for the extended interview and for converting refusals. A reduction in time for the extended interview can have a direct benefit if there is willingness on the part of respondents to agree to participate.

As was discussed in Section 4.3.2, minor issues with the CATI program caused delays that provided respondents with a chance to break off the call. Due to the time it took for the sampling algorithm to pick the proper household respondent, there was a slight pause between the completion of the screener and the start of the extended questionnaire. Following this pause was the introduction to the extended interview, which informed respondents that they could expect to spend as much as 30 minutes on the phone in order to complete the interview. Although a workaround was instituted for situations when the respondent to the screener was known to be the extended respondent, this pause provided an opportunity for respondents to break off. Again, the length of the extended interview had a negative impact on cooperation.

Finally, the content of the Health Communication (HC) section, the first section of the extended interview, was cognitively demanding for the respondents. Although well-crafted, this section had subtle, and sometimes confusing, follow-up questions, which perhaps wore down respondents. Also, interviewers reported many respondents' commenting that they felt the questions were repetitious. The break-offs that did occur were most prevalent following the HC section.

6. REFERENCES

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.

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.

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

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

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

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

Oh, H., and Scheuren, F. (1983). Weighting adjustments for unit response. In W.G. Madow, I. Olkin, and D.B. Rubin (Eds.), Incomplete Data in Sampling Surveys, Vol. II: Theory and Annotated Bibliography. New York: Academic Press.

The American Association for Public Opinion Research (AAPOR). (2000). Standard definitions: Final dispositions of case codes and outcome rates for surveys. Lenexa, KS: AAPOR.

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.

APPENDIX A

HINTS

HEALTH INFORMATION NATIONAL TRENDS SURVEY

MAIN STUDY INTERVIEW INSTRUMENT (ENGLISH)

July 2003

Post-Analysis Version

NATIONAL CANCER INSTITUTE (NCI)

The Privacy Act requires us to tell you that we are authorized to collect this information by Section 411.285a, 42 USC. You do not have to provide the information requested. However, the information you provide will help the National Cancer Institute's ongoing efforts to promote good health and prevent disease. There are no penalties should you choose not to participate in this study.

The information we collect in this study is in accordance with the clearance requirements of the paperwork Reduction Act of 1995. We may not conduct or sponsor, and you are not required to respond to, a collection of information unless it displays a valid control number from the Office of Management and Budget in the Federal Government. We estimate that it will take you between 20 and 30 minutes to answer our questions in this interview. This includes the time it takes to hear the instructions, gather the necessary facts, and complete the interview. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-xxxx)

OMB # 0925 - 0507
Expiration Date: 8/31/03

Table of Contents

This page left blank intentionally.

HOUSEHOLD ENUMERATION (HE)

HE-0. [PHONE NUMBER]
Autodial

USE AUTODIALER	1	
BYPASS AUTODIALER	2	

[SINTRO_1] Hello, the National Cancer Institute is conducting a nationwide research study. This is strictly a scientific study; we are not selling anything.

HE-1. Are you a member of this household and at least 18 years old?
HHMember18+

YES	1	(HE-3)
NO	2	
BUSINESS ADDRESS	3	(END STATEMENT)

HE-2. May I speak to a member of this household who is at least 18 years old?
HHAdultAvailable

AVAILABLE	1	(SINTRO_1)
NOT AVAILABLE	2	(MAKE APPOINTMENT)
THERE ARE NONE	3	(END STATEMENT)

HE-3. Is this phone number used for...
HHAdultPhoneUse
IntroADPhoneUse

home use,	1	
home and business use, or	2	
business use only?	3	(END STATEMENT)
GO TO RESULT	GT	(RESULT)

[HE-4 INTRO] This research study is about people's knowledge and beliefs about health issues such as cancer as well as how they get the information they need. You may have received a letter from the National Cancer Institute introducing this study. I have some questions to see whether you or any member of your household may be eligible to participate in this survey.
SCQHE4Intro

HE-4. Including yourself, how many people aged 18 or older currently live in this household?
[IF NEEDED: "Include people who usually stay in this household, but are temporarily away on business, vacation, or in the hospital. Do not include persons who are away on full-time active military duty with the armed forces, students living away from home in their own apartment, or any other family member who may be in a nursing home or other institution."]
SCQHE4NumberOfAdults

|___|___|
OF ADULT HH MEMBERS

BOX HE-0

IF THERE ARE NO ADULT HH MEMBERS, GO TO END STATEMENT.
IF HH WAS SAMPLED FROM A HIGH MINORITY STRATUM, CONTINUE WITH HE-5. OTHERWISE, GO TO BOX HE-2.

HE-5. We want to be sure that people of diverse ethnic backgrounds are adequately represented in the study, so I need to ask if {you are/any of the [FILL # FROM HE-4] adults in your household are} Black or Hispanic?

SCQHE5AnyBlack
SCQHE5AnyHispanicOrLatino

YES 1
NO 2

BOX HE-1
IF MINORITY HH, CONTINUE WITH BOX HE-2. OTHERWISE, SUB-SAMPLE NON-MINORITY HH AT 0.648 RATE. IF HH NOT SAMPLED, THEN GO TO END STATEMENT. OTHERWISE (IF HH SAMPLED), CONTINUE WITH BOX HE-2.

BOX HE-2
IF ONLY 1 ADULT IN HH, GO TO HE-9. OTHERWISE, RUN RESPONDENT SELECTION ALGORITHM. IF 2 ADULTS IN HH, GO TO HE-9. OTHERWISE IF RESPONDENT WAS SAMPLED, GO TO HE-9. OTHERWISE IF MORE THAN 2 ADULTS IN HH AND RESPONDENT WAS NOT SAMPLED, CONTINUE WITH HE-6.

HE-6. The computer has randomly determined that one of the [HE-4 answer minus 1] adults other than yourself should be selected for the rest of the interview. To help us select this person, do you know who has had the most recent birthday among these adults?

SCQHE6KnowBirthdays

YES 1
NO 2 (HE-8)

HE-7. Other than yourself then, which adult has had the most recent birthday?
[A FIRST NAME IS SUFFICIENT IF IT UNIQUELY IDENTIFIES THE HH MEMBER. IF NEEDED "We need some way to ask for this person should we need to call back. If you prefer, just give me that person's gender and age."]

SCQHE7NameRecentBirthday

NAME: _____

OR

GENDER: MALE .....1 AND AGE: |___|___|
FEMALE.....2

GO TO HE-10.

HE-8. So that the computer can choose someone to interview, please tell me the first names and ages of the [FILL # FROM HE-4 MINUS 1] adults currently living in this household. Please do not include yourself.
[IF NEEDED: "Include people who usually stay in this household, but are temporarily away on business, vacation, or in the hospital. Do not include persons who are away on full-time active military duty with the armed forces, students living away from home in their own apartment, or any other family member who may be in a nursing home or other institution."]

[IF NOT OBVIOUS, ASK: "Is {NAME} male or female?"]

SCQNameHHM
SCQAskAge
SCQGender

[IF R ANSWERS DK OR RF TO IDENTIFYING HH MEMBERS, EXIT INTERVIEW.]

FIRST NAME	GENDER	AGE
_____	MALE ..... 1 FEMALE..... 2	|___|___|
_____	MALE ..... 1 FEMALE..... 2	|___|___|
_____	MALE ..... 1 FEMALE..... 2	|___|___|

BOX HE-3

RUN SELECTION ALGORITHM ON HH MEMBERS LISTED IN HE-8 TO SELECT EXTENDED RESPONDENT. THEN, GO TO HE-10.

HE-9. {What is your first name?/The other adult in the household has been selected to participate in the next part of the study. What is the other adult's name?}

SCQHE9NameSelectedPerson
SCQHEYourGender
SCQHEYourAge

SCQHE9AnameOtherAdult
SCQHEPersonGender
SCQHEPersonAge

[PROBE FOR INFORMATION THAT UNIQUELY IDENTIFIES THE HH MEMBER SELECTED.]

NAME: _____

GENDER: MALE .....1 AND AGE: |___|___|
FEMALE.....2

HE-10. Besides the number I called, do you have other telephone numbers in your household that are for regular telephone usage? Please exclude telephone numbers that are dedicated for business use, faxes, or modems and all cell phones.

MorePhones

YES 1
NO 2

BOX HE-4

IF EXTENDED RESPONDENT = SCREENER RESPONDENT,
GO TO XINTRO_1.
OTHERWISE, CONTINUE.

HE-11. {(HH MEMBER) has been selected to participate in the next part of the study.} May I speak to (HH MEMBER}?

AVAILABLE	1	(XINTRO_1)
NOT AVAILABLE	2	(MAKE APPOINTMENT)

[XINTRO_1] {Hello, my name is [INTERVIEWER'S NAME] and I am calling for a nationwide research study being conducted for the National Cancer Institute. This study is about people's knowledge and beliefs about health issues such as cancer as well as how they get the health information they need. You may have received a letter from the National Cancer Institute introducing this study.} You have been selected to participate in this important research study. Your participation is voluntary, and you can refuse to answer any questions or withdraw from the study at any time. However, all information obtained will be kept confidential. The answers you give us will be used to improve the way that health communicators speak to the public about cancer prevention. The interview will take about 30 minutes depending on your answers.
IMGXintroToInterview

HE-12. [AUTOFILL. ASK ONLY IF NECESSARY: First, what is your age?]
SPAge

|___|___|___| (HE-14)
AGE

DK	999	(HE-13)
RF	998	(HE-13)

HE-13. Are you…
SPAgeRange

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

HE-14. [AUTOFILL. ASK ONLY IF NECESSARY: {Are you/Is NAME} male or female?]
SPGender

MALE	1
FEMALE	2

GO TO NEXT SECTION.

END STATEMENT. Thank you for your time, those are all of the questions that I have for you today. If you have questions about cancer or just want some information about cancer, you can call 1-800-4-CANCER or go to the National Cancer Institute's web site at: www.cancer.gov

HEALTH COMMUNICATION

Provider-Patient Interaction

Media Exposure

Information Seeking

Channel Credibility, Reliance

Internet Usage: General

This page left blank intentionally.

PROVIDER-PATIENT INTERACTION

I am going to ask you questions about people you talk with about your health.

HC-1. Not including psychiatrists and other mental health professionals, is there a particular doctor, nurse, or other health professional that you see most often?

HC1UsualProvider

YES	1	
NO	2	(HC-3)

HC-2. What kind of health professional do you see most often—a doctor or nurse or some other health professional?

HC2ProviderType

DOCTOR	1
NURSE	2
CHIROPRACTOR	3
PHYSICIAN'S ASSISTANT	4
NURSE PRACTITIONER	5
OTHER (SPECIFY) _____	91

HC2ProviderType_OS

HC-3. During the past 12 months, not counting times you went to an emergency room, how many times did you go to a doctor or other health care professional to get care for yourself? [IF NONE, ENTER 95.]

HC3FreqGoProvider

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

HC-4. During the past 12 months, how often did doctors or other health care providers [FILL PROVIDER ACTIVITY]? Would you say always, usually, sometimes, or never? (How often did they [FILL PROVIDER ACTIVITY]?)

PROVIDER ACTIVITY	ALWAYS	USUALLY	SOMETIMES	NEVER
a. listen carefully to you HC4aProviderListen	1	2	3	4
b. explain things in a way you could understand HC4bProviderExplain	1	2	3	4
c. show respect for what you had to say HC4cProviderRespect	1	2	3	4
d. spend enough time with you HC4dProviderTime	1	2	3	4
e. involve you in decisions about your health care as much as you wanted HC4eProviderInvolve	1	2	3	4

MEDIA EXPOSURE

The next questions are about how often you use radio, newspapers, and other media.

HC-5. Do you or anyone in your household have cable or satellite TV?
HC5Cable

YES..... 1
NO 2

HC-6. On a typical weekday, about how many hours do you…
[IF LESS THAN ONE HOUR A DAY, ENTER 1.]

HOURS

a. watch television? .....|___|___|
HC6aWatchTV
b. listen to the radio? .....|___|___|
HC6bListenRadio

HC-7. In the past seven days, how many days did you...

DAYS

a. read a newspaper? ..... |___|
HC7aReadNewspaper
b. read a magazine? ..... |___|
HC7bReadMagazine

HC-8. How much attention do you pay to information about health or medical topics [FILL MEDIA SOURCE]? Would you say a lot, some, a little, or not at all? (How about [FILL MEDIA SOURCE]?)
[CODE "DON'T USE" AS "NOT AT ALL".]

	A LOT	SOME	A LITTLE	NOT AT ALL
a. on TV HC8aAttendTV	1	2	3	4
b. on the radio..... HC8bAttendRadio	1	2	3	4
c. in newspapers..... HC8cAttendNewspaper	1	2	3	4
d. in magazines..... HC8dAttendMagazines	1	2	3	4
e. on the Internet..... HC8eAttendInternet	1	2	3	4

INFORMATION SEEKING

Now, I want to ask you questions about looking for information on cancer. Please consider all sources of information such as the Internet, the library, friends, and health care professionals.

HC-9. Have you ever looked for information about cancer from any source?

HC9SeekCancerInfo

YES	1
NO	2

HC-10. Excluding your doctor or other health care provider, has someone else ever looked for information about cancer for you?

HC10OtherSeekCancerInfo

YES	1	
NO	2	(BOX HC-1)

HC-11. Who was that?
[IF MORE THAN ONE PERSON HAS CONDUCTED A SEARCH ON BEHALF OF THE SP, PROBE FOR THE PERSON WHO CONDUCTED THE MOST RECENT SEARCH.]

HC11WhoSeekCancerInfo

SPOUSE	1
OTHER FAMILY MEMBER	2
FRIEND	3
CO-WORKER	4
INFORMATION SPECIALIST (E.G., LIBRARIAN)	5
CANCER ORGANIZATIONS	6
OTHER (SPECIFY) _____	91

HC11WhoSeekCancerInfo_0S

BOX HC-1

IF RESPONDENT HAS NOT LOOKED FOR INFORMATION FROM ANY SOURCE
ON CANCER, NEITHER INDIVIDUALLY NOR THROUGH SOMEONE ELSE,
THEN GO TO HC-15.
OTHERWISE, CONTINUE.

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).

HC-12. About how long ago was that?
[ENTER NUMBER.]
HC12LastLookCancerInfo
|___|___|
NUMBER

[ENTER UNIT.]
HC12LastLookCancerInfo_Unit
|___|
UNIT

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

BOX HC-2

IF SP HAS NOT LOOKED FOR INFORMATION ABOUT CANCER FOR SELF, GO TO HC-15.
OTHERWISE, CONTINUE.

HC-13. The most recent time you looked for information on cancer, where did you look first?
[IF SP HAS MADE MORE THAN ONE SEARCH, PROBE FOR THE MOST RECENT SEARCH.]
HC13WhereLookCancerInfo

BOOKS 1
BROCHURES, PAMPHLETS, ETC. 2
FAMILY 3
FRIEND/CO-WORKER 4
HEALTH CARE PROVIDER 5
INTERNET 6
LIBRARY 7
MAGAZINES 8
NEWSPAPERS 9
RADIO 10
TELEPHONE INFORMATION NUMBER (1-800 NUMBER) 11
CANCER ORGANIZATIONS 12
TELEVISION 13
OTHER (SPECIFY) _____ 91

HC13WhereLookCancerInfo_OS

HC-14. What type of information were you looking for in your most recent search?
HC14WhatLookCancerInfo

_____

CHANNEL CREDIBILITY, RELIANCE

HC-15. Imagine that you had a strong need to get information about cancer. Where would you go first?

HC15WhereGoFirst

BOOKS	1
BROCHURES, PAMPHLETS, ETC.	2
FAMILY	3
FRIEND/CO-WORKER	4
HEALTH CARE PROVIDER	5
INTERNET	6
LIBRARY	7
MAGAZINES	8
NEWSPAPERS	9
RADIO	10
TELEPHONE INFORMATION NUMBER (1-800 NUMBER)	11
CANCER ORGANIZATIONS	12
TELEVISION	13
CANCER RESEARCH/TREATMENT FACILITIES	14
OTHER (SPECIFY) _____	91

HC15WhereGoFirst_OS

HC-15A. Overall, how confident are you that you could get advice or information about cancer if you needed it? Would you say…

HC15AConfidenceGetCancerInfo

very confident,	1
somewhat confident,	2
slightly confident, or	3
not confident at all?	4

HC-16. People get information about cancer, including how to prevent it and find it early, from many sources. I am going to read you a list of information sources. For each one, please tell me whether you might like to get information about cancer that way, assuming it was free. (How about...)

		YES	NO	
a.	By e-mail or the Internet? HC16aPersonalComputer	1	2	(HC-16b)
	a1. Is there an Internet site you especially like? HC16a1InternetSiteLike	1	2	(HC-16b)
	a2. Which one? (SPECIFY) _____ HC16a2WhichSiteLike			
b.	watching a video cassette? HC16bVideoTape	1	2	
c.	listening to an audio cassette? HC16cAudioTape	1	2	
d.	reading materials created for you based on your personal lifestyle and family history? HC16dPersonalizedPrint	1	2	
e.	using an interactive computer CD-ROM that lets you select the information you want? HC16eCDROM	1	2	
f.	receiving a telephone call from a health care professional who could talk with you and answer your questions? HC16fTelephonecall	1	2	
g.	meeting in person with a health care professional? HC16gInPerson	1	2	
h.	reading a book, magazine, or other publication? HC16hPublication	1	2	
i.	Is there any other way you'd like to get information about cancer? (SPECIFY) _____ HC16iAnotherSource HC16iAnotherSource_OS	1	2	

HC-17 WAS DELETED

HC-18. How much would you trust the information about cancer from [FILL SOURCE]? Would you say a lot, some, a little, or not at all? (How about from [FILL SOURCE]?)

		A LOT	SOME	A LITTLE	NOT AT ALL
a.	a doctor or other health care professional HC18aTrustDoctor	1	2	3	4
b.	family or friends HC18bTrustFriendsFamily	1	2	3	4
c.	newspapers HC18cTrustNewspaper	1	2	3	4
d.	magazines HC18dTrustMagazines	1	2	3	4
e.	the radio HC18eTrustRadio	1	2	3	4
f.	the Internet HC18fTrustInternet	1	2	3	4
g.	television HC18gTrustTelevision	1	2	3	4

BOX HC-3

IF RESPONDENT HAS NOT LOOKED FOR INFORMATION FROM ANY SOURCE
ON CANCER, NEITHER INDIVIDUALLY NOR THROUGH SOMEONE ELSE,
THEN GO TO HC-20.
OTHERWISE, CONTINUE.

HC-19. Based on the results of your overall search for information on cancer, tell me how much 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. You wanted more information, but did not know where to find it. Would you say you strongly agree, somewhat agree, somewhat disagree, or strongly disagree? HC19aWantMoreInfo	1	2	3	4
b. It took a lot of effort to get the information you needed. (Would you say you...)..... HC19bAlotEffort	1	2	3	4
c. You did not have the time to get all the information you needed. (Would you say you...)..... HC19cNoTime	1	2	3	4
e. You felt frustrated during your search for the information. (Would you say you...)..... HC19eFrustrated	1	2	3	4
g. You were concerned about the quality of the information. (Would you say you...)..... HC19gConcernedQuality	1	2	3	4
j. The information you found was too hard to understand. (Would you say you...) HC19jTooHardUnderstand	1	2	3	4
k. You were satisfied with the information you found. (Would you say you...)..... HC19kSafisfied	1	2	3	4

HC-19d, HC-19f, HC-19h and HC-19i WERE DELETED

INTERNET USAGE: GENERAL

[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.]

Next, I'm going to ask about your usage of the Internet.

HC-20. Do you ever go on-line to access the Internet or World Wide Web, or to send and receive e-mail?
HC20UseInternet

YES	1	(HC-22)
NO	2	

HC-21. Which of the following, if any, are the reasons you do not access the Internet?

	YES	NO
a. Because you are not interested. HC21aNotInterested	1	2
b. Because it costs too much. HC21bCannotAfford	1	2
c. Because it is too complicated to use. HC21cTooComplicated	1	2
d. Because you do not think it is useful. HC21dNotUseful	1	2

GO TO HC-34.

HC-22. Do you ever go on-line to use the Internet from home?
HC22InternetHome

YES	1	
NO	2	(HC-24)

HC-23. When you use the Internet at home, do you mainly access it through . . .
HC23AccessInternet

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

HC23AccessInternet_OS

Although some of these questions may seem repetitious, they are all important to this study.

HC-24. In the past 12 months, did you use the Internet, whether from home or somewhere else,…

	YES	NO
a. to look for health or medical information for yourself? HC24aHealthInfoSelf	1	2
b. to look for health or medical information for someone else? HC24bHealthInfoOther	1	2

BOX HC-4

IF DID NOT LOOK FOR HEALTH INFO FOR SELF OR SOMEONE ELSE ON INTERNET,
THEN GO TO HC-26.
OTHERWISE, CONTINUE.

HC-25. In the past 12 months, how often did you use the Internet to look for health or medical information for {yourself (or) someone else}? Was it about . . .

HC25FreqInternetHealth

once a week,	1
once a month,	2
every few months, or	3
less often?	4

HC-26. 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. Bought medicine or vitamins on-line? HC26aBuyMedicine	1	2
b. Participated in an on-line support group for people with a similar health or medical issue? HC26bSupportGroup	1	2
c. Used e-mail or the Internet to communicate with a doctor or a doctor's office? HC26cTalkDoctor	1	2
e. Looked for health or medical information? HC26fBookmarkSite	1	2
g. Done anything else health-related on the Internet? (SPECIFY) _____ HC26hAnythingElseOnline HC26hAnythingElseOnline_OS	1	2

HC-16d AND HC-16f WERE DELETED

HC-27. When was the last time you used the Internet to look for information about health or medical care? Was it within the last week, within the last month, within the last year, over a year ago, or never?

HC27LastOnlineHealth

within the last week,	1	
within the last month,	2	
within the last year,	3	
over a year ago, or	4	
never?	5	(HC-34)

HC-28. The last time you used the Internet to look for information about health or medical care, was this . . .

HC28WhereLastOnline

from home,	1
from work, or	2
from someplace else? (SPECIFY) _____	91
AT FRIEND'S/RELATIVE'S HOUSE	3
FROM SCHOOL	4
FROM LIBRARY	5

HC28WhereLastOnline_OS

HC-29. Have you ever visited an Internet web site to learn specifically about cancer?

HC29InternetForCancer

YES	1	
NO	2	(HC-34)

HC-30. Did you get to this web site by doing an Internet search, or did you find out about it some other way?
HC30HowInternetForCancer

THROUGH AN INTERNET SEARCH	1
FOUND IT SOME OTHER WAY	2

HC-31. Overall, how useful was the cancer-related information you got from the Internet? Would you say . . .
HC31UsefulInternetForCancer

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

HC-32. In the past 12 months, how often have you used the Internet to look for advice or information about cancer? Would you say . . .
HC32FreqInternetForCancer

about once a week,	1
once a month,	2
every few months, or	3
less often?	4

HC-33. Why did you decide to use the Internet to look for information about cancer? Anything else? [CODE ALL THAT APPLY.]
HC33WhyUseInternet

A LOT OF INFORMATION AVAILABLE	10
COULD GET INFORMATION IMMEDIATELY	11
CONVENIENT	12
FREE/INEXPENSIVE	13
COULD INVESTIGATE ANONYMOUSLY	14
EASILY ACCESSIBLE/QUICK	15
MOST CURRENT/RELIABLE	16
OTHER (SPECIFY) _____	91

HC33WhyUseInternet_OS

OTHER COMMUNICATION

HC-34. Now, I'm going to read you a list of organizations. Before being contacted for this study, had you ever heard of...

	YES	NO
a. the National Institutes of Health? HC34aNIH	1	2
b. the American Cancer Society? HC34bACS	1	2
c. the Cancer Information Service? HC34cCIS	1	2
d. the National Cancer Institute? HC34dNCI	1	2
e. the 1-800-4-Cancer information number? HC34eCancerHotline	1	2
f. the United States Center for Cancer Prevention Research? HC34fUSCenter	1	2

CANCER KNOWLEDGE & SCREENING

Cancer History

General Cancer Knowledge

Colon Cancer

Breast Cancer

Cervical Cancer

Prostate Cancer

This page left blank intentionally.

Now, I would like to ask you about your personal experience with cancer.

CH-1. Have you ever been told by a doctor that you had cancer?

CH1EverHadCancer

YES	1
NO	2 (CH-4)

CH-2. What type of cancer was it, or in what part of the body did the cancer start?
[CODE ALL THAT APPLY.]

CH2TypeofSpsCancer

BLADDER CANCER	10
BREAST CANCER	11
CERVICAL CANCER (CANCER OF THE CERVIX)	12
COLON CANCER	13
ENDOMETRIAL CANCER (CANCER OF THE UTERUS)	14
HEAD AND NECK CANCER	15
LEUKEMIA/BLOOD CANCER	16
LUNG CANCER	17
LYMPHOMA	18
MELANOMA	19
OTHER SKIN CANCER	20
ORAL CANCER	21
OVARIAN CANCER	22
PANCREATIC CANCER	23
PHARYNGEAL (THROAT) CANCER	24
PROSTATE CANCER	25
RECTAL CANCER	26
RENAL (KIDNEY) CANCER	27
CODE NOT USED	28
BONE	29
STOMACH	30
HEART	31
BRAIN	32
INTERNAL ORGANS	33
CODE NOT USED	34
THYROID	35
OTHER (SPECIFY) _____	91

CH2TypeofSpsCancer_OS

CH-3. At what age or in what year were you first told that you had cancer?
[ENTER UNIT.]

CH3WhenDiagnosedCancer_Unit

|___|
UNIT

AGE	1
YEAR	2

[ENTER {AGE/YEAR}.]

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

CH3WhenDiagnosedCancer_Age
CH3WhenDiagnosedCancer_Year

CH-4. Have any of your brothers, sisters, parents, children, or other close family members ever had cancer?
[IF INDICATE “DOESN’T HAVE FAMILY,” CODE AS “NO FAMILY.”]

CH4FamilyEverHadCancer

YES	1
NO	2
NO FAMILY	3

GENERAL CANCER KNOWLEDGE (CK)

These questions are to see what people remember about public health messages. If you do not know the answer to a question, that's okay, just tell me you don't know.

RELATIVE PREVALENCE

[RANDOMLY ASSIGN RESPONDENTS TO RECEIVE RESPONSE OPTIONS IN CURRENT OR REVERSE ORDER.]

CK-1. Which of the following do you think causes the most deaths each year in the United States? Would you say . . .

CK1CauseMostDeaths

auto accidents,	1
cigarettes,	2
guns,	3
alcohol, or	4
drug use?	5

CK-2 & CK-3 WERE DELETED.

PREVENTABILITY

CK-4. Can you think of anything people can do to reduce their chances of getting cancer? Anything else?
[CODE ALL THAT APPLY.]

CK4ReduceCancer

EAT BETTER/BETTER NUTRITION	10	
GET SCREENED FOR CANCER/GET TESTED	11	
DON'T SMOKE/QUIT SMOKING	12	
EXERCISE/EXERCISE MORE	13	
STAY OUT OF THE SUN/WEAR SUNSCREEN	14	
DON'T DRINK ALCOHOL/DRINK LESS ALCOHOL	15	
GET A CHECK-UP/GO TO THE DOCTOR	16	
REDUCE STRESS, REST, GET ENOUGH SLEEP	17	
ENVIRONMENTAL, CHEMICALS, POLLUTION, 2ND HAND SMOKE, PESTICIDES, WEAR PROTECTIVE GEAR	18	
SELF EXAMS, BODY CHANGE AWARENESS	19	
DON'T DO DRUGS	20	
HEALTHY LIFESTYLE, POSITIVE ATTITUDE, GOOD STATE OF MIND SAFE SEX, MEDITATION, YOGA, MODERATION, DO NOT HAVE MULTIPLE PARTNERS	21	
EDUCATION, WELL INFORMED, AWARENESS, BOOKS, RESEARCH	22	
KNOWLEDGE OF FAMILY HISTORY, GOOD GENES, GENETIC TESTING	23	
REDUCE WEIGHT, MAINTAIN HEALTHY WEIGHT	24	
RELIGION, PRAYER	25	
LIMIT EXPOSURE TO CARCINOGENS, REDUCE TOXINS, AVOID CAUSES OF CANCER	26	
OTHER (RECORD UP TO 8 SPECIFIES) _____	91	
NO/NOTHING	95	(BOX CK-3)
DK	99	(BOX CK-3)
RF	98	(BOX CK-3)

CK4ReduceCancer_1OS - _8OS

BOX CK-1

IF CK-4 = 10 (EAT BETTER), THEN ASK CK-5.
OTHERWISE, GO TO BOX CK-2.

CK-5. What specific changes should people make in their eating habits to reduce their chances of getting cancer? [CODE ALL THAT APPLY.]

CK5EatReduceCancer

EAT LESS FAST FOOD	10
EAT LESS FAT	11
EAT LESS RED MEAT	12
EAT MORE FRUITS	13
EAT MORE VEGETABLES	14
EAT MORE FIBER	15
STOP DRINKING ALCOHOL/REDUCE ALCOHOL	16
EAT A BALANCED DIET/ALL FOOD GROUPS/FOLLOW FOOD PYRAMID	17
DRINK MORE WATER	18
LESS PROCESSING/CHEMICALS/PRESERVATIVES/ADDITIVES	19
LESS SUGAR/SWEETS/SODAS	20
EAT LESS/HEALTHY/WATCH WEIGHT/LOSE WEIGHT	21
EAT ORGANIC/NATURAL/HOMEGROWN FOODS	22
RESEARCH FOODS/READ LABELS	23
EAT LESS SALT/SODIUM	24
WATCH/LOWER CHOLESTEROL	25
TAKE VITAMIN & MINERAL SUPPLEMENTS/HERBAL SUPPLEMENTS/SPECIFIC VITAMIN-FOOD RECOMMENDATIONS	26
OTHER (SPECIFY) _____	91

CK5EatReduceCancer_OS

BOX CK-2

IF CK-4 = 11 (GET SCREENED), THEN ASK CK-6.
OTHERWISE, GO TO BOX CK-3.

CK-6. You said people should get tested for cancer. What kinds of tests do you have in mind? [CODE ALL THAT APPLY.]

CK6TestsReduceCancer

CLINICAL BREAST EXAM	10
COLONOSCOPY/SIGMOIDOSCOPY	11
DIGITAL RECTAL EXAM	12
MAMMOGRAM	13
PAP TEST	14
PSA TEST	15
STOOL BLOOD TEST/FECAL OCCULT BLOOD TEST	16
X-RAY	17
MRI/CT SCAN	18
BLOOD TEST	19
BREAST SELF EXAM	20
OTHER (SPECIFY) _____	91

CK6TestsReduceCancer_OS

BOX CK-3

IF HAD CANCER, THEN GO TO CK-10.
OTHERWISE, CONTINUE.

CK-7 WAS DELETED.

PERSONAL RISK

CK-8. How likely do you think it is that you will develop cancer in the future? Would you say your chance of getting cancer is . . .

CK8ChanceGetCancer

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

CK-9. How often do you worry about getting cancer? Would you say . . .

CK9WorryGetCancer

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

DESIRE TO CHANGE/BARRIERS TO CHANGE

CK-10. Is there anything about your behavior or your lifestyle that you would like to change to reduce your chances of getting cancer? Anything else?
[CODE ALL THAT APPLY.]

CK10ChangeBehavior

EAT BETTER/BETTER NUTRITION	10	
GET SCREENED FOR CANCER/GET TESTED	11	
DON'T SMOKE/QUIT SMOKING	12	
EXERCISE/EXERCISE MORE	13	
STAY OUT OF THE SUN/WEAR SUNSCREEN	14	
DON'T DRINK ALCOHOL/DRINK LESS ALCOHOL	15	
GET A CHECK-UP/GO TO THE DOCTOR	16	
REDUCE STRESS, REST, GET ENOUGH SLEEP	17	
ENVIRONMENTAL, CHEMICALS, POLLUTION, 2ND HAND SMOKE, PESTICIDES, WEAR PROTECTIVE GEAR	18	
SELF EXAMS, BODY CHANGE AWARENESS	19	
DON'T DO DRUGS	20	
HEALTHY LIFESTYLE, POSITIVE ATTITUDE, GOOD STATE OF MIND SAFE SEX, MEDITATION, YOGA, MODERATION, DO NOT HAVE MULTIPLE PARTNERS	21	
EDUCATION, WELL INFORMED, AWARENESS, BOOKS, RESEARCH	22	
KNOWLEDGE OF FAMILY HISTORY, GOOD GENES, GENETIC TESTING	23	
REDUCE WEIGHT, MAINTAIN HEALTHY WEIGHT	24	
RELIGION, PRAYER	25	
LIMIT EXPOSURE TO CARCINOGENS, REDUCE TOXINS, AVOID CAUSES OF CANCER	26	
OTHER (RECORD UP TO 8 SPECIFIES) _____	91	
NO/NOTHING	95	(BOX CK-6)
DK	99	(BOX CK-6)
RF	98	(BOX CK-6)

CK10ChangeBehavior_1OS - _8OS

BOX CK-4

IF CK-10 = 10 (EAT BETTER), THEN ASK CK-11.
OTHERWISE, GO TO BOX CK-5.

CK-11. What specific changes should you make in your eating habits to reduce your chances of getting cancer? [CODE ALL THAT APPLY.]

CK11ChangeDiet

EAT LESS FAST FOOD	10
EAT LESS FAT	11
EAT LESS RED MEAT	12
EAT MORE FRUITS	13
EAT MORE VEGETABLES	14
EAT MORE FIBER	15
STOP DRINKING ALCOHOL/REDUCE ALCOHOL	16
EAT A BALANCED DIET/ALL FOOD GROUPS/FOLLOW FOOD PYRAMID	17
DRINK MORE WATER	18
LESS PROCESSING/CHEMICALS/PRESERVATIVES/ADDITIVES	19
LESS SUGAR/SWEETS/SODAS	20
EAT LESS/HEALTHY/WATCH WEIGHT/LOSE WEIGHT	21
EAT ORGANIC/NATURAL/HOMEGROWN FOODS	22
RESEARCH FOODS/READ LABELS	23
EAT LESS SALT/SODIUM	24
WATCH/LOWER CHOLESTEROL	25
TAKE VITAMIN & MINERAL SUPPLEMENTS/HERBAL SUPPLEMENTS/SPECIFIC VITAMIN-FOOD RECOMMENDATIONS	26
OTHER (SPECIFY) _____	91

CK11ChangeDiet_OS

BOX CK-5

IF CK-10 = 11 (GET SCREENED), THEN ASK CK-12.
OTHERWISE, GO TO BOX CK-6.

CK-12. You said you should get tested for cancer. What kinds of tests do you have in mind? [CODE ALL THAT APPLY.]

CK12GetTested

CLINICAL BREAST EXAM	10
COLONOSCOPY/SIGMOIDOSCOPY	11
DIGITAL RECTAL EXAM	12
MAMMOGRAM	13
PAP TEST	14
PSA TEST	15
STOOL BLOOD TEST/FECAL OCCULT BLOOD TEST	16
X-RAY	17
MRI/CT SCAN	18
BLOOD TEST	19
BREAST SELF-EXAM	20
OTHER (SPECIFY) _____	91

CK12GetTested_OS

BOX CK-6

RANDOMLY ASSIGN HALF OF RESPONDENTS TO GET LIST A AND HALF TO GET LIST B IN CK-13.
LIST A INCLUDES ITEMS a, e, g, i, k, m, & o.
LIST B INCLUDES ITEMS b, d, h, j, l, n, r & s.

ONLY FEMALES, SHOULD GET "k" AND "l" IN CK-13.

CK-13c, CK-13f, CK-13p, CK-13q WERE DELETED

CK-13. I'm going to read you some things that may affect a person's chances of getting cancer. Do you think that [FILL EXPOSURE] increase(s) a person's chances of getting cancer a lot, a little, or not at all or do you have no opinion? (How about [FILL EXPOSURE]?)

	A LOT	A LITTLE	NOT AT ALL	NO OPINION
a. smoking..... CK13aSmoking	1	2	3	4
b. eating a high-fat diet..... CK13bHighFatDiet	1	2	3	4
d. exposure to the sun..... CK13dExposureToSun	1	2	3	4
e. pesticides or food additives..... CK13eFoodAdditives	1	2	3	4
g. not eating much fiber..... CK13gLowFiber	1	2	3	4
h. not eating many fruits and vegetables..... CK13hFewFruitsVegetables	1	2	3	4
i. stress CK13iStress	1	2	3	4
j. drinking a lot of alcoholic beverages CK13jAlcohol	1	2	3	4
k. being hit in the breast..... CK13kHitBreast	1	2	3	4
l. having many sexual partners CK13lManySexPartners	1	2	3	4
m. having a family history of cancer CK13mFamilyHistory	1	2	3	4
n. being a particular race or ethnicity CK13nRaceEthnicity	1	2	3	4
o. not getting much exercise CK13oLittleExercise	1	2	3	4
r. pollution..... CK13rPollution	1	2	3	4
s. radon..... CK13sRadon	1	2	3	4

CK-14. Tell me how much you agree or disagree with the following statements, or if you have no opinion.

	STRONGLY AGREE	SOMEWHAT AGREE	SOMEWHAT DISAGREE	STRONGLY DISAGREE	NO OPINION
a. It seems like almost everything causes cancer. Would you say you strongly agree, somewhat agree, somewhat disagree, strongly disagree, or you have no opinion?..... CK14aEverythingCausesCancer	1	2	3	4	5
b. There's not much people can do to lower their chances of getting cancer. (Would you say you…) CK14bCannotLowerChances	1	2	3	4	5
c. There are so many different recommendations about preventing cancer, it's hard to know which ones to follow. (Would you say you...) CK14cTooManyRecommendations	1	2	3	4	5

[RANDOMLY ASSIGN RESPONDENTS TO RECEIVE RESPONSE OPTIONS IN CURRENT OR REVERSE ORDER.]

CK-15. Which type of cancer do you think will cause the most deaths in {women/men} this year in the U.S.? Would you say . . .

CK15CancerCauseMoreDeaths

lung cancer, ...	1
{breast/prostate} cancer, ...	2
colon cancer, ...	3
{cervical/testicular} cancer, or	4
skin cancer?.....	5

COLON CANCER (CC)

BOX CC-1

IF RESPONDENT HAS HAD COLON CANCER,
GO TO NEXT SECTION.
OTHERWISE, CONTINUE.

The next questions are about colon cancer.

PERSONAL RISK

CC-1. 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 . . .

CC1ChanceColonCancer

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

CC-2. Compared to the average {man/woman} your age, would you say that you are . . .

CC2RelativeChanceColonCancer

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

CC-3. How often do you worry about getting colon cancer? Would you say . . .

CC3FreqWorryColonCancer

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

SCREENING KNOWLEDGE AND BEHAVIORS

CC-4. Can you think of any tests that detect colon cancer? Anything else?
[CODE ALL THAT APPLY. IF R DOES NOT KNOW NAME OF TEST, ENTER DESCRIPTION OF TEST IN OTHER, SPECIFY.]

CC4TestsForColonCancer

BARIUM ENEMA	10
BIOPSY	11
STOOL BLOOD TEST/FECAL OCCULT BLOOD TEST	12
COLONOSCOPY	13
DIGITAL RECTAL EXAM	14
PROCTOSCOPY	15
SIGMOIDOSCOPY	16
LOWER GI	17
MRI/SCANS/CAT SCANS	18
BLOOD TEST	19
OTHER (SPECIFY) _____	91
NO/NOTHING	95

CC4TestsForColonCancer_OS

The following questions are about the stool blood test, also known as a Fecal Occult Blood Test, 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.

CC-4A. Have you ever heard of a fecal occult or stool blood test?
[IF NEEDED: You smear a small amount of stool on cards at home and send the cards back to the doctor or lab.]
CC4AEverHeardStoolBlood

YES	1	
NO	2	(CC-15A)

BOX CC-2

IF RESPONDENT IS 45 YEARS OLD OR OLDER AND HAS TALKED TO HEALTH PROFESSIONAL IN PAST YEAR,
CONTINUE WITH CC-5.
IF RESPONDENT IS 45 YEARS OLD OR OLDER AND HAS NOT TALKED TO HEALTH PROFESSIONAL IN PAST YEAR , GO TO CC-6.
OTHERWISE, GO TO CC-14.

CC-5. During the past 12 months, did a doctor, nurse, or other health professional advise you to do a stool blood test using a home test kit?
CC5DoctorAdviseStoolBlood

YES	1	
NO	2	

CC-6. Have you ever done a stool blood test using a home test kit?
CC6HadStoolBlood

YES	1	
NO	2	(CC-10)
DK	9	(CC-11)
RF	8	(CC-11)

CC-7. When did you do your most recent stool blood test using a home kit to check for colon cancer?
CC7WhenStoolBlood

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-8 WAS DELETED.

CC-9. You said your most recent stool blood test was {INSERT TIME FRAME FROM CC-7}. How long before that stool test was your last one?
CC9StoolBloodLast3Years

A YEAR AGO OR LESS BEFORE	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS BEFORE	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS BEFORE	3
OVER 5 YEARS BEFORE	4
NONE BEFORE THE MOST RECENT	5

BOX CC-3

IF RESPONDENT HAD A FOBT MORE THAN 1 YEAR AGO, CONTINUE.
IF DK OR RF WHEN MOST RECENT FOBT WAS, GO TO CC-11.
OTHERWISE, GO TO CC-13.

CC-10. Is there any particular reason why you haven't done a home stool blood test {yet/in the past year}? [CODE ALL THAT APPLY.]

CC10WhyNotStoolBlood

NO REASON	10	
DIDN'T NEED/ DIDN'T KNOW NEEDED THIS TEST	11	
DOCTOR DIDN'T ORDER IT/ DIDN'T SAY I NEEDED IT	12	
HAVEN'T HAD ANY PROBLEMS/NO SYMPTOMS	13	
PUT IT OFF/ DIDN'T GET AROUND TO IT	14	
TOO EXPENSIVE/NO INSURANCE/COST	15	
TOO PAINFUL, UNPLEASANT, OR EMBARRASSING	16	
HAD ANOTHER TYPE OF COLON EXAM	17	
DON'T HAVE DOCTOR	18	
NEVER HEARD OF IT/NEVER THOUGHT ABOUT IT	19	(CC-15A)
HAD STOOL BLOOD TEST DONE AT DOCTOR'S OFFICE	20	
AGE/THOUGHT THEY WERE TOO YOUNG	21	
OTHER (SPECIFY) _____	91	

CC10WhyNotBloodStool_OS

CC-11. Have you thought about doing {a/another} home stool blood test?

CC11ThoughtBloodStool

YES	1	
NO	2	(CC-14)

CC-12. Would you say that . . .

CC12PlanBloodStool

you plan to do one,	1	
you <u>don't</u> plan to do one, or	2	(CC-14)
you're undecided?	3	(CC-14)

CC-13. When do you expect to do your next home stool blood test?

CC13WhenNextBloodStool

A YEAR OR LESS FROM NOW	1	
MORE THAN 1 BUT NOT MORE THAN 2 YEARS FROM NOW	2	
MORE THAN 2 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	

CC-14. At what age are people supposed to start doing <u>home</u> stool blood tests? [IF R SAYS "When a doctor says to," PROBE FOR AN ESTIMATE OF THE AGE. RECORD "95" IF R DOES NOT GIVE AN AGE.]

CC14AgeBloodStool

|___|___|
AGE

WHEN A DOCTOR/HEALTH PROVIDER SAYS TO	95	

CC-15. In general, once people start doing home stool blood tests, about how often should they do them? [PROBE FOR GENERAL GUIDELINE ASSUMING NO PRIOR PROBLEMS.]

CC15FreqBloodStool

MORE OFTEN THAN ONCE A YEAR	1
EVERY 1 TO < 2 YEARS	2
EVERY 2 TO < 3 YEARS	3
EVERY 3 TO < 5 YEARS	4
EVERY 5 TO < 10 YEARS	5
10 YEARS OR MORE	6
ONLY WHEN THERE IS A PROBLEM	7
DEPENDS ON AGE	8
DEPENDS ON RESULTS OF PREVIOUS TESTS	9
WHEN DOCTOR/HP SAYS TO	95
OTHER (SPECIFY) _____	91

CC15FreqBloodStool_OS

CC-15A. Have you ever heard of a sigmoidoscopy or a colonoscopy?

CC15AEverHeardSigCol

YES	1	
NO	2	(CC-26)

BOX CC-4

IF RESPONDENT IS 45 YEARS OLD OR OLDER,CONTINUE.
OTHERWISE, GO TO CC-24.

[Just to review], a sigmoidoscopy and a colonoscopy are both tests that examine the bowel by inserting a tube in the rectum. The difference is that during a sigmoidoscopy, you are awake and can drive yourself home after the test; however during a colonoscopy, you may feel sleepy and you need someone to drive you home.

CC-16. Have you ever had …

	YES	NO
a. a sigmoidoscopy? CC16aHadSigmoidoscopy	1	2
b. a colonoscopy? CC16bHadColonoscopy	1	2

BOX CC-4A

IF RESPONDENT HAD EITHER A SIGMOIDOSCOPY OR A COLONOSCOPY, THEN GO TO BOX CC-5.
OTHERWISE, CONTINUE.

CC-17. Did a doctor, nurse, or other health professional ever advise you to get …

	YES	NO
a. a sigmoidoscopy? CC17aAdviseSigmoidoscopy	1	2
b. a colonoscopy? CC17bAdviseColonoscopy	1	2

BOX CC-5

IF RESPONDENT HAS HAD A SIGMOIDOSCOPY OR COLONOSCOPY, CONTINUE. IF DK OR RF TO ANSWER WHETHER HAD SIGMOIDOSCOPY OR COLONOSCOPY, GO TO CC-21.
OTHERWISE, GO TO CC-20.

CC-18. When did you have your most recent {sigmoidoscopy/(or) colonoscopy} to check for colon cancer?

CC18WhenSigCol

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	

CC-19. You said your most recent {sigmoidoscopy/ (or) /colonoscopy] was {INSERT TIME FRAME FROM CC-18}. How long before that {sigmoidoscopy/(or)/colonoscopy} was the last one?

CC19SigColLast10Years

A YEAR OR LESS BEFORE	1	
MORE THAN 1 BUT NOT MORE THAN 5 YEARS BEFORE	2	
MORE THAN 5 BUT NOT MORE THAN 10 YEARS BEFORE	3	
OVER 10 YEARS BEFORE	4	
NONE BEFORE MOST RECENT	5	

BOX CC-6

IF RESPONDENT HAD A SIGMOIDOSCOPY OR COLONOSCOPY MORE THAN 10 YEARS AGO, CONTINUE.
OTHERWISE, GO TO CC-23.

CC-20. Is there any particular reason why you haven't gotten a {sigmoidoscopy/(or) colonoscopy} {yet/in the past 10 years}?
[CODE ALL THAT APPLY.]

CC20WhyNotSigCol

NO REASON	10	
DIDN'T NEED/ DIDN'T KNOW NEEDED THIS TEST	11	
DOCTOR DIDN'T ORDER IT/ DIDN'T SAY I NEEDED IT	12	
HAVEN'T HAD ANY PROBLEMS/NO SYMPTOMS	13	
PUT IT OFF/ DIDN'T GET AROUND TO IT	14	
TOO EXPENSIVE/NO INSURANCE/COST	15	
TOO PAINFUL, UNPLEASANT, OR EMBARRASSING	16	
HAD ANOTHER TYPE OF COLON EXAM	17	
DON'T HAVE DOCTOR	18	
NEVER HEARD OF IT/NEVER THOUGHT ABOUT IT	19	(CC-26)
AGE/THOUGHT THEY WERE TOO YOUNG	21	
OTHER (SPECIFY) _____	91	

CC20WhyNotSigCol_OS

CC-21. Have you thought about getting {a/another} {sigmoidoscopy/(or) colonoscopy}?

CC21ThoughtSigCol

YES	1	
NO	2	(CC-24)

CC-22. Would you say that . . .

CC22PlanSigCol

you plan to get one,	1	
you don't plan to get one, or	2	(CC-24)
you're undecided?	3	(CC-24)

CC-23. When do you expect to have {your next/a} {sigmoidoscopy/(or) colonoscopy}?

CC23WhenNextSigCol

A YEAR OR LESS FROM NOW	1
MORE THAN 1 BUT NOT MORE THAN 5 YEARS FROM NOW	2
MORE THAN 5 BUT NOT MORE THAN 10 YEARS FROM NOW	3
OVER 10 YEARS FROM NOW	4
AM NOT PLANNING TO HAVE ANOTHER	5
IF I HAVE SYMPTOMS	6
WHEN DOCTOR/HEALTH PROVIDER RECOMMENDS	7

CC-24. At what age are people supposed to start having sigmoidoscopy or colonoscopy exams?
[IF R SAYS "When a doctor says to," PROBE FOR AN ESTIMATE OF THE AGE. RECORD "95" IF R DOES NOT GIVE AN AGE.]

CC24AgeSigCol

|___|___|
AGE

WHEN A DOCTOR/HEALTH PROVIDER SAYS TO	95

CC-25. In general, once people start having sigmoidoscopy or colonoscopy exams, about how often should they have them?
[PROBE FOR GENERAL GUIDELINE ASSUMING NO PRIOR PROBLEMS.]

CC25FreqSigCol

MORE OFTEN THAN ONCE A YEAR	1
EVERY 1 TO < 2 YEARS	2
EVERY 2 TO < 3 YEARS	3
EVERY 3 TO < 5 YEARS	4
EVERY 5 TO < 10 YEARS	5
10 YEARS OR MORE	6
ONLY WHEN THERE IS A PROBLEM	7
DEPENDS ON AGE	8
DEPENDS ON RESULTS FROM PREVIOUS TESTS	9
WHEN DOCTOR/HEALTH PROVIDER RECOMMENDS	95
OTHER (SPECIFY) _____	91

CC25FreqSigCol_OS

DETECTION/CURABILITY

CC-26. I am going to read you a few statements people might make about getting checked for colon cancer. For each, tell me how much you agree or disagree, or if you have no opinion.

	STRONGLY AGREE	SOMEWHAT AGREE	SOMEWHAT DISAGREE	STRONGLY DISAGREE	NO OPINION
a. Arranging to get checked for colon cancer would be easy for you. Would you say you strongly agree, somewhat agree, somewhat disagree, strongly disagree, or you have no opinion? CC26aCheckColonEasy	1	2	3	4	5
b. You are afraid of finding colon cancer if you were checked. (Would you say you…)..... CC26bCheckColonAfraid	1	2	3	4	5
c. Getting checked regularly for colon cancer increases the chances of finding cancer when it's easy to treat. (Would you say you…) CC26cCheckColonDetectEarly	1	2	3	4	5
d. Getting checked for colon cancer is too expensive. (Would you say you…)..... CC26dCheckColonTooExpensive	1	2	3	4	5

CC-27. When would you say the risk of colon cancer is highest, when you're . . .
CC27RiskColonCancer

under 40 years old, ...	1
between 40 and 60 years old, or.....	2
over 60 years old?	3

This page left blank intentionally.

BREAST CANCER (BC)

BOX BC-1

IF RESPONDENT IS FEMALE AND HAS NOT HAD BREAST CANCER, CONTINUE WITH BC-1.
OTHERWISE, GO TO NEXT SECTION.

The next questions are about breast cancer.

PERSONAL RISK

BC-1. How likely do you think it is that you will develop breast cancer in the future? Would you say your chance of getting breast cancer is . . .

BC1ChanceBreastCancer

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

BC-2. Compared to the average woman your age, would you say that you are . . .

BC2RelativeChanceBreastCancer

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

BC-3. How often do you worry about getting breast cancer? Would you say . . .

BC3FreqWorryBreastCancer

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

SCREENING KNOWLEDGE AND BEHAVIORS

BOX BC-2

IF RESPONDENT IS 35 YEARS OR OLDER AND HAS TALKED TO HEALTH PROFESSIONAL IN PAST YEAR, CONTINUE WITH BC-4.
IF RESPONDENT IS 35 YEARS OR OLDER AND HAS NOT TALKED TO HEALTH PROFESSIONAL IN PAST YEAR, GO TO BC-5.
OTHERWISE, GO TO BC-13.

BC-4. A mammogram is an x-ray of each breast to look for breast cancer. During the past 12 months, did a doctor, nurse, or other health professional advise you to get a mammogram?

BC4DoctorAdviseMammogram

YES	1
NO	2

BC-5. {A mammogram is an x-ray of each breast to look for breast cancer.} Have you ever had a mammogram?

BC5HadMammogram

YES	1	
NO	2	(BC-9)
DK	9	(BC-10)
RF	8	(BC-10)

BC-6. When did you have your most recent mammogram to check for breast cancer?

BC6WhenMammogram

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-7 WAS DELETED.

BC-8. You said your most recent mammogram was {INSERT TIME FRAME FROM BC-6}. How long before that mammogram was the last one?

BC8MammogramLast6Years

A YEAR OR LESS BEFORE	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS BEFORE	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS BEFORE	3
OVER 5 YEARS BEFORE	4
NONE BEFORE MOST RECENT	95

BOX BC-3

IF RESPONDENT HAD A MAMMOGRAM
MORE THAN 2 YEARS AGO, CONTINUE.
IF DK OR RF TO ANSWER WHEN HAD LAST MAMMOGRAM,
GO TO BC-10.
OTHERWISE, GO TO BC-12.

BC-9. Is there any particular reason why you haven't had a mammogram {yet/in the past 2 years}? [CODE ALL THAT APPLY.]

BC9WhyNotMammogram

NO REASON	10	
DIDN'T NEED/ DIDN'T KNOW NEEDED THIS TEST	11	
DOCTOR DIDN'T ORDER IT/ DIDN'T SAY I NEEDED IT	12	
HAVEN'T HAD ANY PROBLEMS/NO SYMPTOMS	13	
PUT IT OFF/ DIDN'T GET AROUND TO IT	14	
TOO EXPENSIVE/NO INSURANCE/COST	15	
TOO PAINFUL, UNPLEASANT, OR EMBARRASSING	16	
HAD ANOTHER TYPE OF BREAST EXAM	17	
DON'T HAVE DOCTOR	18	
NEVER HEARD OF IT/NEVER THOUGHT ABOUT IT	19	(NEXT SECTION)
AGE/THOUGHT THEY WERE TOO YOUNG	21	
OTHER (SPECIFY) _____	91	

BC9WhyNotMammogram_OS

BC-10. Have you thought about getting {a/another} mammogram?

BC10ThoughtMammogram

YES	1	
NO	2	(BC-13)

BC-11. Would you say that . . .

BC11PlanMammogram

you plan to get one,	1	
you don't plan to get one, or	2	(BC-13)
you're undecided?	3	(BC-13)

BC-12. When do you expect to have your next mammogram?

BC12WhenNextMammogram

A YEAR OR LESS FROM NOW	1
MORE THAN ONE BUT NOT MORE THAN 2 YEARS FROM NOW	2
MORE THAN 2 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

BC-13. {A mammogram is an x-ray of each breast to look for breast cancer.} At what age are women supposed to start having mammograms?
[IF R SAYS "When a doctor says to," PROBE FOR AN ESTIMATE OF THE AGE. RECORD "95" IF R DOES NOT GIVE AN AGE.]

BC13AgeMammogram

|___|___|
AGE

WHEN A DOCTOR/HEALTH PROVIDER SAYS TO	95

BC-14. In general, once women start having mammograms, about how often should they have them?
[PROBE FOR GENERAL GUIDELINE ASSUMING NO PRIOR PROBLEMS.]

BC14FreqMammogram

MORE OFTEN THAN ONCE A YEAR	1
EVERY 1 TO < 2 YEARS	2
EVERY 2 TO < 3 YEARS	3
EVERY 3 TO < 5 YEARS	4
EVERY 5 TO < 10 YEARS	5
10 YEARS OR MORE	6
ONLY WHEN THERE IS A PROBLEM	7
DEPENDS ON AGE	8
WHEN DOCTOR/HP RECOMMENDS	95
OTHER (SPECIFY) _____	91

BC14FreqMammogram_OS

This page left blank intentionally.

CERVICAL CANCER (CV)

BOX CV-1

IF RESPONDENT IS FEMALE AND HAS NOT HAD CERVICAL CANCER, CONTINUE WITH CV-1.
OTHERWISE, GO TO NEXT SECTION.

The next questions are about cervical cancer.

CV-1. A Pap smear is a test for cancer of the cervix. Have you ever had a Pap smear?
CV1HadPapSmear

YES	1	
NO	2	(NEXT SECTION)

CV-2. When did you have your most recent Pap smear to check for cervical cancer?
CV2WhenPapSmear

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
MORE THAN 5 YEARS AGO	4

CV-3 WAS DELETED.

CV-4. You said your most recent Pap smear was {INSERT TIME FRAME FROM CV-2}. How long before that Pap smear was the last one?
CV4PapSmearLast6Years

A YEAR OR LESS BEFORE	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS BEFORE	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS BEFORE	3
MORE THAN 5 BEFORE	4
NONE BEFORE MOST RECENT	95

CV-5. Have you had a hysterectomy?
[IF NEEDED: A hysterectomy is an operation to remove the uterus (womb).]
CV5Hysterectomy

YES	1	(NEXT SECTION)
NO	2	

CV-6. When do you expect to have your next Pap smear?
CV6WhenNextPapSmear

A YEAR OR LESS FROM NOW	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS FROM NOW	2
MORE THAN 2 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

This page left blank intentionally.

PROSTATE CANCER (PC)

BOX PC-1

IF RESPONDENT IS MALE AND HAS NOT HAD PROSTATE CANCER, CONTINUE WITH PC-1.
OTHERWISE, GO TO NEXT SECTION.

The next questions are about cancer affecting the prostate gland in men.

PERSONAL RISK

PC-1. How likely do you think it is that you will develop prostate cancer in the future? Would you say your chance of getting prostate cancer is . . .

PC1ChanceProstateCancer

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

PC-2. Compared to the average man your age, would you say that you are . . .

PC2RelativeChanceProstateCance

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

PC-3. How often do you worry about getting prostate cancer? Would you say . . .

PC3FreqWorryProstateCancer

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

SCREENING KNOWLEDGE AND BEHAVIORS

A Prostate-Specific Antigen test, also called a PSA test, is a blood test used to check men for prostate cancer.

PC-3A. Have you ever heard of a PSA or prostate-specific antigen test?

PC3AEverHeardPSATest

YES..... 1
NO ..... 2 (NEXT SECTION)

BOX PC-2

IF RESPONDENT IS 35 YEARS OLD OR OLDER AND HAS TALKED TO A HEALTH PROFESSIONAL IN THE PAST YEAR, CONTINUE WITH PC-4.
IF RESPONDENT IS 35 YEARS OLD OR OLDER AND HAS NOT TALKED TO A HEALTH PROFESSIONAL IN THE PAST YEAR, GO TO PC-5.
OTHERWISE, GO TO NEXT SECTION.

PC-4. During the past 12 months, did a doctor, nurse, or other health professional advise you to get a PSA test?
PC4DoctorAdvisePSATest

YES	1	
NO	2	
HAD BLOOD TEST, BUT DK IF CHECKED PSA	3	

PC-5. Have you ever had a PSA test?
PC5HadPSATest

YES	1	
NO	2	(NEXT SECTION)
HAD BLOOD TEST, BUT DK IF CHECKED PSA	3	(NEXT SECTION)
DK	9	(NEXT SECTION)
RF	8	(NEXT SECTION)

PC-6. When did you have your most recent PSA test to check for prostate cancer?
PC6WhenPSATest

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

PC-7. You said your last PSA test was {INSERT TIME FRAME FROM PC-6} How long before <u>that</u> PSA test was the last one?
PC7PSALast5Years

A YEAR OR LESS BEFORE	1
MORE THAN 1 BUT NOT MORE THAN 2 YEARS BEFORE	2
MORE THAN 2 BUT NOT MORE THAN 5 YEARS BEFORE	3
OVER 5 YEARS BEFORE	4
NONE BEFORE MOST RECENT	95

PC-8 THROUGH PC-12 WERE DELETED

THE SKIN CANCER SECTION WAS DELETED

PRIMARY CANCER RISK BEHAVIORS

Tobacco Use

Fruits and Vegetables

Exercise

Overweight/Obesity

This page left blank intentionally.

TOBACCO SCREENER

Now, I'd like to ask you about your use of tobacco.

TU-1. Have you smoked at least 100 cigarettes in your entire life?
[IF NEEDED: 5 Packs = 100 Cigarettes.]
TU1Smoke100

YES	1	
NO	2	(BOX TU-4)

TU-2. Do you now smoke cigarettes . . .
TU2SmokeNow

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

TU-3. 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.]
TU3SmokeDayAlways

|___|___|
NUMBER OF CIGARETTES

GO TO TU-5.

TU-4. On the average, when you smoked during the past 30 days, about how many cigarettes did you smoke a day?
[IF NEEDED: 1 Pack = 20 Cigarettes.]
[IF LESS THAN ONE A DAY, ENTER 0. IF 76 OR MORE, ENTER 76.]
TU4SmokeDaySometimes

|___|___|
NUMBER OF CIGARETTES

CURRENT SMOKERS

TU-5. Would you say that . . .
TU5PlanQuitSmoking

you plan to quit smoking,	1
you don't plan to quit, or	2
you're undecided?	3

BOX TU-1

IF SMOKE EVERY DAY OR SOME DAYS, GO TO TU-9.
OTHERWISE, CONTINUE.

FORMER SMOKERS

TU-6. About how long has it been since you last smoked cigarettes?
TU6WhenQuitSmoke

LESS THAN 1 MONTH AGO	1
1 MONTH TO LESS THAN 3 MONTHS AGO	2
3 MONTHS TO LESS THAN 6 MONTHS AGO	3
6 MONTHS TO LESS THAN 1 YEAR AGO	4
1 YEAR TO LESS THAN 5 YEARS AGO	5
5 YEARS TO LESS THAN 15 YEARS AGO	6
15 OR MORE YEARS AGO	7

TU-7. On the average, when you smoked, about how many cigarettes did you smoke a day?
[IF NEEDED: 1 Pack = 20 Cigarettes.]
[IF LESS THAN ONE A DAY, ENTER 0. IF 76 OR MORE, ENTER 76.]
TU7SmokeDayFormer

|___|___|
NUMBER OF CIGARETTES

ALL SMOKERS

TU-8 DELETED

TU-9. I am going to read you some statements people might make about smoking. For each, tell me how much you agree or disagree, or if you have no opinion?

	STRONGLY AGREE	SOMEWHAT AGREE	SOMEWHAT DISAGREE	STRONGLY DISAGREE	NO OPINION
a. Exercise can undo most of the effects of smoking. Would you say you strongly agree, somewhat agree, somewhat disagree, strongly disagree, or you have no opinion?..... TU9aExercise	1	2	3	4	5
b. Vitamins can undo most of the effects of smoking. (Would you say you...) TU9bVitamins	1	2	3	4	5
c. There's no risk of getting cancer if someone only smokes a few years. (Would you say you...)..... TU9cFewYears	1	2	3	4	5
d. Whether a person gets lung cancer depends more on genes than anything else. (Would you say you...) TU9dGenes	1	2	3	4	5

BOX TU-2

IF NO LONGER SMOKE, GO TO TU-12.
OTHERWISE IF SMOKE EVERY DAY OR SOME DAYS, CONTINUE.

TU-10. What type of cigarette do you now smoke most often—a regular, light, ultra light, or some other type?
TU10TypeCigarette

REGULAR/FULL-FLAVOR	1	(TU-12)
LIGHT/MILD	2	
ULTRA-LIGHT	3	
NO USUAL TYPE	4	(TU-12)
MEDIUM	5	
SOME OTHER TYPE (E.G., OMNI, ECLIPSE, ETC.)_____	91	

TU10TypeCigarette_OS

TU-11. What is the main reason why you now smoke {lights/ultra-lights/this type of cigarette}?..Is it . . .
TU11WhySmokeType

a way to reduce the health risks of smoking,	1
to try to quit smoking,	2
because of the taste, or	3
for some other reason? (SPECIFY) _____	91
NOT AS STRONG/LESS NICOTINE/LIGHTER/LOW IN TAR	4
HABIT/ADDICTED	5
NO REASON	6
STRESS/RELAX/NERVES	7
THAT'S WHAT I'VE ALWAYS SMOKED/WHAT I LIKE	8
COST/CHEAPER	9

TU11WhySmoketype_OS

TU-12. If a new cigarette were advertised as less harmful than current cigarettes, how interested would you be in trying it? Would you say . . .
TU12WouldTryLessHarmfulCig

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

BOX TU-3

IF RESPONDENT STOPPED SMOKING 5 OR MORE YEARS AGO, GO TO BOX TU-4.
OTHERWISE, CONTINUE.

TU-13. Tobacco companies have recently introduced new types of cigarettes that are claimed to have fewer harmful chemicals or carcinogens. These have names like Eclipse, Accord, Advance, and Omni. Have you ever tried one of these products?
TU13TriedLessHarmfulCig

YES	1
NO	2

TU-14. Tobacco companies have also recently introduced new types of smokeless tobacco products. These have names like Arriva, Exalt, and Revel. Have you ever tried one of these products?
TU14TriedSmokelessTobacco

YES	1
NO	2

BOX TU-4

IF RESPONDENT HAS HAD LUNG CANCER,
GO TO NEXT SECTION.
OTHERWISE IF CURRENT OR FORMER SMOKER, RANDOMLY
ASSIGN HALF TO GET TU-15 AND HALF TO GET TU-16.
OTHERWISE, GO TO TU-17.

PERSONAL RISK

TU-15. How likely do you think it is that the average {male/female} cigarette smoker will develop lung cancer in the future? Would you say that {his/her} chance is . . .

TU15ChanceLungCancerGeneral

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

GO TO TU-17.

TU-16. 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 . . .

TU16ChanceLungCancerSpecific

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

DETECTION/CURABILITY

TU-17. Overall, how many people who develop lung cancer do you think are cured? Your best guess is fine. Would you say . . .

TU17FreqCuredLungCancer

less than a quarter,	1
about a quarter,	2
about half,	3
about three-quarters, or	4
nearly all?	5

BOX TU-5

IF DO NOT SMOKE, CONTINUE.
IF CURRENT SMOKER, RANDOMLY SELECT HALF TO ANSWER
TU-18 AND HALF TO ANSWER TU-19.

TU-18. Would you say the average smoker has about the same lung cancer risk as a non-smoker, a little higher lung cancer risk than a non-smoker, twice the non-smoker's risk, 5 times the non-smoker's risk or 10 or more times the non-smoker's risk?

TU18RelChanceLungCancerGen

ABOUT THE SAME AS A NON-SMOKER,	1
A LITTLE HIGHER THAN A NON-SMOKER,	2
TWICE AS HIGH AS A NON-SMOKER,	3
5 TIMES HIGHER THAN A NON-SMOKER,	4
1O OR MORE TIMES HIGHER THAN A NON-SMOKER	5

GO TO NEXT SECTION.

TU-19. Would you say you have about the same lung cancer risk as a non-smoker, a little higher lung cancer risk than a non-smoker, twice the non-smoker's risk, 5 times the non-smoker's risk, or 10 or more times the non-smoker's risk?

TU19RelChanceLungCancerSpecifi

ABOUT THE SAME AS A NON-SMOKER,	1
A LITTLE HIGHER THAN A NON-SMOKER,	2
TWICE AS HIGH AS A NON-SMOKER,	3
5 TIMES HIGHER THAN A NON-SMOKER,	4
1O OR MORE TIMES HIGHER THAN A NON-SMOKER	5

This page left blank intentionally.

FRUITS AND VEGETABLES (FV)

These questions are about how often you ate or drank different kinds of foods during the past month, 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.

FV-1. During the past month, how often did you eat fruit? Include fresh, canned, or frozen fruit.
[ENTER NUMBER. IF NEVER, ENTER 95]

|___|___|___| |___|
TIMES UNIT

FV1Fruits

[ENTER UNIT.]

FV1Fruits_Unit

PER DAY 1
PER WEEK 2
PER MONTH 3
NEVER 95

FV-2. During the past month, how often did you drink 100% fruit juice? Include orange, apple, and grape juices. Do not include fruit drinks like Kool-Aid or Hi-C.
[ENTER NUMBER. IF NEVER, ENTER 95]

FV2FruitJuice

|___|___|___| |___|
TIMES UNIT

FV2FruitJuice_Unit

[ENTER UNIT.]
PER DAY 1
PER WEEK 2
PER MONTH 3
NEVER 95

FV-3. During the past month, how often did you eat vegetables? Include things like salad, cooked dried beans, corn, and broccoli.
[ENTER NUMBER. IF NEVER, ENTER 95]

FV3Vegetables

|___|___|___| |___|
TIMES UNIT

FV3Vegetables_Unit

[ENTER UNIT.]
PER DAY 1
PER WEEK 2
PER MONTH 3
NEVER 95

FV-3a. During the past month, how often did you eat potatoes? Do not include things like fried potatoes, french fries or rice.
[ENTER NUMBER. IF NEVER, ENTER 95]

FV3aPotatoes

|___|___|___| |___|
TIMES UNIT

FV3aPotatoes_Unit

[ENTER UNIT.]
PER DAY 1
PER WEEK..... 2
PER MONTH 3
NEVER 95

FV-4. 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.]

FV4NumberServings

|___|___|
SERVINGS

EXERCISE (EX)

The next few questions are about exercise, recreation, physical activities, or anything you do each day to increase the amount that you move other than during your regular job duties.

EX-1. During the past month, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?

EX1AnyExercise

YES	1	
NO	2	(EX-3)

EX-2. At least once a week, do you engage in regular activity such as brisk walking, jogging, bicycling, or another activity long enough to work up a sweat?

EX2Sweat

YES	1	
NO	2	

EX-3. Can exercise help to lower the chances of getting some types of cancer or does exercise not make much difference?

EX3ExerciseLowerCancer

LOWERS CHANCE OF CANCER	1	
DOESN'T MAKE DIFFERENCE	2	(NEXT SECTION)

EX-4. Which type or types of cancers, or don't you know?
[CODE ALL THAT APPLY.]

EX4ExerciseLowerWhichCancer

ALL TYPES OF CANCER	10
BLADDER 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
LEUKEMIA/BLOOD CANCER	17
LUNG CANCER	18
LYMPHOMA	19
MELANOMA	20
OTHER SKIN CANCER	21
ORAL CANCER	22
OVARIAN CANCER	23
PANCREATIC CANCER	24
PHARYNGEAL (THROAT) CANCER	25
PROSTATE CANCER	26
RECTAL CANCER	27
RENAL (KIDNEY) CANCER	28
BONE	29
STOMACH	30
HEART	31
BRAIN	32
INTERNAL ORGANS	33
MOST TYPES OF CANCER	34
THYROID	35
OTHER (SPECIFY)	91

ExerciseLowerWhichCancer_OS

This page left blank intentionally.

OVERWEIGHT/OBESITY (HW)

The next questions are about your health now. Please try to answer them as accurately as you can.

HW-1. About how tall are you without shoes?
[ENTER FEET.]
HW1Height_Feet
|___|___|
FEET

[ENTER INCHES. ROUND FRACTIONS OF INCHES DOWN TO WHOLE INCH.]
HW1Height_Inches
|___|___|
INCHES

HW-2. About how much do you weigh without shoes?
[ROUND FRACTIONS UP TO WHOLE NUMBER.]
HW2Weight
|___|___|___|
POUNDS

HW-3 WAS DELETED

This page left blank intentionally.

HEALTH STATUS & DEMOGRAPHICS

This page left blank intentionally.

HEALTH CONDITION

HS-1. In general, would you say your health is...

HS1GeneralHealth

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

DEPRESSION

Now, I am going to ask you some questions about feelings you may have experienced over the past 30 days.

HS-2. 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 HS2aSad	1	2	3	4	5
b.	nervous HS2bNervous	1	2	3	4	5
c.	restless or fidgety HS2cRestless	1	2	3	4	5
d.	hopeless HS2dHopeless	1	2	3	4	5
e.	that everything was an effort HS2eEffort	1	2	3	4	5
f.	worthless HS2fWorthless	1	2	3	4	5

BOX HS-1

IF RESPONDENT HAD ANY OF THE FEELINGS IN HS-2 ALL, MOST, OR SOME OF THE TIME, CONTINUE.
OTHERWISE, GO TO HS-5.

HS-3. We just talked 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 . . .

HS3DepressionInterfere

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

HS-4 WAS DELETED

HEALTH INSURANCE

Now, I would like to ask you about your health care coverage.

HS-5. Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?

HS5HealthInsurance

YES	1
NO	2

I have a few final questions about you and your household.

DM-1. Are you currently . . .
DM1MainActivity

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-2. Are you . . .
DM2MaritalStatus

Married,	1
Divorced,	2
Widowed,	3
Separated,	4
never been married, or	5
a member of an unmarried couple?	6

DM-3. Are there any children in your household under the age 18?
DM3aChildrenUnder5Years

YES	1
NO	2

DM-4. Are you Hispanic or Latino?
DM4Hispanic

YES	1
NO	2

DM-5. Which one or more of the following would you say is your race? Are you . . .
DM5Race

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

American Indian or Alaska Native,	10
Asian,	11
Black or African American,	12
Native Hawaiian or other Pacific Islander, or	13
White?	14

DM-6. What is the highest grade or year of school you completed?
DM6Education

NEVER ATTENDED SCHOOL OR ONLY ATTENDED KINDERGARTEN	1
GRADES 1 THROUGH 8 (ELEMENTARY)	2
GRADES 9 THROUGH 11 (SOME HIGH SCHOOL)	3
GRADE 12 OR GED (HIGH SCHOOL GRADUATE)	4
COLLEGE 1 YEAR TO 3 YEARS (SOME COLLEGE OR TECHNICAL SCHOOL)	5
COLLEGE 4 YEARS OR MORE (COLLEGE GRADUATE)	6

DM-7. Is your annual household income from all sources...

		YES		NO	
a.	less than $25,000? DM7aIncome25Thousand	1		2	(GO TO E)
b.	less than $20,000? DM7bIncome20Thousand	1		2	(END2)
c.	less than $15,000? DM7cIncome15Thousand	1		2	(END2)
d.	less than $10,000? DM7dIncome10Thousand	1	(END2)	2	(END2)
e.	less than $35,000? DM7eIncome35Thousand	1	(END2)	2	
f.	less than $50,000 ($35,000 to less than $50,000)? DM7fIncome50Thousand	1	(END2)	2	
g.	less than $75,000 ($50,000 to less than $75,000)? DM7gIncome75Thousand	1	(END2)	2	
h.	$75,000 or more? DM7hIncomeOver75Thousand	1		2	

END STATEMENT 2. Those are all of the questions that I have for you. Thank you for your time. If you have questions about cancer or want some information about cancer, you can call 1-800-4-CANCER or go to the National Cancer Institute's web site at: www.cancer.gov

APPENDIX B

Experiment on Effects of Incentives on Response Rates

In the HINTS study, we carried out an experiment on the effects of small monetary incentives on response rates. We experimented with two types of incentives: a 'pre-incentive' and a 'refusal conversion incentive', with the monetary amount in both cases being $2. In all cases, these incentives were only sent to the mailable numbers in the RDD sample, i.e., those telephone numbers for which we had address information. The pre-incentive was sent to households before the initial telephone contact. The refusal conversion incentive was only sent to households which had refused participation at least once. The design was a randomized design: the mailable numbers were randomly partitioned into four groupings, with the first 'NN' grouping receiving neither incentive, the 'YN' grouping receiving a pre-incentive only, the 'NY' grouping receiving a refusal conversion incentive only, and the 'YY' grouping receiving both a pre-incentive and a refusal conversion incentive (the latter only if they refused at least once).

The experiment was carried out on the whole of Wave 1: a total of 16,280 telephone numbers. Table 1 below presents the breakdown of Wave 1 into experimental groups. Note that 42.4% of the numbers in the sample were mailable (had addresses): a total of 6,905 telephone numbers. These were randomly assigned to the four experimental groups.

Table 1. Counts of telephone numbers in experimental groups.

Mailable status	Experimental group	Sample count	Percent of total
Mailable	YY	1,727	
Mailable	YN	1,726	
Mailable	NY	1,726	
Mailable	NN	1,726	
Mailable	Total	6,905	42.4%
Nonmailable		9,375	57.6%
Total		16,280	100.0%

At the end of the experiment , we computed (using the same response rate formulas) screener response rates, extended interview response rates, and overall response rates (screener response rate times extended interview response rate) for each of the experimental groups separately. These are given in Table 2 below.

Table 2. Response rates by experimental group.

Pre-incentive cell	Refusal conversion cell	Screener response rate	Extended interview response rate	Overall response rate
No pre-incentive	No RC incentive	57.7%	57.2%	33.0%
No pre-incentive	$2 RC incentive	57.8%	60.7%	35.1%
No pre-incentive		57.8%	59.0%	34.0%
$2 pre-incentive	No RC incentive	59.9%	64.0%	38.3%
$2 pre-incentive	$2 RC incentive	62.4%	65.0%	40.6%
$2 pre-incentive		61.2%	64.5%	39.4%

For screener response, extended interview response, and overall response we tested three null hypotheses on the two-by-two tables generated by the four experimental groups:

- No pre-incentive effect: response rate for the no pre-incentive cells equals the response rate for the $2 pre-incentive cells;
- No refusal conversion effect: response rate for the no refusal conversion cells equals the response rate for the $2 refusal conversion cells;
- No interaction: response rate for the off-diagonal cells equals the response rate for the on-diagonal cells.

We tested this by computing a contrast of response rates (two cells having a minus sign and two cells having a plus sign), generating an estimate of the difference in response rates, and the standard error of that difference. The test of significance was a two-sided test, assuming approximate normality. The standard error was computed using the replicate weights. Tables 3, 4 and 5 present the results for screener response rates, extended interview response rates, and overall response rates respectively.

Table 3. Two-sided tests of null hypotheses: screener response rates.

Null hypothesis	Estimated effect	Chi-square	Two-sided p-value
No preincentive additive effect	-3.4%	6.29	0.0121
No refusal additive effect	-1.3%	0.91	0.3397
No interaction effect	1.2%	0.76	0.3820

Table 4. Two-sided tests of null hypotheses: extended interview response rates.

Null hypothesis	Estimated effect	Chi-square	Two-sided p-value
No preincentive additive effect	-5.5%	7.98	0.0047
No refusal additive effect	-2.3%	1.35	0.2452
No interaction effect	-1.2%	0.41	0.5234

Table 5. Two-sided tests of null hypotheses: overall response rates.

Null hypothesis	Estimated effect	Chi-square	Two-sided p-value
No preincentive additive effect	-5.4%	14.42	0.0001
No refusal additive effect	-2.2%	2.32	0.1277
No interaction effect	0.1%	0.00	0.9598

As can be seen, the null hypothesis of no refusal conversion additive effect and of no interaction was accepted in all cases. The null hypothesis of no preincentive effect was rejected in all cases.

The experiment has yielded the significant result that a small monetary pre-incentive sent in advance of telephone contact has a positive effect on both screener response and extended interview response, for an effect on estimated effect on overall response of 5.4% (34.0% overall response in the no pre-incentive group and 39.4% overall response in the $2 pre-incentive group). The 95% confidence interval for this estimated effect of 5.4% is [2.6%,8.2%].