The Intersection of Response Propensity and Data Quality in the National Health Interview Survey (NHIS)

James M. Dahlhamer
National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782

Abstract
Research suggests that the factors influencing the decision to participate in a survey may also influence the respondent's motivation and ability to respond to survey questions. If response propensities are positively correlated with respondent effort during the interview, the participation of reluctant respondents may reduce the quality of estimates. Thus, understanding the associations and common causes of nonresponse and measurement error is essential for reducing total survey error and designing high quality surveys.

Using contact history, sample frame, Census 2000, and limited interview data, we estimate a response propensity model for the sample adult module of the 2010 National Health Interview Survey. Response propensities from the model are categorized into quintiles. We then present various quality indicators (e.g., item nonresponse, response consistency) by response propensity quintile. Factors that may affect both response propensity and data quality are explored.

Key Words: Survey nonresponse, response propensity, measurement error, data quality

1. Introduction[1]

With continuing declines in survey response rates and the potential resulting harm to survey estimates, considerable attention has been devoted to the relationship between response propensity and nonresponse bias. Much less attention, however, has been given to the relationship between response propensity and survey measurement error. To the extent that individuals' response propensities are positively correlated with their level of effort during the response process, securing the participation of reluctant individuals will increase measurement error and reduce the quality of estimates.

Cannell and Fowler (1963) were among the first to identify a link between participant reluctance and measurement error. Respondents who participated at the end of the interview period provided less accurate reports of hospital stays than those who responded earlier. Cannell and Fowler attributed this finding to low respondent motivation, what Krosnick (1991) later termed satisficing—providing the minimum response necessary to allow the interview to proceed. Among other indicators of quality, Krosnick (1991) suggested that satisficing results in increased item refusals or "don't

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

know" responses, the choosing of more socially desirable responses, and less complete responses to open-ended questions.

More recent research has also demonstrated a response propensity-measurement error link. Blair and Chun (1992) found that converted refusers were more likely than initial cooperators to provide item "don't know" and "refused" responses, and to produce interviews of shorter duration. Mason, Lesser and Traugott (2002) reported item-missing data in a quarter of converted refusal cases compared to just 11% of nonrefuser cases, while Triplett et al. (1996) identified higher rates of item nonresponse among converted refusers in a time-diary survey. Furthermore, when asked to record all activities from a 24-hour period, converted refusers recorded 5.5% fewer activities than initial cooperators.

Similar results have been reported in studies analyzing householder statements of reluctance made during survey introductions (Campanelli et al., 1996; Couper, 1997; Dahlhamer et al., 2008). Respondents stating "too busy" or "not interested" were more likely to break off the interview, produce item refusals and don't knows, produce shorter item and section times, and were less likely to consent to future survey participation. Each of these studies concluded that statements made at the doorstep convey important information about a respondent's likely level of engagement in and commitment to the interview.

A limitation of much of this research is the reliance on simple dichotomous indicators of participant reluctance. Two recent studies utilized predicted probabilities from logistic regressions of survey participation to explore the response propensity-measurement error nexus. Olson (2006) examined the separate impact of contact and cooperation propensities on estimates of length of marriage and time since divorce. Securing less cooperative respondents increased measurement error, while the inclusion of hard-to-contact respondents led to reductions in measurement error and total survey error. Fricker and Tourangeau (2011) identified linear trends between nonresponse propensity quintiles and item nonresponse, missing diary reports, and rounding of activity durations. In sum, data quality decreased as the probability of nonresponse increased.

In this paper, we explore the relationship between response propensity and data quality among sample adult participants in the 2010 National Health Interview Survey. A response propensity score (predicted probability from a logistic regression) was assigned to each participant, participants were divided into propensity quintiles, and bivariate and multivariate analyses examining the relationship between response propensity and data quality were performed. Where evidence of covariation emerged, we explored possible common causal factors underlying the relationship. For guidance, we drew from three theoretical approaches to survey participation: social integration/social capital, topic interest, and busyness/discretionary time (see Groves and Couper, 1998, for a thorough discussion of these theoretical approaches). As Fricker and Tourangeau (2011) argued, higher levels of social capital could activate stronger norms of survey cooperation, and those norms could also influence respondents' willingness to engage in more careful processing of survey questions. Similarly, interest in the survey topic may dispose individuals to agree to a survey request and also stimulate careful processing of the survey items. And finally, time stress could produce a general disinclination to both participate and to respond accurately. Identifying and statistically controlling for shared explanatory variables would provide a means for reducing or eliminating bias in key survey estimates.

2. Data and Methods

2.1 National Health Interview Survey

The National Health Interview Survey (NHIS) is a multi-purpose survey of 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). It has been in the field virtually continuously since 1957. Utilizing a multistage, clustered sample design, with oversampling of black, Hispanic, and Asian persons, the NHIS 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 each data collection year.

Data are collected by roughly 750 trained interviewers with the U. S. Census Bureau using computer assisted personal interviewing (CAPI). In 2010, interviews were conducted in 34,329 households, yielding data on 89,976 persons.

The core survey instrument contains four main modules: Household Composition, Family, Sample Child, and Sample Adult. For the household composition module, a household respondent provides basic sociodemographic information on all members of the household. Within each family, the family module is completed by a family respondent who provides health information on each member of the family. Additional health information is subsequently collected from one randomly selected adult (the “sample adult”) aged 18 years or older and from the parent or guardian of one randomly selected child under age 18 (the “sample child”).

2.2 Examining the Response Propensity-Data Quality Link

Data on interviewed 2010 NHIS families, including county-level 2000 Decennial Census, sample frame, contact history, and responses to the family module, were used in a logistic regression analysis of participation among eligible sample adults (n=35,153). By focusing on interviewed families, a variety of measures could be constructed for both responding and nonresponding sample adults.[2] The analysis examining the link between response propensity quintiles and data quality was limited to participating sample adults (n=27,157). All analysis was weighted using the sample adult base weight and performed in SUDAAN (Research Triangle Institute, 2005) to account for the complex sample design.

2.2.1 Model of Sample Adult Participation

To assign a response propensity score to each sample adult participant we fitted a logistic regression model of sample adult participation. The dependent variable was defined as “interview” (n=27,157; 77.3%) versus “noninterview” (n=7,996; 23.7%). Ideally, we would have explored contact propensities among eligible sample adults, and then explored cooperation propensities conditional on contact. While we have information on

[2] It is important to note that the final family response rate for 2010 was 78.7%. Thus, we have no information on adults from the 21% of families that failed to participate.

nonparticipation for many sample adults, the information is not easily codable into traditional nonresponse categories.

Table 1 presents the significant predictors ($p < .05$) from the model of sample adult participation, grouped by theoretical approach. Overall, the model explained approximately 24% of the variance in participation. Table 2 presents the participation rates by response propensity quintile (all eligible sample adults). Participation rates ranged from a low of 41.4% in the lowest propensity quintile to 97.3% in the highest propensity quintile. To facilitate the examination of a possible response propensity-data quality link, we then divided the **responding** sample adults into response propensity quintiles.

Table 1. Significant Predictors from Model of Sample Adult Participation

Theoretical Approach	Predictor
Social Integration/Social Capital	
	Age
	Sex
	Education
	Marital status
	Born in the U.S.
	Total family income
	Children in the family
Busyness/Discretionary Time	
	Householder mention of time constraints
	Employment status
Topic Interest	
	Functional limitation
	Delayed care due to cost in past 12 months
	Received care in past 2 weeks
	Injury/poisoning episode in past 3 months
Social Environment and Paradata Covariates	
	Region of residence
	Item nonresponse in family interview
	Sample adult also family respondent
	Mode of family interview
	Number of noncontacts prior to first contact
	Mention of hard refusal concerns
	Mention of
	Mention of gate-keeping concerns
	Case reassigned to different interviewer

Table 2. Participation Rates by Response Propensity Quintile

Propensity Quintile	Percent Participated in Sample Adult Interview
1 (low)	41.4
2	71.6
3	83.2
4	92.8
5 (high)	97.3

2.2.2 Associations between Response Propensity and Data Quality

We examined three indicators of data quality: item nonresponse, response inconsistency, and the completeness or length of open-ended responses. A dichotomous indicator of item nonresponse measured whether or not the adult interview contained 10 or more missing responses to 158 questions asked of all sample adults. In addition to don't know and refused responses, not ascertained outcomes were also included in the measure. Since the analysis is limited to participating sample adults, not ascertained on a question indicates that the sample adult proceeded far enough into the interview to receive a sufficient partial disposition code, but did not fully complete the interview, including the item in question.

The indicator of response inconsistency utilized questions on seeing or talking to doctors/health care professionals. Sample adults are first asked if they had seen or talked to various providers in the past 12 months (we focused on 11 providers for this measure). The question structure is as follows: "DURING THE PAST 12 MONTHS, that is since {12 month reference date}, have you seen or talked to any of the following health care providers about your own health? …<provider>." Later in the interview, respondents receive the following question: "About how long has it been since you last saw or talked to a doctor or other health care professional about your own health? Include doctors seen while a patient in the hospital." Combining the first two response options for this question covers the past 12 months. We defined an inconsistent response as one in which the sample adult reported seeing or talking to one of the 11 specific providers in the past 12 months, but at the more global question said they had not seen or spoken to a doctor or other health care professional in the past 12 months.

Finally, our measure of completeness of open-ended responses was based on four questions that ask about a current job, a most recent job, or the job held the longest by the sample adult. Interviewers are asked to record the responses verbatim. The recorded information is later used to categorize the respondent's job into detailed occupation and industry codes for subsequent analysis. By respondents providing more detailed responses, respondent jobs can be coded into more detailed industry and occupation categories. To create our measure, we summed the number of characters contained in the recorded responses to the four questions.

To assess the relationship between data quality and response propensity, these three quality indicators were examined across the response propensity quintiles. Then, two-tailed t-tests (conducted at the .05 level) were performed to compare quintile-specific estimates. When a significant association was identified, we then explored the extent to which controlling for possible common causal variables impacted the association. For the item nonresponse and response inconsistency indicators, we performed a series of logistic regressions. For the measure of response completeness we performed a series of OLS regressions. Covariates included in these models were the significant predictors from our model of sample adult participation (see Table 1). For each of the three data quality indicators, the first model included just the response propensity measure. The second model included the response propensity measure and the social integration/social capital measures. The third model included the response propensity measure and the busyness/discretionary time measures, while the fourth model included the response propensity measure and the topic interest measures. The fifth and final model included all measures from the previous four models along with region of residence and a set of paradata measures.

3. Results

Figure 1 presents the percent of sample adult interviews with 10 or missing responses by response propensity quintile. Sample adults in the lowest response propensity quintile (quintile 1) had the highest rate of 10 or more missing responses (16.5%). The rate of missing fell considerably to 9.0% for propensity quintile 2, and then declined more gradually to a low of 2.9% for the highest response propensity quintile (quintile 5). The estimate for the lowest response propensity quintile (quintile 1) was significantly different from the estimate for each of the other four quintiles, and significantly different

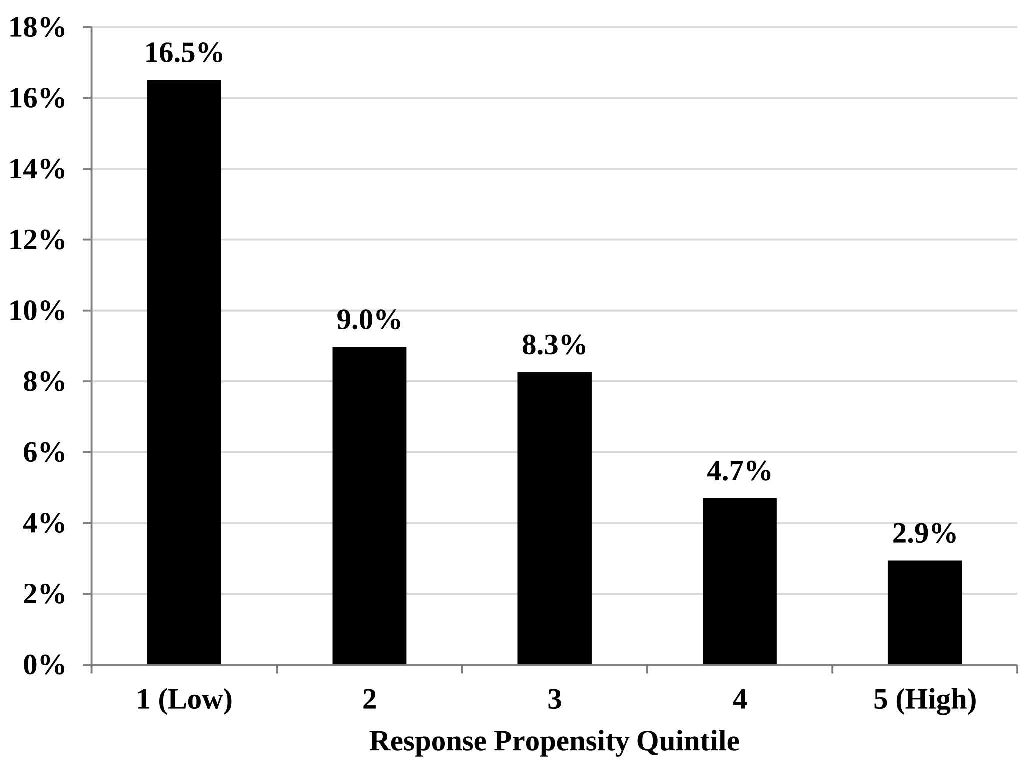

Figure 1. Percent of Sample Adult Interviews with 10 or More Missing (don’t know, refused, not ascertained) Responses (158 items), by Response Propensity Quintile: NHIS, 2010

from the estimate (6.4%) for the other four quintiles combined. Consistent with the bulk of prior research, the lowest response propensity adults, or those most difficult to recruit, produced the highest rates of item nonresponse.

Figure 2 presents the percent of sample adult interviews with inconsistencies in responses to questions on talking to or seeing a doctor or other health professional in the past 12 months. Here a pattern is less discernible than that observed for item nonresponse. The estimates for response propensity quintiles 1-4 are similar, ranging from 4.8% to 5.5%. Comparisons of estimates among these four quintiles revealed no significant differences. However, the highest response propensity quintile (quintile 5) had the lowest inconsistency rate at 2.8%. The estimate for the highest (quintile 5) response propensity adults was significantly different from each of the estimates for the other four quintiles, and significantly different from the estimate (5.1%) for adults in the other four quintiles combined. To summarize, the highest response propensity adults, or those easiest to recruit, were less likely to produce inconsistent responses to questions on provider consultations in the past 12 months.

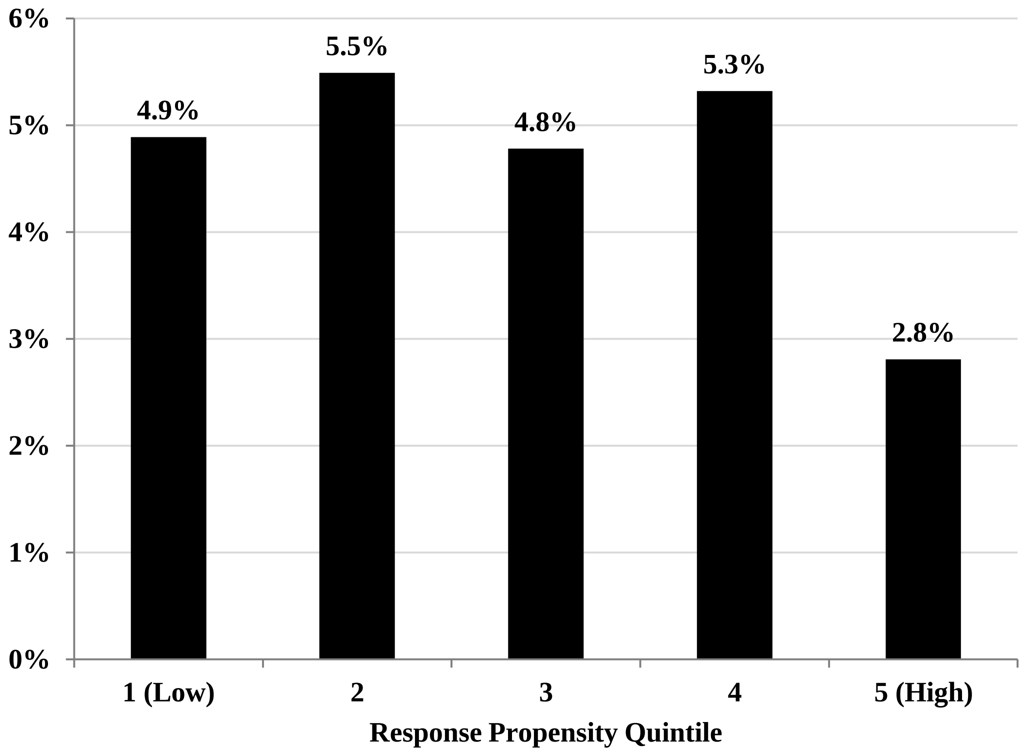

Figure 2. Percent of Sample Adult Interviews with an Inconsistency in Responses to Questions on Talking to/Seeing a Doctor or Other health Professional in the Past 12 Months, by Response Propensity Quintile: NHIS, 2010

Figure 3 presents the average characters recorded per industry/occupation description among sample adults who ever worked, by response propensity quintile. While the differences are not large, there is a clear trend of increasing characters per description when moving from the lowest response propensity quintile (quintile 1; mean=64.8

characters) to the highest response propensity quintile (quintile 5; mean=72.9 characters). As anticipated, lower response propensity adults were less verbose with information on their jobs. As observed with item nonresponse, lower response propensity adults (quintile 1) provided significantly shorter industry/occupation descriptions than adults in each of the other four propensity quintiles, and significantly shorter descriptions than adults in the other four quintiles combined (mean=70.1).

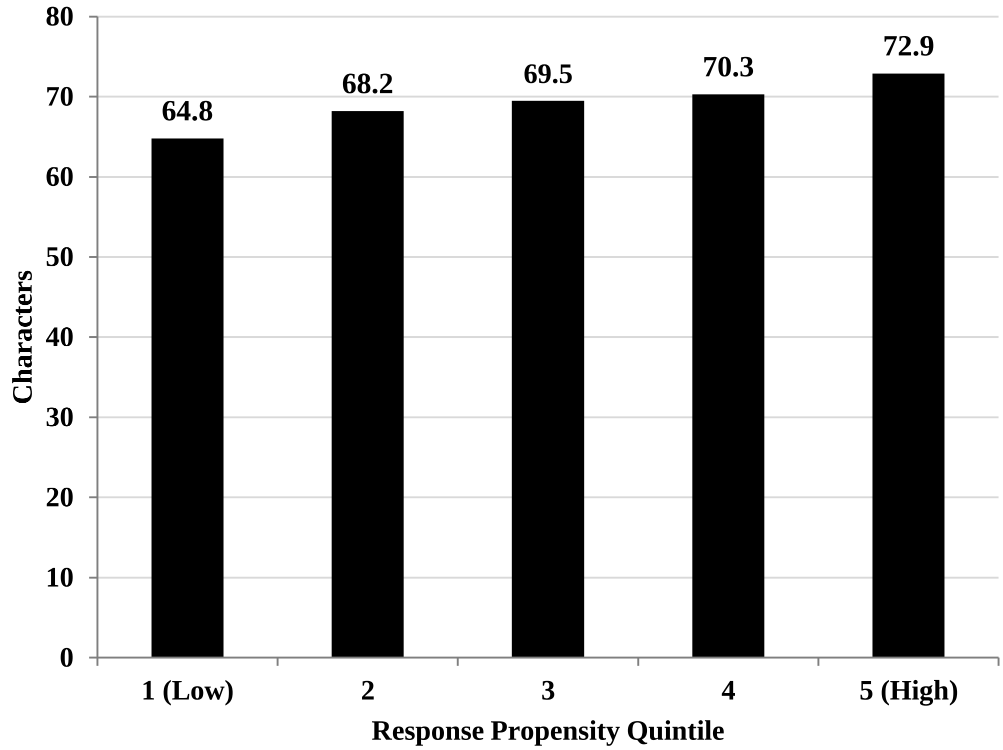

Figure 3. Average Characters per Occupation/Industry Description among Sample Adults Who Ever Worked, by Response Propensity Quintile: NHIS, 2010

In sum, we identified significant associations between response propensity quintiles and each of the three measures of data quality. Next, we attempted to eliminate those relationships by controlling for possible common causal variables in a series of multivariate analyses. As stated previously, identifying and statistically controlling for common causal variables would provide a means for reducing bias in key survey estimates.

Table 3 presents odds ratios for response propensity quintiles from a series of logistic regressions predicting 10 or more missing responses. The unadjusted odds ratios for response propensity quintile (model 1) validate the graphical analysis presented in Figure 1. Compared to the highest (5th) response propensity quintile, sample adults in the lowest (1st) response propensity quintile had significantly greater odds (Unadjusted Odds Ratio=6.54) of 10 or more missing responses to questions (158 total) in the sample adult interview. Surprisingly, the addition of social integration/social capital measures in the second model, and topic interest measures in model 4, enhanced rather than attenuated this association. The addition of busyness/discretionary time indicators (model 3) attenuated the relationship somewhat, but the association remained significant. More importantly, the association between response propensity and 10 or more missing responses remained significant in the full model (model 5).[3]

Table 3. Odds Ratios for Response Propensity Quintiles from Logistic Regressions Predicting 10 or More Missing (don't know, refused, not ascertained) Responses (158 items): NHIS Sample Adults, 2010

Propensity Quintile	Model 1: Unadjusted	Model 2: Social Isolation/ Social Capital	Model 3: Busyness/ Discretionary Time	Model 4: Topic Interest	Model 5: Full Model
1: Low	***6.54***[1]	***9.68***	***2.64***	***7.81***	***2.31***
2	***3.26***	***4.54***	***1.95***	***3.83***	***1.91***
3	***2.98***	***3.65***	***1.74***	***3.35***	***1.71***
4	***1.63***	***1.93***	***1.33***	***1.82***	***1.32***
5: High (reference)	1.00	1.00	1.00	1.00	1.00

[1] Odds ratios in bold and italics indicate $p < .05$.

Table 4. Coefficients for Response Propensity Quintiles from OLS Regressions Predicting Length (number of characters) of Industry and Occupation Descriptions: NHIS Sample Adults, 2010

Propensity Quintile	Unadjusted	Social Isolation/ Social Capital	Busyness/ Discretionary Time	Topic Interest	Full Model
1: Low	***-8.09***[1]	***-7.43***	***-8.86***	***-7.75***	1.76
2	***-4.74***	***-4.35***	***-5.58***	***-4.50***	0.64
3	***-3.44***	***-3.29***	***-4.03***	***-3.26***	0.63
4	***-2.64***	***-2.36***	***-3.28***	***-2.51***	-0.12
5: High (reference)	0.00	0.00	0.00	0.00	0.00

[1] Coefficients in bold and italics indicate $p < .05$.

[3] Note that we do not present results for our measure of response inconsistency. A series of logistic regressions yielded results consistent with those reported for the indicator of item nonresponse.

Table 4 presents coefficients for response propensity quintiles from a series of OLS regressions predicting the length of industry/occupation reports. Here again, the unadjusted coefficients for response propensity quintile (model 1) support the graphical analysis presented in Figure 3. Compared to adults in the highest (5th) response propensity quintile, adults in each of the other four propensity quintiles reported significantly shorter industry and occupation descriptions. Consistent with the results for 10 or more missing responses, separate models (2-4) including social integration/social capital measures, busyness/discretionary time measures, and topic interest measures had little impact on the association between response propensity and length of industry/occupation reports. However, once all covariates, including region of residence and a set of paradata measures, were entered into a full model, the effect of response propensity was reduced to nonsignificance. This suggests the potential to jointly address nonresponse and measurement error in industry and occupation codes.

4. Discussion

Like much of the previous research, we identified a relationship between response propensity and data quality. Low response propensity adults produced more item nonresponse and shorter industry/occupation descriptions, while high response propensity adults produced fewer inconsistent responses to survey questions. What are the implications of these findings for field work? Where do we draw the line on expending resources to secure the participation of difficult/reluctant households, families, and individuals? If nonresponse error is not significantly increased by excluding low propensity sample units, and these cases also contribute to increases in measurement error, then we may be able to focus resources elsewhere, potentially on other error-reduction techniques.

Findings from previous research on nonresponse bias in NHIS sample adult estimates have been mixed. One study, using simple dichotomous indicators of hard-to-contact and potential refusal cases, revealed little in the way of biasing effects of high effort interviews (Simile and Dahlhamer, 2006). A more recent study using response propensities from a model of sample adult participation suggested that the nonparticipation of low response propensity adults may be biasing estimates toward a higher prevalence of chronic and acute health conditions and greater access to and utilization of healthcare resources (Dahlhamer and Simile, 2009). If the former study is accurate, then the prescription for field work seems clear: spend less time and money on securing more difficult sample adult interviews. Otherwise, the likely outcome is an overall increase in total survey error due to increases in measurement error. The picture is less clear if the more recent study is accurate. More careful analysis of the trade-offs between reductions in nonresponse error and increases in measurement error would be warranted.

The findings from this study and similar research also have implications for level of effort analyses of nonresponse bias, especially those just mentioned for the NHIS. When examining respondent means, it is possible that measurement error in reluctant cases may conceal nonresponse bias (Fricker and Tourangeau, 2011). If lower response propensity individuals produce noisier data (increasing the variance of the statistic), then detecting differences between these individuals and higher propensity respondents becomes more difficult.

As a final note, there were limitations to this research that need to be addressed going forward. First, the model of sample adult participation was built using data collected from the 79% of families that participated in the NHIS. We know nothing about the sample adults from the 21% of families that did not participate. How that affects our results is unknown. Second, most of our indicators of data quality are indirect measures at best. The length of industry/occupation descriptions, for example, tells us nothing about the actual experiences in coding those descriptions and producing usable industry and occupation codes. Was there a higher rate of non-codable data among low response propensity adults? Were we forced to use more summary-level codes with this group? Future research should address these questions. And finally, the analysis presented here would have benefitted from an exploration of interviewer error. Confronted with reluctant and/or rushed respondents, interviewers may take shortcuts resulting in error that we have attributed to respondents. Plans are to re-fit the multivariate analyses of participation and data quality within a multi-level framework, enabling us to tease out the error contributions of interviewers.

References

Blair, J. and Y. I. Chun. 1992. “Quality of Data from Converted Refusers in Telephone Surveys,” Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL.

Campanelli, P., P. Sturgis, and N. Moon. 1996. “Exploring the Impact of Survey Introductions,” *Proceedings of the Joint Statistical Meetings*, pp. 1054-1059.

Cannell, C. F. and F. J. Fowler. 1963. “Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study,” *Public Opinion Quarterly*, Summer: pp. 25-264.

Couper, M. P. 1997. “Survey Introductions and Data Quality,” *Public Opinion Quarterly*, 61: 317-338.

Dahlhamer, J. M. and C. M. Simile. 2009. “Subunit Nonresponse in the National Health Interview Survey (NHIS): An Exploration Using Paradata.” *Proceedings of the Joint Statistical Meetings*, pp. 262-276.

Dahlhamer, J. M., C. M. Simile, and B. Taylor. 2008. “Do You Really Mean What You Say? Doorstep Concerns and Data Quality in the national health Interview Survey (NHIS).” *Proceedings of the Joint Statistical Meetings*, pp. 1484-1491.

Fricker, S. and R. Tourangeau. 2011. “Examining the Relationship between Nonresponse Propensity and Data Quality in Two national Household Surveys.” *Public Opinion Quarterly*, 74, pp. 934-955.

Groves, R. M. and M. P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.

Krosnick, J. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology*, 5: 213-36.

Mason, R., V. Lesser, and M. W. Traugott. 2002. “Effect of Item Nonresponse on Nonresponse Error and Inference.” R. Groves et al. (eds.), *Survey Nonresponse*. New York: John Wiley & Sons, Inc, pp. 149-161.

Olson, K. 2006. “Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias.” *Public Opinion Quarterly*, 70: 737-758.

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

Simile, C. M. and J. M. Dahlhamer. 2006. “Evaluating High Effort Measures for Studying Nonresponse Bias.” *Proceedings of Statistics Canada Methodology Symposium 2005.*

Triplett, T., J. Blair, T. Hamilton, and Y. C. Kang. 1996. “Initial Cooperators Vs. Converted Refusers: Are There Response Behavior Differences?” *Proceedings of the Joint Statistical Meetings*, pp. 1038-1041.