

Searching for Alternatives to a Random Digit Dial Telephone Interview – Redesigning the National Household Education Surveys

Chris Chapman

National Center for Education Statistics
Institute of Education Sciences

Mary Hagedorn

Westat, Inc.

FCSM Conference - Washington DC
November 4, 2009

This discussion is intended to promote the exchange of ideas among researchers and policy makers. The views expressed during discussion and in these handouts are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education.

I. Overview and Background of the National Household Education Surveys Program

Redesign efforts for the National Household Education Surveys Program (NHES) are the focus of this paper. To help readers understand why particular new design options are being pursued, we discuss the original purpose and design of NHES when it was planned and first fielded in the late 1980s and early 1990s. Next, a review of problems facing the original NHES design are presented, followed by a discussion of the solutions currently under consideration. The paper concludes with factors we will be considering in relation to pilot test results and how they relate to the 2011 field test. The pilot test is currently in the field.

The National Household Education Surveys Program (NHES) is intended to provide the United States Department of Education (ED) trend data on important topics that are best studied through surveys conducted directly with households. Prior to the development of NHES, ED's National Center for Education Statistics (NCES - the primary statistical agency within ED), did not have a data collection program in place that contacted households regularly. Most of the work done by NCES then and now involves surveys that start with samples of state and local education agencies, elementary and secondary schools, or colleges and universities. These surveys are central to NCES' mission, but are not well suited for studying many important education topics such as the education needs and activities of populations not attending traditional elementary and secondary schools, populations that are not attending colleges and universities, and topics better addressed through interviews with parents.

Prior to the development of NHES, NCES relied on items inserted into surveys conducted by other agencies to collect some data from households. Perhaps the best example of this kind of work is the School Enrollment Supplement to the October Current Population Survey (CPS). NCES paid for several items in this supplement every October, and continues to do so. However, the primary focus of surveys conducted by agencies outside of NCES, such as the CPS, is not on education topics. As such, NCES typically cannot develop extensive sets of questions since the questionnaires are largely composed of items from sponsoring agencies that address topics other than education.

Beginning in the early 1980s, NCES began exploring options for developing its own household survey that could be fielded on a regular basis. Various design approaches were considered and it was decided to move forward with a survey that would rely on telephone interviews with households. Sampling would rely on random digit dialing (RDD) techniques to select the numbers to be called, and interviews would be conducted by telephone interviewers with the assistance of computer assisted telephone interview (CATI) technology.

The decision to pursue this data collection model was based on several different factors. First, NCES wanted data from the survey to be used to develop trends lines on various populations and topics. As such, the cost of the survey could not be high or NCES could not afford the frequent data collections needed for trend line development. Second, NCES anticipated that the questionnaires themselves would need to be relatively complex to study the topics to be addressed by NHES. For example, NCES anticipated developing instruments to collect detailed information about the early educational experiences of preschool children including information on the range of early care and education providers with whom they interacted. Third, NHES would be used to study at least two populations simultaneously (the initial designs focused on preschool children and high school dropouts) covering a wide age spectrum, and would necessarily be composed of distinct questionnaires targeted for different samples.

Consideration was given to personal interviews because the interaction between interviewers and respondents would allow NCES to field relatively complex questionnaires. Interviewers could help address respondent questions and help them properly navigate the instruments. Personal interviews had relatively good response rates and procedures for conducting them were well developed. However, the costs of personal interviews were prohibitive given NCES' budget and the intent to use NHES to develop trend lines requiring annual or biennial data collections.

Design work also included consideration of using mail-out surveys as the mode of collection for NHES. Data collection costs were relatively low for this mode of data collection, meaning that NCES could afford this approach on a regular basis. However, mailing frames for national household surveys were not well developed at the time. In addition, NCES opted not to pursue this approach because of concerns about response rates and a desire to have some interviewer-respondent interaction to facilitate the use of complicated questionnaires.

To meet the goal of designing NHES to be conducted on a regular and frequent basis, with relatively low costs, and with relatively high response rates, NCES decided to pursue a telephone interview approach. Telephone interviewing had the advantage of an interviewer available to encourage participation, address respondent questions during the interview, and probe out-of-range responses. In addition, computer-assisted interviewing (discussed further in the next section) permitted the fielding of more complex instruments than self-administration would allow. Because the interaction would be by phone, NCES would not incur the prohibitive cost of sending interviewers to sampled households. As a result, NCES could afford to conduct the study every year to two. Telephone interviews also tended to have higher response rates than mail-out surveys at the time, and by the 1980s land-line telephones were in over 90 percent of U.S. households and approaches to limiting the remaining possible coverage bias were available (Massey and Botman 1988).

A large-scale field test was conducted in the fall of 1989 to determine if NHES would be successful as a telephone interview survey of households based on samples derived from land-line telephone number lists. Results indicated that the telephone interview approach provided sufficiently high response rates that the data could be used with limited concerns about possible nonresponse bias, provided nonresponse adjustments were applied during the weighting procedures (Brick, Celebuski, and Collins, 1992). Similarly, at least for children ages 3-5 and youth ages 14-21 (important populations for NHES), coverage bias was not a significant concern and could be corrected for in the weighing approaches available for the study (Brick and Burke 1992). Subsequent analyses of data collected through the first full-scale NHES data collection in 1991 provided similar results in terms of possible nonresponse bias, and in terms of possible coverage bias for the civilian, non-institutionalized adult population.

Trend line development from NHES began with the first national data collection in 1991.¹ That year, NCES used NHES to produce estimates for preschool children and children just entering into formal education in early grades, along with adults engaged in educational activities (for a list of reports generated from the NHES data collection series, please visit <http://nces.ed.gov/pubsearch/getpubcats.asp?sid=004>). With relatively minor adjustments to the sampling and data collection techniques used for the 1991 NHES collections, NHES was fielded approximately every other year from 1991 through 2007.

Sampling and Data Collection Procedures for NHES:1991 through NHES:2007

More detail about the specific sampling approaches and data collection techniques is needed to help understand changes being considered for NHES into the future. The initial sample design recommended by Westat was a variation of the Mitofsky-Waksberg RDD approach (Brick and Waksberg 1991). The approach was modified in 1995 to a list-assisted RDD approach (Collins, Brick, Loomis, Gilmore, and Chandler 1996). Both approaches were based on sampling land-line telephone numbers starting with frames of 100 banks with working phone numbers associated with them. A 100 bank is composed of numbers sharing the first eight digits of a telephone number: the three digit area code, three digit prefix, and the next two digits of a standard ten digit telephone number. As cell-phones became more prevalent, efforts were taken to remove telephone exchanges assigned to cellular phones with the intent being to not dial such phones. This decision was made primarily to avoid situations where respondents had to pay cell phone charges for our voluntary survey. As it became possible for consumers to migrate their telephone numbers when they changed services and service providers, some cell phones could appear in the samples, but interviewers were instructed not to conduct interviews on cell phones. The frame remained focused on land-line phone numbers.

Because some of the NHES topical surveys were designed to collect data about individuals who were relatively rare across households such as preschool children, decisions were made at the outset of the study design to bundle several topical collections during any given NHES collection. The purpose of bundling was to gain data collection efficiencies from the large number of household contacts needed to develop representative samples of relatively rare populations. If a person with the relatively rare characteristic did not exist in the household, the household might

¹ The topical focus shifted between the field test and the start of fully representative data collections in 1991. Several other topics, including adult education and parent and family involvement in education, replaced the focus on high school dropouts. Insufficient numbers of dropouts were captured in the field test to support nationally representative estimates without significantly larger sample sizes.

still be eligible for the other surveys in the NHES collection that year. Typically, at least one topical component focused on preschool children (relatively low percentages of households include such children) and this was combined with topical components focusing on elementary, middle, and high school students and adults.

What this means in operational terms is that the collections were typically done in two stages. First, a screening questionnaire was asked of a random adult household member. The screener collected information about household member age, school enrollment status, grade of enrollment for children, educational attainment, and several demographic characteristics. Within-household sampling to determine who would be the focus of second stage topical surveys was then done while the interviewer was on the phone with the screener respondent. To reduce burden, further subsampling was often implemented so that no more than two persons in a household were selected as subjects of topical surveys; it was possible for the same parent respondent to answer more than one interview if two children were selected.

Immediate within-household sampling was greatly facilitated by the use of computer-assisted telephone interviewing (CATI), which was used from the field test of NHES in 1989 through the most recent collection in 2007. Information provided by the screener respondent was entered into the CATI program by telephone interviewers and automated procedures programmed into CATI handled within household sampling. That is, CATI determined which topical interviews a household would receive based on eligibility of household members and applied the sampling algorithm to select person(s) within the household; this was accomplished while the interviewer was still on the phone with the screener respondent. By automatically executing within-household sampling during a single telephone call, CATI helped avoid errors in sampling that can occur when interviewers manually select second stage survey respondents. Response rates were also increased because we already had a respondent willing to complete interviews by phone talking with an interviewer, and could often complete the second stage topical interview with this person or another selected person without the need to try to call back into the household at a later time.

The use of CATI also made it possible to field questionnaires with a large number of complex skip patterns and reduced potential data entry errors associated with coders or scanners transferring respondent provided information into databases. By facilitating the use of complex skip patterns, CATI allowed NHES topical surveys to include a wide range of related topics that would not be suitable for all children in a particular research area of interest. For example, NHES:2007 included the Parent and Family Involvement in Education Survey (PFI). One important aspect of that topical survey was collecting information about school choice options considered and pursued by the sampled children's parents. The questions included a series of questions about homeschooling. Homeschooled children represent about 3 percent of all students in grades K-12 (Bielick 2009). Skip patterns allowed us to ask the homeschooling questions just of the homeschooling parents, and to avoid asking them questions specific to more traditional school settings.

An additional benefit of using CATI was that some interviews shared common sections (for examples, questions about household characteristics). The CATI programs kept track of which sections had been administered so that they could be skipped in a subsequent interview. This feature served to reduce redundancy and response burden when more than one household member was selected for a topical survey.

II. Increasing Problems with Telephone Interviews and RDD Sampling

Like most telephone surveys, NHES experienced declining response rates over time. Overall unit response rates for NHES started in the high 60 percent to high 70 percent range for the 1991 and 1993 collections and gradually declined over time through 2005 (U.S. Department of Education 1997, and Van de Kerckhove et al. 2008). Uncharacteristically sharp declines between 2005 and 2007 and the very low overall response rate experienced in 2007 are what ultimately led NCES to pursue a redesign of the NHES study. As response rates fell (table 1), increasingly expensive data collection options and data evaluation options were implemented to help check the slide of the response rates and to evaluate the potential for nonresponse bias.

Table 1. – Unit Response Rates and Overall Unit Response Rates for NHES: 1991 through 2007

Interview Time and Name	Unit Response Rate	Overall Unit Response Rate
NHES:1991		
Screeners	81.0	81.0
Early Childhood Education	94.5	76.5
Adult Education	84.7	68.6
NHES:1993		
Screeners	82.1	82.1
School Readiness	89.6	73.6
School Safety and Discipline (parents)	89.6	73.6
School Safety and Discipline (6 th – 12 th graders) ¹	83.0	68.1
NHES:1995		
Screeners	73.3	73.3
Early Childhood Program Participation	90.4	66.3
Adult Education	80.0	58.6
NHES:1996		
Screeners	69.9	69.9
Parent/Family Involvement in Education & Civic Involv.	89.4	62.5
Youth Civic Involvement ¹	76.4	53.4
Adult Civic Involvement	84.1	58.9
NHES:1999		
Screeners	74.1	74.1
Parent Interview	88.3	65.4
Youth Interview ¹	76.0	56.3
Adult Education	81.6	60.4
NHES:2001		
Screeners	69.2	69.2
Early Childhood Program Participation	86.6	59.9
Before and After School Programs and Activities	86.4	59.7
Adult Education	77.2	53.4
NHES:2003		
Screeners	64.6	64.6
Parent and Family Involvement in Education	83.3	53.8
Adult Education for Work-Related Reasons	76.2	49.2
NHES:2005		
Screeners	66.9	66.9
Early Childhood Program Participation	84.4	56.4
After School Programs and Activities	84.1	56.3
Adult Education	71.2	47.6
NHES:2007		
Screeners	52.8	52.8
Parent and Family Involvement in Education	74.1	39.1
School Readiness	77.0	40.7
Adult Education for Work-Related Reasons	Not completed	Not completed

¹Youth were interviewed independently in these collections.

Steps taken to improve the response rates in NHES accumulated such that approaches taken in earlier years were typically carried forward as new approaches were added.² Throughout the study's history, we have always tried two refusal conversion attempts³ at the screening with at least some cases, and at least one at the topical survey-level. Prior to the 1995 study, cases that gave a firm refusal during the first two screener contacts were not contacted for a final refusal conversion attempt. Hostile refusals were never recontacted. Beginning in 1995, households with two firm refusals were attempted one more time for the screener interviews. In the topical surveys, prior to 1995 one refusal conversion attempt was made provided the refusal was not hostile or firm. We began recontacting households that firmly refused a topical survey beginning in 1995. Second refusal attempts were not integrated into the topical surveys until 1999. Also beginning with the 1995 collection, we began leaving messages on answering machines regarding the purpose of the survey and toll-free numbers to call to participate.

Various mailing strategies also began to be introduced in the 1995 collection. That year, we began sending refusal conversion letters to households before trying a refusal conversion call.⁴ Mailing activities were augmented in 1996 when we started to send letters to households explaining the purpose of the survey before we started calling the sampled households. Building on work such as that done by Singer, Van Hoewyk, and Maher (2000), we began experimenting with pre-telephone contact monetary incentives in the advanced mailings and incentives for a refusal conversion at the screener and topical survey stages beginning with NHES:2003. The experiments did show that relatively small monetary incentives of \$2 during an initial mailing combined with \$2 for the first refusal followup contact, or \$5 included in the first refusal followup contact led to higher response rates at the screening level. At what point incentives were sent, either pre-contact or at refusal conversion, did not have much relationship to increased response rates. Since sending \$5 to just those households that refused initial contacts to complete a screener interview was the most cost effective of the strategies tested, that approach was taken for NHES:2005. The strategy changed to a \$2 advanced-letter incentive and a \$2 refusal conversion incentive for NHES:2007 to help ensure as equitable a use of incentives for sampled households as possible. For more information on the incentive experiment, see Brick et al. (2006). Apart from testing monetary incentives, different mailing strategies were also tested including the use of Priority Mail, Federal Express mailings, and first class mail approaches.

Falling response rates, and increased efforts needed to complete cases that did participate led to increasing average costs per sampled households in the NHES collections. Average per-household costs also increased because increasingly costly nonresponse analyses needed to be undertaken to evaluate the potential for nonresponse bias as response rates fell. The most significant example of this was a special bias study that was conducted as part of NHES:2007 (Van de Kerckhove 2008).⁵ Results from the study indicated that nonresponse bias was not a significant problem in the data, particularly after weighting. However, it did suggest that, without weighting adjustments, some estimates in NHES:2007 would exhibit undercoverage related biases in terms of some estimates from the School Readiness Survey.

The results need to be tempered with important considerations. Nonresponse bias analyses are helpful and informative, but there may be biases in the data that are not detectable because of limitations of the information available against which to conduct the tests. In the case of the NHES bias study, the response rate for the nonresponse bias sample itself was relatively low and may have suffered from similar biases as the main NHES collection itself. Extant sources of data against which NHES can be compared often lack the kinds of information

² One exception was a test of calling back households listed as having problems communicating. This was attempted in 1995, and was discontinued as it did not help boost response rates.

³ A "refusal conversion attempt" refers to the release of a case, and not only one call attempt. In many cases, multiple calls were made before contacting a household to attempt refusal conversion.

⁴ We did not pursue this approach in 1996, but reintroduced and kept the approach with the 1999 collection.

⁵ To conduct this study, we drew a sample of 7,500 cases independent of the main study sample. These cases were subject to both telephone and in-person contact strategies. Households that did not respond by phone or for which telephone numbers could not be found were contacted in person. This approach allowed us to evaluate both nonresponse bias to telephone survey operations and coverage bias related to households without landline phones.

NHES is intended to study. Results from such comparisons with data sources like CPS are informative and generally show that NHES does not have a nonresponse bias problem. However, while comparisons can be made on key demographic characteristics, extant studies lack education-related measures that are central to NHES design. For example, while comparisons with CPS show that NHES:2007 tends to have the same or very similar distributions on important household, family, and child characteristics such as household income and grade of enrollment for school-aged children (Hagedorn et al. 2009), similar comparisons on NHES topics like the number of children whose parents help them with homework, or the number of preschool children who are cared for by relatives are not possible because CPS lacks these data. Similar considerations need to be kept in mind when evaluating the undercoverage bias analyses that are possible for NHES.

NCES and Westat were comfortable with releasing the two surveys in NHES:2007 that had overall response rates in the 40 percent range. As noted, across a range of variables that could be tested, the data did not exhibit problems with nonresponse bias or coverage bias that could not be corrected through various weighting approaches. However, NCES was sensitive to the limitations of the nonresponse and coverage bias analyses, and was very concerned about trends that forecast increasing nonresponse to telephone surveys and undercoverage in landline sampling frames. Even in 2007, the response rates were so low with a third NHES survey (estimated at approximately 30 percent), Adult Education for Work-Related Reasons (AEWR), that NCES and Westat opted to stop AEWR data collection midway through the study.

III. New Design

Concerns with the NHES design increased significantly with the 2007 data collection. That collection exhibited an unprecedented decline in overall response rates for the NHES series with overall response rates falling over 10 percentage points when compared to the 2005 collection (Hagedorn et al. 2009 and Hagedorn et al. 2006). As a result, NCES cancelled a request for proposals (RFP) to conduct an end-of-decade 2010 NHES following the RDD and CATI approach that had been used for NHES. Over the summer and fall of 2007, NCES developed a new RFP that focused on generating ideas for how to shift NHES away from its traditional design to a new design that would consider mixed mode collection strategies and still maintain the main purpose of NHES – collecting trend data directly from households about educational experiences of household members. Collection was originally slated for early 2010 to minimize time between the last NHES collection in 2007 and when new NHES collections could be fielded. The new design would focus on two traditional topical surveys regularly fielded as part of NHES: the Parent and Family Involvement in Education Survey (PFI) that focuses on family involvement in the education of sampled students in kindergarten through twelfth grade, and the Early Childhood Program Participation Survey (ECPP) that focuses on the educational and care experiences of children who are not yet in kindergarten. The redesign explicitly stated that the resulting data would not be used for official statistics, but to inform how the next nationally representative NHES study would be conducted.⁶ As with the original NHES design parameters from the 1980s, cost considerations precluded personal contacts and in-person interviews. The RFP included a requirement for an evaluation of the effects of shifting from an RDD-based survey approach to new approaches proposed by bidders.

Westat, Inc. won the contract for the NHES redesign in the summer of 2008. Several design features of the original proposal should be mentioned here. The primary approach proposed by Westat focused on the use of self administered mail surveys as opposed to CATI driven phone survey modes. Given overall NHES:2007 response rates in the 40 percent range and recent work on mail survey strategies both within the firm and by researchers such as Dillman (e.g., Dillman 2007), researchers at Westat projected that mail-mode surveys could boost NHES response rates in a cost effective manner. Westat anticipated being able to achieve overall response rates of over 50 percent using less expensive approaches than the RDD approach traditionally used for NHES. Because of extensive changes to the original proposal resulting from a shift in the data collection schedule (see below for further details on the schedule change) and technical review panel recommendations, detailed information on the original proposal is not provided here.

⁶ NCES anticipated that proposals would recommend several different data collection methods none of which would have sufficient sample size to produce detailed estimates by various typically produced with NHES data.

One aspect of the original proposal that does require some discussion is evaluation of changes in estimates associated with changes in data collection mode from a RDD phone survey to a survey that, at least initially, is a self administered mail survey. A traditional bridge study would include a subsample of cases that would be treated using the same data collection approaches as the traditional NHES design. Results from this subsample would then be compared against the rest of the sample that would experience different data collection treatments, in this case mail survey materials and procedures. Given significant declines in overall response rates experienced in NHES:2007 compared to NHES:2005, and rapidly decreasing coverage rates for landline based RDD sampling approaches (see Keeter et al. 2007 for a good discussion of projected landline coverage rates), we were not certain that a bridge sample fielded in 2010 and following traditional NHES data collection approaches would be comparable to past NHES collections. As such, an RDD bridge study sample would not necessarily provide a good basis of comparison to new data collection approaches focusing on mail mode survey strategies and was not incorporated into the final design as a result. We will study unusually high or low participation rates in key measures traditionally measured with NHES to determine if we have changes that are out of range given our experience with NHES from 1991-2007 as an RDD study. This is not an ideal approach, but is arguably the best available given problems with the traditional RDD telephone interview approach in terms of the NHES collections.

Shifting Time Frames and Design Parameters

The RFP and original design proposal called for a large scale field test of new data collection approaches to be fielded in early 2010. Results from the test would then be used to guide how NHES collections would be conducted in the near future. At the time of award, the Office of Management and Budget (OMB) issued directives prohibiting household data collections during the first 9 months of 2010 to avoid federal data collections possibly interfering with the Decennial Census of 2010. Plans were then put in place by NCES and Westat to move the field test to 2011. A smaller scale pilot test would be fielded in the fall of 2009 to test possible alternative designs proposed by the NHES methodology technical review panel (TRP) members. In early January 2009, Westat convened the NHES methodology TRP to advise the project on how we might optimally address the competing design goals of increasing response rates, decreasing undercoverage rates, and maintaining comparability over time in NHES data without increasing costs significantly.

The TRP consisted of experts in a range of data collection modes used to conduct household surveys. Figure 1 provides a list of the TRP members. Members on the TRP were provided with a history of the previous NHES design along with the design options proposed by Westat during the summer of 2008. Extensive discussion focused on costs and benefits of maintaining the traditional NHES data collection approach with possible augmentation of the sampling frame data to include cell phones. The primary reasons that some members provided for this continuation was that a) even with relatively low response rates, NHES was not exhibiting significant nonresponse bias, b) that possible coverage bias issues could be addressed through approaches to augment the landline number frame with cell phone numbers, c) questionnaires would need to be changed to simplify them for mail out survey modes, d) within household sampling would be more difficult using self-administered mail out questionnaires, and e) trend lines could be more readily maintained if the data collection mode was not changed significantly.

Of particular concern in the self-administered survey strategies, were approaches available to conduct within household sampling. As with the traditional NHES approach, it is important for the new NHES collection approaches to be able to limit the number of topical surveys that households would be expected to complete. Because of concerns about errors respondents would make in terms of self selecting for topical surveys, Westat proposed a two stage mail out design whereby households would first fill out a screener survey and then mail it back to Westat. Within household sampling would be done by Westat who would then mail the appropriate topical surveys back to households. The sampled child's name (or some characteristics of the child if names were not provided in the screener) would be printed in the instructions and at the first question in the survey. Apart from reducing within household sampling errors, this approach would also allow Westat to determine how many of what kinds of topical surveys to mail to households, thus limiting the need to mail a large amount of materials (many to ineligible households) in an initial mailing.

Figure 1. – NHES Redesign Methodology Technical Review Panel Members and Other Experts Who Reviewed and Revised NHES Data Collection Approaches

Participant Name and Role	Affiliation as of January 31, 2009
Technical Review Panel Members	
Nancy Bates	U.S. Census Bureau
Paul Beatty	National Center for Health Statistics
Johnny Blair	Abt Associates
Stephen Blumberg	National Center for Health Statistics
Mick Couper	University of Michigan
Don Dillman	Washington State University
Bob Groves	University of Michigan
Scott Keeter	Pew Center for People and the Press
Kristen Olson	University of Nebraska
Roger Tourangeau	University of Michigan
Clyde Tucker	Bureau of Labor Statistics
Gordon Willis	National Cancer Institute
Westat and NCES Staff	
Stuart Kerachksy	Acting Commissioner, NCES
Marilyn Seastrom	Chief Statistician, NCES
Mike Brick	Vice President, Westat
Chris Chapman	Project Officer for NHES, NCES
Mary Hagedorn	Project Director of NHES, Westat
Jill Montaquila	Principal Investigator for NHES, Westat
Andrew Zukerberg	Senior Technical Advisor for NHES, NCES
Laura Lippman	Vice President, Child Trends
Lina Guzman	Senior Technical Advisor for NHES, Child Trends

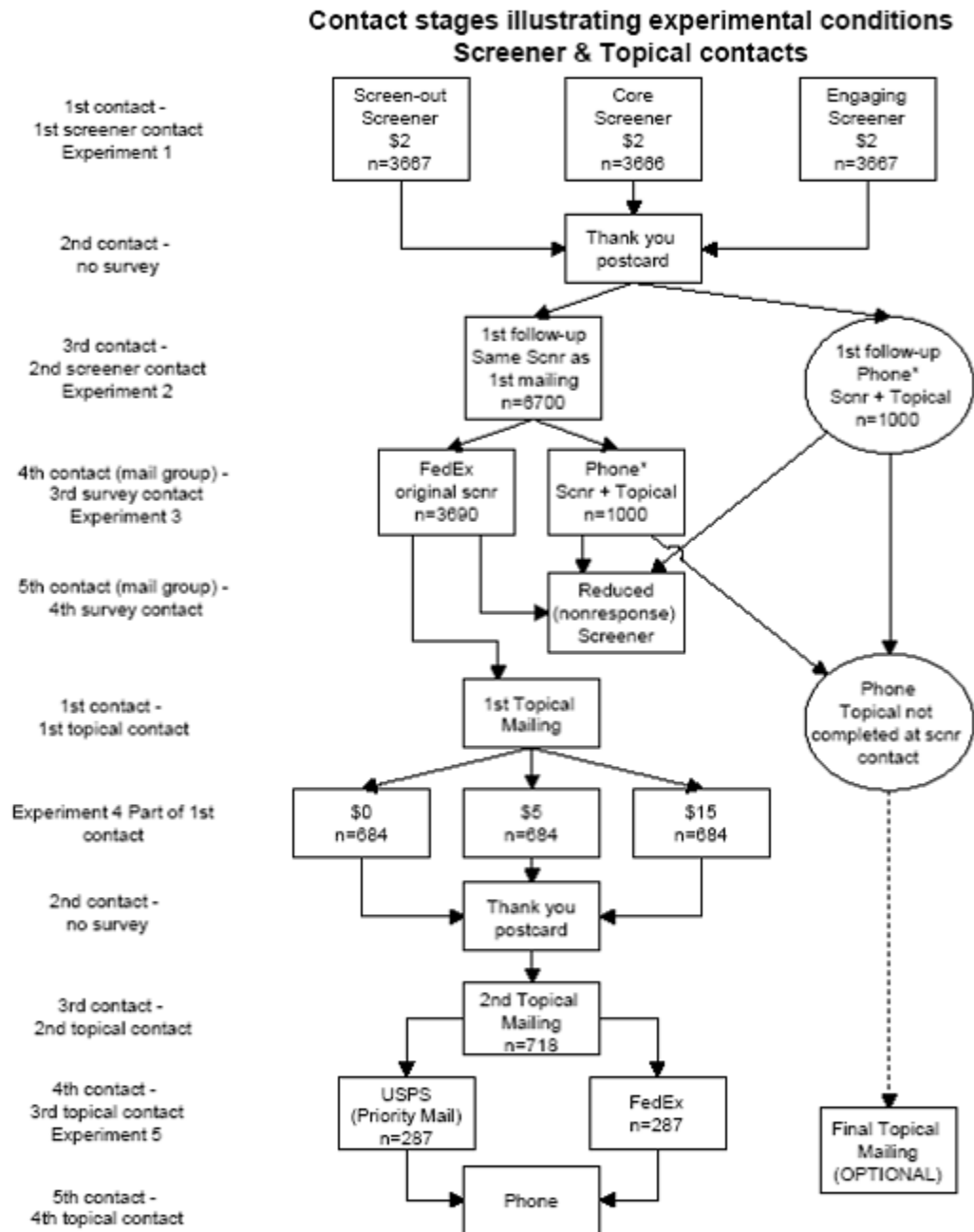
Several different points were raised about the limitations associated with maintaining the traditional NHES approach and the panel ultimately decided that moving away from the traditional RDD telephone survey approach should be studied at least in the 2009 pilot test. In general, these points were similar to those that motivated NCES to seek a redesign. Given that NHES experienced an overall response rate decline of over 10 percentage points over a 2 year period and had experienced a long decline in rates more generally, expectations were that response rates would continue to decline increasing the risk of nonresponse biases. While coverage issues associated with shifts from landline phones to cell phones might be addressed by augmenting landline telephone number frames with cell-phone numbers, response rates for cell-phone interviews tended to be even lower than for landline interviews which could exacerbate potential nonresponse bias (see for example, Link et al. 2007).

Ultimately, the TRP determined that shifting data collection modes away from telephone surveys based on RDD samples to self administered mail out surveys was worth pursuing. However, the TRP did recommend keeping a telephone survey option open for the main field test in 2011 as a possibility depending on the success of approaches studied in the pilot test.⁷

After careful deliberation, and follow-up with TRP members after the January meeting, several different options were developed for evaluation with the fall 2009 pilot test. The range of options that could be tested in the pilot was restricted by the need to keep the sample size to approximately 12,000 households given existing resources available for the work. Figure 2 summarizes the design of the 2009 fall pilot test and highlights several different key experimental conditions.

⁷ If an RDD phone survey approach is needed for the field test in 2011, we can build on extensive experience from past NHES collections to develop an efficient design model without the need for extensive pilot testing.

Figure 2. – NHES 2009 Pilot Test Contact Strategies and Experimental Treatments: Screener and Topical Surveys



* Experiment 6: For these groups, Topical surveys attempted by phone without prior Topical mail contact, a random 50% of each group will be offered a promised incentive of \$5 for completing the Topical survey. The other 50% will not be offered any incentive.

NOTE: Other than the first contact (Experiment 1) sample sizes, which will be fixed by design, all remaining sample sizes are estimated. Sample sizes exclude the 800 addresses that will be part of the 'targeted sample'.

It is important to note here that NCES and Westat do not plan to release the pilot or field test data for public use and there are no current plans to use the resulting data to produce official estimates related to the populations that are the focus of the tests. The purpose of the tests is to help identify a data collection approach that can replace the RDD and CATI approach typically used in NHES. Sample sizes needed to test various options will not be sufficiently large to produce stable estimates and funding is not available for significantly larger samples.

Pilot Test Experiments

All households will receive a \$2 incentive in the mailing with the screening instrument. This level of incentive has been shown to boost response rates in previous NHES experiments (Brick et al. 2006). The incentive and screening instrument will be accompanied by an introductory letter explaining the purpose and importance of the survey. It will be addressed to “{CITY NAME} Resident” and tailored to match where the respondent lives.

The first tests will focus three different screening instrument designs. One instrument will include a series of interesting substantive questions about education related topics along with questions needed to determine eligibility for second stage topical surveys. This is referred to as the “Engaging Screener” in figure 2. A second instrument will include only items needed for topical survey screening, which include a few questions about the age, enrollment status, and grade of enrollment for children in the household along with information about phone numbers that we can call for follow-up contacts and a few questions about household composition. This is the “Core Screener” in figure 2. A third instrument be similar to the Core Screener, but will have an introductory question that will facilitate households without children quickly completing the instrument without having to work past three or four questions. This is the “Screen-out Screener” in figure 2.

The test will focus on whether the Engaging Screener helps boost response rates over screeners that contain little, if any, education-related content. The assumption being tested here is that including some substantive questions about education will engage respondents and improve screener-stage response. Both Engaging Screener and Core Screener approaches will be compared against the Screen-out Screener instrument. The Screen-out instrument allows households without eligible children to quickly identify themselves and to easily complete the questionnaire. This “screen-out” model could help increase overall response rates by convincing those without eligible children to fill out a very short series of questions and to respond. NHES has traditionally had lower response rates for adult-focused as opposed to child-focused surveys, and our assumption is that part of the reason is that households without adults assume a survey from the U.S. Department of Education must be about children and therefore are less likely to respond. The screen-out model would help these households understand that we need their participation and that participation has a very low level of burden for them. The problem with this approach is that it may not convince households with children to respond so it needs careful review.

The pilot test will also consider how to approach households that do not respond to the initial mailing as shown in “Experiment 2” in figure 2. One week following the initial mailing, a “thank-you/reminder” postcard will be sent to all households, thanking them for returning the questionnaire and asking them to please respond if they have not yet done so. After a few weeks, if no response is obtained, two treatments will be tested. Most cases will receive a second set of self-administered paper screening instruments. A subset of cases for which we have obtained telephone numbers matched to the original mailing address will be contacted by phone to complete the screener with an interviewer using CATI versions of the screening instruments.

Among households mailed a second set of screening materials, if no response is received, a third follow-up to complete the screener questionnaires will be attempted (see experiment 3 in figure 2). One group of nonrespondents will have a third set of materials mailed to them via FedEx. The purpose of the FedEx as opposed to resending materials regular mail in U.S. Department of Education envelopes (the approach taken in the first two mailings) is to draw the attention of possible respondents to the materials. The second set of nonrespondents who stayed in the “mailing path” will be contacted by phone to complete the screening materials with a telephone interviewer using CATI programs.

All cases who do not respond after three attempts to complete a screening instrument, regardless of the approaches taken to contact the household, will receive a postcard size “reduced screener.” This questionnaire will ask 5 or 6

questions needed to help with the nonresponse evaluation. They will not be contacted again after the “reduced screeners” are sent.

The next set of experiments focus on the topical surveys. For cases that complete the screener by mail, a fourth experiment will test the effects of offering varying levels of respondent incentives ranging from no incentive to \$15 (see experiment 4 in figure 2). Although \$15 might be considered a rather high incentive for a general household survey, it is worth testing in the Pilot Test to gauge how much of an effect it might have. If significant, it might be cost effective. A thank-you/reminder postcard will be sent one week after the initial topical mailing.

All cases that do not initially respond to the mailed topical questionnaire, will be sent a second set of topical questionnaire materials. Nonrespondents to this second mailing will be included in a fifth experiment. Here, we plan to vary how the third topical survey material mailing is sent. One subset of nonrespondents will be sent the package via USPS Priority Mail and the second group will be sent materials via FedEx. The primary reason for the test is to determine if past experience with mailing attempts to contact households that had not responded to prior telephone contacts holds true when the primary data collection mode is self-administered mail out surveys. Tests done as part of NHES:2005 on the screening interviews suggest that contacting households that had twice not responded to screener interview attempts by phone had higher response rates on a third contact attempt if they had been sent follow-up informational material via FedEx as opposed to Priority Mail (Hagedorn et al. 2006). All cases that do not respond to this final mailing will be contacted by phone for one attempt to complete the interview via CATI applications.

For households that complete screening instruments by telephone interview, a topical survey experiment will also be conducted (see footnotes to figure 2). This sixth experiment will be used to determine if promised incentives of \$5 for completing the topical interview by phone have a sufficiently positive effect on response rates to warrant using in larger scale data collection operations. Half of the households slated for topical interviews to be done by phone will be offered \$5 if they complete the interview and half will not be offered any monetary remuneration. Past experience suggests that a prepaid incentive should significantly boost response rates, but we want to test the promised incentive under this new design approach⁸. Provided we have resources and time⁹ to follow up with households that refuse to do the topical interview by phone or that cannot be contacted by phone, we will mail these households topical interviews.

Not shown in figure 2 are two other tests that will be implemented during the pilot study. First, we will test the utility of using address sources that have a lot of rich frame data in them. The purpose would be to determine if we could better target households with children before general mailing occurs thus reducing mailing costs and potentially improving response rates. The test will be conducted on a subsample of 800 cases that will be drawn from sample frames that contain extensive household characteristic data. The test will focus on how well frame data correlate with what respondents in sampled households tell us about those households and in the analysis of nonresponse. Second, we will use part of the pilot test sample to develop an oversample of addresses in areas with relatively high concentrations of Hispanics. This sample will be used to test bilingual forms that will initially be mailed to only households in these areas, though households can request such forms if they are outside of the high-Hispanic population areas.

Other Design Considerations for the Pilot Test

Results from the pilot test will drive the design of more focused tests for a larger scale field test in early 2011. At least two important aspects of the main field test are important to note. First, we decided to focus most of the pilot test on the screening process and how to link screening information to the second-stage topical surveys. Without

⁸ While some literature such as Church (1993) indicates that promised incentives are not as effective as prepaid incentives in mail surveys, the fact that a prior prepaid cash incentive will be sent at the screener stage may increase the efficacy of this promised incentive at the topical stage.

⁹ We need to be out of the field in December of 2009 so make sure NHES does not interfere with 2010 Census operations.

good response rates to the screener and clear approaches to linking sampling results from the screeners to the subsequent topical surveys, the content of the topical surveys would not be very useful. Also, because of time constraints between the award of the contract in September of 2008 and the need to request OMB approval in time to start the pilot test in September 2009, the modification of the topical surveys to reflect the shift from a phone interview to a self-administered survey was compressed. As such, we planned to put more effort into the topical survey designs after the design of the pilot test was worked out. We will begin convening content review panels for the PFI and ECPP in late 2009 or early 2010.

In the interim, we did make some significant changes that will be included in the pilot test, and are working through the instruments in detail to determine how clear they will be in a self-administered questionnaire mode. The key changes made to the topical instruments that will be part of the pilot test include a simplification and removal of a lot of the skip patterns that had been built into the CATI programs. One skip pattern change was so significant that it led to the development of a third topical survey instrument. As noted above, homeschooled children are part of the PFI sample. However, they represent approximately 3 percent of all children in grades K-12 and many questions about them would not be appropriate for children in public and private schools and vice versa. CATI programs seamlessly addressed this issue by skipping homeschooled children around half of the PFI items, and routing children in public and private schools around homeschool questions. These skip pattern rules are extensive and would likely confuse respondents so a decision was made to produce a separate PFI homeschooler instrument and a separate PFI instrument for children in public and private K-12 schools.

A second significant instrument change made for the pilot test was the removal of looping sequences in the questionnaires that asked respondents to provide a lot of similar data for multiple programs and activities of the same kind, e.g. multiple center-based care and education programs for a sampled child. Such a simplification is not new to NHES, having been used at least once while NHES was an RDD phone survey back in NHES:1999. That study provided a wealth of information that was largely comparable to previous NHES collections.

Associated with the skip pattern issue is a third significant change that is related to the PFI instrument for public and private school children. CATI had allowed an interactive lookup program to be incorporated into interviews with parents that allowed interviewers to identify the sampled child's school during the interview process. Such information is useful in terms of being able to link children a wealth of school-level data available from extant sources like the Common Core of Data and the Private School Survey. This lookup process will not be possible for cases completing the self-administered mail surveys for the pilot test. To address the issue, lists of schools near the child's home will be identified and the PFI instrument will be tailored to include the list of these schools for respondents to choose from when filling out the questionnaire. Screening information on whether a child attends public or private school will be used to determine what list of schools to include in the topical instrument.

IV. Next Steps

The pilot test is scheduled to run from early September through mid December 2009. Results from the pilot test will be evaluated through early 2010 to determine which of the approaches tested look most promising for consideration in the 2011 field test. The 2011 field test will be much larger, approximately 106,000 households, to facilitate comparisons of estimates from the field test to past NHES collections while also permitting different experiments to be conducted during the field test.¹⁰ Without the pilot test results in hand, it is difficult to provide detailed information about what specifically the 2011 field test will include or look like. However, there are several issues that will need to be addressed that can be discussed at this point.

First, if none of the mailing strategies evaluated in the pilot are successful, we will need to have alternative designs ready to field. As noted, one of these will be some form of an RDD telephone interview design. Most of that design would be borrowed from previous NHES collections. Some features that would need to be included would be

¹⁰ To reiterate, data from the pilot test and field test are not intended to produce official estimates and will not be released to the public. The need to test multiple approaches and data collection modes means that samples from any given experimental condition will not be sufficiently large to produce stable estimates, and approaches are sufficiently different that combining data from the different test conditions is not advisable.

incorporation of cell-phone numbers into the sampling frame, and consideration of how to devise a sample to facilitate follow-up in other modes including, possibly, in-person interviews. This is a relatively costly design approach in comparison to the self-administered, or primarily self-administered, mail surveys in the pilot test. We have experience with most aspects of such a design, so it is not part of the 2009 pilot test. Another strategy that could be developed further is one in which a copy of the screener and each topical survey is sent to each sampled household to test the feasibility of respondent-administered within-household sampling. While not ideal in terms of possibly discouraging respondents with a large volume of material and in terms of respondents incorrectly applying within-household sampling rules, the approach might result in higher response rates since it requires just one household contact to determine who is eligible for which topical survey and to have those surveys filled out. On the other hand it would require sending the materials to all sampled addresses when only about 37 percent will have an eligible household member.

Second, follow-up design work for the field test will need to be completed by August 2010 in order for OMB to have enough time to comment on and clear the data collection request. Such a schedule would allow us to conduct the NHES field test starting in January 2011 and wrapping up in April of 2011. These dates correspond to the traditional collection period for NHES and are important to keep in mind. Most of the PFI questions focus on activities during the current school year. The primary reason for this is to avoid recall issues associated with asking about school-related events from previous school years. If we start the collection much before January, we run the risk of parents and families not having had opportunities to participate in activities addressed in PFI. If we start the collection much after January, we run the risk of not being able to run cases through all possible contact strategies at the screener and then at the topical survey levels before the end of the school year. Many of the PFI questions do not make sense outside of the context of a school year. With these goals and deadlines in mind, content review panels will start being convened late in 2009 or early 2010.

Third, additional work on the topical surveys between the pilot test and the field test is anticipated. We were able to greatly simplify the surveys to make them more appropriate for self-administered modes. We will examine the responses for navigation patterns, substantive response patterns, item response rates, and respondent use of the scannable form. Part of this work may focus on evaluating who the respondent should be. Traditionally, NHES has been filled out by the parent or guardian most knowledgeable about the sampled child's education. To help increase response rates, this criterion was eased in the pilot such that any parent or guardian who is knowledgeable about the sampled child's education can respond. It would be useful to directly study the effect of selecting different respondents.

Fourth, after the field tests, several other evaluations will need to occur and these are scheduled for development through the summer of 2011. One evaluation will be a comparison of the field test PFI and ECPP data against data from the last time these two studies were fielded, 2007 and 2005 respectively. The purpose is to determine if the new NHES collection approaches provide estimates that can be included with data from prior NHES collections to develop meaningful trend lines. The time lags are not ideal, but they can be accounted for by considering rates of change between the 2007 PFI and 2003 PFI (to mirror the 4-year period between the 2007 PFI and the field test) and between the 1999 Parent Interview and 2005 ECPP data (to mirror the 6-year period between the 2005 ECPP and the field test). This work will focus on detecting estimates that changed significantly between the field test and prior collections to determine if they make sense given previous trends and rates of change. As noted, an ideal bridge study to test mode effects will not be possible. Directly related to this work will be the evaluation of different weighting strategies for the new data collection approach. New weighting approaches will be evaluated both in terms of helping to maintain comparability with past NHES collections, and in terms of producing estimates that match current population totals and estimates.

Finally, NCES will need to evaluate the results, with Westat's recommendations, to determine if NHES can continue with any of the approaches tested and if so, on what schedule. Ideally, this evaluation would happen relatively quickly and be completed in 2011. Given contract award and OMB clearance schedules and assuming that a clear model emerges from the 2011 field test that requires only minimal changes for a full-scale national collection, a 2013 collection would be possible. This would only be feasible if NCES can make this kind of decision by late summer 2011.

References

- Bielick, S. (2009). *1.5 Million Homeschooled Students in the United States in 2007*. (NCES 2009-030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Brick, J.M. and Burke, J. 1992 (NCES 92-101). National Household Education Survey Technical Report #2: Telephone Undercoverage Bias of 14- to 21-Year-Olds and 3- to 5-Year-Olds. NCES, Washington DC.
- Brick, J.M., Celebuski, C., and Collins. M. 1992 (NCES 92-099). National Household Education Survey Technical Report #1: Overview of the NHES Field Test. NCES, Washington DC.
- Brick, J.M., Hagedorn, M.C., Montaquila, J., Roth, S.B., and Chapman, C. (2006). *Monetary Incentives and Mailing Procedures in a Federally Sponsored Telephone Survey*. (NCES 2006-066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Brick, J.M., and Waksberg, J. 1991. *Avoiding Sequential Sampling with Random Digit Dialing*. Survey Methodology.
- Church, A. 1993. Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly, Vol. 57, No. 1*.
- Collins, M., Brick, J.M., Loomis, L., Gilmore, S., and Chandler, C. 1996 (NCES 96-825). National Household Education Survey of 1995: Early Childhood Program Participation Data File User's Manual. NCES, Washington, DC.
- Dillman, D. (2007). *Mail and Internet Surveys: The Tailored Design Method, Second Ed*. John Wiley & Sons. Hoboken, New Jersey.
- Hagedorn, M., Roth, S.B., Carver, P., Van de Kerckhove, W., and Smith, S. (2009). *National Household Education Surveys Program of 2007: Methodology Report*. (NCES 2009-047). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Hagedorn, M., Montaquila, J., Carver, P., O'Donnell, K., and Chapman, C. (2006). *National Household Education Surveys Program of 2005: Public-Use Data File User's Manual, Volume I, Study Overview and Methodology*. (NCES 2006-078). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Keeter, S. Kennedy, C., Clark, A., Thompson, T., and Mokrzycki, M. (2007). What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population. *Public Opinion Quarterly, Vol. 71, No. 5*.
- Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborne, and A. H. Mokdad. (2007). Reaching the U.S. Cell Phone Generation: Comparison of Cell Phone Survey Results With an Ongoing Landline Telephone Survey. *Public Opinion Quarterly, Vol. 71, No. 5*.
- Massey, J. and Botman, S. "Weighting Adjustments for Random Digit Dialed Surveys," in Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls, W., and Waksberg, J. eds. Telephone Survey Methodology. 1988. John Wiley & Sons. New York, New York.
- Singer, E., Van Hoewyk, J., and Maher, M.P. (2000). Experiments with Incentives on Telephone Surveys. *Public Opinion Quarterly*, 64: 171-188.

U.S. Department of Education. National Center for Education Statistics. National Household Education Survey. *An Overview of Response Rates in the National Household Education Survey: 1991, 1993, 1995, and 1996*. NCES 97-948, by J. Michael Brick, Mary Collins, and Kathryn Chandler. Washington, DC: 1997.

Van de Kerckhove, W., Montaquila, J.M., Carver, P. R., and Brick, J.M. (2008). *An Evaluation of Bias in the 2007 National Households Education Surveys Program: Results From a Special Data Collection Effort* (NCES 2009-029). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.