

**Evaluation of respondent and interviewing debriefing techniques on questionnaire development methods for health provider-based surveys**

*Catharine W. Burt, Ed.D.*

*Susan M. Schappert, M.A.*

*CDC's National Center for Health Statistics*

*ABSTRACT: This study compares results from interviewer and respondent debriefing and item nonresponse rates from a pilot test during the questionnaire development phase with item nonresponse rates from the redesigned data collection forms for a records-based physician survey. During the forms development phase, a pilot test was held with a convenience sample of physicians, and item nonresponse rates were calculated. In addition, respondents and U.S. Census Bureau field representatives (FR's) involved in the pilot test provided feedback through structured interviews on the ease of completing the new forms and their understanding of the instructions. Two different encounter forms were put into production in 2001 based on results from the pilot test. One form retained some of the items that appeared to be problematic during the debriefing while a shorter form excluded many of the problem items. Results indicated that both the pilot test item nonresponse rates and debriefing results were useful in predicting the 2001 item nonresponse rates.*

While questionnaire development for household surveys has used think-aloud cognitive testing and pilot test methods for the last 20 years, establishment-based surveys were less likely to use such techniques for refining data items. In an establishment survey, the data elements are less likely to come from the respondent's memory and more likely to come from records kept by the

business. In addition, these surveys are also more likely to involve multiple respondents providing information about the establishment, and terms used in the collection effort may require more detailed descriptions than those used in person-based surveys. Therefore, methods of questionnaire development for establishment surveys have historically used retrospective respondent interview techniques to determine the availability of information and the clarity of instructions.

Survey organizations now use a broader range of tools to test and refine new questionnaire items including focus groups, review by methodological or subject area experts, vignettes or mock records, pretests and pilot tests, and respondent debriefing interviews; however, few evaluate the usefulness of such techniques. The purpose of this study is to describe and evaluate how well respondent debriefing interview results and item nonresponse rates in the pilot test predicted item nonresponse rates in the national sample for a records-based physician survey. The extent to which these methods were predictive of actual item nonresponse in the 2001 survey will indicate the value of such questionnaire development techniques for records-based surveys.

The National Ambulatory Medical Care Survey (NAMCS), conducted by CDC's National Center for Health Statistics, collects data on the provision and utilization of ambulatory medical care in physician offices.<sup>1</sup> The survey utilizes a multistage probability sample design. Data are collected from physicians or their designees using encounter forms (Patient Record forms, or PRFs) which have been designed specifically for use in this ambulatory medical care setting. Using PRFs, participants provide information for a sample of patient visits during an assigned

reporting period. The survey is redesigned every 2-4 years to add, delete, or modify existing questions on the Patient Record forms. The redesign process for the 2001 survey solicited input on new and modified survey content from primary and specialty medical care association representatives, experts in subject matter and survey methodology, and health services researchers nationwide. The NAMCS has traditionally used a single-sided, approximately letter-sized survey instrument. However, there was considerable demand for expanded survey content, leading to development of a double-sided legal-size instrument for the 2001 survey. While the longer form accommodated the recommendations of expert advisors, concerns were raised about its potential effects on survey and item nonresponse as well as feasibility of data collection. To address these competing perspectives, NCHS undertook a pilot test prior to the start of the 2001 survey year.

#### Questionnaire development method:

In March and April of 2000, a pilot test of new survey questions and forms was conducted for the NCHS by the U.S. Census Bureau, which has been the data collection agent for the NAMCS for more than 10 years. Census Field Representatives from four Regional Offices (ROs) conducted the study. The aim of the test was to evaluate the content and format of new survey items. Consequently, Census Representatives were instructed to select a convenience sample of participants familiar with survey procedures. Each RO was responsible for securing eight physician practices that participated in either the 1998 or 1999 panel, for a total of 32 pilot test

participants. Authorization for the pilot test fell under a blanket OMB approval for content development and testing.

Each Census Field Representative (FR) met with physicians to complete an induction questionnaire and provide instruction on filling out the test Patient Record form. Standard NAMCS reporting periods of 1 week were not used. Instead, physicians, their designated staff, or (in a limited number of cases) field representatives completed 10 Patient Record forms each, using the test survey forms. Approximately 350 Patient Record forms were completed as part of the pilot study. When the forms were completed, the FR returned and conducted structured debriefing interviews with each person completing one or more Patient Record forms. In general, the debriefing reviewed most, but not all, of the items on the form to find out how well the user understood the question being asked, problems experienced providing the answer (e.g., availability of patient and visit information), and suggestions for improvements. Items evaluated in prior years were not necessarily re-evaluated in this pilot test. In cases where the FR completed the PRFs for the physician through abstraction from medical records, another FR conducted the debriefing interview with the FR abstractor. A total of 35 physicians or physician designees participated in the debriefing interviews, but, because not all of them had responsibility for each item on the PRF, the number responding to the various debriefing questions ranged between 26 and 35. Some examples of typical debriefing questions are shown in Figure 1.

Question 9. Refer to item 3(b), **Primary Expected Source of Payment for this Visit**

a. Did you work on this item?

9 Yes

9 No, *skip to question 10*

b. Is this information generally recorded in:

9 a) the patient's medical chart

9 b) the patient's billing chart

9 c) other record, please explain

9 d) this information is generally not available (*skip to question 10*)

c. Was it clear that only one expected source of payment was to be recorded?

9 Yes

9 No, please explain

d. For patients with multiple expected sources of payment, how did you determine the patient's primary source?

e. This item contains an acronym which may not be familiar to all survey participants. Can you tell me what, if anything, SCHIP means to you?

f. Did you have any difficulty providing information for this item?

9 Yes, please explain

9 No

Figure 1. Excerpt from NAMCS Pilot Test Debriefing Questionnaire

The debriefing questionnaires were edited by the FR's who conducted the interviews as well as by the RO clerical staff before being transmitted to the Census Bureau for data entry.

Participating physicians were allowed to charge the Census Bureau \$100 for services rendered.

The RO's reported that they felt the cash reimbursement had little or no impact on the medical staff's decision to participate or the quality with which they completed the work.

Two types of evaluations were conducted as part of the pilot: analysis of data from the debriefing questionnaires and item nonresponse rates from the completed Patient Record forms. In interpreting our results, it was important to keep in mind that the most qualified FR's in the best performing Regional Offices chose their most cooperative physicians to participate in the pilot test. Real-world situations were expected to produce poorer results than were obtained here.

Results of the debriefing questionnaire showed that 20 percent or more of the respondents flagged the following items as generally unavailable (with item response format shown in parentheses):

- Was a co-payment made for this visit? (3 check boxes)
- What is the amount of the expected revenue for this visit? (6 check boxes including None, <\$50, <\$50-\$99, \$100-\$149, \$150 or more, N/A - Contractual Agreement)
- List up to 3 CPT-4 codes for this visit (write-in fields)
- Initial Measurements (write-in fields for temperature, blood pressure, and weight)

Except for patient's blood pressure, all of these items were dropped from the 2001 NAMCS survey instruments based on these pilot test results.

In the small number of cases (n<10) where an FR completed PRFs through abstraction of medical data and was then debriefed by another FR, problems with availability were also cited for these items: primary expected source of payment, patient checklist (does patient now have...?), does patient use tobacco, time spent with physician, and time spent with other

providers. All of these were retained for 2001, but the patient checklist and time spent with other providers were included only on a longer version of the PRF that was given to a split panel of physicians.

The following additional items were flagged by 20 percent or more of participants as being difficult to obtain:

- Strength of medication provided or prescribed at the visit (write-in)
- Was medication from insurance formulary list? (checkbox)
- Surgical procedures ordered or provided (3 check boxes and 3 write-in fields)

Strength of medication and formulary list items were dropped for 2001. The surgical procedures item was shortened to 2 write-in fields but sub-item checkboxes were retained to indicate if the procedure was ordered/scheduled or performed at the visit. The longer surgical procedures item was retained in essentially the same format on a longer version of the PRF that was given to a split panel of physicians in the 2001 national survey.

Among the group of FR abstractors, results were similar to those reported above, but difficulties were also cited with type of insurance plan, is visit related to alcohol use, cause of injury or poisoning, route of medication, regimen of medication (i.e. dosage and frequency), whether medication was a new prescription, diagnostic/screening services, and has anyone in practice seen patient before and, if yes, how many visits in the past year. Of these items, three were subsequently dropped for 2001 (visit related to alcohol, route of medication, regimen of

medication), while type of insurance plan and whether medication was a new prescription were included only on the longer version of the PRF that was administered to a split panel of physicians.

The debriefing questionnaire also identified items that were unclear to survey respondents and required additional instructions or definitions. These included the following (with the accompanying problem shown in parentheses):

Was medication from an insurance formulary list? (unfamiliar with formulary list)

Does patient use tobacco? (unclear that all forms of tobacco should be included)

In addition, 37 percent of physicians/physician designees (n=33) recommended that the diagnostic/screening services item check box categories be rearranged in alphabetical order. Similar comments were made for the check box categories for the items on Counseling and Education and Other Therapy.

Results from the debriefing questionnaire were compared with pilot test item nonresponse rates on the more than 350 Patient Record forms that were completed by the test panel. Nonresponse rates were 20 percent or higher for the following items:

– Patient's occupation (for occupational illness or injury visits)

– Was a co-payment made at this visit?

- List up to 3 CPT codes for this visit.
- Initial measurements (temperature, blood pressure, weight)
- Drug strength, route, regimen, which diagnosis drug is for
- Time spent with other providers

Nonresponse rates were between 10 and 19 percent for these additional items:

- Patient's race
- Is visit related to an occupational injury or illness?
- Has anyone in practice seen patient before? If yes, how many visits in past year?
- Does physician share care for this patient with another physician?
- Type of insurance plan

Results from the pilot test were instructive and resulted in removal of a number of problematic items from the 2001 NAMCS. However, there were concerns that the convenience sample was too small to yield definitive results on other items. For these reasons, a decision was made to split the 2001 NAMCS sample into two panels of about 1,300 physicians each. One panel would receive a single-sided legal size page (the "A" form) with slightly shorter item content than that of the 1999-2000 survey instrument. The other panel would receive the "B" form, a double-sided legal size page with significantly more detail than the "A" form. Although each instrument covers the same broad topic areas, the longer "B" form includes five questions not included on the shorter form, and asks for more detail on many of the other items. Physicians were randomly assigned to one of the two panels, except in the rare case where multiple

physicians were sampled from a single group practice. The split panel design, together with the results of the pilot test, allowed NCHS to evaluate the effectiveness of debriefing questionnaires as a methodology for identifying problematic survey items in establishment-based surveys.

Development evaluation method:

In order to evaluate whether the pilot test with debriefing questionnaires to respondents and interviewers is a useful tool for testing and refining records-based survey items, we compared the pilot test results with the observed 2001 item nonresponse rates. The results from over 11,000 form B NAMCS records were tabulated and item nonresponse rates computed. The comparison was limited to the 11 items that remained on the form after the pilot test and related debriefing results (see table 1). Three pilot test indicators were used in the analysis: 1) item nonresponse rates; and, from the debriefing results, 2) the percent of respondents who indicated that the requested information was unavailable, and 3) the percent who reported difficulty obtaining the requested information. Three Spearman rank correlations were computed between the three pilot test indicators and the observed item nonresponse rates in the national survey (criterion variable).

Results:

Unweighted item nonresponse rates were computed using the records from form B of the 2001 NAMCS patient encounter form. Table 1 shows the observed item nonresponse rates from the

national survey and compares them with the results from the debriefing questionnaires and the pilot test. Results from the debriefing questionnaire on whether the information was generally unavailable were fairly predictive of the production item response (rank correlation = .795). The pilot test results were similarly predictive (rank correlation = .750). Results from the debriefing questionnaire about whether the respondent had difficulty providing the requested information were not very predictive of actual item nonresponse (rank correlation = .191).

#### Discussion:

The use of a pilot test with associated debriefing questionnaire appears to be a fairly valid approach to refine and weed out items for an establishment-based survey. Where responses to the debriefing questionnaire indicated that data would be generally unavailable, data collection for that survey item revealed the same pattern of data availability. In other words, item nonresponse in the national survey was particularly high (greater than 10 percent) for those items that were flagged as being unavailable in at least 9 percent of the debriefing interviews. The pilot test item nonresponse rates were also fairly predictive of national survey item nonresponse with one exception. For the item "Have you or anyone in your practice seen this patient before?" the pilot test predicted that the item would be unavailable 10% of the time while the national survey nonresponse rate for this item was only 1.2%.

In contrast, where the debriefing questionnaire indicated that the item would be difficult to obtain, this information was not very predictive of nonresponse. In other words, just because the

information was harder to obtain, respondents were not more likely to omit the information from the form. It may help explain why the average time to complete the Patient Record appeared more than we had anticipated and led to increased field costs when Census Bureau field representatives had to abstract the data. The difficulty item may help explain some of the variation in item nonresponse rates when the FR does the abstraction, but analysis indicated that the difficulty item did not predict FR item nonresponse any differently than it did total item nonresponse ( $r=.191$ , data not shown). FR item nonresponse and total nonresponse were highly correlated ( $r=.973$ ).

From a subjective viewpoint, the debriefing questionnaire provided us insight into how respondents were completing some of the items. This gives us good information when we attempt to interpret findings from the data. For example, we gained information on the method by which respondents tried to complete the duration item (time spent with a provider other than the physician). We learned that often the respondent uses the scheduled time for the patient, then parcels out how much of it might have been spent with the physician, another provider, or both. Results from the debriefing questionnaire also helped us to improve the item instructions in those cases where it was unclear what information was being requested.

The extent of the predictive ability of the pilot test and debriefing questionnaire for predicting good item response may be underestimated in this study because we curtailed the range of responses on the predictive variables by deleting items from the Patient Record form that indicated extremely large item nonresponse or other problems from the debriefing questionnaire.

For example we deleted items about expected revenue and CPT-4 procedure codes because they were judged unavailable by 35 and 25 percent of pilot test respondents, respectively.

In conclusion, we believe that the pilot test with debriefing questionnaire to respondents and field representatives is an excellent way to determine which items might prove problematic when the survey hits the field in production. This method allows us to weed out items that might not perform well and to improve our item instructions.

References:

1. Cherry DK, Woodwell DA. National Ambulatory Medical Care Survey: 2000 Summary. Advance data from vital and health statistics; No. 328. Hyattsville, Maryland: National Center for Health Statistics. 2002.

Table 1: Comparison of questionnaire design testing with 2001 production item nonresponse rates: NAMCS

NAMCS item	Debriefing questionnaire (% generally unavailable)	Debriefing questionnaire (% had difficulty providing info)	Pilot test item nonresponse %	2001 production item nonresponse %
Seen patient before?	0	9	10	1.2
Counseling and education?	6	17	2.8	3.6
Primary expected source of payment?	6	22	6	4.6
Surgical procedures ordered or performed?	6	37	3.6	5.3
If seen before, how many visits in the past 12 months?	0	14	3	6.9
Whether new prescription?	0	15	...	8.2
Do other physicians share care for this patient?	9	17	11	14.0
Does patient now have (chronic condition)?	15	0	...	15.3
Type of insurance plan?	9	19	14	16.5
Use tobacco?	15	17	32	35.4
Time spent with other provider?	17	24	25	45.6
<i>Spearman Rank Correlation w/ production item nonresponse</i>	0.795	0.191	0.750	1.0

... data not available