

## **B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS**

### **B.1 Respondent Universe and Sampling Methods**

The HINTS target population is all adults aged 18 or older in the civilian noninstitutionalized population of the United States. The sample design for HINTS 2007 consists of two samples with each sample being selected from a separate sampling frame. One sample will be a list-assisted random digit dial (RDD) sample selected from all telephone exchanges in the United States, following the design of HINTS 2003 and HINTS 2005. This will result in a nationally representative sample of telephone households. The second sample is new and comprises addresses selected from a list based on U.S. Postal Service (USPS) administrative records. Questionnaires will be mailed to the address sample, and telephone interviewers will follow up with mail nonrespondents. Appendix L discusses in detail the HINTS 2007 sample design.

For the RDD sample, an unstratified sample of 59,020 telephone numbers will be selected, with an expected yield of 3,500 completed interviews. A reserve sample of 29,510 telephone numbers will also be sampled and set aside to be used in case expectations are not met. During the administration of the RDD household screener, one adult will be sampled within each household and recruited for the extended interview. Refusals to the screener will be subsampled for refusal conversion. The expected overall response rate for the HINTS 2007 RDD sample is 21 percent. Section B.3 discusses the RDD response rates in previous HINTS cycles.

For the address sample, a stratified sample with total sample size of 6,944 addresses will be selected from two strata: one containing a high concentration of minority adults and the other containing a low concentration. The high-minority stratum will be oversampled by 50 percent to increase the yield of blacks and Hispanics. All adults in the household will be asked to complete a survey. Households that do not return one or more completed questionnaires will be assigned to telephone followup. The expected overall response rate for the address sample, including telephone followup, is 26 percent. The sample design will yield approximately 7,457 completed interviews: 3,500 from the RDD sample and 3,957 from the address sample. The expected survey-level response rate, including both RDD responses and address-sample responses, is  $7,457 / (3,500 / 0.21 + 3,957 / 0.26) = 23$  percent.

Section L.3 of Appendix L describes how data collected from both the RDD sample and the address sample will be used to calculate composite estimates. Our goal is to be able to generate 95-percent confidence intervals for composite estimates of proportions that are no wider than  $\pm 4$  percentage points, for the domain of all adults and for the domains of Hispanic adults and black adults. Table L-3 of Appendix L presents Current Population Survey (March 2005 supplement) estimates of adults within the domains of interest, with the expected number of completed questionnaires from the RDD sample. Table L-5 presents the expected the number of completed questionnaires from the address sample.

## **B.2 Procedures for the Collection of Information**

This section describes the data collection procedures to be used in HINTS 2007. The discussion includes stratification, sample selection, RDD-interviewer training and monitoring, RDD screening and extended-interview procedures, address-sample data collection procedures, and estimation.

### **B.2.1 Statistical Methodology for Stratification and Sample Selection**

#### **B.2.1.1 RDD sample**

The list-assisted RDD method is a random sample of telephone numbers from all working banks in U.S. telephone exchanges (Tucker, Casady, and Lepkowski, 1993). A *working bank* is a set of 100 telephone numbers (e.g., telephone numbers with area code 301 and first five digits 294-44) with at least one listed residential number.<sup>3</sup> The list-assisted method has been used in most RDD surveys in recent years. The within-household sample involves asking the respondent how many adults are in the household, identifying the adults in a nonintrusive way and then sampling one adult. HINTS 2007 will use the same approach for sampling one adult (18+) per eligible household that was developed for HINTS 2003 (Rizzo, Brick, and Park, 2004) and also used for HINTS 2005. Use of this screening methodology minimizes the number of screener questions that are asked of a respondent and also minimizes the intrusiveness of the questions for the majority of households, while still accomplishing a valid probability sample.

The RDD sample will involve two types of subsampling. First, 13.2 percent of the nonmailable numbers (numbers for which no address information can be obtained) will be deselected. These numbers will not be assigned to initial screening. The second type of subsampling is that 49.7 percent of the initial screener refusals and noncontacts (numbers for which there is human contact with nonhostile refusal or noncontact at the first round of screening calls) will be deselected. The first type of subsampling was conducted for HINTS 2005 but the second type of subsampling was not. Appendix L discusses these two types of subsampling in detail.

#### **B.2.1.2 Address Sample**

The sampling frame for the address sample will be a database of addresses used by Marketing Systems Group (MSG). Our decision to use this database as a sampling frame is the result of an evaluation study conducted by Link and colleagues (2005). The sampling unit will be an individual address. Following the selection of the address sample, the AUTOMATCH computer program will be used to compare the address sample with the addresses of telephone numbers assigned to the mailable stratum of the RDD sample. Addresses in both the address sample and the RDD sample will **not** be contacted by RDD telephone interviewers. (Telephone numbers assigned to the nonmailable stratum will not have addresses,

---

<sup>3</sup> All numbers are part of the sampling frame, whether listed as residential or not, as long as they are in working banks.

so they cannot be tested for membership in the address sample.) Envelopes containing mailed questionnaires will be marked “Do Not Forward” so that address changes will not provide multiple opportunities for a household to be selected for the address sample.

Two strata will be used to sample the addresses—one with a high concentration of minority adults and the other with a low concentration. Appendix L provides additional details about strata construction and how the address sample will be allocated to the two strata. Unlike the RDD sample, all adults in the household at a sampled address will be asked to complete a questionnaire. Hence, the mail sample is a stratified cluster sample, in which the household is the cluster. Our decision to not subsample the adults in sampled households is the result of an evaluation study conducted by Battaglia and colleagues (2005).

### **B.2.2 Problems Requiring Special Sampling Procedures**

No unusual problems requiring special sampling procedures have been identified.

### **B.2.3 Periodic Data Collection to Reduce Burden**

The National Cancer Institute (NCI) has elected to conduct HINTS biennially, rather than annually, to reduce costs and respondent burden.

### **B.2.4 RDD Interviewer Training and Monitoring**

If possible, interviewers will be selected from Westat’s current pool of interviewers; however, additional interviewers may be hired as needed. Any new hires will participate in Westat’s general interviewer training and basic computer-assisted telephone interview (CATI) training. All interviewers will be required to successfully complete Westat Telephone Research Center’s (TRC’s) new automated contact procedures module, which trains and tests interviewers on contact procedures and coding. Experienced interviewers usually complete this program very quickly, leaving more time to focus on project-specific issues. The remainder of training will focus on information specific to HINTS 2007. Training will include modeling interviews and role plays. Westat staff will carefully monitor the performance of all interviewers before allowing them to begin interviews with sample persons and will provide additional training as needed. We estimate that HINTS 2007 training will require approximately 12 hours.

During the pilot of the mail and RDD survey and the first few weeks of the field period, TRC team leaders and other project staff will focus on monitoring activities of interviewers to identify any problems or a need for retraining. Team leaders will also evaluate the interviewers in terms of their refusal avoidance abilities. Informal meetings of interviewers and team leaders will be held to discuss reasons for nonresponse and to disseminate the nonresponse conversion or avoidance strategies that have been most successful in converting HINTS refusals. Interviewers who are most successful in nonresponse

conversion will be assigned to a special team that will focus primarily on refusal conversion work. In addition, a percentage of live interviews will be monitored during the field period in accordance with Westat's standard operating procedures to provide interviewers with feedback on their performance and to provide additional training as necessary.

#### **B.2.5 RDD Screening Procedure**

The RDD sample will be address-matched using the commercial services of Telematch and Acxiom so that advance letters (see Appendix M) can be sent to potential respondents. A prepaid incentive of \$2 will be included with this letter, as indicated from the results of the HINTS 2003 experiment, as well as other research conducted by Westat (Cantor, et al., 2003). It is anticipated that between 45 percent to 60 percent of the sampled telephone numbers will have an address match.

All screeners will be administered over the telephone using CATI. The objectives of the screening interview are to find residential households and to select eligible persons for the extended interview. The screener asks the respondent how many adults live in the household, identifies the adults in a nonintrusive way (i.e., avoids asking for names), and then samples one adult. The screening method minimizes the number and intrusiveness of screener questions required to sample a member of the household. Some RDD households will not be reached for screening and others will refuse to participate. We will send a refusal conversion letter (see Appendix M) to the households among this group for which we have addresses. This letter will explain the purpose of the study and the importance of their participation.

#### **B.2.6 RDD Extended-Interview Procedure**

The majority of the content required to answer the research questions for HINTS 2007 is contained in the extended interview. For the RDD sample, after a respondent is selected from the household, he or she will be asked to complete this portion of the interview. Some sampled persons selected during the screener will not be reached to complete the extended interview, and others will refuse to participate. Two weeks after initial contact, refusal conversion letters will be sent to people for whom we have address matches (see Appendix M). This letter will explain the purpose of the study and the importance of their participation. Also included in this letter will be a \$5 incentive to complete the survey. Then, an interviewer will follow up with a telephone call to attempt to do the interview. If this is not successful, a followup will be attempted 2 weeks later.

#### **B.2.7 Address-Sample Data Collection Procedures**

The address sample will be sent advance letters explaining the study (see Appendix N). One week after the advance letter is sent, the household will be mailed a packet containing the instruments. The advance letter will be short and concise.

Within 1 week of the advance letter, each sampled household will receive a packet of four identical questionnaires with a request that a questionnaire be completed by each adult member of the household. Letters and mailing labels will be computer-generated from a file containing identifying information. Identification labels (ID numbers) will contain a household identifier and an individual survey identifier. All mailed material will be marked “Do Not Forward.” If no completed questionnaires have been received from a household within 2 weeks of the mailing of the instruments, a reminder postcard will be sent to the household. If no completed questionnaires have been received from a household within 2 weeks of the mailing of the reminder postcard, replacement questionnaires will be mailed to the nonresponding questionnaire. Households that have not responded within 2 weeks of the second mailing of the instrument and for which we have a telephone number will be contacted by telephone to complete the CATI instrument. Procedures for selecting a household member to complete the CATI extended interview will be identical to those employed in the RDD section of the study.

### **B.2.8 Tracking Respondents**

A series of production and management reports will be generated daily and weekly during the field period. These reports provide information on response rates, cooperation rates, production to date in terms of total interviews, and problems encountered during the course of data collection. The results of these reports will be reported weekly to the Project Officer, as well as in the monthly project progress report. NCI will be provided both hard-copy and electronic copies of all reports. Additional reports will be provided as requested by NCI.

### **B.2.9 Spanish Interviews**

A Spanish version of the RDD instrument will be developed. RDD participants requesting telephone interviews in Spanish will be connected to a Spanish interviewer and administered the appropriate instrument.

A Spanish version of the mail survey will not be developed. The introductory letter sent before initial mailing and the refusal-conversion letter in the followup mailing will contain a sentence written in Spanish stating that if the selected household would like the instrument administered in Spanish they should call a 1-800 telephone number that will connect them to the Westat TRC. Households requesting a Spanish interview will follow the RDD protocol, where only one adult will be sampled and interviewed.

### **B.2.10 Estimation**

Composite estimation will be used to calculate estimates of means, proportions, and totals based on the HINTS data collected from both the RDD sample and the address sample. Section L.3 of Appendix L provides additional detail about the calculation of composite estimates and associated composite sampling

weights. Replicate weights for composite estimates will be produced that will allow for the computation of consistent variance estimators. These weights are based on the jackknife method, in which the sampled telephone numbers and addresses are assigned to groups based on the HINTS sample design, with each replicate weight corresponding to the dropping of one group. Nonresponse and poststratification adjustments will be replicated so that the jackknife variance estimator correctly accounts for these adjustments. Stratification information necessary to compute linearization variance estimates will also be available using software packages such as SUDAAN.

Our goal is to be able to generate 95-percent confidence intervals for composite estimates of proportions that are no wider than  $\pm 4$  percentage points, for the domain of all adults and for the domains of Hispanic adults and black adults. Table L-7 of Appendix L presents maximum standard errors and the associated half-widths of 95-percent confidence intervals for composite estimates of proportions in the race/ethnicity domains of interest. As can be seen, the standard errors are in the range 0.82 percent to 1.90 percent, giving 95-percent confidence intervals with half-widths in the range 1.61 percent to 3.72 percent, as desired.

### **B.3 Methods to Maximize Response Rates and Address Nonresponse**

This design recognizes the increasing difficulty with response rates, coverage and expense of RDD surveys. Response rates for RDD surveys have been declining for a number of years and there is no sign that this decline has stopped (e.g., Battaglia, et. al., 2005; Curtin, et al., 2005). In addition, the increasing penetration of mobile telephones is deteriorating the RDD sample frame and even making it more difficult to contact those with a landline telephone. This is especially true for younger respondents (Brick, et al., 2004). Previous administrations of HINTS reflect this trend. In 2003, the combined response rate was approximately 33 percent. This rate dropped to 20 percent in 2005. The primary reason for the drop was a 20-percentage point drop in the screening response rate (55% to 34%). The response rate for the extended interview remained stable at 60 percent across the two surveys.<sup>4</sup>

To compensate for nonresponse and coverage, both HINTS 2003 and 2005 have been adjusted for nonresponse and have been poststratified to national totals for age, gender, race/ethnicity, and education. When comparing the final HINTS estimates for health-related variables to the National Health Interview Survey, it is apparent that HINTS respondents report being less healthy, as indicated by higher rates of reported cancer, people rating themselves as not being in good health, and having more psychological and/or emotional issues (e.g., nervousness, hopeless, sad, restless).

---

<sup>4</sup> In 2003, the screener response rate was 55 percent and the extended response rate was 60 percent, resulting in a final response rate of 33 percent (.55 x .60). In 2005, the screener response rate was 34 percent and the extended response rate was 60 percent, resulting in a final response rate of 20 percent (.34 x .60).

In reaction to the above trend, HINTS 2007 has made two changes to its design. One change is the use of a dual-frame, dual-mode approach. In addition to the RDD survey, a mail survey will be conducted using the USPS sampling frame (see discussion in previous section). The intent is to directly address the increasing migration of landline telephone to mobile-only telephone households in a cost-effective manner. The poststratification factors for HINTS 2005 indicated a substantial shortfall for young people age 18 to 35. The use of a mail survey is seen as a way to address this segment of the population.

Recent research by Link and colleagues (2004, 2005, 2006) suggests that use of a mail survey, with appropriate followup, can achieve a higher response rate than RDD alone. This design has a number of advantages. One is that using two modes in sequence (mail and telephone) there is the potential for improving response rates over a design that exclusively relies on RDD. If the mail survey is substantially less expensive than the RDD, then future HINTS surveys can develop designs that optimally mix the different modes to maximize data quality within the HINTS cost constraints. The use of the USPS frame allows coverage of mobile-only telephone users, and those without a telephone. There is also the possibility of improved measurement for a number of characteristics (e.g., those subject to social desirability bias). Moving to an address-based frame also leaves open the opportunity to implement other modes if they are found to be appropriate (e.g., in-person; web).

The second change in design is to introduce several new methods to maximize the response rate for the RDD telephone survey. Among these methods is to assess the messages that HINTS has been using over the previous two administrations. One theme that has emerged from focus groups that have been conducted is that previous emphases on cancer-specific messages in letters and introductions may be one reason why young people, as well as more healthy people, are less likely to participate on the survey. As will be discussed below, Westat will use this qualitative research to refine the HINTS advance materials and introductions to appeal to a wider sector of the general population.

### **B.3.1 Maximizing Response Rates—RDD**

Nonresponse can occur at either the initial screening attempt or the subsequent interview attempt with a subsampled member of the household. Steps to minimize nonresponse are built into the study protocol. These steps include:

- **Household Advance Letters with a \$2 Incentive.** Advance materials with an incentive will be sent to all households for which an address can be obtained. The advance letters were developed through a series of focus groups and will be tested in a pilot survey (see Appendix M).
- **Experienced, Well-Trained Interviewers.** Interviewer training will focus on gaining cooperation in the first minute or so of the initial contact with a potential respondent. Emphasis will be placed on sending a general health survey message to potential respondents.
- **Effective Call Scheduling.** To maximize the contact rate, we will use a calling algorithm that handles all dimensions of call scheduling, including time zone (respondent and interviewer); skill

level of interviewer; special needs of the case (e.g., non-English language); call history; and priority of case handling.

- **Subsampling Refusals for Conversion.** To be able to focus resources on getting as high a response rate as possible, we will only follow up a subsample of those that refuse to complete the screening interview.
- **Refusal Conversion Letters.** For the subsample of the refusing households that are followed up (with an address available), a refusal conversion letter will be sent. The refusal letter will address some of the main reasons for refusals and will contain a \$5 incentive (see Appendix M).

From these efforts, we are projecting a 35-percent response rate for the RDD screener and a 60-percent response rate for the extended interview. This will result in a response rate of approximately 21 percent (.35 x .60). These response rates will be calculated as both weighted and unweighted rates.

### **B.3.2 Maximizing Response Rates—Mail**

Steps to minimize nonresponse are also built into the mail study protocol. As mentioned earlier, the study will take proactive measures to help ensure that high response rate goals are met. These include the following:

- **Household Advance Letters.** Advance materials will be sent to all households. The advance letters will describe the study's goals and objectives and will give assurances of confidentiality. Letters will be sent to households approximately 1 week before the household is mailed the survey (see Appendix N).
- **Multiple Followup for the Mail Survey.** If a survey is not received from a designated household 2 weeks after they are sent, a postcard reminder will be sent. If a survey has not been received 2 weeks after the postcard, a final remailing of the surveys will be sent. Pilot tests will be used to assess the effectiveness of a small incentive and/or use of express delivery for the final mailing.
- **Pilot Testing of the Mail Survey.** As discussed below under the pilot studies, an experiment will test the effectiveness of the above procedures, along with testing whether the length of the survey has an effect on the response rate. Results from this will be used as input into the final design of the instrument.

For the mail survey, we expect that 25 percent of the households will return at least one survey. An additional 12 percent of households that do not complete a mail questionnaire will complete a telephone interview. This would yield a final response rate from the mail frame of 26 percent. These rates are based on results from the recent pilot studies reported by Link and colleagues (2006), from which the procedures for HINTS have been adapted. Relative to the work by Link and colleagues, we have assumed a lower mail response rate. The response rate for the telephone followup is based on what is expected from the refusal conversion effort from the RDD survey.<sup>5</sup>

---

<sup>5</sup> Link and colleagues report a 28-percent response rate to their mail survey. We have assumed a 20-percent response rate (3,500/17,499; see discussion of mail survey in previous section). The final rate of 26 percent was computed by adding in the additional returns from the telephone followup of the mail survey nonrespondents (457 completes x 2.52 for subsampling one person per household = 1,152 weighted completes). The final response rate of 26 percent was estimated by [(3,500+1,152)/17,499] x 100.



### **B.3.3 Addressing Nonresponse**

Sample weights will be provided for each completed interview to allow for unbiased estimation of national percentages. The sample weights are products of the base weight, nonresponse adjustments, and a poststratification adjustment. The *base weight* is the reciprocal of the probability of selection of each sampled adult. The *nonresponse adjustments* are designed to reduce the potential bias caused by differences between the responding and nonresponding population and are equal to the reciprocals of weighted response rates within carefully selected response cells. The *poststratification adjustment* modifies the nonresponse-adjusted base weights to the most recent Current Population Survey totals of adults by race/ethnicity, age, region of the country, and other demographic factors. This adjustment has the effect of reducing variance. To understand the differences in response rates and coverage across the two frames, HINTS 2007 will compare response rates and respondent characteristics across the two frames. The specific analyses that we plan on conducting are provided in Appendix O.

### **B.4 Test of Procedures or Methods to be Undertaken Proposal**

For HINTS 2007, we are conducting the following activities to test the procedures for HINTS 2007.

#### **B.4.1 Pilot Mail Survey**

The mail survey will be piloted to test the instrument and the survey processes that will be used for the full field effort. The pilot will aid in understanding how the HINTS instruments and procedures for the mail mode will work in a national setting. The primary questions to be addressed by the pilot will be (1) what is the response rate for the mail survey; (2) how does the response rate vary by the length of the questionnaire, inclusion of an incentive, and the use of express delivery; (3) what types of respondents fill out the survey by demographic, socioeconomic and health-related characteristics; and (4) what is the data quality of the surveys, including the completeness of data and adherence to skip instructions? The pilot is scheduled to occur several months before the full data collection to allow time to make changes to the full study.

The pilot will have three embedded experiments to test procedures to be used on the main study: (1) the length of the survey (two conditions), (2) the use of a \$2 incentive at the second mailing (two conditions) and (3) the use of express mail at the second mailing (two conditions). The sample size of 640 households will detect observed differences of 5 percent to 6 percent as statistically significant when comparing within each experimental condition.

#### **B.4.2 Pilot Advance Materials**

These experiments will be carried out by calling households to complete a short HINTS interview that will take approximately 5 minutes. The interview will contain the screener and a small sample of

questions from the main HINTS interview. Two different introductions and letters will be tested against one another with samples of approximately 350 households in each experimental group (total sample size of 700). A similar design will be used to test two prenotification letters. Refusal conversion will not be done because the methodology is primarily concerned with obtaining initial cooperation within the household.

### **B.4.3 Field Experiments**

Within the mail mode of the field study, an experiment will be conducted to determine the effect of an incentive and the methods used to deliver the questionnaires at the last mailing. This will be a continuation of the experiment described above for the pilot study, except with a larger sample size. If the pilot provides definitive results, we will implement them in the field study.

The results of HINTS 2007 will be analyzed to assess the relative merits of the dual frame approach from both a measurement (e.g., response rate, data quality) and cost perspective. These analyses will also inform analysts on the best way to use the data to take advantage of the strengths of the design. The methodological analysis will focus on the following questions: (1) what are the differences in the coverage and unit response rate of the mail and telephone surveys; (2) what are the differences in the item response rates of the mail and telephone surveys; (3) do the two designs produce different estimates; (4) can differences in estimates be attributed to the frame, response rates or mode of interviewing (e.g., visual on mail survey versus aural on the telephone); and (5) what are the relative costs of conducting the mail and telephone surveys?

The dual-frame multimode design of HINTS 2007 allows one to address each of the above questions by directly comparing the mail and telephone surveys. This design also provides the ability to bridge the HINTS trends if significant mode effects are found. Appendix P provides a discussion of how the above research questions will be addressed, along with the methodology that will be used.

### **B.5 Individuals Consulted on Statistical Aspects and/or Analyzing Data**

A number of individuals were critical in developing the research plan, the conceptual framework, survey questions, and sampling strategies underlying HINTS. Many of the same individuals will be involved with analysis of HINTS data once those data are collected. A list of these individuals can be found in Appendix Q.