

The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality

Jennifer Hess, Jeffrey Moore, Joanne Pascale, Jennifer Rothgeb, and Catherine Keeley, U.S. Census Bureau^{1/}
Jennifer Hess, US Census Bureau, HHES 1445-3, Washington, DC 20233

Key words: questionnaire design, household-level screening questions, nonresponse, data quality

1. Introduction and Background

Designers of household demographic surveys face a multitude of questionnaire design options, each of which offers a mix of costs and benefits. One such option is the use of person-level questions to assess the characteristics of interest: Does John have a disability? Is Susan covered by health insurance? Does Robert receive Food Stamps? Such surveys generally conduct person-level interviews for all eligible household members, returning to the “top” of the interview and repeating the entire question sequence for each eligible household member in turn. An alternative is the household-level approach, in which a household screener is asked to determine if anyone in the household has the characteristic of interest: “Does anyone in the household have trouble seeing...?” Follow up questions to determine who has the characteristic are asked only if the answer to the household screener is positive. To distinguish this from the traditional person-level approach, we term this the “household-level” approach.

The person-level approach has a long history, perhaps because of its ease of administration in a paper and pencil interview. This advantage is disappearing, however, with the increasing use of computer-assisted survey instruments, which enable fairly smooth administration of a household-level design. There is also evidence of problems with the person-level design – e.g., perceived tedium and burden, and improper implementation (Hess, Rothgeb, Zukerberg 1997; Hess and Rothgeb 1998). While there may be important benefits of a household-level design, there is still a concern that the use of household-level questions increases the risk of missed events and circumstances, and consequently results in under-reporting. We implemented the experimental study that is the focus of this paper to gather quantitative evidence to inform this questionnaire design decision.

2. Research Methods and Procedures

The research presented here was embedded in the initial launch of the Census Bureau's Questionnaire Design Experimental Research Survey (QDERS), a special survey developed by Bureau staff for conducting

questionnaire design research “off-line” from the agency's ongoing production surveys. Fielded in April 1999, QDERS included several experiments on alternative questionnaire design strategies for collecting information about disabilities, health insurance, transfer program income sources, asset ownership and income, and within-household relationships. This paper focuses on the person-level/household-level component of the 1999 QDERS experiment.

QDERS was a split-sample controlled experiment, using paper and pencil questionnaires in a telephone interview with a single household respondent. We used a nationally representative (excluding Alaska and Hawaii) RDD sample, with independent samples for each of the two treatments. The response rate was 44% for the person-level treatment and 48% for the household-level treatments (excluding cases of unknown eligibility). Refusals accounted for approximately half of all nonresponse – 32% for the person-level treatment and 27% for the household-level treatment.

3. Evaluation Methodologies

Our analysis of the person/household experiment employs multiple evaluation methodologies, including: a comparison of survey estimates for the characteristics of interest, item nonresponse rates, response reliability, behavior coding of interviewer-respondent interactions, interview length and unit nonresponse, and interviewers' assessments. The analyses are limited to the 908 interviewed households containing two or more persons (2,948 persons in total), since the two questionnaire formats were identical in single person households. Space limitations severely constrain our presentation of results, only a few of which are summarized here – see Hess, et al., 2000 for complete details.

4. Results

4.1 Survey Estimates

The left-hand columns of Table 1 show the survey-estimated “prevalence” rates produced by the two questionnaire treatments. In general, estimates of demographic characteristics (except school enrollment) and of receipt of income from government programs (except Social Security) appear to be unaffected by household-level/person-level design differences. The

same cannot be said for the other three topic areas – functional limitations, health insurance coverage, and asset ownership. Although we find only one significant difference for individual items concerning functional limitations, in combination some clear patterns emerge: the person-level approach produced higher estimates of persons with any functional limitation, persons with a severe limitation, and households containing at least one “limited” person. Due in great measure to its significantly higher yield of employer/union-based health plan coverage, the person-level approach produced more insured persons overall than the household-level approach, and thus lower estimates of the uninsured²⁴. Asset ownership, however, shows the clearest effect of the person/household treatments, where both multiple significant effects and the overall pattern of differences indicate consistently higher asset ownership rates among households interviewed with the person-level approach.

The absence of information about the true characteristics of QDERS households prevents definitive conclusions about data quality differences. For asset ownership, which tends to suffer from under-reporting (Moore, Stinson, and Welniak, 1999), a “more is better” conclusion is certainly defensible. The same cannot be said for the other topics. The most we can conclude regarding functional limitations and health insurance coverage is that if these topical areas also suffer from under-reporting in surveys, then the QDERS results suggest that the household-level approach may result in more under-reporting than the person-level approach.

4.2 Data Quality

The right-hand columns of Table 1 compare the response reliability effects of the two questionnaire treatments as measured by the index of inconsistency²⁵. For four of the five topic areas – demographic characteristics, functional limitations, program income sources, and asset ownership – the results indicate a tendency toward greater reliability for the household screener interview approach, a difference which is particularly evident in the latter two areas. The general pattern clearly does not hold for health insurance, where the individual items provide some evidence that the person-level approach yields more reliable data. The apparent superiority of the person-level design applies only to specific types of coverage – reliability estimates do not differ for a summary insured/not insured measure. Differences in item nonresponse are for the most part trivial (except for asset ownership, where the person-level nonresponse rates are significantly elevated); similar non-effects characterize our behavior coding assessment, which does not suggest any inherent superiority of one

design over the other (see Hess et al., 2000, for item nonresponse and behavior coding details).

4.3 Other Evaluations

Interview length Timing data confirm the increased efficiency of the household-level interview. The average duration of a QDERS person-level interview was 14.7 minutes, versus 12.0 minutes for household-level interviews ($t=2.03$, 32df, $p=.05$). Person-level interviews took approximately 5.9 minutes per person to complete, compared to 4.6 minutes per person for the household-level treatment, a 28% increase ($t=1.79$, 32df, $p<.10$).

Interviewers' Evaluations Interviewers evaluated both questionnaire treatments via a debriefing questionnaire. Items yielding significant differences indicate a clear preference for the household-level design. They include: (a) a 7-point, “boring/repetitious –engaging/ NOT repetitious” scale; interviewers rated the household-level questionnaire significantly less “boring” than they did the person-level questionnaire (average scores = 4.3 and 2.9, respectively; $F=11.4$, $p<.005$); (b) a 7-point scale with regard to whether the instruments “worked very poorly” (1) or “worked very well” (7) in large (4+ person) households; interviewers rated the household-level form significantly superior to the person-level form (average scores = 5.3 and 2.1, respectively; $F=54.7$, $p<.001$); and (c) a similar 7-point scale with regard to questionnaire performance in “households with reluctant/unenthusiastic respondents;” here again the household-level form emerged as the clear favorite (average scores = 4.2 and 2.9, respectively; $F=11.1$, $p<.005$).

Unit Nonresponse An even more compelling indicator, perhaps, of interviewers' attitudes toward the two instruments can be found in their behavior. We noted earlier the higher response rate for the household-level treatment and a lower rate of refusals. The latter finding, especially, seems telling. Refusals almost always occur in the first few seconds of the interaction, well before the design of the interview has any chance to affect a respondent's desire to participate (Groves and Couper, 1998). Thus, we would expect the two interview treatments to present interviewers with equivalent levels of initial reluctance on the part of respondents. The fact that interviewers failed more often on the person-level side to convert initial reluctance into a completed interview suggests that interviewers invested less effort in persuasion, perhaps because they were less eager to conduct that type of interview. This is conjecture, of course, but it does conform to the logic of the situation (i.e., the limited direct impact that instrument design can

have on one-time respondents), and is consistent with other research which finds similar effects (e.g., Moore and Moyer 1996).

5. Conclusions

As is often the case with complex experimental studies – especially those which, like this one, offer a broad range of findings using a broad range of evaluation dimensions across a broad range of topics – the results of the QDERS person/household experiment do not lend themselves to easy generalization. There is some evidence that the use of household screeners increases the risk of under-reporting, but we find this evidence only for the summary measures of functional limitations, coverage by employer/union-based health insurance plans, and – perhaps most clearly – asset ownership. For other types of characteristics the two treatments produced estimates without clear or consistent differences. While our results suggest that the person-level approach might increase the completeness of reporting for some topic areas, our response reliability measures suggest that this improvement may come at a cost of decreased reliability (health insurance coverage being the notable exception). Again, item nonresponse and behavior coding results did not suggest that either the household- or the person-level version was superior.

Naturally, responsible survey designers would want to choose questionnaire design features that minimize respondent burden, increase interviewing efficiency, reduce refusals and overall nonresponse, and which appeal most to interviewers. On these dimensions, results from our study suggest that the household-level approach is preferable. We cannot, however, conclude that the household-level approach is preferable across the board in light of other data quality indicators. We detect very little evidence suggesting that the use of a household screener would cause any problems for the items on demographic characteristics; and for program income sources we find evidence to recommend it, even apart from its efficiency/burden benefits. For functional limitations and asset ownership, however, there is some indication that the household-level approach risks data quality to an extent that increased efficiency and reliability may not be sufficient compensation. Finally, for health insurance we find fairly strong evidence that the household-level approach results in under-reporting of employer-based plans, but caution that this finding could be an artifact of the particular QDERS design, not the general household-level approach.

On the whole, we view the QDERS experience as a positive step toward providing concrete data about the

costs and benefits of using household-level screening procedures in household-based surveys. At the same time, we are not blind to QDERS' limitations – chief among them its low response rate, lack of validating information, non-automated format, and limited number of content areas – and the restrictions they impose on our ability to draw firm conclusions.

The mix of results from this study prompt us to consider several alternative avenues of research to determine why the two design strategies produced different and sometimes conflicting results. One area of research might address the different cognitive processes people use when responding to person-level versus household-level questions. Collecting information about memory organization and knowledge structures as they relate to these two designs – e.g. how the different designs are understood, how respondents decide who does or does not possess a specific characteristic of interest, and what determines how the respondent decides which individuals to report – would be useful for gaining a better understanding of the different results. For example, higher reports from the person-level approach may relate to saliency. That is, individuals who marginally possess the characteristic of interest may be reported in the person-level approach because of the direct person-by-person questioning, whereas they are not reported in the household approach because the household screener wording “Does anyone in the household...” is insufficient to prompt reports of individuals who only marginally possess the characteristic of interest. Research into the cognitive processes used in responding to the two approaches – such as cognitive interviews, respondent debriefings, and reconciled reinterviews – is needed to determine whether issues such as saliency merit further investigation.

Our results also suggest that the household-level approach may be more appropriate for some content areas than for others. Research examining how the two approaches compare with various topics and question characteristics might help determine when one design may be better suited than the other. We suggest conducting experiments whereby the two approaches are tested and question content and characteristics are varied along dimensions such as subjectivity/objectivity, content sensitivity, prevalence of characteristic of interest, concept clarity, and knowledge level and awareness. The previously noted efficiencies of the household screening questions may be enhanced or reduced depending on the question characteristic or content area.

Lastly, we view the addition of validating data as an especially promising component of the next stage of

research. One type of validation study, a one-directional record check approach (i.e. sample people with known characteristics and interview them), may be sufficient to address the primary substantive concern, namely: Does the household-level screening approach increase the risk of under-reporting?

If more rigorous research continues to suggest that use of household-level screening designs increases the risk of under-reporting errors, then survey methodologists will face two important research challenges. First, we will need to develop and test good theories to explain why household screening designs are acceptable for some content areas and not for others. And second, we will need to find ways to refine those designs to maintain their efficiencies and other benefits, while at the same time improving the accuracy of respondents reports.

References

- American Association for Public Opinion Research. 1998. "Standard Definitions - Final Dispositions of Case Codes and Outcome Codes for RDD Telephone Surveys and In-Person Household Surveys." (<http://www.aapor.org/ethics/stddef.html#code>)
- Groves, R. and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Hess, J., J. Rothgeb, and A. Zukerberg. 1997. "Survey of Program Dynamics: Pretest Evaluation Report." Unpublished U.S. Census Bureau report prepared for the SPD Steering Committee.
- Hess, J., and J. Rothgeb. 1998. "Summary Report: 1998 SPD Taped Interviews." Unpublished U.S. Census Bureau report prepared for the SPD Steering Committee.
- Hess, J., J. Moore, J. Pascale, J. Rothgeb, and C. Keeley. 2000. "The Effects of Person-level vs. Household-level Questionnaire Design on Survey Estimates and Data Quality." Unpublished U.S. Census Bureau paper prepared for the annual meetings of the American Association for Public Opinion Research, Portland, OR, May 18-21, 2000.
- McGuinness, Richard. 1997. "Reinterview report: Response Variance in the 1995 Food Security Supplement." Report prepared by the U.S. Bureau of the Census, Demographic Statistical Methods Division/QAEB for the U.S. Department of Agriculture Food and Consumer Service, Alexandria, VA.
- Moore, J. and L. Moyer. 1998. "Questionnaire Design Effects on Interview Outcomes." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp 851-856.
- Moore, J., L. Stinson, and E. Welniak. 1999. "Income Reporting in Surveys: Cognitive Issues and Measurement Error." Chapter 10 in Sirken, M., D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds.), *Cognition and Survey Research*, New York: John Wiley & Sons, 1999.

Notes:

1. U.S. Census Bureau, Statistical Research Division/Center for Survey Methods Research, Washington, DC 20233-9150. This article reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.
2. In the household-level treatment, interviewers first identified all policyholders in the household and then identified dependents on those policyholders' plans. Detailed behavior coding results indicate that often interviewers did not probe sufficiently to identify all policyholders in the household; therefore, follow-up questions to identify dependents on those plans were not asked. This failure to probe for all policyholders may have been associated with the hard-copy design; an automated instrument that displayed the entire household roster and controlled the flow of questions may have aided interviewers in proper administration of the policyholder question. The apparent underreporting of employer-based plans in the household-level design, then, could be an artifact of the particular QDERS hard-copy design, and not of the household-level approach in general.
3. The index of inconsistency estimates the ratio of simple response variance to the combined total of sampling variance and simple response variance for a survey item. A low index indicates high reliability and a high index indicates low reliability. As a rule of thumb, the Census Bureau considers an index of less than 20 as low response variance (high reliability); an index between 20 and 50 as moderate response variance and one over 50 as high response variance (low reliability) (see McGuinness 1997).

Table 1: Estimated Prevalence Rates and Reliability for Demographic Characteristics, Functional Limitations, Health Insurance Coverage, Program Income Sources, and Asset Ownership

ANALYSIS SUMMARY: PERSON-LEVEL vs. HH-LEVEL QUESTION FORMAT (EXCLUDES 1-PERSON HHs)	ESTIMATED RATE (% yes for all persons 15+)		INDEX OF INCONSISTENCY	
	Person-level (n=1,110)	HH-level (n=1,152)	Person-level (n=715)	HH-level (n=740)

Demographic Characteristics

Usually live here? (includes kids) (% no)	1.0	<	1.2	86.4	<	89.7
Hispanic origin? (includes kids)	7.9	<	8.0	15.3	>	9.4
Ever served in US armed forces?	15.4	>	14.4	8.2	<	12.9
Currently enrolled in school?	21.5*	>	18.7*	22.1***	>	15.4***

Functional Limitations – individual items

Difficulty seeing newsprint?	5.2	<	5.3	60.1	>	46.7
Difficulty lifting/carrying 10lbs?	8.5**	>	6.0**	35.8	<	36.2
Difficulty walking 1/4 mile?	9.3	>	7.9	28.8	>	24.9
Difficulty climbing stairs?	6.8	>	6.5	35.0	>	33.3
Difficulty hearing normal conversation?	5.8	>	5.1	48.4	>	47.5
Use special aids?	5.4	<	5.8	13.5	<	21.9

Functional Limitations – summary measures

% of persons with any limitation	20.2***	>	16.2***	40.4**	>	28.3**
% of persons with severe limitation	17.3**	>	14.0**	45.6	<	50.4
% of persons with:						
1 limitation	12.1***	>	8.5***	53.7**	>	41.8**
2 limitations	3.2	>	3.0			
3+ limitations	4.9	>	4.7			
% of households w/1+ “limited” person	15.2**	>	12.1**	33.3**	>	22.2**

p<.10*; p<.05**; p<.01***

ANALYSIS SUMMARY: PERSON-LEVEL vs. HH-LEVEL QUESTION FORMAT (EXCLUDES 1-PERSON HHs)	ESTIMATED RATE (% yes for all persons 15+)		INDEX OF INCONSISTENCY	
	Person-level (n=1,110)	HH-level (n=1,152)	Person-level (n=715)	HH-level (n=740)

Health Insurance Coverage

Plan thru employer/union?	75.1***	>	65.3***	22.8	<	25.8
Directly purchased plan?	9.4	<	10.8	42.7	<	47.6
Plan of non-household-member?	3.8	<	4.5	32.1	>	32.0
Medicare?	9.7	<	11.2	0.0***	<	9.8***
Medicaid?	5.9	<	8.0	11.0**	<	32.1**
Military plan?	4.7	>	3.8	4.4**	<	40.1**
Other plan?	1.9	>	1.2	100.6	<	101
% uninsured (constructed item)	6.6***	<	12.6***	40.1	>	31.1

“Program” Income sources

Receive worker’s compensation?	1.4	<	1.7	52.4	>	44.5
Receive unemployment benefits?	3.1	>	2.4	38.2	>	32.3
Receive Social Security?	13.8***	<	17.4***	12.7	>	9.4
Receive vets pension/compensation?	2.6	>	2.0	39.6***	>	27.0***
Receive SSI?	1.9	>	1.5	50.9	>	36.3
Receive Food Stamps?	2.6		2.6	37.7***	>	22.5***
Receive AFDC/welfare/public assist.?	1.3	>	1.0	67.3	>	44.9

Asset Ownership

Interest-earning checking account?	49.2***	>	42.5***	55.2**	>	46.1**
Savings account?	69.3***	>	60.6***	39.6	>	38.7
CDs?	17.2	>	15.0	47.9	>	44.5
Mutual funds?	19.4	>	17.8	45.6	>	41.9
Stocks?	19.4	<	19.8	44.6*	>	34.7*

p<.10*; p<.05**; p<.01***