

RESPONDENT SELECTION PROCEDURES WITHIN HOUSEHOLDS

Alan Roshwalb, University of Michigan

Personal interview surveys often involve the selection of a probability sample of households (Although personal interviews are also used to conduct establishment surveys, this paper deals exclusively with household surveys). In some surveys, information is to be obtained for all eligible household members. In others, information is to be obtained only for a single designated respondent who is chosen at random from each household.

In either case, the use of a proxy (substitute) respondent is considered. The proxy respondent can possibly reduce the difficulties in contacting a desired but unavailable respondent and the subsequent costs of repeated callbacks. Or for economy, one adult member of a household could answer for all other members. With the desire to reduce costs, one becomes more concerned with the interaction of expenses, sampling errors, non-response errors, and response errors. Therefore, a respondent rule is considered to select the individual or individuals who will supply the required information and control some of the previously stated issues.

For surveys which involve a designated respondent, two respondent rules are most often used.

- DR1) The designated respondent is interviewed. If he/she is competent but unavailable, repeated callbacks are made until contact is established.
- DR2) If the designated respondent is not present or not competent (deaf, senile, ill) a proxy respondent (another responsible related adult household member) is chosen. Variants of this rule involve different definitions of the permissible proxy respondents.

For surveys in which responses for more than one eligible (i.e., adult) household member are desired, one of several possible respondent rules is chosen.

- HR1) Every adult member of the household is to respond personally.
- HR2) One adult member of the household is to answer for every member of the household.
- HR3) Self-response is obtained for all adult family members present at the time of the interview. For those not present, information is sought from any responsible related adult, including adult children, siblings or other distant relatives. Unrelated adults respond for themselves.

- HR4) Self-response is obtained for all family members present at the time of the interview. Information about absent adults is obtained only from a spouse or parent. Unrelated adults respond for themselves.

Methodological investigations of the effects of using various respondent rules have focused on two basic areas. The first involves the differences in callbacks needed to contact the desired respondent; the second involves differences in the quality of the information obtained, i.e. response errors.

Since use of a proxy respondent increases the chances of obtaining information for a specified individual on the first visit to their household, respondent rules other than DR1 or HR1 ordinarily require fewer callbacks. Thus, surveys using them may be less costly to administer. Respondent rules other than DR1 or HR1 require additional coding on the questionnaire and data records to flag the use of a proxy. Relatively, the additional cost of this flag would be minor. In the Charlotte pretest of the National Health Survey and subsequent follow-up studies (22,16,17), the average number of callbacks needed to complete the interview was about .5 higher for households interviewed using HR1 than for those interviewed using a proxy respondent rule (either HR3 or HR4). In the Charlotte pretest, this was equivalent to a 26% difference (2.4 vs. 1.9) and in the follow-up studies, the difference was 36% (2.53 vs. 1.86).

In a Statistics Canada study, Singh and Tessier (25) conclude that the implementation of a non-proxy reporting procedure on a regular basis would involve a cost increase of 29%. The data were not available to provide comparable callback ratios for the proxy and non-proxy procedures. Only fieldwork related data were collected in this study--no estimates of differences in the quality of the data were obtained. This study will be discussed further later.

Although the use of a proxy respondent can reduce the number of callbacks and the costs of a survey, the respondent of interest does not respond for him/herself. It is commonly felt that the information obtained from a proxy respondent is not as reliable as the information received directly from the designated respondent. Several reasons are presented to explain this loss of reliability; 1) Often no communication is held between a proxy and designated respondent, whether intentional or unintentional, about an event of interest. 2) If the event of interest is of relatively little importance to a proxy respondent, the proxy respondent's recall could be low. 3) Recall of events is more accurate for the participant than the observer of the event.

Notice, the use of proxies for children is allowed in each of the respondent rules. Although the worth of proxies is being questioned, but the use of proxies for children is an accepted procedure. Proxies for children have been reported to perform better than proxies for adults, but proxies for children perform worse than adult self reporters (3).

The North Dakota Health Survey (8) compared the single designated respondent rules DR1 and DR2. Separate groups of households were assigned to procedures DR1 and DR2 while nonresponders in DR2 were also asked to complete a self-administered questionnaire. In most cases the wife was the proxy respondent. No direct questions regarding health were asked on the self-administered questionnaires. Examples of questions on the self-administered questionnaires were occupational history, number of cigarettes smoked, and the presence of chest pains. The answers from the self-administered questionnaires were compared with answers to the same questions received from the proxy respondent. Generally, there was agreement in all of these areas. The general conclusion from the comparison of households assigned to DR1 and DR2 is that "the use of household respondents will probably result in no less disease reported in total and for broad categories than would be the case if each adult were to be interviewed for himself."

Cartwright (6), in a health survey, examined the effect of a single respondent for an entire household (HR2). Data collected from wives about their husbands were compared with information obtained from the husbands themselves. Illnesses were reported for 5 times as many men in the self-response group as in the proxy response group. An interesting sidelight emerged from the study of the group of wives who were interviewed, those who were household respondents (reporting on the illnesses of several household members) reported fewer illnesses for themselves than those who answered questions only about themselves. Although the control for this comparison was not perfect, the author speculates that being asked about a number of people is likely to decrease the number of illnesses reported for themselves.

The Baltimore Health Survey (19) also used a single household respondent, any responsible related adult member of the household. A subsample of individuals was given a physical exam by doctors at John Hopkins Hospital. The results of this exam were compared with the data from the survey questionnaires. Only 22% of all chronic conditions diagnosed in the clinical examinations were matched by the interviewer reports. Limiting the analysis to self-reporters on the questionnaire increased the match rate by 30%, and many factors other than proxy reporting were posited to account for the low match rate.

The 1952 San Jose Study (7) studied HR2 a responsible related adult was chosen to be the respondent for the entire household. A subsample of 118 non-respondents was chosen and reinterviewed independently of the initial interview. The results of this procedure showed

that illness reporting for adult non-respondents was subject to considerable variability and bias.

In a study published by the Office of Federal Statistical Policy (23) a single household respondent fourteen years or older was used to measure labor force information within their household. The results obtained by using this procedure were significantly different from those employing a self-respondent rule (HR1) in the household. In the same study, the use of an advance form containing important labor force questions was investigated. The form was mailed and each adult member of the household was asked to fill the form out personally. At the time of the interview, the responses on the advance form were transcribed onto a questionnaire by the interviewer. The remaining questions were answered by a household respondent. The results of the procedure using the advance form and the self-response procedure (HR1) were comparable, but significantly different from the proxy response procedure (HR2).

There has been a great deal of study on respondent rule HR3 all available adults respond for themselves while proxies respond for children and for those adults unavailable or incapable of responding. This is the respondent rule employed in the National Health Interview Survey (NHS) and other U.S. government surveys such as the Income Survey Development Program (ISDP).

An early report was the Charlotte, N.C. pretest of the NHS (21) which studied the third respondent rule (HR3). Half of the interviews were conducted using HR3, while the other half used HR1. The results seemed to show a tendency to higher reporting of disease by the self-reporters. Due to high sampling variability and response variability, the researchers could not make definitive conclusions about the bias resulting from the use of a proxy respondent. Nisselson suggested that a more restrictive rule (i.e., HR4) should be used in the NHS only a parent or spouse should be allowed to answer for an absent adult.

In a series of studies about hospitalizations and reporting visits to doctors (2,3,4), the NHS rule (HR3) was compared to other rules, as well as records from hospitals, clinics and doctors. It was generally concluded that self-reporters were more accurate than proxy reporters. However, the eligible population from which the sample was drawn could be considered different enough to restrict the generality of the results to the NHS population.

In a series of papers by Kovar, one in conjunction with Wright (18) and the other with Wilson (17), the utility of the respondent rule in the NHS was investigated. A control group was selected using the standard NHS respondent rule. For the experimental group every adult in the household answered for him/herself while proxy respondents were requested for children under 18 and for adults incapable of being interviewed (HR1). Self-reports yielded significantly higher rates of illness, disability, and out-patient

utilization (when one-tailed tests were used) based on 6 out of 10 objective measures (18). When measuring perceived health status of members within households, the use of the respondent rule HR3 is quite good (17). There were differences between the two respondent rules when considering the four perceived health states (excellent, good, fair, poor). However, when the "excellent" and "good" categories and the "fair" and "poor" categories were combined, the difference between the two respondent rules were minimal. Use of proxy responses was judged to be effective for measuring the direction of perceived health, but dubiously effective for measuring strength or degree of perceived health.

Koons (16) reported reinterviews of the respondents to the NHS. All the reinterviews were self-respondents while the original interviews used the standard NHS rule (HR3). It was assumed that the effect due to the time lag between interview and reinterview was the same whether the original reporter was a self or proxy respondent. Thus, the difference between first interview and reinterview for a proxy respondent is additively the time effect and effect due to the proxy respondent. The proxy respondent was not randomly assigned in the original interview which hindered the analysis. Nevertheless, the results indicate the rate of chronic conditions or illnesses is less due to the proxy effect.

In a recent article, Groves and Mathiowetz (9) examine NHS rules in the context of telephone interviewing of households. Contrary to other results, they found higher reporting rates for proxy respondents than for self respondents. They explain these unusual results to the use of the telephone, the possible social stigma of males having to declare physical illnesses, and a seasonal effect due to the interviewing period.

The California Health Survey (21) also used the third household respondent rule. After the initial survey, a subsample (Post Enumeration Survey, PES) of the original sample was chosen for reinterview independently of the first interview. In the PES every respondent was interviewed in person. The distribution in the PES of adult self-reporters, adults who previously had a proxy respondent report, and children was the same as in the original CHS. The results suggested that adult reporters were fairly consistent in their responses about themselves. There was evidence of under-reporting by the proxy respondent.

In another health survey (7), the results showed that self-respondents reported 50% more clinically diagnosed conditions than proxy respondents. The same study compared the number of reported diagnosed conditions with those found in the clinical exam. Nineteen percent of the clinically diagnosed conditions were reported for an individual by a proxy respondent, whereas 29% of the clinically diagnosed conditions were reported by respondents themselves.

An amendment to the general Hawaii Health Survey (15) was designed to study the adequacy of data collected from a proxy respondent. Three

hundred pairs of individuals were asked to report independently on the smoking, drinking and dieting habits for himself and for his partner. Of the 300 pairs, 281 were husband-wife relationships. The remaining 19 pairs were other familial relationships. The results of this study led to a conclusion that surrogates, usually wives, can respond well to questions about personal habits of their partner. In addition, the authors concluded that for large scale interview studies where it is not practical or feasible to callback adults not at home, proxy respondents can be used. Interviewing the spouse should not compromise the data.

The authors of several of the previously reviewed articles suggest stricter respondent rules (i.e., allowing only a spouse or parent to answer for an absent household member) would reduce reporting error. There is no general consensus by the researchers that the use of a proxy respondent rule provides accurate results. Asking a household respondent general information about perceived health, labor force status, or dietary habits of another individual can produce good results. Less conclusive are results from studies dealing with household respondents' knowledge of specific information about illnesses, conditions, particular snack foods, or degree of perceived feelings.

A feasibility study was conducted by Statistics Canada in 1974 to investigate the possibility and problems of using a non-proxy reporting procedure on the Labour Force Survey (LFS). A small sample (450 households in three sections of the country) was involved in the six-month experiment. The households were not part of the regular LFS sample, and the interviewing was done in weeks which did not overlap the ongoing LFS survey period. Comparisons with the LFS are generally comparable, and Singh and Tessier (25) report that 1) nonresponse rates, and particularly refusal rates, were much higher for the experimental than for the the LFS sample; 2) the proxy rate was reduced by 61% in the experimental group (proxies were accepted as a last resort alternative to getting no information at all)\*\*the rate of reduction was higher for households containing three or fewer persons, and lower for households with four or more persons; 3) implementation of a non-proxy reporting procedure on a regular basis would involve a cost increase of 29%\*\*the data were not available to provide comparable callback ratios for the proxy and non-proxy procedures. Only fieldwork related data were collected in this study\*\*no estimates of differences in the quality of the data were obtained.

In contrast, the U.S. Bureau of the Census recently conducted some research which focused on the estimates produced by different types of respondents. In a Methods Test Panel conducted in conjunction with the Current Population Survey (CPS), an experiment was set up to test the effect of three different variables: 1) mode of interview\*\*telephone vs. personal visit; 2) assignment of interviewers\*\*maintaining the same interviewer throughout the eight-visit survey period vs. alternating interviewers for the sample

households; 3) type of respondent\*\* CPS household vs. self response vs. a random designation of a household respondent.

The three treatments were interpenetrated in a 2x2x3 experimental design, and approximately 1500 interviews were conducted monthly between May 1978 and November 1979 in four diverse areas of the country. According to an interim report (24) using data from June 1978 through March 1979, a complex pattern of interactions, which occurs between levels of the three treatment variables, prevents direct comparison of the levels of a variable. Thus, the authors cannot make a general statement about the differences between the unemployment estimates produced by the three types of respondents. The significant interactions are specified, however; for example, self response to a telephone interview produces a significantly higher estimate of the unemployment rate than does a personal visit with a household respondent (for others, see Roman and Woltman(24)).

The use of proxy respondents can decrease unit nonresponse and/or the cost of obtaining information from a unit, but at the same time can increase the rates of item nonresponse (individual questions). In the ISDP 1978 panel, incidence of nonresponse for individual income questions were greater for proxy than for self reporters (12).

A proxy procedure can increase information about unavailable respondents or even hard\*core refusals. Highly specific or personal information obtained from a proxy respondent about an individual is considered suspect, while more general or less personal information is more reliable. It seems that a proxy respondent rule creates the situation where an imputation procedure for item nonresponse, such as a hot deck, can be used. Large data sets are usually necessary for these imputation schemes (13). It was suggested that self\*respondents' data should be used as donors for imputing data sets containing proxy respondents (12).

While there is always the danger of a bias when using a household respondent, there is an acceptance of such procedures. Alternatives to the use of proxy respondents, such as advance forms or self\*administered questionnaires, are also subject to biases. It has been suggested that to attempt to control the possible biases, more complex respondent rules for determining the proxy respondent should be studied. Reinterviewing could be used to help correct response variability and biases (23). Bailor and Brooks (23) used the difference between self\*reporters and proxy spouse reports as an estimate for adjustment of the response error due to a proxy in the initial interview. The difference between self\*reporters and proxy non\*spouse reports were not used as estimates for adjustment, therefore the adjustment made from proxy spouse reports should cast suspicion on the results.

A caveat should be noted concerning some of the results reported in this literature review. Ratios obtained in 1956 or even 1973 for the number of callbacks necessary to complete

interviews using the self vs. proxy respondent rules may not be relevant to current data collection. According to the 1981 Statistical Abstract of the United States, 22.5% of the U.S. households in 1980 contained only one person, an increase of 5.5% since 1970. Slightly over half (54%) of all households in 1980 consisted of either one or two persons.

This trend of decreasing household size may effect callback differences between the two procedures. One person households would require the same number of callbacks regardless of which respondent rule was employed. From this viewpoint, one might speculate that differences between the rules would decrease as household size decreases. Callback ratios may not be the only factor to consider in evaluating cost differences. However, the Canadian experience (25) showed that more evening work, which might entail more overtime, was necessary to interview non\*proxy respondents. Also, assignment sizes were reduced; if done on a regular basis, this might affect the number of interviewers employed and thus the cost of the survey.

The impact of decreasing household size on the quality of the data collected by the various respondent rules depends somewhat on changing household composition. The percentage of two person households has remained relatively stable over time, but their composition has not. Roomates of the same or opposite sex may not be able to provide as much information as a spouse. Cartwright (6) in 1951 concluded that men provide better data for themselves than are provided by their wives. Conditions today would seem even more favorable to self rather than proxy response.

Comparisons of respondent rules are awkward. The magnitudes of the biases introduced by a proxy respondent are difficult to measure and are not constant from question to question and survey to survey. For each respondent rule, the components of cost are different and have to be adapted. One approach to compare respondent rules would be to fix the cost available for the design, let the sample sizes vary, and judge the worth of the rules with respect to the lowest error.

It is apparent that proxy respondents are widely used in practice and an attempt to understand their behavior has been made. The costs of using a proxy respondent rule can reduce the total cost of a survey, but the effects of proxy respondents on the quality of data is in dispute. Generally, the results of studies ascertaining the effects of proxy respondents are unfavorable, while in some instances proxy respondents can report as effectively as self\*respondents. The determining factors, on the effectiveness of a proxy respondents, are the nature of the information desired and the relationship of the proxy respondent to the designated respondent. Many of the reports are themselves unreliable due to the lack of statistical design for testing the effect of proxy respondents, excessive recall periods for reinterviews, and questionable ability to generalize from one population to another.

Nisselson (22) stated "as a result of this lack of conclusiveness in the evidence available, the extra cost of interviewing all adults for themselves was not considered a good investment for the national (Health Interview) survey." Current household composition may have a profound effect on the choice and use of respondent rules. Then as today, the propriety of the use of proxy respondents cannot be conclusively judged at this time.

#### Footnotes

- 1 A commonly known procedure for objectively selecting a single respondent within a household is presented by Kish (14).
- 2 Information about children below a predetermined age is always obtained from a parent or adult in charge of care for the child.

#### References

1. Aquilino, Robert (1971), "Methods Test Phase III: Third Report on the Accuracy of Retrospective Interviewing and Effects of Change in Respondents on Labor Force Data," unpublished memorandum, Washington, D.C., Bureau of the Census.
2. Cannell, Charles F. and Floyd J. Fowler (1963). "A Study of the Reporting of Visits to Doctors in the National Health Survey." Research Report. Survey Research Center. Ann Arbor, Michigan.
3. Cannell, Charles F. and Floyd J. Fowler (1965). "Comparison of Hospitalization Reporting in Three Survey Procedures." Series 2, No. 8 of Vital and Health Statistics.
4. Cannell, Charles F., Gordon Fisher, and Thomas Bakker (1965). Reporting of Hospitalization in the Health Interview Survey. Series 2, No.6 of Vital and Health Statistics.
5. Cannell, Charles F., Kent H. Marquis, and Andre Laurent (1977), A Summary of Studies of Interviewing Methodology, Washington, D.C., U.S. Government Printing Office. (Vital and Health Statistics, Series 244>Data Evaluations and Methods Research, No. 69.)
6. Cartwright, Ann (1951), "The Effect of Obtaining Information from Different Informants on Family Morbidity Inquiry." Applied Statistics, 6 (1), 15428.
7. Elinson, Jack and Ray E. Trussel (1957), "Some Factors Relating to Degree of Correspondence for Diagnostic Information as Obtained by Household interview and Clinical Examinations." American Journal of Public Health, 49.
8. Enterline, Philip E., and Katherine G. Capt (1957), "A Validation of Information

Provided by Household Respondents in Health Surveys." American Journal of Public Health, 45, 2084212.

9. Groves, Robert M., and Nancy A. Mathiowetz (1981). "An Experiment to Measure the Effects of Respondent Rules on Health Survey Responses." Proceedings of the Survey Research Methods Section of the American Statistical Association, 3464351.
10. Haases, Kenneth W. and Ronald W. Wilson (1972), "The Study Design of an Experiment to Measure the Effects of Using Proxy Responses in the National Health Survey." Proceedings of the Social Statistics Section of the American Statistical Association, 2894293.
11. Jones, Charles and Robert Aquilino (1970), "Methods Test Phase III: Second Report on the Accuracy of Retrospective Interviewing and Effects on Nonresponse on Labor Force Status." Unpublished Memorandum, Washington, D.C., Bureau of the Census.
12. Kalton, Graham, Daniel Kasprzyk, and Robert Santos (1981). "Issues of Nonresponse in the Survey of Income and Program Participation." Current Topics in Survey Sampling, eds. D. Krewski, R. Platek, and J.N.K. Rao. Academic Press, New York.
13. Kalton, Graham (1981). Compensating for Missing Survey Data. Survey Research Center, Institute for Social Research. Ann Arbor, Michigan.
14. Kish, Leslie (1949). "A Procedure for Objective Respondent Selection Within the Household." Journal of the American Statistical Association, 44, 3804387.
15. Kolomel, L.N., Hirohula, T. and Nomura, A. (1977), "Adequacy of SurveyData Collected from Substitute Respondents." American Journal of Epidemiology, 106, 4764484.
16. Koons, David A. (1973). Quality Control and Measurement of Non-sampling Error in the Health Interview Survey, Washington, D.C., Bureau of the Census, U.S. Government Printing Office, 20421. (Vital and Health Statistics Series 244>Data Evaluation and Methods Research, No. 54.)
17. Kovar, Mary Grace and Ronald W. Wilson (1976), "Perceived Health Status4How Good is Proxy Reporting." Proceedings of the Social Statistics Section of the American Statistical Association, Part II, 4954500.
18. Kovar, Mary Grace and Robert A. Wright (1973), "An Experiment With Alternate Respondent Rules in the National Health Interview Survey." Proceedings of the Social Statistics Section of the American Statistical Association,
19. Krueger, Dean E. (1957), "Measurement of Prevalence of Chronic Disease by Household

- Interviews and Clinical Evaluation." American Journal of Public Health, 47, 953-960.
20. Martin, Jean and Bob Butcher (1981), "The Quality of Proxy Information & Some Results from a Large Scale Study." Survey Methodology Bulletin, 12, 30-31.
  21. Mooney, H. William (1962), "Methodology in Two California Health Surveys." Public Health Monograph No. 70, PHS Pub. 942.
  22. Nisselson, Harold and Theodore D. Woolsey (1959), "Some Problems of the Household Interview Design for the National Health Survey." Journal of the American Statistical Association, 54, 69-87.
  23. Office of Federal Statistical Policy and Standards (1978), An Error Profile: Employment as Measured by the Current Population Survey. (Prepared by Camilla A. Brooks and Barbara A. Bailar, U.S. Bureau of the Census), Washington, D.C., U.S. Government Printing Office, 17-18.
  24. Roman, Anthony M. and Henry F. Woltman (1980), "The Methods Test Panel: Analysis of the Unemployment Rate. An Interim Report," Unpublished memorandum, Washington, D.C., U.S. Bureau of the Census.
  25. Singh, M.P. and R. Tessier (1975), Methods Test Panel Phase III: Feasibility Test for Complete Non-Proxy Procedure, Methodology and Analysis, Ottawa, Statistics Canada.
  26. Williams, Louis E. (1969), "Methods Test Panel Phase III: First Report on the Accuracy of Retrospective Interviewing and Effects of Nonself Response on Labor Force Status." Unpublished memorandum, Washington, D.C., U.S. Bureau of the Census.