

RESEARCH REPORT SERIES
(Survey Methodology #2007-17)

**A Valuable Vehicle for Question Testing
in a Field Environment:
The U.S. Census Bureau's Questionnaire Design
Experimental Research Survey**

Jennifer M. Rothgeb

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: June 13, 2007

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

A Valuable Vehicle for Question Testing in a Field Environment: The U.S. Census Bureau's Questionnaire Design Experimental Research Survey¹

Jennifer M. Rothgeb
U.S. Census Bureau

I. Introduction

Survey methodologists within the U.S. Census Bureau conduct questionnaire design research, including questionnaire pretesting and evaluation. Typically, the pretesting research is conducted using cognitive interview methods. Frequently, however, we want to expand that research by conducting “split-sample” field experiments to compare different questionnaire designs, such as different question wording, question sequencing, etc.

In the past, the only available option has been to piggyback onto one of the demographic surveys like the Current Population Survey (CPS) or the Survey of Income and Program Participation (SIPP) which typically presents many constraints (time, procedural, managerial). Usually the lead time for production surveys is too long, researchers are not allowed much control when experiments are piggybacked onto production surveys, and the fact is, that production surveys don't want experiments attached to them. Because of the lack of available field surveys in which to do question testing, in 1998 researchers in the Statistical Research Division (SRD) proposed to establish an independent omnibus demographic household survey intended solely for research purposes. We proposed to call the survey the Questionnaire Design Experimental Research Survey (QDERS). To contain costs, we proposed that the QDERS use a nationally representative Random-Digit-Dialing (RDD) sample, be conducted from one of the Census Bureau's centralized telephone centers, be no longer than 15 minutes in duration, and produce approximately 800 -1000 completed interviews. The proposal stated that involved researchers would be responsible for questionnaire development and specifications, developing and administering interviewer training, monitoring the survey's progress, developing the SAS data file and conducting analysis. When presented to senior management within the Census Bureau's research area, the proposal was well received, viewed as an exceptional opportunity for researchers, and provided generous funding.

As well as getting internal funding, in order for QDERS to be implemented, it was necessary to get “buy-in” from the Telephone Center Coordination Staff at Headquarters and the Hagerstown Telephone Center (HTC), which is one of the Census Bureau's centralized calling centers. Once our proposal was posed to these groups, they enthusiastically supported the project.

¹ Disclaimer: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

At the Census Bureau, survey instruments administered through computer-assisted interviewing (CAI) are authored in a group outside the survey methods research area. Researchers do not possess knowledge of the authoring language to program the QDERS instruments. Authoring resources are costly and scarce. In addition, much lead time is usually required for CAI instrument authoring. Given that the production surveys sometimes experience a scarcity of authoring resources, we knew allocation of any available resources would obviously be given to the production surveys rather than a research survey. Since the QDERS project had minimal funding available for instrument authoring (and a short lead time until field implementation) we did not think it was feasible to use a CAI instrument for QDERS.

II. Fielding QDERS

As previously stated, the purpose of QDERS is to serve as a vehicle for SRD researchers to have an opportunity to independently conduct questionnaire design field experiments in a timely manner and within a flexible environment. The flexibility of the HTC staff was instrumental in the success of QDERS. The HTC staff has been very generous in allowing involved researchers free rein over the project. QDERS researchers identify staffing requirements for different interviewer groups for various questionnaire treatments, balancing groups by interviewer experience, skill level, gender, tenure and experience with RDD surveys. The HTC staff permits the researchers to develop the interviewer training package and administer training without getting input or feedback from the HTC staff. QDERS researchers also request that interviewers in different assigned groups not discuss the questionnaire treatment on which they are working at a specific time during QDERS. To avoid interviewer effects, interviewers are rotated among all questionnaire treatments so all interviewers are exposed to all experimental treatments for the same duration of time during the fielding of QDERS.

The QDERS researchers are also permitted to determine the case management parameters such as how many call attempts will be made to an unreached number, whether refusal conversions calls will be made, etc. In addition, the HTC staff permits the QDERS researchers to develop an interviewer debriefing protocol and conduct the debriefing without HTC involvement and without supervisory staff present during the debriefing.

Provided below is some basic information about the first two implementations of QDERS in 1999 and 2000.

A. QDERS 1999

QDERS 1999 included two treatment groups and a four-week data collection. Twenty-two interviewers participated in QDERS and the interviewers were “flipped” midway through, so all interviewers administered both versions of the questionnaire, but were not exposed to both versions at the same time. Paper and pencil questionnaires were used because resources were not available for authoring of a computer-assisted-telephone-interview (CATI) survey instrument.

QDERS 1999 experiments included examination of the following issues:

- The effects of person-level versus household level questionnaire design strategies on survey estimates and data quality;
- Methodological issues in measuring the uninsured; and,
- Using alternative question strategies to reduce income nonresponse.

QDERS 1999 received additional funding from another division that allowed us to double the planned sample size and also afforded a reinterview to be conducted to provide a reliability measure. Our target was to have 1800 completed interviews (900 of each treatment) and to have 900 completed reinterviews (450 of each treatment.) We used an RDD sample of 5400 sample telephone numbers to produce 1291 completed household interviews. (We had been advised that we needed three sample telephone numbers for each completed interview desired.) The 1999 response rate ranged from 36 to 46 percent (using AAPOR RDD response rate standards), depending on whether cases with unknown eligibility were included in the denominator. In 1999 we did not attempt any refusal conversions. Our targeted number of completed interviews fell short of our goal. We suspected this was largely due to the lack of refusal conversion attempts. However, because our analyses focused on relative differences between treatment groups, the low response rate was not as large of a concern as it would be had this been a production survey. In addition to the 1291 interviews, we reinterviewed over 900 households.

B. QDERS 2000

QDERS 2000 had four treatment groups, an eight-week data collection period (divided into four ten-day interview periods) and 24 interviewers (split into four groups). In addition to rotating the interviewers among questionnaire versions, we also decided to use sample replicates so that new sample could be released at the beginning of each new interview period. This allowed each group of interviewers to begin work on a different questionnaire treatment using a fresh sample and without any sample cases remaining from the earlier interview periods.

Because of the increased complexity of the QDERS questionnaire design and the number of experimental treatments, we used CATI instruments in QDERS 2000. Also, some of the production surveys for which the experiments were relevant are automated surveys and it was more methodologically sound to conduct experiments using the same mode of interview as that used in the production environment. (We did not have financial resources available for instrument authoring but we were able to acquire the resources through bartering of our research services in exchange for authoring.)

Experiments in QDERS 2000 included examining:

- Question ordering issues related to health insurance;
- Question wording experiments to facilitate pretesting evaluation research;
- Topic-based income reporting versus person-based reporting; and,
- An interviewer training experiment (refusal aversion training).

As with QDERS 1999, we received additional funds in order to double the sample size for QDERS 2000. Our targeted number of completed interviews was 2000 interviews (500 for each treatment.) Due to the low response rate in 1999, we decided to make two revisions to how we approached QDERS 2000. First, since the suggested ratio of 3:1 sample cases to completed interviews had proved inadequate, we decided to increase the sample telephone numbers to 8000 in the hopes of reaching our goal of 2000 completed interviews (a 4:1 ratio). Second, based on our experience in 1999, we decided to devote resources to refusal conversion attempts in an effort to boost response rates.

It is worth noting that within one year, QDERS expanded in terms of the types of experiments included. It grew from including only questionnaire design experiments in 1999 to including experiments focused on interviewer training and another on the evaluation of pretesting techniques. In such a short time, researchers were realizing the multiple utility that QDERS permitted.

With QDERS 2000, the sample of 8000 telephone numbers produced 1862 completed household interviews. The response rate ranged from 42 to 52 percent (using AAPOR standards), depending on whether cases with unknown eligibility were included in the denominator. Refusal conversion attempts were made for all households for which an initial refusal was obtained. As part of one of the QDERS experiments, refusal aversion training was provided to some interviewers. No doubt this contributed to the higher response rate. We were disappointed that our goal of 2000 completed interviews was not reached, but we were encouraged that we came much closer to meeting that goal than we had a year earlier.

III. Benefits of QDERS

QDERS has proved to be a valuable tool by which survey researchers can conduct methodological research. Many more controlled split-sample experiments are conducted at the Census Bureau now than ever before. Researchers realize that when they need to follow up some laboratory research with field testing, they now have an available vehicle by which to continue their research. The availability of QDERS has served to stimulate researchers to further develop their research ideas. QDERS has also resulted in more collaboration between survey methodologists and content experts and between researchers inside and outside of the Census Bureau. The availability of QDERS as an independent research vehicle has prompted organizations external to the Census Bureau to provide funding to increase QDERS sample size to provide enough power for certain experiments. Census Bureau researchers have produced more journal articles, book

chapters, and conference papers about question design and survey methodology than would have been possible without QDERS.

Some of the QDERS 1999 and 2000 experiments have resulted in the introduction of revised question design and new approaches to interviewer training in some production surveys. One of the experiments within QDERS that was used to evaluate pretesting techniques demonstrated that pretesting does appear to reduce measurement error.

QDERS was conducted again in 2002 and 2003. Preparations are underway for QDERS 2004.

IV. QDERS Publications and Conference Papers

The details and results of the QDERS 1999 and 2000 experiments would consume too much space for this paper. For readers interested in specific experiments, I refer you to the publications below which are products of the QDERS project. The main results of the specific experiments are also provided.

Forsyth, B., Rothgeb, J., and Willis, G., (2004). "Does Question Pretesting Make a Difference? An Empirical Test Using a Field Survey Experiment." Forthcoming in Presser, S., Rothgeb, J., Couper, M., Lessler, J., Martin, E., Martin, J., and Singer, E., in *Methods for Questionnaire Evaluation and Testing*. New York: Wiley Interscience. Main results: Pretesting results based on expert review, questionnaire appraisal, and cognitive interviews predict actual problems in field survey outcomes. There were mixed results regarding whether questionnaire revisions based on pretest results yield improved survey outcomes.

Hess, J., Moore, J., Pascale, J., Rothgeb, J., and Keeley, C., (2002). "The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality." *Public Opinion Quarterly*, Winter 2001, University of Chicago Press. Main results: Mixed results were obtained. There was some evidence that the use of household screeners increases the risk of under reporting, but it was specific to particular content areas. There was evidence suggesting the person-level approach might increase the completeness of reporting for some topic areas, but response reliability measures suggest this improvement may come at a cost of decreased reliability.

Hess, J., Rothgeb, J., Moore, J., Pascale, J., and Keeley, C., (2001). "Measures of Functional Limitations: The Effects of Person-level vs. Household-level Questionnaire Design" in Barnartt, S., and Altman, B. (Eds.), *Exploring Theories and Expanding Methodologies: Research in Social Science and Disability*, Volume 2, Oxford, England: Elsevier Science Ltd.

Main results: Findings indicate some evidence that the use of a household-level design results in lower survey estimates than the person-level design. However the household-level approach produces somewhat more reliable data and also results in a shorter interview than the person-level design.

Landreth, A., and O'Brien, E., (2001). "Respondents' Understanding of the Vague Economic Concept "Cash": A Comparative Study." Presented at the annual meeting of the American Association of Public Opinion Research. Montreal, Canada.

Main results: Findings indicated that respondents use the vague economic term "cash" more literally than intended in questions about residential finance. The residential finance questions focused on understanding how context may improve respondents' interpretations of vague economic terms and phrases in factual questions.

Mayer, T., and O'Brien, E., (2001). "Interviewer Refusal Aversion Training to Increase Survey Participation" in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Alexandria, VA.

Main results: First contact cooperation rates increased in the range of 3 to 7 percentage points for interviewers who participated in the Refusal Aversion Training and as much as 14 percentage points over time compared to interviewers who did not receive the training.

Moore, J. and L. Loomis, (2001). "Reducing Income Nonresponse in a Topic-Based Interview." Paper presented at the annual meeting of the American Association of Public Opinion Research, Portland, OR.

Main results: Immediately preceding the income questions, using a brief statement which acknowledges respondents' reluctance to report income and emphasizes the non-personal, statistical uses of the data was found to eliminate the income item nonresponse disadvantage for the topic-based interview treatment and in some cases reverses the effect.

Pascale, J. (2002). "A Quantitative and Qualitative Assessment of the Data Quality of Health Insurance Measurement Methodologies." Paper presented at the International Conference on Improving Surveys, Copenhagen, Denmark.

Main result: Studies that focused on public health coverage found that respondents generally could report status (insured vs. uninsured) but had much more difficulty identifying the particular source of coverage.

Pascale, J. (2001). "Measuring Private and Public Health Coverage: Results from a Split-Ballot Experiment on Order Effects." *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Alexandria, VA.

Main results: Results showed that Medicaid recipients do not mistakenly report their coverage as some type of private plan when the usual plan sequence (private first) is used. There is evidence that the plan type that is presented first in the series may suffer under reporting relative to plan types that are asked later in the series.

Pascale, J. (2000). "Alternative Questionnaire Design Strategies for Measuring Medicaid Participation." Paper presented at the 128th Annual Meeting of the American Public Health Association.

Main results: When testing two key features (reference period and unit of measurement) of health insurance questions, no statistically significant difference in Medicaid reporting was found for the overall sample. However, among large households (4 - 6 persons)

within the “calendar year” treatment, the “household-level” estimate was almost twice the “person-level” estimate (11.5 percent versus 6.2 percent.)

Pascale, J. (1999). “Methodological Issues in Measuring the Uninsured” in *Proceedings of the Seventh Health Survey Research Methods Conference (PHS01-1013.)* Department of Health and Human Services. Washington, D.C.

Main results: Findings indicate that “household-level” “calendar year” questionnaire results in under-reporting of health plans relative to the “person-level” questionnaire. This finding is driven entirely by significantly lower reports of employer-based plans in the household-level design.

Pascale, J., (1999). “Effects of Questionnaire Format and Reference Period on Measuring the Uninsured.” Paper presented at the 127th Annual Meeting of the American Public Health Association.

Main results: Results show no difference between the “current” and “calendar-year” treatments among the “household-level” questionnaires. Among “person-level” treatments the expected gap between the “current” and “calendar year” versions was observed.