

THE STUDY DESIGN OF AN EXPERIMENT TO MEASURE THE EFFECTS OF USING
PROXY RESPONSES IN THE NATIONAL HEALTH INTERVIEW SURVEY

Kenneth W. Haase and Ronald W. Wilson, National Center for Health Statistics

Although the first census was taken in this country in 1790, the utilization of an ongoing national survey procedure designed to obtain a large amount of detailed information about a specific area of interest, such as health, is still a rather new phenomenon. The Current Population Survey is only 32 years old. The Health Interview Survey conducted by the National Center for Health Statistics, although first implemented on an ad hoc basis in 1935, has only been an ongoing operation since 1957. For this reason there are still many things we do not know about the techniques of collecting valid and reliable information from a household respondent on topics which often can be quite complex.

Important public health decisions made at the national and local levels are often influenced by data obtained in the Health Interview Survey. These decisions can affect the lives of large segments of the population. In the process of improving both the validity and reliability of the data collection techniques, we must first try to better understand the types and magnitude of the errors and biases that exist within these procedures.

In assessing the validity of the statistics derived from the Health Interview Survey, of major concern has been the effect of the respondent rules which allow adult respondents to report for other family members. It is generally accepted that for most types of personal information, such as health, the best informant will, with a few exceptions, be the person to whom that information is related. The adoption of the present respondent rules was based on the concept that, while the validity of the data may be improved by implementing a "self-respondent only" procedure where all adults are interviewed for themselves, the greatly increased cost and decreased efficiency resulting from such a change would more than offset any gain.

Over the years, a number of special studies have been conducted to evaluate the validity of specific types of information obtained in the Health Interview Survey.^{1/} These studies often involve the validation of information obtained in the interview with records from selected health facilities where the respondents were patients (such as hospitals or clinics). Most of these studies tend to indicate that the most important problem is related to underreporting. There is a tendency for some people not to report all of their experiences either because of memory loss or for some other reason. Further analysis of these studies also indicate that the degree of underreporting for selected health variables tends to be more severe when the information is obtained through a proxy respondent rather than from the person himself.^{2/}

The evidence on reporting differences resulting from self or proxy respondents derived from these and other studies^{3/} is far from conclusive. This is due to a number of reasons, including the small sample size used in some of these studies, the fact that most of these studies were not specifically designed to investigate this specific phenomenon and the fact that these studies often included very select population groups which made inference to the general population difficult.

From April to June 1972, the Division of Health Interview Statistics conducted a study designed to measure the degree to which the use of proxy respondents affects the national statistics. The findings from this study could have a wide range of implications far beyond the National Health Survey and could affect the Census Bureau and other survey activities.

First, it might be helpful to look at some of the survey data which are related to this respondent problem. Data on approximately 40 percent of all adults in the Health Interview Survey sample are obtained from proxy respondents, that is, the information about them is obtained from another family member. If we look at the rates of selected health indices we find the following differences between data obtained from self and proxy respondents.

Table 1 presents the annual rate of restricted activity days for persons 17 years of age and over. The restricted activity statistic is derived from questions that ask about cutting down on the things a person usually does.

There are more than 17 days of restricted activity per year for each adult in the United States. However, there is considerable variation in these rates depending upon whether the information was obtained directly from the person himself or through another family member. Self-respondents report a rate of almost 20 days of restricted activity per year as compared to only 13 days when data are obtained by proxy respondents. The higher rates for self-respondents are found entirely among persons age 17-64.

For the 65 year and over age group, there is a reversal in the rates according to respondent status. The 29 days reported for the self-respondents is lower than the 35 days reported by the proxy respondents.

Table 2 shows the number of days of bed disability per person 17 years of age and over per year according to respondent type and age. Since by definition all bed days are also restricted activity days, we would, therefore, expect a somewhat similar pattern for these two sets of data. The overall rate for self-respondents is higher than the rate for proxy respondents, although this difference is not quite as great

as that observed for total restricted activity. For each of the age groups shown, a similar pattern is observed; the self-respondent rates for persons under 65 years of age are higher than the rates obtained by proxy respondents. Also, the same reversal is observed for persons over 65 years of age, with more than twice as many days of bed disability reported by proxy respondents as compared to self-respondents.

Table 3 presents the number of physician visits per person per year by respondent status and age. During 1970, there were five physician visits for each adult in the United States; whereas the self-respondent rate is 5.6 visits per person per year, the proxy respondents report only 4 visits per year. The self-proxy difference is most pronounced for persons under 65 years of age. The reversed pattern for persons 65 years and over that was observed for the disability data is not present in the physician visit data.

Estimates from the Health Interview Survey indicated that during 1970 there were 1.6 acute conditions for each adult in the United States. The same pattern for the two respondent groups emerges in Table 4 that was observed for physician visits.

Table 5 presents the annual number of short-stay hospital discharges per 100 persons 17 years of age and over. Again, as in all previous tables a higher rate for self-respondents than proxy respondents and that the younger age groups are the major contributors to this difference.

Summarizing the findings for these selected health indices, the rates for adult self-respondents are higher than those obtained by proxy respondents. The pattern held consistently for persons under 65 years of age. The pattern for persons over 65 was not consistent, with proxy respondents reporting higher rates for several variables. Although not shown here, there were also sex differences in the pattern for the aged. While these variations in the general pattern of higher rates for self-respondents are interesting in and of themselves, an indepth analysis of these variations are not essential to the main purpose of this paper.

What can be inferred from these data? The three most obvious hypotheses that can be developed to account for these differences are:

1. These differences are due to different degrees of error in reporting by proxy as compared to self-respondents.
2. These differences are due to differences within these two population groups and, therefore, reflect true differences and not reporting error.
3. These differences are due to a combination of both factors, reporting error and the select nature of the population groups involved.

The possibility that true differences exist between self-respondents and persons for whom data is obtained from another family member can be elaborated further. Whether data is gathered from a self or proxy respondent depends on whether or not the subject is available at the time of that interview. However, the person's availability may well be a consequence of his health status. Since most of the interviews are conducted during the work week and within the daylight hours, a person who might otherwise be expected to be at work or away from home for other activities but who is at home at the time of interview might well have a health problem and can conceivably inflate these health statistics for the self-respondent groups. The observed higher rates for persons over 65 for whom data were obtained by proxy could be explained by the fact that when these older persons are ill they are more likely to be bed-ridden or hospitalized and, therefore, data must be obtained by proxy.

To know how much of the differences in the rates for these health indices are due to reporting error as compared to true population differences, the following questions must be answered: First, what would be the rates of these selected health indices for persons responded for by proxy if they had responded for themselves? If this question is answered adequately, it is necessary to know how much of these differences in health indices are due to the use of proxy respondents and how much of these are due to differences in health characteristics between the persons at home and those not at home at the time of the interview.

If the difference in reporting bias between self and proxy respondents is significantly large, there would be a second question to be answered. Is the proxy respondent's relationship to the person for whom he is responding an important factor in these differences? Are there certain groups of respondents that can more adequately respond for another family member? For example, is a woman responding for her husband a better respondent than a woman responding for her father-in-law?

A third question that this study was designed to answer is how much will a change in the respondent rules cost, both in terms of dollars and other factors such as the nonresponse rate?

Finally, there is one important question this study will not answer, that is, "What type of respondent reports the most accurate information?" We would need to tie the present study to health records in order to determine the accuracy of reporting.

The remainder of this paper will describe the study design. During the second quarter of the 1972 data collection year, 6 of the 13 weeks were selected for the use of special respondent rules. Under the regular respondent rules, any responsible adult can respond for another related household member.

During the 6 weeks of the test, interviewers were instructed to follow their regular schedule in making their initial contact at a household. They then determined which adults were at home and obtained information from them as a self-respondent. Appointments were made to return and interview as self-respondents all adults not home at the time of the initial interview. However, a special notation was made to indicate that these persons would have been responded for by proxy under the normal survey rules. Thus, under the regular rules we have persons responded for by self and proxy determined by whoever is at home, while under the special rules we have all self-respondents, but some who would normally have been responded for by proxy under the old rules. If it was impossible to obtain a self-interview, rather than lose information, data were obtained by proxy, but this residual group will be deleted from the major analysis.

Table 6 further illustrates the study design and the proposed analysis plans.

The first two columns represent respondent status under the usual respondent rules. Column 3 represents those adults at home at the time of the initial interview. They would have been self-respondents under the usual procedure and will also be self-respondents under the experimental procedure. Column 4 represents those adults who were not at home during the initial interview and, therefore, would have been responded for by proxy under the usual procedure but will be interviewed for themselves during the experimental procedure.

This table illustrates how the various hypotheses discussed above can be tested. If the earlier observed differences between self and proxy respondents resulted entirely from reporting error, then Column 4 would approximate Column 1 and 3 and Column 4 would be greater than Column 2. If the different rates reflect true differences in the two groups, then Column 4 would approximate Column 2 and be less than Columns 1 and 3. If, however, both these factors account for the different rates, which is probably the case, then the effect of each can be calculated. Column 4 minus Column 2 represent the difference due to reporting and Column 3 minus Column 4 would represent the difference due to population difference. If a significant part of these self-proxy differences are a result of reporting error on the part of proxy respondents, it is necessary to determine if certain types of proxy respondents report more adequately than others. For this analysis the proxy columns (2) and (4) will be divided into spouse and other respondents.

What are the implications of this study? In terms of the Health Interview Survey, the first benefit that can be derived from this study is a much better understanding of how the national statistics are being affected by the utilization of the present respondent rules. Regardless of whether any changes are made in the respondent rules, this knowledge would greatly benefit all future analysis of Health Interview Survey sta-

tistics. If it is found that the self-proxy differences are primarily the result of true population differences, then probably no change in the respondent rules would be necessary.

If, however, it is found that the present respondent rules create a very significant bias in the statistical product, alternative procedures must be developed. These procedures could include being more discriminating in who can be a proxy respondent or a decision that only self-respondents are permissible. If the latter course of action is taken, this could possibly affect the entire household interview procedure. For example, if a self-respondent procedure is adopted it might be necessary to interview only a sample of the persons in the assigned households rather than all household members as is presently done.

REFERENCES

- (1) National Center for Health Statistics, Vital and Health Statistics, PHS Pub. 1000, Series 2, Public Health Service, Washington, U. S. Government Printing Office: "Measurement of Personal Health Expenditures," No. 2, (June 1963); "Health Interview Responses Compared With Medical Records," No. 7, (July 1965); "Interview Responses on Health Insurance Compared With Insurance Records," No. 18, (August 1966); "Interview Data on Chronic Conditions Compared With Information Derived from Medical Records," No. 23, (May 1967); "Development and Evaluation of an Expanded Hearing Loss Scale Questionnaire," No. 37, (April 1970); "Optimum Recall Period for Reporting Persons Injured in Motor Vehicle Accidents," No. 50, (April 1972).
 - (2) Series 2, Nos. 6 and 7
 - (3) Ballweg, John A., "Husband-Wife Response Similarities on Evaluative and Non-Evaluative Survey Questions," Public Opinion Quarterly, 33 (2) (Summer 1969), 249-254.
- Cartwright, Ann, "The Effect of Obtaining Information from Different Informants on a Family Morbidity Inquiry," Applied Statistics 6 (1), (March 1957), 18-25.
- Elinson, Jack, and Trussell, Ray E., "Some Factors Relating to Degree of Correspondence for Diagnostic Information as Obtained by Household Interview and Clinical Examinations," American Journal of Public Health, 49, (March 1957), 311-321.
- Enterline, Philip E., and Capt, Katherine G., "A Validation of Information Provided by Household Respondents in Health Surveys," American Journal of Public Health, 49, (Jan-June 1959), 205-212.
- Feldman, Jacob J., "The Household Interview as a Technique for the Collection of Morbidity Data," Journal of Chronic Diseases, 11, (May 1960), 534-554.

Fisher, Gordan, "A Discriminate Analysis of Reporting Errors in Health Interviews," Applied Statistics, 11 (3), (November 1962), 148-163.

Scanzoni, John, "A Note on the Sufficiency of Wife Responses in Family Research," Pacific Sociological Review, 8, (Fall 1965), 109-115.

Nisselson, Harold, and Woolsey, Theodore D., "Some Problems of the Household Interview Design for the National Health Survey," Journal of the American Statistical Association, 54, (March 1959), 69-87.

Weiss, David J., and Dawis, Rene V., "An Objective Validation of Factual Interviewing Data," Journal of Applied Psychology, 44 (6), (1960), 381-385.

Table 1. Number of days of restricted activity per person 17 years of age and over per year by age and respondent type: United States, 1970

Age of Subject	Respondent Type		
	Total	Self Respondent	Proxy Respondent
	Number of days per person per year		
All persons 17+	17.2	19.6	12.8
17-44 years	12.0	14.4	9.0
45-64 years	20.0	23.3	14.0
65+ years	30.7	29.2	35.2

Table 2. Number of days of bed disability per person 17 years of age and over per year by age and respondent status: United States, 1970

Age of Subject	Respondent Type		
	Total	Self Respondent	Proxy Respondent
	Number of days per person per year		
All persons 17+	7.0	7.5	6.2
17-44 years	4.9	5.6	4.0
45-64 years	7.5	8.5	5.8
65+ years	13.8	10.9	23.3

Table 3. Number of physician visits per person 17 years of age and over per year by age and respondent status: United States, 1970

Age of Subject	Respondent Type		
	Total	Self Respondent	Proxy Respondent
	Number of visits per person per year		
All persons 17+	5.0	5.6	4.0
17-44 years	4.6	5.4	3.5
45-64 years	5.2	5.6	4.6
65+ years	6.3	6.3	6.0

Table 4. Incidence of acute conditions per person 17 years of age and over per year by age and respondent status:
United States, 1970

Age of Subject	Respondent Type		
	Total	Self Respondent	Proxy Respondent
	Number of conditions per person per year		
All persons 17+	1.6	1.7	1.5
17-44 years	1.9	2.1	1.7
45-64 years	1.3	1.4	1.1
65+ years	1.0	1.1	0.9

Table 5. Number of discharges from short-stay hospitals per 100 persons 17 years of age and over per year by age and respondent status: United States 1970

Age of Subject	Respondent Type		
	Total	Self Respondent	Proxy Respondent
	Number of discharges per 100 persons per year		
All persons 17+	16.5	19.6	11.6
17-44 years	15.7	20.7	9.3
45-64 years	14.7	15.9	12.6
65+ years	23.4	23.1	24.4

Table 6. Example of the type of table format that can be used to analyze the findings from the Respondent Rules Study

Health Indices	Present Respondent Rule Procedure		Experimental Respondent Rule Procedure	
	Self	Proxy	Self Self	Proxy Self
	(1)	(2)	(3)	(4)
	<p>HO₁: Self-proxy differences entirely due to reporting bias Then: Col. (4) should equal Cols. (1) and (3) Col. (4) should be greater than Col. (2)</p> <p>HO₂: Self-proxy differences entirely due to population differences Then: Col. (4) should equal Col. (2) Col. (4) should be less than Cols. (1) and (3)</p> <p>HO₃: Both factors involved Then: Col. (4) minus Col. (2) should equal the difference due to reporting Col. (3) minus Col. (4) should equal the true differences in population</p>			