Minimizing Survey Error through Interviewer Training: New Procedures Applied to the National Health Interview Survey (NHIS)

James M. Dahlhamer[1], Marcie L. Cynamon[1], Jane F. Gentleman[1], Andrea L. Piani[2], and Michael J. Weiler[2]

[1]National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782
[2]U. S. Census Bureau, 4600 Silver Hill Road, Washington, DC, 20233

Abstract
Minimizing survey error requires adherence to the accepted principles and best practices of survey research. Interviewers can be a significant source of error that is difficult to control. Ensuring that interviewers execute their jobs properly requires that they be well-trained, monitored, and provided feedback. In this paper, we discuss new procedures for training interviewers with the National Health Interview Survey (NHIS). An in-person survey, the NHIS has in the past used a decentralized system for training its field staff. In 2010, roughly 650 experienced interviewers attended one of four "refresher" training sessions to be briefed on the survey's content and interviewing procedures. Using a total survey error framework, the goal was to have a core of very able instructors deliver a uniform message with the intent of achieving consistent application of established interviewing protocols across interviewers and sites. An initial assessment of the new training procedures reveals buy-in from interviewers, but preliminary short-term pre-training/post-training comparisons of performance and data quality indicators suggest only minor (though largely positive) impacts.

Key Words: interviewer training, interviewer performance, data quality

1. Introduction[1]

Survey interviewers can be a significant source of survey error (Biemer and Lyberg, 2003; Fowler and Mangione, 1990). This can occur in a variety of ways including outright falsification of data, inappropriate probing, data entry errors, or otherwise failing to comply with survey procedures. Prior research has also demonstrated that interviewer training can have a significant, beneficial effect on the reduction of interviewer error. Fowler and Mangione (1990) explored the impact of training length (half a day, two days, five days, and 10 days) on a number of outcome measures including the percentage of total variance of survey statistics associated with interviewers (intra-interviewer correlation coefficient or rho), and the percentage of interviewers rated as excellent or satisfactory on reading questions as worded, probing appropriately, recording answers to open and closed questions, and engaging in nonbiasing interpersonal behavior. Two and

---

[1] The findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of the National Center for Health Statistics, Centers for Disease Control and Prevention, or the U. S. Census Bureau.

five day training sessions significantly reduced the amount of interviewer variance in several survey statistics, while training sessions of two days or more resulted in a significantly greater percentage of interviewers who read questions as worded, probed appropriately, and engaged in nonbiasing interpersonal behavior.

Billiet and Loosveldt (1988) conducted a field experiment to measure the effects of interviewer training on the quality of responses obtained during in-person interviews. Interviewers in the study received either a three-hour briefing or five three-hour training sessions. The longer-trained interviewers produced lower item nonresponse, produced more complete recording of responses to open-ended questions, were more likely to read instructions and questions as worded, and were more likely to probe and to probe appropriately.

More recent studies have assessed training modules designed to increase survey participation. Groves and McGonagle (2001) performed two experiments, one involving the Current Employment Statistics survey and the other the U. S. Census of Agriculture. In the first experiment, all participating interviewers received a one-and-a-half day training workshop focusing on the general principles of refusal avoidance, and how to respond to and counter various types of respondent concerns (e.g., time and burden concerns, government concerns). Comparisons of pre-training period and post-training period (both one-and-a-half months in length) interviewer cooperation rates revealed improvements, especially among lower performing interviewers. In the second experiment, one set of interviewers received a similar training workshop to that described above, while a set of control interviewers received no training. Pre-training/post-training interviewer cooperation rate comparisons revealed significantly greater gains in cooperation among the interviewers attending the training workshop. O'Brien et al. (2002) assessed the impact of a similar training module on cooperation rates among interviewers with the National Health Interview Survey (NHIS). Consistent with the Groves and McGonagle (2001) results, interviewers receiving the refusal avoidance training produced greater gains in cooperation rates post-training (roughly four months in length), compared to pre-training (approximately one-and-a-half months in length) than did a set of control interviewers (received no special training).

In this paper, we make a preliminary assessment of the effectiveness of a new set of procedures applied to interviewer "refresher" training[2] with the National Health Interview Survey (NHIS). Moving to a centralized training format with the utilization of very capable instructors and a heavy emphasis on data quality, four one-and-a-half-day training sessions were held during the first two weeks of January 2010. Following the training, we set out to address the following research questions:

How well received was the training by interviewers?

What impact did the training have on performance and data quality?

---

[2] Refresher training is usually held once a year for interviewers who already have some amount of NHIS interviewing experience (more description is provided in section 2.2). It can be distinguished from NHIS initial training, which is mandatory for interviewers new to the U. S. Census Bureau or new to working on the NHIS. With few exceptions, initial training is taken just once.

In the next section, we provide a brief description of the NHIS, followed by a description of past refresher trainings and the new procedures applied to the 2010 interviewer training. In section 3, we present results of interviewer evaluations completed at the close of each training session; this is meant to address the first of our two research questions. In section 4, we attempt to address the second of our research questions by presenting pre-training/post-training comparisons of 25 survey performance and data quality indicators. Section 5 concludes with a summary of findings, a discussion of next steps for the training evaluation, and recommendations for future training.

2. Description of the NHIS and Interviewer Refresher Training

2.1 The National Health Interview Survey

The NHIS is an annual, multi-purpose health survey and the principal source of information about the health of the civilian, noninstitutionalized, household population of the United States. Conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), the NHIS utilizes a multi-stage, clustered sample design, with oversampling of black, Hispanic, and Asian persons. The survey produces nationally representative data on health insurance coverage, health care access and utilization, health status, health behaviors, and other health-related topics. The microdata are released on an annual basis, approximately six months after the end of data collection.

Roughly 650 interviewers with the U. S. Census Bureau conduct the in-person interviews (some telephone follow-up is allowed[3]) using computer assisted personal interviewing (CAPI). Interviewing is continuous throughout the year, with the exception of the two weeks set aside for interviewer refresher training. The core survey instrument contains four main modules: household composition, family, sample child, and sample adult. A household respondent provides demographic information on all members of the household in the household composition module. For each family within a household, the family module is completed by one family respondent who provides sociodemographic and health information on all members of the family. Additional health information is collected from one randomly selected adult (the "sample adult") aged 18 years or over, and from the parent or guardian of one randomly selected child (the "sample child") under age 18 (if there are children in the family). In addition to the core survey modules, supplemental questions on special topics, co-sponsored by other government agencies, are added to the NHIS questionnaire each year.

2.2 Interviewer Refresher Training

Interviewer refresher training has traditionally included two components: an interviewer self-study module (including descriptions of new survey content and practice interviews in preparation for the upcoming year), and classroom training. The classroom component was usually held in the first two weeks of January, just prior to the start of data collection for the calendar year. Prior to 2010, as many as 30 classroom training sessions were

---

[3] Once a personal visit contact has occurred, telephone followup is permissible if a personal visit followup is not possible. At the end of an interview, interviewers are asked to report which main sections (household composition, family, sample child, sample adult), if any, were conducted primarily by telephone.

conducted each January, with multiple sessions and locations in each of the 12 Census Regional Offices. The training, with an extensive focus on survey content new for that year, was usually conducted by Regional Office survey supervisors or Senior Field Representatives (experienced interviewers) working from a verbatim training script developed at Census Headquarters. The length of training sessions varied somewhat by year, but averaged two to two-and-a-half days in length.

It was common for NCHS and Census Headquarters staff to observe a number of the training sessions. Upon return from training, staff from both agencies would compile their notes and often hold a debriefing. Over time these observations revealed variations in the knowledge and skill sets of trainers, which translated into variations in the quality of training. Furthermore, this decentralized approach to training fostered deviations from the training agenda. Hence, inconsistent coverage of core materials across the Regional Offices became a concern. To further hamper the effectiveness of training, NCHS budget shortfalls forced the cancellation of the classroom component in 2003, 2008, and 2009. In parallel, staff at NCHS began to observe increases in the number of interviews with excessive "don't know" and "refused" responses, along with inordinately short interview times and shortcutting of questions, deliberate interviewing of wrong sample persons, and other violations of interview protocols. Together, concerns over the quality of training and collected data prompted a review of and revision to refresher training procedures for 2010.

While the self-study module remained largely the same in format, the classroom component of the 2010 refresher training represented a clear break from past procedures. A major logistical change involved the reduction of roughly 30 training sessions to four. Each of four training sessions (two held in Atlanta, one in San Antonio, and one in Tucson) involved roughly 150-175 interviewers and staff members from up to four of the 12 Census Regional Offices. The move to a centralized format was designed to place NHIS subject matter experts from NCHS, Census Bureau Headquarters, and Census Bureau Regional Offices in front of field staff. The experts gave presentations and handled all question and answer sessions. In addition, the centralized format enabled the same speakers, with some departures, to speak at all four training sessions. This ensured the delivery of a more consistent, standardized message.

The classroom training placed a heavy, up-front emphasis on data quality. An early presentation set the tone, outlining various sources of survey error and documenting the roles of survey designers and interviewers in reducing error and maintaining quality throughout the survey process. Additional presentations on the opening day focused on topics such as reading questions as worded, the importance of collecting contact history data for assessing and monitoring data quality, the performance and data analysis tool (PANDA) used to monitor case-level data quality and interviewer performance, best practices for completing quality interviews in difficult situations, how reinterview helps in improving performance, and others. Overviews of new survey content were presented in the afternoon of the first day, and practice interviews were conducted on the second day of training.

3. Interviewer Evaluations of Refresher Training

3.1 Data

Interviewers and other trainees from the 12 Census Regional Offices were asked to complete a training evaluation form at the close of their training session. The training form included roughly 75 questions. The vast majority of questions were closed-ended and designed to capture information on the usefulness of the training materials (e.g., binders containing handouts, practice interviews) and presentations, including pre-classroom training materials, for conveying key concepts; the adequacy of coverage of key concepts and time allotments to various presentations and topics; the usefulness of training materials and presentations in providing information helpful for securing survey participation; an overall rating of the training; and other topics such as ratings of various characteristics of the training site. A handful of open-ended questions were also included to capture trainees' comments, including their likes and dislikes from the training. Of the 674 trainees who attended one of the four training sessions, 539 or 80.0% completed an evaluation form.
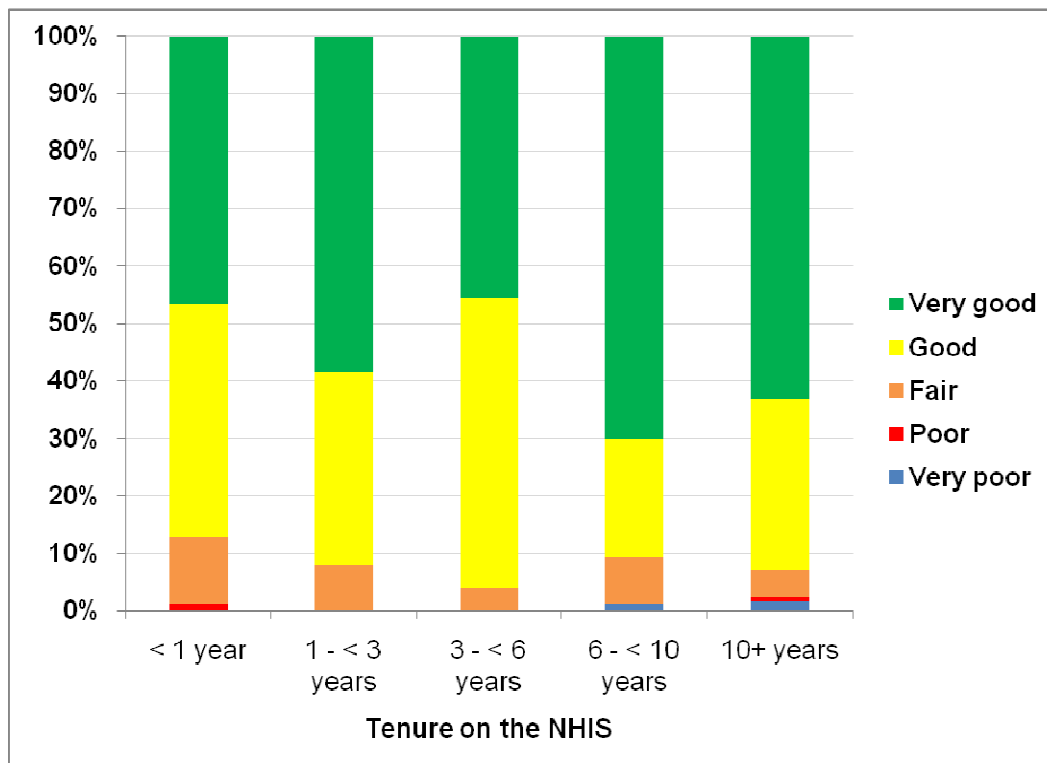

Figure 1: Overall Rating of 2010 NHIS Refresher Training by NHIS Tenure

3.1 Results

Figure 1 presents trainee responses to the question "Overall, how would you rate this year's NHIS training?" Responses are broken out by tenure on the NHIS.[4] Overall, regardless of tenure, roughly 90% of interviewers rated the training as very good or good.

_____
[4] NHIS interviewers may have worked on other surveys prior to and/or while working on the NHIS.

This is an encouraging figure considering the changes in training format and the vast logistical undertaking that characterized the centralized approach. Interesting results also emerge by tenure. Compared to trainees with less than 6 years of NHIS experience, longer-tenured trainees (6+ years) were more likely to rate the training as very good (66.0% versus 50.3%; p < .01). We hypothesize that the longer-tenured trainees had experienced multiple refresher trainings in the past and were better positioned to assess and rate the new training format.[5] If we are correct in our assumptions, the findings by tenure further boost our confidence in the revised training procedures.

Table 1 presents the top five responses or themes to emerge from two open-ended questions on the evaluation form: "What did you like most about this year's NHIS training?" and "What did you like least about this year's NHIS training?" By far, the most prevalent response to the "like most" question was the ability to meet and interact with interviewers from other Census Regional Offices. This was mentioned by roughly

Table 1. Top Five Responses/Themes Provided by Trainees (n=539) to the Open-Ended Questions "What did you like the most about this year's NHIS training?" and "What did you like least about this year's NHIS training?"[1]

Question and Top Five Open-Ended Responses/Themes	Number of Trainees	Percent of Trainees
What did you like most about this year's NHIS training?:		
Interacting with other interviewers and meeting interviewers from other Census Regional Offices	131	24.3
Having Census Bureau Headquarters staff and the NCHS subject matter experts presenting the information and answering questions	58	10.8
Cell Phone Data presentation	50	9.3
Nice training site/enjoyed being off-site	45	8.3
I learned the reasons/purpose for questions and supplements and can better answer respondent questions	45	8.3
DID NOT PROVIDE A RESPONSE	89	16.5
What did you like least about this year's NHIS training?:		
Presenters read verbatim from screens and handouts	39	7.2
Did not like holding questions until the end/Not enough question and answer time	35	6.5
The training was not long enough	25	4.6
Cramped work space	23	4.3
Too much time was spent on practice interviews; only need to go over new material	23	4.3
DID NOT PROVIDE A RESPONSE	122	22.6

SOURCE: U. S. Census Bureau (2010).
[1] Coding of the open-ended questions was performed by staff with the Methods Research Branch, Field Division, U. S. Census Bureau.

---

[5] As noted earlier, interviewer refresher training was not held in 2003, 2008, and 2009. It is plausible that several of the trainees with less than three years of NHIS experience had never participated in refresher training until 2010.

24% of trainees. This is clearly an artifact of the centralized approach, and not a surprising response, since we anticipated that interviewers would enjoy the opportunity to swap stories, tips, and experiences with colleagues from other regions. The second most prevalent "like," mentioned by 11% of trainees, was having Census staff and NCHS subject matter experts giving the presentations and answering questions. This was an encouraging response as it spoke directly to an intended goal of the revised format: have qualified, knowledgeable speakers conduct the training, clearly communicate key survey concepts and interview protocols, and do so in a consistent, standardized manner. The third and fifth most prevalent responses to the "most like" question reinforce the effectiveness of this approach. The "Cell Phone Data presentation" was intended to address interviewer questions and confusion over asking questions on cell phone usage at the beginning of the survey. Many interviewers had felt the questions hindered their ability to maintain respondent participation. After hearing the presentation by a leading NCHS expert in the field, interviewers left the training with a better understanding of the questions and with the confidence and ability to address respondent concerns. This was echoed more generally with the fifth most prevalent response: "I learned the reasons/purpose for questions and supplements."

As with all trainings, there were certain elements that were not as well received by the trainees. Table 1 also presents the top five responses to the question "What did you like least about this year's NHIS training?" There are two observations that are quite telling. First, the number of trainees who did not provide a response to this question was considerably greater than the number of trainees who did not provide a response to the "like most" question. And second, no single response to the "like least" question was mentioned by more than 7.2% of trainees. The top response, however, is somewhat of a concern. Just over 7% of trainees were critical of some presentations being read directly from slides and/or handouts. Having subject matter experts conduct the training does not ensure that the styles of the presentations are of high quality or engaging, something that will need to be addressed for the 2011 refresher training. The trainees were also critical of holding questions until designated question-and-answer sessions (mentioned by 6.5% of trainees). This is not surprising, but unavoidable given the need to ensure that all training material was covered in a timely manner over the one-and-a-half-day sessions. Other dislikes mentioned by the trainees included "the training was not long enough" (mentioned by 4.6% of trainees), "cramped work space" (mentioned by 4.3% of trainees), and "too much time was spent on practice materials; only need to go over new material" (mentioned by 4.3%). These are concerns that can be addressed in preparation for the 2011 refresher training.

4. Pre-Training/Post-Training Comparisons of Survey Performance and Data Quality Indicators

While the interviewer evaluations were positive, the question remained as to whether the training, with its heavy emphasis on data quality, translated into performance and quality improvements in the field. In this section, we present preliminary pre-training/post-training comparisons for a set of survey performance and data quality indicators.

4.1 Data and Analysis

For the comparisons, the pre-training period covered calendar quarter four (October, November, and December) of the 2009 data year, while the post-training period covered

the last two weeks of January (immediately following the refresher training) and all of February and March of 2010 (calendar quarter one of 2010). The sample sizes for the two periods were 23,460 cases (pre-training) and 17,012 cases (post-training).

The unit of analysis for the comparisons is the "case,"[6] which is the equivalent of a family unit in the NHIS. Therefore, each record on the data file represents a family. The analysis was limited to cases from 592 interviewers who worked in both the pre-training and post-training periods. It is important to note that at the time of this analysis, information was not available on which interviewers attended the refresher training. Given the total number of attendees, we estimate that over 80% of the 592 interviewers with pre-training and post-training workloads attended the training. Nonetheless, the comparisons were influenced by data from interviewers who did not attend the training.

The analysis consisted of comparing pre-training indicator estimates to post-training indicator estimates. Two-tailed t-tests were used to determine if the differences between the pre-training and post-training figures were statistically significant. All analysis was unweighted and performed in SUDAAN (Research Triangle Institute, 2005) to account for the complex sample design.

4.1.1 Survey Performance and Data Quality Indicators

In total, we compared pre-training and post-training estimates for 25 performance and data quality indicators. The first set of indicators, constructed from survey paradata (data about the data collection process), included the response rate (AAPOR Response Rate 6), cooperation rate (AAPOR Cooperation Rate 2), rate of first contact attempts by telephone[7], rate of first contact attempts during weekday evening hours,[8] rate of first contact attempts in the third (and last week) of the interview period,[9] contact rate at first contact attempt, first contact cooperation rate, and percentage of cases with no contact history data (also referred to as paradata).

The last two sets of indicators were based on data collected in the household composition module and the family interview. The first set included the percentage of family interviews administered primarily by telephone; the break-off rate in the family interview; the use of fake names or aliases in place of real names;[10] item nonresponse

---

[6] For a participating household, the household respondent answers questions in the household composition module, which includes a rostering of all household members. Each unrelated family in a household becomes a case and is interviewed separately. Roughly 98% of households contain one family.

[7] As noted, the NHIS is an in-person interview survey. The interview protocol is to make the first contact with a sample household in person before use of the telephone is permissible.

[8] Research has consistently shown weekday evening hours to be among the best times to make contact with sample households (Groves and Couper, 1998; Dahlhamer et al., 2006). Since an NHIS interview assignment period always starts on a Monday, making initial contact attempts during weekday evening hours is a recommended contact strategy.

[9] An interview assignment period in the NHIS is 17 days in length. Waiting until the third week of the interview assignment period to make the first contact attempt on a household significantly reduces the available time to secure participation and complete all four main interview modules.

[10] NCHS maintains an active data linkage program whereby survey records are linked to administrative records. Name, of course, is a significant match variable. Interviewers are strongly encouraged to collect the real names of all household members, but respondents have the option to refuse.

rates for questions on a home telephone number, total family income for the prior calendar year, and the family respondent's earnings from the prior calendar year; percentage of cases with any item nonresponse among a set of cell phone questions; and the percentage of cases with any item nonresponse to family-level questions (asked of all families) within the family interview (excluding the total family income question).

The final set of indicators, which also used paradata, focused on the average time per question (in seconds) for a set of cell phone questions, the total family income question, the entire family interview, and the following sections of the family interview: the health status and limitations section, the injuries and poisonings section, the health care access and utilization section, the health insurance section, the sociodemographic section, and the income and assets section. We log transformed the time measures to correct for highly right-skewed distributions.

4.1.2 Comparisons of Pre-Training and Post-Training Sample Compositions

Because no attempt was made to randomize cases or interviewers into a treatment (receive the new training procedures) or control (receive old refresher training procedures) group, it was important to assess the extent of equivalence between the pre-training and post-training samples (after removing data from interviewers who did not work both the pre-training and post-training periods). Comparisons were performed (table not shown) for several variables including Census region of residence, metropolitan statistical area (MSA) status, Census Regional Office, a set of family-level measures from completed family interviews (e.g., total family income, own or rent residence, total number of persons in the family), and a set of family interview respondent characteristics (e.g., age, sex, race/ethnicity). In total, 62 comparisons were performed with only one significant difference emerging between the compositions of the pre-training and post-training samples.[11] The lack of significant differences between the pre-training and post-training sample compositions bolsters our confidence that any significant improvements we observe in the performance and data quality indicators (pre-training to post-training) may be attributable to the training.

4.2 Results

Table 2 presents the results of the indicator comparisons. For each indicator, we present the pre-training estimate and standard error and the post-training estimate and standard error. The final column of Table 2, labelled "Imp.", provides an indication if the post-training estimate was an improvement, whether or not significant, over the pre-training estimate. If an improvement was observed, "yes" appears in the "Imp." column for that indicator.

In total, only 4 of the 25 performance and data quality indicator comparisons produced significant differences.[12] Of the four significant findings, two involved paradata-based indicators. We observed a significant decline in the rate of first contact attempts made during the third and final week of the interview period (from 1.9% pre-training to 1.2%

---

[11] The pre-training period included a significantly higher percentage of cases where the family interview respondent was between the ages of 18-24.

[12] While more significant improvements in performance and data quality were anticipated, the direction of change in the indicators from pre-training to post-training indicated some level of improvement for 19 of the 25 indicators.

Table 2. Case-Level Analysis[1] of Survey Performance and Data Quality Indicators, Pre-training (Quarter 4, 2009) and Post-training (Quarter 1, 2010) Periods: National Health Interview Survey (unweighted)

Indicator	# of Cases	Pre-training Estimate (S.E.)[2]	Post-training Estimate (S.E.)[2]	Imp.[3]
Response rate	26,248	80.3% (0.72)	81.1% (0.68)	Yes
Cooperation rate	24,923	84.9% (0.54)	84.9% (0.55)	No
First contact attempts by Telephone	40,022	3.9% (0.43)	3.7% (0.39)	Yes
First contact attempts during weekday evening hours	39,933	31.9% (1.14)	33.5% (1.06)	Yes
First contact attempt in third week of interview period	39,766	1.9%* (0.22)	1.2%* (0.19)	Yes
Contact rate at first contact Attempt	26,124	43.9% (0.66)	42.3% (0.72)	No
First contact cooperation rate	24,778	62.8%* (0.77)	59.4%* (0.78)	No
Cases with no contact history Records	40,473	1.1% (0.15)	1.1% (0.18)	No
Family interview administered primarily by telephone	21,139	20.2% (0.94)	19.3% (0.99)	Yes
Breakoff rate in family Interview	21,911	1.7% (0.17)	1.6% (0.18)	Yes
Use of fake names in place of real names	21,165	2.7% (0.21)	2.5% (0.24)	Yes
Telephone number nonresponse	21,165	3.8% (0.37)	3.6% (0.32)	Yes
Total family income question nonresponse	20,778	22.4% (0.69)	21.1% (0.66)	Yes
Family respondent earnings nonresponse	13,700	16.5% (0.70)	15.6% (0.76)	Yes
Any item nonresponse among cell phone questions	21,165	1.7% (0.19)	1.4% (0.23)	Yes
Any item nonresponse in family interview (main screener questions, excluding total family income question)	21,164	8.1% (0.33)	7.7% (0.43)	Yes
Natural log of mean seconds per question:				
Cell phone section	21,164	2.17 (0.01)	2.18 (0.01)	Yes
Family interview	21,165	2.17 (0.01)	2.18 (0.01)	Yes
Family health status and limitations section	21,162	2.07 (0.02)	2.08 (0.02)	Yes
Family injury and poisoning Section	21,163	2.39* (0.02)	2.44* (0.02)	Yes
Family access and utilization Section	21,161	2.32 (0.02)	2.31 (0.01)	No
Family health insurance section	21,163	2.21 (0.02)	2.24 (0.01)	Yes

Table 2. (continued)

Indicator	# of Cases	Pre-training Estimate (S.E.)[2]	Post-training Estimate (S.E.)[2]	Imp.[3]
Family sociodemographic Section	21,164	2.09 (0.01)	2.09 (0.01)	No
Family income and assets Section	20,894	1.85 (0.02)	1.86 (0.02)	Yes
Total family income question	20,780	2.58* (0.02)	2.64* (0.02)	Yes

* $p < .05$ for two-sided t-test comparing pre-training and post-training results.
[1] The "case," which is the equivalent of a family unit in the NHIS, is the unit of analysis. As an example, the case-level response rate was calculated by taking the number of fully complete and sufficient partial interview cases or families and dividing by the number of eligible cases or families.
[2] S.E. = Standard Error
[3] Imp. = Improvement. A "yes" in the improvement column indicates that the post-training estimate is an improvement, whether or not significant, over the pre-training estimate.

post-training). For an NHIS case, the interview assignment window lasts 17 days. A first contact that occurs earlier in the interview period is more likely to have a successful outcome than one that occurs later.[13]

The second paradata-based measure for which we observed a significant difference pre-training to post-training was the first contact cooperation rate. Unlike the previous finding, however, the significant difference marked a reduction in performance, with the post-training first contact cooperation rate being significantly lower (59.4%) than the pre-training first contact cooperation rate (62.8%). Why we observed this decline is unclear. Possible explanations include a shift of first contacts, pre-training to post-training, into day-time combinations less amenable to securing participation, and/or interviewers stopping interviews (and scheduling return visits) when confronted with interviewing conditions conducive of poor data quality (e.g., non-attentive respondent, loud noise and distractions, etc.). The latter interpretation may constitute improvement, especially if the final outcomes for the post-training period were better quality interviews.

The remaining significant findings involved our time indicators. First, the log-transformed measure of the average time per question (in seconds) for the family injury and poisoning section significantly increased from pre-training to post-training. Similarly, the log transformed measure of the average time (in seconds) spent on the total family income question significantly increased from pre-training to post-training. Both represent improvements and, we hypothesize, are directly tied to specific training materials and presentations. As noted in the introduction, inordinately brief interview times and concerns over the short-cutting of questions were among the factors providing the impetus for a revision in training procedures. Considerable training time, including a presentation by a staff member with the NCHS Questionnaire Design Research Laboratory, was devoted to the importance of reading questions as worded.

[13] Though difficult to discern with the available data, it is also possible that the significant finding for this indicator represents improved recording of contact histories by interviewers, rather than a true reduction in the percentage of first attempts in the third week of the interview period.

5. Discussion

For 2010, several important changes were made to the NHIS interviewer refresher training. Chief among them was a move from a highly decentralized (~30 training sessions in the 12 Census Regional Offices) to a highly centralized format (4 training sessions held in 3 locations). Again, this proved advantageous in that NCHS and Census Bureau subject matter experts were able to conduct the training, ensuring a consistent and high quality delivery of information. In addition, the training began with a series of presentations on data quality, and maintained a quality assurance perspective throughout.

Feedback from interviewers, via evaluation forms completed at the close of each training session, was clearly favorable and reinforced the decision to move to a centralized format. As such, the decision has been made to adopt the centralized format for the upcoming 2011 NHIS refresher training. Trainees also provided critical and insightful feedback on the training, information that is being used to prepare for 2011 refresher training. In particular, greater effort will be made to ensure that presentations are consistently engaging throughout the training session.

Positive evaluations aside, analysis to date has revealed few significant impacts of the revised training procedures on performance and data quality. There are a number of possible explanations. First, this was an observational study or natural experiment. As we noted previously, neither cases nor interviewers were randomized into a treatment or control group. How a formal experimental design would have altered our results is unclear. However, it would have been too expensive to conduct both, and, cost aside, there was clear consensus, given data quality concerns, to have all interviewers participate in the new training.

Second, and as noted previously, at the time of the analysis we were unable to identify interviewers who did not participate in training. Again, we estimate that 80% or more of the interviewers who contributed cases to this analysis did attend the training. It is possible, however, that the interviewers who did not attend contributed a disproportionate number of cases to the analysis and had a significant impact on the post-training indicator estimates. If the training had the intended, beneficial effect on performance and data quality, the inclusion of these cases could certainly mitigate our ability to identify significant differences.

Finally, it is important to note that the training occurred during a period of performance improvement, potentially making it more difficult to observe significant, positive effects of the training procedures. As early as 2008, NCHS and Census staff began working on a system of performance and data quality monitoring, both at the interviewer and case level. Fully implemented, this system has proved highly effective, best evidenced by a reduction of interviews removed from NHIS data files, prior to release, due to data quality problems.

For a more thorough training evaluation, a number of future steps are being planned. First, a list of training attendees has recently been made available, enabling a replication of the analysis presented in this paper, but limited to attending interviewers and their cases. In addition, we are currently exploring pre-training/post-training comparisons of the performance and data quality indicators at the interviewer level. The analysis is consistent with a one-group pre-test/post-test design where we have a pre-training observation (e.g., pre-training cooperation rate) and a post-training observation (post-

training cooperation rate) on each interviewer. Paired sample t-tests and Wilcoxson signed-rank tests are being used to test for significant differences in pre-training and post-training indicator estimates. So far, and consistent with the case-level analysis, these tests have yielded few significant findings.

Beyond these analyses, the best approach to this evaluation may involve multi-level models where sample cases are nested within interviewers. Such an analysis would enable the estimation of interviewer effects while controlling for case-level characteristics as well as changes in interviewer case loads from pre-training to post-training. Additionally, and data permitting, propensity score matching techniques could be explored. These techniques would allow for quasi-experimental contrasts between interviewers in naturally occurring "treatment" and "control" groups, but who display similar likelihoods of experiencing the treatment based on their observed characteristics (Rosenbaum and Rubin, 1985). More formal consideration of these evaluation techniques will also be considered for the 2011 refresher training.

References

Biemer, P. P. and L. E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons, Inc.

Billiet, J. and G. Loosveldt. 1988. "Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly*, 52: 190-211.

Dahlhamer, J. M., B. J. Stussman, C. M. Simile, and B. Taylor. 2006. "Modeling Survey Contact in the National Health Interview Survey." *Proceedings of Statistics Canada Symposium 2005*.

Fowler, F. J. and T. W. Mangione. 1990. *Standardized Survey Interviewing*. Thousand Oaks, CA: Sage Publications.

Groves, R. M. and M. P. Couper. 1998. Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, Inc.

Groves, R. M. and K. A. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics*, 17(2): 249-265.

O'Brien, E. M., T. S. Mayer, R. M. Groves, and G. E. O'Neill. 2002. "Interviewer Training to Increase Survey Participation." Pp. 2502-2507 in the *Proceedings of the Joint Statistical Meetings – Section on Survey Research Methods*.

Research Triangle Institute. 2005. *SUDAAN 9.0 Language Manual*. Research Triangle Park, NC: Research Triangle Institute.

Rosenbaum, P. R. and D. B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician*, 39(1): 33-38.

U. S. Census Bureau. 2010. "2010 NHIS Refresher Training Evaluation." Washington, DC: Methods Research Branch Memorandum No. 10 – 05.