Supporting Statement B For:

Health Information National Trends Survey 4 (HINTS4)

(NCI)

OMB No: 0925-0538, Expiry Date <mark>10/31/2014</mark>

<mark>October 2014</mark>

This submission is a Reinstatement with Changes.

<mark>Yellow Highlights indicate changes from the 2011 submission.</mark>

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

9609 Medical Center Drive
MSC 9761 Room 3E610
Rockville, MD 20850

Telephone: 240-276-6721
E-mail: Hesseb@mail.nih.gov

Table of Contents

LIST OF APPENDICES

Appendix A: History of the Health Information National Trends Survey (HINTS)

Appendices B: B1: Instrument in English

                          B2: Instrument in Spanish

Appendix C: List of HINTS Publications

Appendix D: Privacy Impact Assessment

Appendix E: Comparison of FDA-NCI Survey items to other surveys

Appendix F: List of people consulted

Appendix G: Privacy Act memo

Appendix H: Westat Confidentiality Pledge

Appendix I: NCI OHSR Exemption and Westat IRB Approval Letter

Appendix J: Cover letters and FAQs in English

Appendix K: Cover letters and FAQs in Spanish

Appendix L: New questions to HINTS

Appendix M: References

B.   STATISTICAL METHODS

B.1   Respondent Universe and Sampling Methods

The HINTS target population is all adults aged 18 or older in the civilian non-institutionalized population of the United States. The sample design for HINTS 4 consists of a single-stage stratified sample of addresses selected from a file of residential addresses based on the United States Postal Service (USPS) Computerized Delivery Sequence File (CDSF). The sample will be selected just prior to the data collection 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.   Iannocchione (2011) discusses how improvements in the quality of the CDSF results in nearly perfect coverage rates in more urban areas, though rural areas, and addresses with only P.O. box mail delivery have slightly lower coverage rates. However, the author also notes that for most mail surveys, over-coverage is more of a concern than under-coverage, referencing previous HINTS research by Norman and Sigman (2009) that showed that households with P.O. box addresses had a mean of 1.24 ways to receive mail.

Addresses in the frame will be grouped into four sampling strata based on county-level smoking rates[1] (high, medium-high, medium-low, and low). The number of addresses to be sampled is 13,000 with an expected yield of 4,318 completed interviews. The high and the medium-high strata will be oversampled by 60 percent and 20 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

---

[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.

for this round of HINTS is 33 percent, which is approximately the rate achieved for HINTS 4, Cycle 1.

B.2   Procedures for the Collection of Information

Statistical Methodology for Stratification and Sample Selection

The sampling units for HINTS 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. Four strata will be created for the sampling of addresses based on county-level smoking rates (Section B.1 above).

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 B2-1.

Table B2-1. Profile of the sampling strata

Stratum (smoking rate)	Proportion of frame (percent)	Current Smokers[1]		Former Smokers[1]		Ever Smokers[1]	
		Coverage (percent)	Prevalence (percent)	Coverage (percent)	Prevalence (percent)	Coverage (percent)	Prevalence (percent)
high	17.1	24.4	19.8	16.8	26.6	19.4	46.8
medium-high	24.6	28.2	16.8	24.7	28.7	25.8	45.8
medium-low	45.1	38.6	11.8	43.7	26.1	41.9	38.2
low	13.2	8.8	8.9	14.8	29.2	12.9	38.7

[1] Calculated from HINTS 4 Cycle 1 & 2 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 B2-2 contains the stratum allocations, assumed response rates, and expected number of completed questionnaires. Response rates were computed using HINTS 4 data for Cycles 1 and 2. Table B2-3 contains the expected number of completions by stratum and by analysis domains of interest.

Table B2-2.   Stratum allocations, assumed response rates, and expected completions[1]

	Total	Stratum (smoking rate)			
		High	Medium-high	Medium-low	Low
Allocation rate of sample to strata	100%	27.4%	29.5%	36.1%	7.0%
Number of sampled addresses	13,000	3,566	3,831	4,693	911
Assumed undeliverable rate	11.5%	14.8%	13.8%	10.5%	6.3%
Number of deliverable addresses	11,505	3,038	3,302	4,200	854
Assumed household response rate	37.5%	37.1%	38.4%	38.2%	37.3

					%
Number of responding households	4,318	1,127	1,268	1,605	318

[1] Calculated from HINTS 4, Cycle 1 data.

Table B2-3.   Expected number of completes by stratum and analysis domains of interest

Stratum (smoking rate)	Analysis domain	Proportion of stratum (percent)	Completed questionnaires
high	Current Smokers	19.8	223
	Former Smokers	26.6	300
	Never Smokers	53.2	600
	Ever Smokers	46.8	527
	All	100.0	1,127
medium-high	Current Smokers	16.8	213
	Former Smokers	28.7	364
	Never Smokers	54.2	687
	Ever Smokers	45.8	581
	All	100.0	1,268
medium-low	Current Smokers	11.8	189
	Former Smokers	26.1	419
	Never Smokers	61.8	992
	Ever Smokers	38.2	613
	All	100.0	1,605
low	Current Smokers	8.9	28
	Former Smokers	29.2	93
	Never Smokers	61.3	195
	Ever Smokers	38.7	123
	All	100.0	318
Total	Current Smokers	14.0	654
	Former Smokers	27.2	1,175
	Never Smokers	58.4	2,474
	Ever Smokers	41.6	1,844
	All	100.0	4,318

Data Collection Procedures

There will be four attempted contacts with the household. All households in the sample will receive the first mailing, while only non-responding households will receive subsequent mailings. Most households will receive one survey per mailing (in English), while households that are flagged as potentially Spanish-speaking will receive two surveys per mailing (one

English and one Spanish). This flag will be set for those households in linguistically isolated areas (as defined by the Census Bureau) and those with a Hispanic surname match.

A $2 incentive will also be included with the mailing. All mailed materials will be marked "Do Not Forward." If no surveys 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 surveys have been received within 2 weeks of the mailing of the reminder postcard, replacement questionnaires will be mailed to nonrespondents. Please see **Appendix J** for copies of the cover letters and postcard in English and **Appendix K** for the materials in Spanish.

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. 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.

The goal of weighting is to correct the final estimates for nonresponse and noncoverage biases. 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 previously used for HINTS. 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 previous HINTS weighting, it is anticipated that raking to control totals will be included rather than doing post-stratification.

In addition to the 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. Replicate weights for this HINTS sample will be

generated 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 Version 9.2 of 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**

As noted above, 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 previous HINTS data collections. 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 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, two follow-up surveys 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.
- **Use of USPS express delivery.** The second mailing of the questionnaire will be completed using express delivery. This has been found to increase the response to a mail survey (Dillman, et. al, 2009).

These procedures produced a response rate of, on average, 35%[2] for the first three cycles of HINTS 4.

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.

_____

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

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

The questionnaires have been pretested with 20 cognitive interviews for the English version and 10 cognitive interviews with the Spanish version (OMB #0925-0589-06, Exp. 4/30/2014). The procedures used to administer the survey are based on those for HINTS 4. Given this experience we do not believe there is a need to pre-test the procedures described above.

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

A number of individuals at NCI, FDA, and Westat were critical in developing the research plan, conceptual framework, survey questions, and sampling strategies underlying this round of HINTS. The list of the individuals can be found in **Appendix F**.