

Developing New Methods and Questions for Improving Response and Measurement on Sensitive Questions on the National Health Interview Survey

Peter S. Meyer, National Center for Health Statistics
James M. Dahlhamer, National Center for Health Statistics
John R. Pleis, National Center for Health Statistics

Peter S. Meyer
National Center for Health Statistics
Division of Health Interview Statistics
3311 Toledo Rd., Room 2207
Hyattsville, MD 20782
pmeyer1@cdc.gov

The content of this paper reflects the opinion of the authors, and not that of the Department of Health and Human Services, the Centers for Disease Control and Prevention, and the National Center for Health Statistics.

Abstract

The National Center for Health Statistics (NCHS) conducted a study during the second quarter of the 2006 data collection year of the National Health Interview Survey (NHIS) to test alternative income questions, new wealth questions, and respondent willingness to provide partial Social Security numbers. The field test took advantage of the oversampling feature of the NHIS sample design, which specifies that households that are designated for oversampling (15% of the 45,000 households) and that do not have at least one black, Hispanic, or Asian member are screened out, i.e., excluded from the sample. For this study, the usually-excluded households were interviewed and treated as an independent sample. We asked those respondents to answer the test questions along with some of the standard NHIS questions. The purpose of this study is to test methods for asking sensitive questions and determine the feasibility of using data from “screened out” households in future research.

1. Introduction

This paper introduces a new vehicle for testing survey questions and describes a field test that used the vehicle to test different methods for asking sensitive questions on the NHIS. The vehicle for testing questions was developed by tapping unused sample in the annual NHIS that is the result of the oversampling feature of the NHIS sample design. The questions were identified as being sensitive because they had consistent annual increases in their item nonresponse rates, and this threatened to

invalidate the data collected for those items. The items that we field tested using alternate methods were questions about income, wealth, and Social Security numbers. These types of questions are typically difficult to field because of respondent and field interviewer sensitivity to the subject matter. This paper provides a brief description of the NHIS and its file structure, defines some of the issues associated with these sensitive questions, and describes the research conducted to inform the formulation of questions and methods. The work on this study was spread out over almost five years, and involved staff throughout NCHS and the coordination of multiple funders throughout the Department of Health and Human Services.

2. Overview and File Structure of the NHIS

The NHIS is a cross-sectional in-person household survey and is conducted continuously, with microdata released annually. It is the principal source of information on the health of the civilian noninstitutionalized population of the United States and is one of the major data collection programs at NCHS. NHIS data are collected annually from approximately 35,000 households that include approximately 87,500 individuals. The NHIS collects data on a wide array of health and health related topics that include health status and behaviors, health care utilization, health insurance coverage, demographic information, and socioeconomic status.

The NHIS questionnaire is actually a series of semi-nested questionnaires that use both self

report and proxy response: the Household, Family, Sample Child, and Sample Adult questionnaires. The Household questionnaire is used to determine the number of families in the household and to create a roster of the names, ages, race/ethnicity and other demographic information for all of the individual family members. The Household questionnaire obtains a description of the family (or families). The Family questionnaire follows and asks questions about topics that include health insurance coverage, activity limitations, housing tenure, and income. These questions are asked of the person identified as the family respondent. All members of the household 18 years of age or older who are at home at the time of the interview are invited to participate and respond to the Family questionnaire for themselves, but these are largely proxy responses. The Sample Adult and Sample Child questionnaires ask more specific questions about health conditions and behaviors and are answered, respectively, by a randomly selected adult and by a knowledgeable adult who answers questions about a randomly selected child. In addition to the Core questionnaire, the NHIS has Supplements every year. These are topic-specific questions and usually go into much more depth. Some examples are the 2005 Cancer Control Supplement and the 2006 Complementary and Alternative Medicine Supplement.

3. Sensitive Measures

The item nonresponse rates for income, wealth, and Social Security number questions on the NHIS and other surveys have been increasing for the last several years. The suspected reason for this is that there is combined respondent and interviewer reluctance to broach subjects considered personal in nature. There are specific reasons why these data are important to collect:

- **Income:** The health economics literature points to a relationship between health and income. It is not possible to determine causality because health and income are mutually determined to some degree, but the relationship is still important. Income is currently asked about in the NHIS in terms of family total and then as a series of follow-up questions.
- **Wealth:** Likewise, there is a consistent relationship between health and wealth. There is also a good deal of evidence

that the relationship is net of income and that wealth is a better predictor of health for the elderly than income because the elderly often have higher wealth but lower income relative to younger people. The NHIS does not currently have any specific wealth questions except for a broad question in the Family questionnaire that asks whether the respondent's home is owned or rented.

- **Social Security numbers:** The number of respondents unwilling to provide Social Security numbers is climbing more quickly than the number unwilling to provide income information. This is strongly correlated with the increase in identity theft. The problem with not collecting a respondent's Social Security number is that the data from the interview cannot then be linked to other administrative datasets at a later date. At NCHS, the NHIS is periodically linked to data from the Medicare, Social Security programs, and other administrative data sources. The Social Security number question on the current NHIS asks for the full nine-digit number.

We hypothesized that we could reduce nonresponse rates and reduce the time it took to collect these data if we created new income and Social Security number questions. We also wanted to see how respondents reacted to the new battery of wealth questions.

4. Formulation of the Test Questions

To plan for research on improving these measures, an expert meeting was held and papers were commissioned. We then conducted a review of the literature, a full review of income and wealth items used on other Federal surveys, several rounds of cognitive testing of the questions as they were being developed, and a quantitative study using the National Health Interview Survey, 1993-1995: Family Resources Income and Assets Supplement as a data source. The Family Resources Supplement focused on employment, income from employment and businesses, other income sources including retirement and Social Security, and asset holdings such as cars, houses, businesses, and investment properties. During the quantitative study, we paid special attention to the wealth questions that consistently provided the lowest

item nonresponse and also to the variability in response rates to the different types of income questions.

The results of this research confirmed the need for new income questions and supported the development of a short battery of wealth questions. It was determined from this research that the best strategy for increasing item response on the income questions was to introduce a “bracketing” or “unfolding” approach using broad income ranges in the follow up questions asked after the initial question of total family income was refused. For example, if the respondents answers that they “Don’t Know” the total family income, the next questions will present the respondent with a bracketed choice such as “Is your total family income below \$50,000 or \$50,000 and above.” If the respondent answers the bracketed follow up question, the interviewer asks more follow up questions and attempts to draw the respondent to a more precise estimate of his or her family income. The broad income ranges also allowed for the collection of data that would be more useful analytically than the data that are generated by the current NHIS. For instance, in the current NHIS, income follow up questions, begin by asking the respondent if the family income is below \$20,000, or \$20,000 or above, and in the new questions, the bracketing begins at below \$50,000, or \$50,000 or above. The new cut point is more analytically useful because it approximates the median family income whereas the \$20,000 cut point is arbitrary and does not have any economic significance. We were also able to ask fewer questions on this sensitive subject by broadening the income range and eliminating twelve questions on source of income (see Pleis et al., 2007 for detail).

The development of new questions relating to wealth was a more difficult task because any addition of such questions to the survey is required to be budget neutral and interview time neutral, i.e., costs associated with their inclusion need to be covered, and some questions already on the survey would need to be removed if the wealth questions were to be adopted. Through the quantitative study we determined that the wealth items that had the highest response rates and were the most analytically tractable were total family wealth, home equity, and car ownership. The purpose of asking for home equity values was to provide a proxy for wealth when respondents refused to provide or did not

know their total family wealth. The car ownership question was intended to help identify respondents with the lowest levels of family wealth.

These modified or newly developed questions were tested by the NCHS cognitive laboratory, and changes in wording were made after consultation with staff from NCHS’ Office of Research and Methodology, resulting in two test questionnaires (two versions) modeled on the current Family questionnaire. The first questionnaire included the test income questions, which replaced analogous questions on the current NHIS Family questionnaire, and the new wealth questions. The second questionnaire maintained the current NHIS income questions and added the new wealth questions. Finally, on both questionnaires, a question that asked respondents to provide the last four digits of their Social Security number was included instead of the full nine-digit-number question. Respondents were randomly assigned to one of these questionnaires. This allowed for a split ballot to test the income questions. With this design, neither the wealth questions nor the Social Security number questions had benchmarks for comparison; there were not prior baseline questions in the case of the wealth questions, and there was not enough time to program the instrument to test the nine digit Social Security number against the four digit version.

5. Formulating the Sample

The sample used for the test was created from screened-out cases during the quarter two data collection period of the 2006 NHIS. Screening was introduced in the 1995 NHIS sample design as a cost effective way to oversample black and Hispanic populations and has recently been expanded to include more subgroups. A subset of the sample (15%) in specified geographic areas is designated as screening households: these are eliminated from the interviewed sample if after brief questioning it is determined that there are no black, Hispanic, or Asian persons in the household. There are approximately 3,150 households screened out in this manner in every quarter of the data collection year. Usually, when an interviewer ascertains that a household designated for screening does not have a member of the formerly mentioned racial/ethnic groups the interview is terminated. For this study, the test was conducted by continuing the interview.

The Computer Assisted Personnel Interview (CAPI) software seamlessly switched from the current NHIS to the test Family Questionnaire when a household was screened out of the NHIS. This allowed for a cost effective vehicle for testing the sensitive questions and for more efficient use of the NHIS sample.

6. Preliminary Results

The test questionnaire was fielded nationally between April and July of 2006. The test instrument and the random assignment algorithm performed as anticipated and produced a sample that was evenly split into the randomly assigned control and experimental groups (see Dahlhamer et al., 2007). The hypothesized results appear to have held. The new income items outperformed the old income items in terms of improved response rate and took less time to administer. The four digit Social Security number question on the test questionnaire also performed better than the nine digit Social Security number question on the current NHIS when response rates are compared. It is more difficult to interpret the performance of the wealth questions because they were new, and there were no older questions for comparison. However, in further analysis it appeared that the wealth questions added explanatory power in multivariate analyses of some health related variables. This was especially true for older respondents.

Subsequent to the field test, we conducted a series of focus groups with interviewers to augment the quantitative results. They confirmed that the sensitive questions in the current NHIS were problematic for both the interviewers and the interviewees. The repetitive income source questions and the family income follow up questions, which were eliminated or changed on the test questionnaire, were both noted by many of the participating interviewers as particular problems. It was also noted that recent news of Federal data being stolen and other reports of data theft had substantially and negatively impacted the response rates for the Social Security number questions. The consensus was that the new income and Social Security number questions were superior to those currently on the NHIS and that the method of bracketing employed in the test should be adopted. It was recommended that the wealth questions undergo more testing before they are included.

References

C Moriarity (2007). Sampling and oversampling in the National Health Interview Survey. Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association.

JM Dahlhamer, PS Meyer, and JR Pleis (2007). Questions people don't like to answer: Wealth and Social Security numbers. ASA Proceedings of the 2006 Joint Statistical Meetings.

A Deaton (2002). Policy implications of the gradient of health and wealth. Health Affairs, Vol 21, Issue 2, 13-30.

JR Pleis, JM Dahlhamer, and PS Meyer (2007). Unfolding the answers? Income brackets and income nonresponse in the National Health Interview Survey (NHIS). ASA Proceedings of the 2006 Joint Statistical Meetings.