Analyzing Field Notes Systematically to Better Understand Respondent Participation

Barbara J. Stussman[1] and Catherine M. Simile[2]

[1][2]National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782

1. Introduction

Previous research has found that researchers can benefit from the release of paradata, which are data that describe the interviewing process (Safir et al., 2001). For the 2006 data year, the National Health Interview Survey (NHIS) released its first public use paradata file. This paper focuses on two research topics using textual data from the NHIS paradata file: reasons for incomplete interviews and reasons for conducting interviews or parts of interviews via telephone. Specifically, this paper examines why there are incomplete interviews and why, in a face-to-face survey, some interviews or parts of interviews are completed by telephone. We used qualitative analysis to examine interviewer notes and responses to questions about the circumstances of the interview, and then used these results to improve the efficiency of gathering and analyzing data by redesigning the questions. Finally, we evaluated these redesigned questions. This paper describes the research conducted and steps taken to change field procedures and to make these data available for public use.

Key Words: paradata, textual data, qualitative analysis

2. Methods

2.1 Data Source

The NHIS is a nationally representative household survey of the general health of the civilian, non-institutionalized household population of the United States. It has been in the field continuously since 1957. The U.S. Census Bureau, which serves as data collection agent for the NHIS, has 12 Regional Offices throughout the country, where data collection activities are locally managed. While the NHIS is primarily a face-to-face interview, interviewers also rely on the telephone to administer some interviews or interview sections, usually after an initial face-to-face encounter. Procedures instruct that interviewers must attempt an initial face-to-face contact before interviewing by telephone, although, in some rare circumstances, such as in remote areas of the country, entire interviews may be conducted by telephone without an initial face-to-face encounter. Furthermore, respondents sometimes call interviewers as follow-up to notes the interviewers left at the door and request to have the entire interview administered by phone.

The NHIS is comprised of four main modules. The Household Composition module collects demographic information on all family members; the Family module collects general health information about all family members from a knowledgeable adult; the Sample Adult module collects additional health information from one randomly selected adult; and the Sample Child module (for families that have children) collects additional health information about one randomly selected child aged 17 or younger. A knowledgeable adult answers questions about the sample child. For 2007, the NHIS contained 29,266 eligible households, 29,915 families, 23,393 sample adults and 9,417 sample children. Supplemental modules are added on a yearly basis. The median time for an NHIS interview is about and hour, although this varies depending on the length of the supplements for the particular year. (For more information on the NHIS, please consult the survey documentation available at www.cdc.nchs/nhis.)

There are features of the data collection process that may have an impact on whether an interview is completed and/or whether a telephone follow-up was conducted. For example, because the sample adult module is self-reported, the interviewer has to contact a specific person, even if it requires a return visit to the household. In 40% of cases there are two or more respondents who answer some questions in the NHIS, usually because the sample adult is different from the family respondent. This increases the likelihood of that case being incomplete or of being completed over the telephone. Additionally, the length of the interview can have an impact on respondents' stopping the interview before completion. Analysis performed for this paper included data from the 2002, 2004, 2005, 2007, and 2008 NHIS.

2.2 Textual Data Collection

For both research areas we examined data from the interviewer debriefing questionnaire. This is a case-level questionnaire that interviewers fill in at the conclusion of the interview and includes questions about the data collection process. Textual entries are collected for a variety of topics including reasons for respondent reluctance, interviewer strategies for gaining participation, reasons for nonresponse or incomplete interviews, and reasons for conducting interviews or parts of interviews over the telephone. For the first research topic, reasons for incomplete interviews, interviewers are asked to select the main reason for incomplete interviews from a list. Open-ended responses were collected via an other specify option only if the pre-coded categories provided did not fit the situation. In 2002 (the data year we first analyzed), there were 2,065 textual entries describing reasons for incomplete interviews.

For reasons for conducting interviews over the telephone (our second topic), the interviewers are instructed to record the specific reason why any major section of the interview was conducted primarily by telephone. That is, the reasons for telephone use were collected exclusively from textual entries (no pre-coded categories). The design only allowed for one reason to be given even if more than one section was conducted via telephone. Therefore, except in the cases where only one section is completed by telephone, it is not possible to link the particular reason with the particular section that was conducted by telephone. Nevertheless, it is possible to determine generally the reasons given for why the telephone was used. In 2005, there were 10,461 textual entries describing reasons for conducting interviews over the phone. We analyzed two 5% samples using systematic sampling with a random start, totaling 1,043 entries. Results from samples 1 and 2 were very similar, suggesting that the categories generated were exhaustive.

2.3 Analysis

Our approach to this analysis entailed four steps. First, for step 1, we examined the original questions for their conceptual content and flow. We analyzed the quantitative and qualitative data to learn the interviewers' perspectives about the field situations and to determine whether the questions were collecting the data we want them to collect. Qualitative data were analyzed using the constant comparative method (Strauss and Corbin, 1990). We used this information to determine whether or not we could make the questions more efficient both in terms of data collection and in data production and release. For step 2, we redesigned the questions. These results were used to create new questions that were tested in the field as step 3. Finally, the process came full circle with the analysis of the redesigned questions, step 4. This four-step process was employed for both research areas. In addition to learning about the field situations and the efficiency of data collection, this new analysis also shed light on what we may have lost or gained by our redesign.

3. Results

3.1 Reasons for Incomplete Interviews

3.1.1 Step 1: Analysis of Original Questions on Reasons for Incomplete Interviews

We distinguish between two types of incomplete interviews: those that stop in progress (break-offs) and those where the interviewer needed to get additional information at a later time, but did not (partials, other than break-offs). The latter often occur because the sample adult respondent is not home or unavailable. By analyzing both the quantitative and qualitative data from 2002 questions about incomplete interviews, we discovered both design problems and problems with the pre-coded categories. The original question flow did not allow us to distinguish between reasons for partial interviews and reasons for break-off interviews. This is important because these reasons can be very different. A partial may involve contacting another person who is not available at the time of the initial interview. A break-off involves a person quitting the interview while in progress. Also, in most cases we could not determine in which section the partial or break-off occurred. Knowing in which section the incomplete interview terminated provides additional information as to the reason for an incomplete outcome. For example, if the interview is a partial because the sample adult section is not completed, this provides evidence that obtaining an additional respondent was a problem. Likewise, if a break-off occurs during the income section, this provides evidence that the income section could be too personal.

The 2002 NHIS included six pre-coded choices as well as an other specify category. The pre-coded categories included "No one home, repeated calls," "Language problem," "Too personal," "Too long," "No time,"

"Respondent sick," and "No reason." Our analysis revealed that these pre-coded categories were not sufficient. We found several categories that emerged from the analysis of the textual entries but were not listed in the 2002 pre-coded categories. The category "Respondent had to leave" captures situations where the respondent stops the interview to go somewhere such as picking up children. "Respondent does not like/trust government" captured entries such as "Angry. Government knows everything about us. Give me supervisor's name" and "Respondent said government is wasting money and doing nothing." The category "Respondent's relative (spouse/child) did not want her/him to participate" captured situations where the respondent's parent or spouse told the respondent not to participate. "Computer error" includes situations where the computer was caught in a loop, or otherwise malfunctioned. In addition to identifying these new categories, we found that some categories needed to be reworded or expanded. "Respondent Too Busy" replaced "No time" in order to clarify that this category refers to the respondent being busy and not the interviewer. Also, "or mentally" was added to "respondent physically unable to participate" based on textual entries.

3.1.2 Step 2: Question Redesign, and Step 3: Fielding New Questions

Three design changes to the questions on reasons for incomplete interviews went into the 2004 NHIS instrument. First, the new set of questions contains two paths funneled through a single question that asks, *"Indicate the reason interview is not complete."* The interviewer then picks from two response options: (1) Interview stopped in progress/break-off or (2) Completion before closeout not possible (other than break-off). Closeout refers to the end of the 17-day period during which any particular household can be interviewed. With this design, we can match the specific reasons with the type of incomplete interview, that is, with either a break-off or a partial, other than break-off. Also, with the new design, the next question for both paths is:*"In which section did the break-off occur?" [break-off path] or "Which sections are not complete?" [partial path].* The interviewer then picks from the following modules: (1) Household Composition module (2) Family module, (3) Sample Child module or (4) Sample Adult module. Again, with the new design it is now possible to match specific reasons for incomplete interviews with specific sections of the instrument.

3.1.3 Step 4: Evaluation of Redesigned Questions

As the final step in our analysis process, we examined the pre-coded or quantitative data from the redesigned 2007 questions. With the new design, we now know whether the incomplete interview was a break-off, or a partial, other than break-off. We found that they are split nearly evenly with partials, other than break-offs slightly higher at 52%. We know that the majority (70%) of break-offs are occurring in the Sample Adult module, and that very few (5%) occur in the Household composition module, which is earlier in the interview. Also, we know that interviewers report that for partial interviews, the Sample Adult module is incomplete for 94% of cases. The sample adult section is the final major section of the interview and often requires a different respondent than the Family module. These data are useful for us to learn more about where to target efforts to reduce the partial and break-off rates.

Next we looked at the textual data. We wanted to see how many responses were still falling into the other specify category and also to look at the content of the entries. We found that our effort to reduce the number of entries recorded in the other specify category resulted in a drop from 2,065 entries in 2002 to 827 entries in 2007. Next, we examined these 827 cases to see why we are still getting textual entries. Are the pre-coded categories still insufficient? Are there situations going on in the field that do not fit into the existing categories? For situations entered into the other specify category, are we gaining information, losing information, or obtaining the same amount of information as the pre-coded category? That is, are interviewers using the other specify category in order to give us additional information, rather than using the broad categories provided? Or, conversely, do interviewers use the other specify category to quickly move on to the next item, and give very brief entries?

In order to answer these questions, we backcoded the textual entries from the 2007 reasons for break-off interviews. We examined each entry to determine whether or not it could fit into one of the pre-coded categories, and if so, whether we were getting more or less information from the textual entry than we would if it had been recorded in one of the pre-coded categories. We found that the entries were split evenly between being able to fit into one of the existing pre-coded categories or not fitting into one of the existing categories. Additionally, there were more entries where we would have gained information if they had been put into an

existing category than entries where we would have lost information if they had been put into a pre-coded category (28% vs. 18%, respectively), and a few where we would not gain or lose (5%).

Examples of where we are losing information from the textual entries include "refused to complete interview" or "refusal." It would be more helpful in these cases for the interviewer to choose the pre-coded reason that most closely matches the reason the respondent gave for refusing. If the respondent did not give any reason at all, it would be more helpful for the interviewer to explain that no reason was given in the textual entry. Examples of where we are gaining information from the textual entries include cases where the interviewer provides more detail than what the pre-coded category specifies. Examples include "respondent received business call," "respondent's guests arrived and we had to stop," "late and had to get kids dinner and ready for bed." These could be put into the "respondent too busy" category, but we would not get this level of detail. Examples of the textual entry providing exactly the same information as the pre-coded category include "respondent was sick," "survey too long," and "computer acting strange." All of these have pre-coded categories that nearly exactly match the textual entries.

Textual entries that did not fit into an existing category included situations where a new category is needed (35%) and cases where the interviewer erroneously recorded break-off, rather than partial, other than break-off (15%). The majority of the cases requiring that a new pre-coded category be created were related to the use of cellular or mobile phones. Some examples include "battery died on his phone," "battery of cell phone ran out," "phone went to fax signal, couldn't get her back," and "telephone texting line was dropped after 40 minutes." We will consider adding "phone technology issues" as a category to the 2009 pre-coded reasons for break-off interviews. Finally, that 15% of textual entries were recorded as break-offs but were actually partials, other than break-offs indicates that some interviewers need additional training.

3.2 Reasons for Telephone Follow-up

3.2.1 Step 1: Analysis of Original Questions on Reasons for Telephone Interviews
A number of categories emerged from the analysis of our second research focus, the 2005 textual entries for reasons for a partly or fully telephone interview. The most common category was "Respondent requested, no more information" (42%). This included cases where the interviewer recorded a response such as "Per respondent request" and did not list the reason. For these cases, the reason for the request is unknown. The next most common reason was "Only way to reach Respondent" (17%). Examples include "No response to letters sent or notice on door – last resort" and "Could not find during personal visits. Finally contacted by phone." Another category that emerged was "Respondent did not want interviewer in their house" (8%). Some examples are "initially a refusal. Does not want anyone in home" and "Very leery to let people into apartment. She lives alone." Also, entries that indicated that respondents sometimes call the interviewer and ask to do the interview right then (8%). Examples include "Respondent called and wanted to get it over with" and "Respondent called me and said let's do it." There were a number of other reasons that emerged ranging from 1-6 % of all reasons.

3.2.2 Step 2: Question Redesign, and Step 3: Fielding New Questions
We used the categories that emerged from the textual analysis to create pre-coded categories. The new question was fielded in the 2008 NHIS. It contains eight pre-coded categories and still allows for a textual response if the pre-coded categories do not capture the situation. We combined "R. called and wanted to do interview right then" and "R. busy/ran out of time" into a single category, "R. too busy for home interview." When we developed the pre-coded categories, we purposely left out "Respondent preferred", hoping to force the interviewers to choose a more specific reason. Analysis of the 2008 data revealed a fairly high number still falling into the other specify category (15%).

3.2.3 Step 4: Evaluation of Redesigned Questions
We performed an initial analysis of the 303 responses from the other specify category for quarter 1 of 2008. About 23% of the entries exactly match pre-coded categories, indicating that some interviewers are not using the pre-coded categories, perhaps due to habit since this was the first time pre-coded categories were included. Also of note, 32% of interviewers are still reporting "respondent requested" without giving any more information. This could mean that either the interviewers do not know any more information and really cannot

pick a pre-coded category, or they are still skipping to the other specify category and typing respondent requested out of habit. We will need to explore this further. Also, our assumption that the category "respondent called and wanted to interview right then" should be put into the category "respondent too busy for home interview" was faulty. Judging by the 17% of textual entries that stated "respondent called and wanted to interview right then" it seems that many interviewers see this as a unique category. This is encouraging in that it indicates that interviewers do make distinctions between categories.

4. Conclusions

Judging by our failed attempt at combining categories, it appears that interviewers see clear distinctions between categories, and many prefer to use the open-ended option when the pre-coded category does not closely match the field situation. Additionally, it appears that some interviewers have a tendency to use the open-ended category even when the textual entry is identical to a pre-coded category. For the analyst, pre-coded categories are much quicker and easier to use than the textual data, however, the textual data are very useful in determining the appropriate pre-coded categories. Because interviewers provide short phrases and not narratives, little information is lost by creating categories for their responses. All the data currently available on the public use data file are in numeric form. We are cognizant that because of the need to protect confidentiality of respondents, category indicators are more easily released than raw textual data.

References

Safir, A., Black, T. and Steinbach, R. (2001). Using Paradata to Examine the Effects of Interviewer Characteristics on Survey Response and Data Quality. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

Strauss, A. and Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park: Sage Publications.