

Supporting Statement B For:

Health Information National Trends Survey V (HINTS V)

(NCI)

OMB No: 0925-0538, Expiry Date X/XX/XXXX

August 18, 2016

This submission is a Reinstatement with Changes.

Yellow Highlights indicate changes from the 2015 submission.

Bradford Hesse, Ph.D., HINTS Project Officer
Chief, Health Communication and Informatics Research Branch
National Cancer Institute

9609 Medical Center Drive, 3E610
Bethesda, MD 20892-9760
Telephone: 240-276-6721
E-mail: Hesseb@mail.nih.gov

Table of Contents

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 RESPONDENT UNIVERSE AND SAMPLING METHODS.....1

B.2 PROCEDURES FOR THE COLLECTION OF INFORMATION.....2

B.3 METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NONRESPONSE.....10

B.4 TEST OF PROCEDURES OR METHODS TO BE UNDERTAKEN.....11

B.5 INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS AND INDIVIDUALS COLLECTING
AND/OR ANALYZING DATA.....13

LIST OF APPENDICES

Appendix A:	History of HINTS
Appendix B:	Cover letters and FAQs
Appendix C:	HINTS IV Publications Overview
Appendix D:	Main Study-Draft Cycle 1 Instrument
Appendix E:	Pilot Test Instrument
Appendix F:	Healthy People 2020 Objectives and Sources
Appendices G:	G1: Privacy Impact Assessment (PIA) G2: Privacy Act Memo
Appendix H:	Illustrative List of OMB-Approved Incentives
Appendix I:	Westat Confidentiality Pledge
Appendices J:	J1: NCI OHSR Determination J2: Westat IRB Approval Letter
Appendix K:	Theoretical Framework
Appendix L:	Changes to HINTS instrument for Cycle 1
Appendix M:	References

B. STATISTICAL METHODS

B.1 Respondent Universe and Sampling Methods

The HINTS target population for regular cycles of data collection is all adults aged 18 or older in the civilian non-institutionalized population of the United States. The sample design for HINTS V consists of a series of three single-stage stratified samples of addresses selected from a file of residential addresses based on the United States Postal Service (USPS) Computerized Delivery Sequence File (CDSF). Each sample will be selected just prior to the data collection cycle in which it is to be used. The frame will cover addresses from all zip codes in the 50 states and the District of Columbia.

Addresses in the frame will be grouped into two strata: one containing a high concentration of minority adults and the other containing a low concentration. The number of addresses to be sampled at each cycle is 13,330 with an expected yield of 3,500 completed interviews. The addresses in the high minority stratum will be oversampled by a factor of 2. One adult will be sampled within each household, using the next Birthday method, and recruited for the extended interview. The expected overall response rate for each HINTS V cycle is 34 percent, 26 percent from the high minority stratum and 37 percent from the low minority stratum. These rates are approximately the rates achieved for HINTS IV Cycle 4. Across the three HINTS V cycles, we expect to sample 40,000 addresses and complete 10,500 interviews.

In addition, HINTS V will include a pilot study of a sampling strategy for identifying smokers. The respondent universe and the sampling frame for the pilot study will be same as that for the regular cycles of data collection. The sample design consists of a single-stage stratified sample of addresses where the frame will be grouped into four sampling strata based on county-level smoking rates¹ (high, medium-high, medium-low, and low). The number of addresses to be sampled is 5,741 with an expected yield of 1,652 completed interviews. The high and the medium-high strata will be oversampled by 90 percent and

¹ The county-level smoking rates are based on the 2003 BRFSS small area estimates adjusted by the ratio of the 2011 to the 2003 Behavioral Risk Factor Surveillance System (BRFSS) state smoking rates so that when county rates are aggregated to the state level they are in agreement with the 2011 BRFSS state-level smoking estimates.

45 percent, respectively, to increase the yield of current smokers. One adult will be sampled within each household, using the next Birthday method, and recruited for the extended interview. The expected overall response rate expected for this round of HINTS is 33 percent, which is approximately the rate achieved for HINTS 4 FDA.

B.2 Procedures for the Collection of Information

Statistical Methodology for Stratification and Sample Selection of the Cycles

The sampling units for HINTS V will be household addresses that receive mail. The sampling frame will be a database of addresses used by Marketing Systems Group (MSG) to provide random samples of addresses. All non-vacant residential addresses in the United States present on the MSG database, including post office (P.O.) boxes, throwbacks (i.e., street addresses for which mail is redirected by the United States Postal Service to a specified P.O. box), and seasonal addresses will be subject to sampling. Two strata will be created for the sampling of addresses – one containing a high concentration of minority adults and the other containing a low concentration. The purpose of creating high- and low-minority strata and then oversampling the high-minority stratum is to increase the precision of estimates for minority subpopulations. The increases in precision result from the increase in sample sizes for the minority subpopulations produced by the oversampling.

The two strata will be formed by first using demographic data from the American Community Survey (ACS) to determine the population percentages of Hispanics and African Americans for individual U.S. Census tracts. Addresses will then be matched to Census tracts by their nine-digit ZIP Code. Addresses in Census tracts that have a population proportion for Hispanics or African Americans equaling or exceeding 34 percent will be assigned to the high-minority stratum. All other addresses will be assigned to the low-minority stratum. This stratification procedure is the same as that used to stratify the HINTS III and IV address samples and is described in more detail in Norman and Sigman (2009). A profile of the sampling strata is shown in Table B2-1.

Table B2-1. Profile of the sampling strata

Stratum	Proportion of frame (percent)	Coverage of African Americans and Hispanics (percent)	Prevalence of African Americans or Hispanic in stratum (percent)
High-minority	27.3	69.6	64.3
Low-minority	72.7	30.4	11.5

Each of the three samples will be selected just prior to the data collection cycle in which it is to be used. Each time a stratified sample is selected, an equal probability sample of addresses will be selected within each sampling stratum.

Table B2-2 contains the stratum allocations, assumed response rates, and expected number of completed questionnaires for all three cycles. Table B2-3 contains the expected number of completions by stratum and by analysis domains of interest for all three cycles. Due to the modular nature of HINTS (with core items asked in all cycles and other items asked in fewer cycles), Table B2-4 uses the results in Table B-3 to calculate maximum expected half widths of 95 percent confidence intervals for estimated domain proportions, when the total number of completes for an item that appears in all three cycles (10,500), in two cycles (7,000), and in either the first, second, or third cycle (3,500). Table B2-4 assumes that the design effect due to disproportional allocation, within-household correlation, and weighting adjustments is approximately equal to $1.0 + 1.3^2 = 2.7$, where 1.3 is the observed coefficient of variation of the final weights in the HINTS IV address sample.

Table B2-2. Stratum allocations, assumed response rates, and expected completions for all three cycles

	Total	High-minority stratum	Low-minority stratum
Allocation rate of sample to strata	100%	54.6%	45.4%
Number of sampled addresses	40,070	25,590	14,480
Assumed undeliverable rate ¹	--	13.5%	11.5%
Number of deliverable addresses	34,950	22,135	12,815
Assumed household response rate ¹	--	25.9%	37.2%
Number of completed questionnaires	10,500	5,730	4,770

¹ Calculated from HINTS IV cycle 4 data

Table B2-3. Expected number of completes by stratum and analysis domains of interest for all three cycles

Stratum	Analysis domain	Proportion of stratum (percent) ²	Completed questionnaires
High-minority	Hispanic	34.7	1,989
	Black	29.6	1,697
	Non-Hispanic/Non-Black	35.7	2,047
	All	100.0	5,733
Low-minority	Hispanic	6.6	315
	Black	4.9	234
	Non-Hispanic/Non-Black	88.5	4,219
	All	100.0	4,767

² Based on 2010-2014 ACS data.

Table B2-4. Expected half widths of 95 percent confidence intervals for estimated proportions in race/ethnicity domains of interest

Total number of completed questionnaires	Half width of 95 percent confidence intervals about estimated domain proportions of p=50 percent			
	Hispanics (percent)	African Americans (percent)	Non-Hispanic Whites and other races (percent)	All adults (percent)
10,500	3.35	3.66	2.03	1.57
7,000	4.11	4.49	2.49	1.92
3,500	5.81	6.35	3.52	2.72

Statistical Methodology for Stratification and Sample Selection of the Pilot Study

As with the regular data collection cycles of HINTS V, the sampling units for the pilot study will be household addresses that receive mail. The pilot study will also use the same sampling frame as the regular HINTS sample. The pilot study will stratify the sampling frame into four separate strata based on county-level smoking rates. The pilot study will use the same sampling strata as used for the HINTS IV FDA sample. The four strata will be formed by first using small area estimates of smoking rates at the county level for the 2000 – 2003 time period. Addresses will then be matched to Census Bureau counties by their FIPS Code. Addresses in Census Bureau counties that have high smoking rates (equaling or

exceeding 25.1 percent) will be assigned to the first stratum. Addresses in Census Bureau counties that have medium-high smoking rates (between 21.2 and 25.0 percent) will be assigned to the second stratum. Addresses in Census Bureau counties that have medium-low smoking rates (between 15.0 and 21.1 percent) will be assigned to the third stratum. All addresses in the remaining counties with low smoking rates (less than 15.0 percent) will be assigned to the fourth stratum. A profile of the sampling strata is shown in Table B2-5.

Table B2-5. Profile of the sampling strata

Stratum (smoking rate)	Proportion of frame (percent)	Coverage of Smokers in stratum ¹ (percent)	Prevalence of smokers in stratum ¹ (percent)
High	17.1	24.4	25.0
Medium-high	24.6	30.4	21.6
Medium-low	45.1	38.6	10.9
Low	13.2	8.8	7.1

¹ Calculated from HINTS IV FDA cycle data.

The sample will be selected just prior to data collection. An equal probability sample of addresses will be selected within each sampling stratum.

Table B2-6 contains the stratum allocations, assumed response rates, and expected number of completed questionnaires. Response rates were computed using results from the HINTS IV FDA cycle. Table B2-7 contains the expected number of current smokers who complete a questionnaire by stratum.

Table B2-6. Stratum allocations, assumed response rates, and expected completions for all three cycles

	Total	Stratum (smoking rate)			
		High	Medium-high	Medium-Low	Low
Allocation rate of sample to strata	100%	33.7%	36.5%	29.0%	2.4%
Number of sampled addresses	5,741	1,935	2,093	1,662	137
Assumed undeliverable rate ¹	--	15.8%	14.9%	11.8%	9.3%
Number of deliverable addresses	5,000	1,628	1,780	1,467	125
Assumed household response rate ¹	--	34.6%	32.5%	33.0%	33.2%
Num. of completed questionnaires	1,652	563	578	484	41

¹ Calculated from HINTS IV cycle 4 data

Table B2-7. Expected number of completes by stratum and analysis domains of interest

Stratum (smoking rate)	Proportion of stratum (percent)	Expected Number of Current Smokers
high	25.0	141
medium-high	21.6	125
medium-low	10.9	53
low	7.1	3
All	14.9	321

Data Collection Procedures

Each data collection cycle as well as the pilot will follow a standard mailing protocol. All households in the sample will receive a packet requesting that one questionnaire be completed and returned in the postage-paid return envelope. A \$2 incentive will also be included with the mailing. All mailed materials will be marked “Do Not Forward.” If no survey has 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 surveys have been received within 2 weeks of the mailing of the reminder postcard, a second mailing will be sent. A third mailing will be sent to households that do not respond to the first two mailings. Please see Appendix B for copies of the cover letters and postcard.

Once a household has returned a questionnaire, it will not receive further mailings. If a package is returned as nondeliverable, the household will be removed from future mailings.

Helpdesk Assistance. Respondents will be provided with two toll-free numbers to reach project staff. The primary toll-free number will be provided on all letters and instruments for respondents to call and ask questions about the study or request additional/replacement questionnaires. The other number will be monitored by Spanish-speaking project staff to allow Spanish-speaking respondents to ask questions or request a mailing of the materials in Spanish. All English materials will include reference to the Spanish toll-free number.

Monitoring. A series of production and management reports will be generated daily and weekly during the field period. These reports will provide information on response rates, cooperation rates, and problems encountered during the course of data collection. Reports tracking the data collection process, documenting problems encountered, and offering resolutions or necessary revisions to the process will be prepared on a weekly basis during the field period.

Scanning. Returned hard-copy forms will be scanned using high-speed scanners. Receipt and scan staff will follow written project procedures developed for the handling of incoming hard-copy forms. A supervisor will review any forms that require special handling, for example, if any are too damaged to be scanned as returned.

Estimation

Sample weights and replicate weights will be calculated for each data collection cycle and for the pilot study. Sample weights will permit data users to calculate nationally representative estimates of the population of interest--that is, the adult (18+) non-institutionalized population in the United States--from the collected data. Replicate weights will allow users to compute standard errors for the estimates from the collected data. To analyze data across cycles, data analysts can combine the sample and replicate weights by using the procedure documented in the report (Rizzo, et.al. 2008), which describes how to analyze integrated data from the 2003 and 2005 HINTS surveys. Although this report describes integrating across years, the methodology applies to integrating across data collection cycles as well.

The goal of weighting is to correct the final estimates for nonresponse and noncoverage biases. The same weighting procedure will be used for each data collection cycle and for the pilot study. Weighting will consist of the following steps:

1. Calculating household-level base weights;
2. Adjusting for multiple ways that a household can receive mail;

3. Adjusting for household nonresponse;
4. Calculating person-level initial weights;
5. Calibrating the weights to population counts (also known as control totals).

The initial step in calculating weights is to attach a household-level base weight to each record in the file. The household base weight is the reciprocal of the probability of selecting the household for the survey. Note that if two different addresses would have led to the same household – for example, if a household receives mail via both a street address and a post office box – that household has twice the chance of selection of a household with only one address (and should therefore receive half the normal weight). Thus, an initial adjustment will be made to the base weights of households that have multiple ways of receiving mail (as determined by the answers to a survey question about this).

Next, adjustments for household nonresponse will be made within adjustment cells defined by characteristics that are known for all households in the survey, such as the sampling stratum, U.S. Census Bureau region and, as recommended by Norman and Sigman (2009), the United States Post Office classification of a household's type of mail delivery. A nonresponse adjustment factor will be calculated for each cell as the ratio of the sum of household weights for all eligible households to the sum of the household weights for all responding households. The nonresponse adjustment factor will then be applied to the household weight of each responding household. In this way, the weights of the responding households are “weighted up” to represent the full set of responding and nonresponding households in the adjustment cell.

Each sampled adult in responding households will be assigned an initial person-level weight. The initial person-level weight is calculated by multiplying the nonresponse-adjusted household weight by the reciprocal of the sample person's within-household probability of selection. Since only one adult is

selected from a household, the initial weight for the sampled adult is equal to the nonresponse-adjusted weight times the number of eligible adults in that household. For example, if a household contains three adults and only one adult was selected, the initial weight for the selected adult is equal to the nonresponse-adjusted household weight times three.

Finally, the person-level weights will be adjusted so that weighted counts from the survey match known national totals for selected demographic and health-related variables. The demographic variables will include age, gender, race/ethnicity, and educational attainment. The health-related variables will include health insurance status and cancer diagnoses. This is the same set of variables used for HINTS III and IV. The American Community Survey will be the source of the control totals for demographic variables, and the National Health Information Survey will be the source of control totals for health-related variables. If the survey data differ across categories of one or more of the calibration variables, then calibrating the weights in this way can reduce the variance of resulting estimates. More importantly, calibration will help to compensate for any noncoverage of the address frame, such as for rural areas with simplified addresses that cannot be used for sampling, or for nonresponse bias that is not adjusted for by the nonresponse adjustment procedures performed prior to calibration. As was done for the HINTS IV weighting, the calibration adjustments will be carried out using a raking procedure.

For each set of sample weights, a set of replicate weights will also be created to allow users to compute variances of survey estimates and to conduct inferential statistical analyses. Replication methods work by dividing the sample into subsamples (also referred to as replicates) that mirror the sample design. A weight is calculated for each replicate using the same procedures as used for the sampling weight. That is, the nonresponse and calibration adjustments will be replicated so the jackknife variance estimator correctly accounts for these adjustments. The survey estimate that is calculated for each replicate and variation among the subsample replicates is then used to estimate the variance for the survey estimates. HINTS V will generate replicate weights using the jackknife procedure, in which sampled households are

formed into groups reflecting the sample design and each replicate weight corresponding to dropping one group. The replicate weights can be used with a software package, such as WesVar, SUDAAN, STATA or SAS, to produce consistent variance estimators for totals, means, ratios, regression coefficients, logistic regression coefficients, etc.

In case users are interested in calculating variances using the software package like SUDAAN or SPSS which uses linearization variance estimation procedures, the necessary stratification information will be made available as well.

B.3 Methods to Maximize Response Rates and Address Nonresponse

To compensate for nonresponse and coverage, the estimates will be adjusted for nonresponse and will be poststratified to national totals for age, gender, race/ethnicity, education, health insurance status and cancer diagnosis. This same set of variables was used for HINTS IV. The national totals for health insurance status and cancer diagnosis will be taken from the National Health Interview Survey. These are used based on the observation from prior HINTS surveys that non-respondents tend to be healthier than respondents (Cantor, 2009). Post survey analysis will examine the characteristics of respondents by the relative timing of the returns. For example, methodologists will compare respondent characteristics of early returns received soon after the first mailing compared to those responding near the end of the data collection period to assess the extent to which the mailing strategy successfully engaged the cooperation of different demographic groups.

Steps to minimize nonresponse are 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:

- **Multiple Followups.** 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 second remailing of the surveys will be sent using Priority Mail. **If a survey is still not received, a third survey will be sent.**

■ **Use of \$2 incentive.** As discussed in Part A, we will include a \$2 incentive when the questionnaire is mailed to the household. Prior experiments on HINTS have shown this to have an impact on response rates.

These procedures to minimize nonresponse were used in HINTS IV and produced response rates of 35-40%².

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 person-level weights to the most recent ACS totals of adults by race/ethnicity, age, region of the country, and other demographic factors. This adjustment has the effect of reducing variance.

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

HINTS has traditionally included methodological substudies to guide future data collection efforts. HINTS IV conducted substudies on within-household respondent selection, survey cover variations, survey question formatting options, and multiple variations on the mailing of Spanish materials to potentially Spanish-speaking households. HINTS V will likely include methodological substudies as part of each cycle of data collection. For example, HINTS V may consider testing two or more variations of questionnaire items. The objective is to improve reliability and validity of the data, as well as to simplify questions to reduce burden. We illustrate below examples of questionnaire design issues that would be amenable to a field experiment. This list is not meant to be either definitive or exhaustive. It is intended to provide concrete illustrations of how field experiments could be used to advance the HINTS

² The response rates were based on the AAPOR formula that counts partial interviews as completes and includes interviews, non-interviews and all eligible unknown cases in the denominator (RR2, AAPOR).

research agenda. At the time the OMB package is submitted for particular cycles, the specific experiments that are planned will be submitted.

- **Question Wording.** For any type of self-administered questionnaire, there is a tension between being precise and keeping items simple. Precision usually requires providing more conditions and definitions to the respondent. One possible type of experiment would be to compare alternative wordings, one using precise terminology and the other using more simplified language.

We expect that HINTS will be developing new items related to knowledge, attitudes and behaviors related to health communications. For example, development of scales related to opinions about different cancer communication methods, using a series of items might be created, with alternative wordings resulting from the initial questionnaire development process. these alternatives could be compared in a field test.

- **Open vs. Closed-ended Questions.** HINTS has usually contained a number of questions that included a relatively long list of response alternatives, including where individuals went for health information, what type of cancer the person had, and hearing of cancer tests. Similarly, HINTS has included items with ordinal response categories that asked “how long ago” or “when in the future” something (might) happen. The form of these response alternatives may have an effect on estimates (Schwarz, et al, 1985).
- **Use and Placement of Definitions.** Inevitably, there are technical terms or concepts that cannot be communicated by the question itself. On HINTS IV, for example, the nutrition section included highly visible definitions of the serving sizes. On HINTS III, definitions were provided for stool blood occult tests, sigmoidoscopy and colonoscopy. Alternative forms and displays for these definitions could be tested to measure if respondents are using them.
- **Context and Order Effects.** Many of the items included on HINTS are attitudes, subjective assessments and estimates of “factual” items that are difficult to define (e.g., awareness; communication activities). These items are particularly subject to order and context effects (Tourangeau, et al., 2000). Experimentation might include testing for these types of effects on key HINTS items. With different combinations of items on different questionnaires, it might be important to measure if these have effects on the measures.

As described above, HINTS is also planning to conduct a pilot study of a sampling procedure designed to identify smokers. The results of this pilot will inform future HINTS sampling plans.

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

A number of individuals both within NCI and from other agencies/organizations were critical in developing the research plan, the conceptual framework, survey questions, and sampling strategies underlying HINTS. These individuals, who will also be involved in analysis, included:

NCI

Erik M. Augustson, Ph.D., MPH
Tobacco Control Research Branch
240-276-6774

Kelly Blake, ScD
Health Communication and Informatics Research Branch
240-276-6839

Sylvia Chou, PhD, MPH
Health Communication and Informatics Research Branch
240-276-6954

Robert T. Croyle, Ph.D.
Director, Division of Cancer Control and Population Sciences
240-276-6690

Bradford W. Hesse, Ph.D.
HINTS Project Officer
Chief, Health Communication and Informatics Research Branch
240-276-6721

Annette Kaufman, PhD, MPH
Tobacco Control Research Branch
240-276-6706

William Klein, PhD
Associate Director, Behavioral Research Program
240-276-6972

Benmei Liu, PhD
Statistical Methodology and Applications Branch
240-276-6718

Richard P. Moser, Ph.D.
Science of Research and Technology Branch
240-276-6915

Gordon Willis, PhD
Applied Research Program
240-276-6788

Other government agencies:

Vaishali Patel, PhD MPH
Office of Planning, Evaluation & Analysis
Office of the National Coordinator
Department of Health and Human Services
202-603-1239

David Portnoy PhD, MPH
Center for Tobacco Products
Food and Drug Administration
301-796-9298

Private Organizations:

Lila J. Finney Rutten, Ph.D., M.P.H.
Division of Epidemiology
Mayo Clinic
507-293-2341

Alexandra Greenberg
Division of Epidemiology
Mayo Clinic

Mark Savage
National Partnership for Women & Families
202-986-2600

Westat. The contractor conducting the data collection is Westat. The Westat employees who were consulted on statistical aspects of the design are:

David Cantor, Ph.D.
Principal Investigator
301-294-2080

Terisa Davis, M.P.H.
Project Director
301-294-2864

Lloyd Hicks, M.S.
Sampling Statistician
301-610-4960

Aaron Maitland, Ph.D.
Survey Methodologist
301-251-2299