SURVEY OF INCOME AND PROGRAM PARTICIPATION

Working Paper Series

Proxy Reports: Results from a Record Check Study

No. 9032 137

Jeffrey C. Moore
Bureau of the Census

December 1990

This paper was prepared for the International Conference on Measurement Errors in Surveys, Tucson, Arizona, November 11-14, 1990. I gratefully acknowledge the invaluable contributions of Kent Marquis, both conceptual and technical, to the research described in this paper. The paper has also benefitted from careful reviews by Elizabeth Martin, Nancy Mathiowetz, and Lynn Weidman, to whom I am also grateful. Notwithstanding this wealth of expert assistance, the paper's shortcomings, of course, are solely attributable to me. In addition, the views expressed here are my own, and do not necessarily represent the official views or policies of the Census Bureau.

TABLE OF CONTENTS

PROXY REPORTS:
RESULTS FROM A RECORD CHECK STUDY

1. INTRODUCTION AND OVERVIEW

Among responsible survey organizations, understanding survey data quality is a pervasive and continuing concern. Data consumers--clients--have a right to demand the highest possible quality for given resources, and to know as much as possible about quality limitations which might affect their uses of the data. For those who conduct surveys, information about quality is essential for making optimal decisions about survey design. One key design decision faced by all household surveys is how much effort to expend to obtain data about each sample person from that person himself/herself. Stringent self-response rules can add substantially to survey costs and lengthen data collection periods, but are also generally assumed to yield higher quality data.

This paper examines response error differences between self- and proxy respondents in a major and relatively new Census Bureau household survey--the Survey of Income and Program Participation (SIPP). Standard SIPP procedures call for self-response interviews with all adult members of sampled households if possible, but allow proxy interviews for those not present (or unable or unwilling to be interviewed) at the time of the interviewer's visit. As a result of these respondent rules, a substantial proportion of SIPP interviews are conducted by proxy. The large number of proxy interviews, coupled with concerns about the quality of SIPP data, have led to questions about the contribution of proxies to SIPP data quality problems. If proxies were found to contribute disproportionately to SIPP's quality problems, the survey's designers and sponsors would have to consider imposing more stringent self-response rules.

The evaluation uses a record check against administrative data as the means for assessing response errors. Overall, the investigation finds only weak evidence that proxy reports of program participation are consistently more subject to error than self-reports. The absence of important differences holds up even under conditions which might be expected to have especially adverse effects on proxies--maximum recall difficulty, due to a long recall interval, and a large number of persons for whom to report.

The paper begins with a description of SIPP and the design of the SIPP Record Check Study. This "Background" section also discusses some of the major response error issues of concern to the survey's sponsors and designers, and summarizes briefly what the survey methods research literature has to say concerning the effects of self/proxy response status and response error.

The third section, "Methods," describes the analyses. This section defines the response error characteristics on which we compare self- and proxy respondents, and describes the analytical approaches to the investigation of response error differences.

Section four presents the "Results," which focus on errors in reports of monthly participation in government transfer programs. The major findings can be summarized very simply: for the states, programs, and time periods investigated here we find little evidence of important or consistent response error differences between self- and proxy respondents. The paper concludes with a discussion of possible explanations for these apparently counterintuitive results.

2. BACKGROUND

This section describes the Survey of Income and Program Participation and the SIPP Record Check Study. It discusses self/proxy response quality issues in general, and then highlights some of the key response error issues of concern to SIPP's designers and sponsors, including assumptions about the association of self/proxy response status with error magnitudes and directions.

2.1 The Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is a major Census Bureau survey designed to provide improved information on the economic situation of people and families in the United States. SIPP collects comprehensive longitudinal data on cash and noncash income, eligibility for and participation in state and Federal Government transfer programs, assets and liabilities, labor force participation, and many other related topics. For a detailed description of the SIPP program, see Nelson, McMillen, and Kasprzyk (1985).

SIPP sample housing units are selected to represent the noninstitutional population of the United States. A new panel--currently about 13,000 households--is introduced each February. Each sample household is interviewed by personal visit eight times, at four month intervals, over a 2-1/2 year period. The reference period for each interview is the four months preceding the interview month. At each visit to the household, each person fifteen years of age or older is asked to provide information about himself or herself, but proxy reporting is permitted for household members not available (or unable or unwilling) to self-respond at the time of the visit.

SIPP measures program participation on a monthly basis, using a large number of questions to elicit reports of which programs are providing benefits, and detailed followup questions to elicit reports of the monthly dollar amounts received from them. In the second and subsequent waves of the panel, respondents are reminded of which programs were reported for the sample person in the previous interview[1].

---

[1] For various reasons, final SIPP participation information may differ from the original information (e.g., due to computer editing). In the rare instances where there is a difference, the analyses reported in this paper use the final values.

2.2 The SIPP Record Check Study

The purposes of the SIPP Record Check Study are to provide an evaluation of the quality of the major program participation data gathered in SIPP and to generate ideas for improving the data quality. Elsewhere (Moore and Marquis, 1989) we have described the project in detail. The major aspects of the research--including the record check design, the people, programs, and time periods which comprise the data for the study, and the matching procedures employed--are summarized below.

2.2.1 Basic Record Check Design

The SIPP Record Check uses a "full" rather than a one-directional design, which permits the evaluation of the full range of survey responses--for example, both "yes" and "no" reports of program participation. Marquis (1978) describes the limitations of partial designs (e.g., checking records only for those who report in the survey that they possess the characteristic of interest; or surveying people known to possess the characteristic to see if they report it), which are almost guaranteed to produce biased estimates of survey measurement errors.

2.2.2 Programs

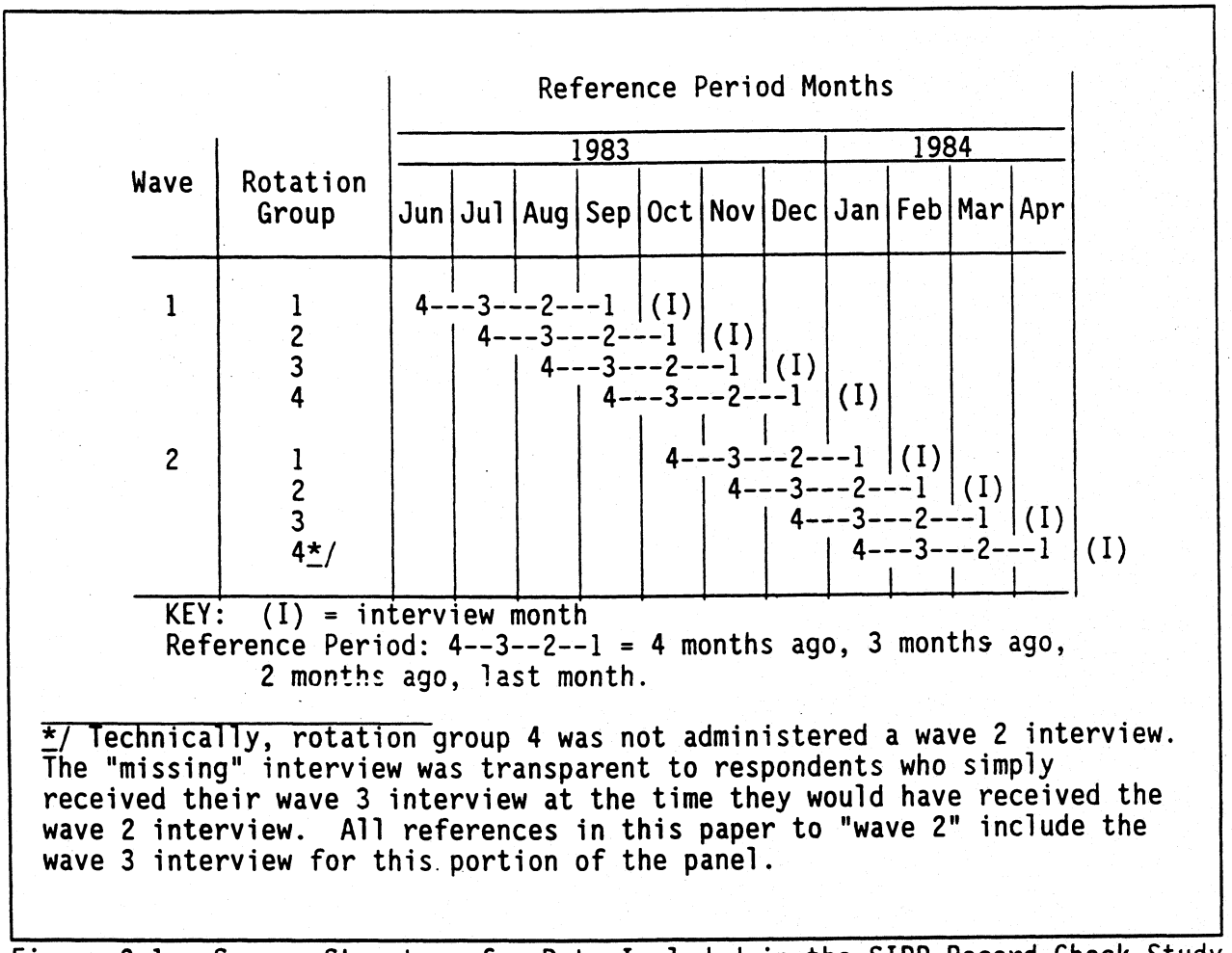
The SIPP Record Check Study obtained program participation records for eight government transfer programs, half administered by the states and half administered by the Federal Government[2]. These programs, and their acronyms are:

State-administered programs:	
Aid to Families with Dependent Children	(AFDC)
Food Stamps	(FOOD)
Unemployment Insurance	(UNEM)
Workers' Compensation	(WORK)
Federally administered programs:	
Federal Civil Service Retirement	(CSRET)
Old Age Survivors Disability Insurance ("social security")	(OASDI)
Supplemental Security Income	(SSI)
Veterans' Pensions and Compensation	(VETS)

From each agency we obtained identifying information (for matching) and monthly benefit receipt information (for response error assessment) for all persons who received income from the target program at any time from May 1983 through June 1984 (see below). The administrative records provide comprehensive coverage of the population in each state, and define program participation and benefits in virtually the same way that SIPP does.

---

[2] We also obtained records on recipients of Pell Grants. Analysis of these data are as yet incomplete, and we do not include them in this report.



*/ Technically, rotation group 4 was not administered a wave 2 interview. The "missing" interview was transparent to respondents who simply received their wave 3 interview at the time they would have received the wave 2 interview. All references in this paper to "wave 2" include the wave 3 interview for this portion of the panel.

Figure 2.1: Survey Structure for Data Included in the SIPP Record Check Study

2.2.3 Time periods

The interview data are from the first two interviews ("waves") of the 1984 SIPP Panel, for which interviewing began in October 1983. Figure 2.1 illustrates the wave, rotation group, interview month, and reference period structure for the survey data. As shown in the figure, the calendar months in the reference periods for the first two interviews for all rotation groups include June 1983 through April 1984. In our analyses, however, we ignore calendar months, and instead refer to the time periods covered by the survey data in terms of SIPP wave and reference month--e.g., wave 1, month 4; wave 1, month 3, etc. This is preferable because of the staggered rotation group structure of SIPP.

2.2.4 States and People

The record check study was conducted in four states: Florida, New York, Pennsylvania, and Wisconsin. These states were selected for convenience, and are not necessarily representative of the larger SIPP sample. The primary selection criteria included the following:

1) a reasonably large SIPP sample;

2) an appropriate, high quality, computerized, comprehensive, and accessible administrative record system for the programs of interest;

3) a willingness to share detailed, individual-level data for purposes of the research; and

4) some geographic diversity.

For the first two waves of the 1984 SIPP Panel the total SIPP sample included about 20,000 interviewed households. Of these, about 5,000 were included in the record check, yielding a total of about 11,000 people.

The analyses reported in this paper do not use all available SIPP sample persons. The major restriction is that the approximately 2,700 children under age 15--who are included as sample persons but not interviewed--are excluded. Other restrictions are as follows:

1) approximately 350 adult sample persons who refused to report their social security number in the survey (SSN refusers) were excluded from the personal identifiers file made available to the study for matching--although we have survey data for these people, we exclude them from our analyses because they were not subjected to matching against the administrative records[3];

2) approximately 500 adult sample persons for whom data reported by self or proxy were not available for all eight months (e.g., deaths, movers, refusers) are excluded from the analysis files; and

3) for the state-administered programs (AFDC, FOOD, UNEM, and WORK) we exclude the New York portion of the sample, about 2,300 cases, because there are some unresolved issues concerning the quality of selected data fields in the available New York administrative files.

For the Federal-level programs, then, the total number of sample persons available for analysis is about 7,550; for the state-level programs about 5,200.

---

[3] Matching procedures occasionally linked an SSN refuser's administrative record(s) to another member of the SSN refuser's household (because of similarities in name, address, etc.). Such SSN refuser matches became apparent through our review of all cases in which a child had been matched to any administrative data (prior to deleting children from the analysis group) and all matches which we had classified as "uncertain." If we judged that, based on available match information (such as name, age, sex, etc.), there was a better match to someone else in the household, someone other than the child or the original "uncertain" match target, then we manually relinked the information to that person. In this way a small number of SSN refusers are reincluded into the analysis group for selected programs (usually not more than two or three per program).

2.2.5 Matching

The record check used computerized matching software developed by the Census Bureau's Record Linkage Research Staff (e.g., LaPlant, 1989; Jaro, 1989), which is based on the theoretical work of Fellegi and Sunter (1969). The major advantages of this system (over, say, a clerical match) are its speed, its ability to process huge data sets, its ability to evaluate a match based on many variables simultaneously, and its ability to resolve, consistently and objectively, possible matches that differed on the value of one or more match variables. We matched on variables that were very likely to uniquely identify people such as their name, address, social security number and date of birth. See Moore and Marquis (1989) for a description of the matching techniques used in the record check[4].

2.3 SIPP Response Error Issues

Concerns about SIPP response quality have focused primarily on two areas: monthly reporting of participation in various transfer programs, and the measurement of month-to-month participation changes.

2.3.1 Monthly Program Participation Reports

Until recently, the primary information on the quality of monthly participation reports has come from "benchmark" comparisons, in which SIPP estimates are aggregated to national totals and compared to official estimates of participation rates and participation change rates from the sponsoring agencies. The aggregated SIPP change estimates have been found to coincide fairly well with the benchmarks (Singh, Weidman, and Shapiro, 1988), but participation rates suggest a general tendency toward negative bias in SIPP which is notable for some programs--Aid to Families with Dependent Children (AFDC), for example (Kasprzyk and Herriot, 1986).

One problem with the benchmark technique is that observed discrepancies can have many causes--e.g., sample differences, nonresponse bias, differences in definitions or concepts--which have nothing to do with response error. The

---

[4] Although they are treated as matches in our analyses, there is uncertainty about whether a few of the matches should have been made. In most cases these matches met the statistical criteria for a match, but our review suggested either that the match should be made to someone else in this family (e.g., someone not included in the match file) or that there was an error in measuring one of the match attributes either in SIPP or the administrative record. Additionally, since we excluded people under age 15 from the analysis, some of whom were legitimate SSI or OASDI program recipients, we relinked any such SSI or OASDI administrative record information to an appropriate adult in the household (e.g., the adult payee for a child social security beneficiary) who should have reported the income for the child in the survey. We also classified these matches as uncertain. By program, the numbers of uncertain matches are as follows: AFDC=14 (out of a total of 173 matches), FOOD=13 (out of 282), UNEM=64 (491), WORK=12 (127), CSRET=1 (69), OASDI=109 (1,617), SSI=19 (126), and VETS=14 (151).

case-by-case record check approach, however, permits a clearer assessment of the role of response error in producing the discrepancies. We (Marquis and Moore, 1990) recently have reported findings from the SIPP Record Check which tend to corroborate the results of the benchmark studies and to confirm the presence of response bias. Our data show a negative net response bias in program participation reports--quite severe in some cases--for seven of the eight programs tested.

An additional shortcoming of benchmark studies is that they are uninformative concerning the total extent of response error in respondents' reports. An aggregated SIPP estimate which agrees quite well with an external benchmark may simply contain a large number of compensating errors. Marquis and Moore (1990) also use the record check results to estimate the total extent of response error for reports of monthly program participation and month-to-month participation change, not just the net effect of the errors on estimates of population means and totals. Such information is essential for understanding response error effects on higher-order estimates such as estimates of association.

2.3.2 Month-to-Month Participation Change Reports

Much attention has been directed toward the quality of SIPP's participation change estimates, probably because an obvious quality problem--the "seam bias" phenomenon--has proved so ubiquitous and easy to document. Many investigators, looking at month-to-month program participation change in SIPP, have found that changes occur with much greater frequency between adjacent months which span two interviews (i.e., at the interview "seam") than between adjacent months within a single interview (Burkhead, 1985a, 1985b, 1985c; Burkhead and Coder, 1985; Coder, 1986; Judkins, 1986; Ryscavage and Feldman-Harkins, 1986; Weidman, 1986). As Czajka (1983), Moore and Kasprzyk (1984), and Hill (1987) have demonstrated, the seam bias effect is not unique to SIPP.

The seam effect represents a clear data quality problem, since there is no reason why month-to-month true change should differ consistently for seam and non-seam month pairs. Our recent record check work (Moore and Marquis, 1989; Marquis and Moore, 1989; Marquis and Moore, 1990) suggests that neither the seam nor the non-seam change estimate is unbiased. A simplified summary of these results is that, in general, SIPP non-seam change estimates tend to underestimate true change, while change estimates at the seam tend to be too high.

2.4 Self/Proxy Status and Survey Response Error

Survey designers have studied the issue of the relative quality of self- and proxy response survey reports for at least three decades, ever since Cartwright (1957) reported the results of a pilot experiment in which husbands' self-reports of their own "illnesses" greatly exceeded the reports of their wives acting as proxies. (Horvitz (1952), five years earlier, had noted in passing a similar result in a large-scale field survey.) Cartwright considered several possible causes of the reporting discrepancy: a husband's illness is of lower "salience" for the wife-proxy, and thus causes her to report less completely; proxy reports might also suffer because of a simple

lack of information; or, the wives might simply have grown weary with the additional reporting load (since each had to report for both herself and her husband) and failed to report their husbands' illnesses out of fatigue.

Traditionally, however, the primary concern about response errors in proxy reports is that proxies have less immediate and complete knowledge of the subject of their reports than do self-respondents (Sudman and Andersen, 1977; Mathiowetz and Groves, 1985), so that a proxy's information retrieval in response to a survey question is more prone to error. In addition, because the decay function is steeper for information less effectively encoded in memory originally, the advantage of self-respondents over proxies is likely to become more pronounced with the passage of time between the to-be-recalled material and the act of retrieval.

Through the years other hypotheses about presumed self/proxy response quality differences have also been added to the list--proxies are more prone to forget or to misplace events in time (Sudman and Bradburn, 1974), but they are less susceptible to bias due to perceived threat or to social desirability pressures (Sudman and Andersen, 1977)--and many more researchers have reported differences in the response profiles of self- and proxy respondents (e.g., Enterline and Capt, 1959; Haase and Wilson, 1972; Kilss and Alwey, 1976; Berk et al., 1982).

However, there is an inferential flaw in the leap from response differences between naturally-occurring self- and proxy respondents to response quality differences, since the observed effects could easily reflect true differences due to self selection. Moore (1988) describes this problem more fully, and also outlines other methodological issues which call into question much of the evidence in the literature which purports to address self/proxy response error differences. His review, and others' as well (e.g., Mathiowetz and Groves, 1985), suggests that the common assumption of quality differences--particularly, of increased response error when proxies are involved--is at best unproven.

2.5 Purpose and Limits of the Current Investigation

Although research support for self/proxy response error differences is perhaps not as uniform and compelling as is generally assumed, the notion's intuitive appeal certainly remains, and when survey designers become concerned about response quality, respondent rules are a natural target. This is true when evidence of response error is circumstantial, and even more so when it is more concrete.

The evidence of response errors in SIPP--possible underreporting bias for participation in some programs, and the problems with participation change estimates--have led to consideration of design changes for SIPP to attempt to reduce errors. Inevitably, discussions of possible design changes focus on SIPP's relatively lax respondent rules, which do not require (or even encourage) interviewers to make repeated visits to households to maximize self-response. As a result, about one-third of SIPP interviews are conducted by proxy, a proportion that has remained quite stable over the entire life of the survey (Bowie, 1989). Despite the growing body of literature that fails

to find evidence that proxies reports are more prone to error than self-reports, the assumption persists that more stringent rules about accepting proxy interviews would improve data quality.

The purpose of this paper is to address this assumption with the response error information available from the SIPP Record Check Study. Do self- and proxy reports of monthly program participation differ overall on the extent to which they are contaminated by response error? If there are no differences in general, can we at least identify certain response situations common in SIPP--for example, an especially long recall interval, or the presence of large numbers of people to be responded for--which are particularly detrimental to proxy reporting? Does the quality of participation change estimates differ for self- and proxy reports? Answers to these questions will certainly help inform SIPP design decisions, and may also assist survey researchers in general to focus survey design resources efficiently to elicit data of the highest possible quality.

The non-experimental nature of this investigation necessitates a word of caution about the conclusions that can be drawn from the results. Self/proxy response status in SIPP is not a randomly assigned experimental treatment, but simply occurs naturally in the execution of survey procedures. . Thus, without strong assumptions the data do not permit ruling out self-selection bias as a possible explanation for any observed effects. (As noted, Moore (1988) offers a detailed discussion of this issue.)

In practical terms, this limits what we can conclude from the differences we observe here about the likely impact of different respondent rules. If proxy data are more prone to error, it may be because of some set of characteristics which affect both the likelihood of self-response and the likelihood of error. If this were the case, then the self-reports of those originally interviewed by proxy would be expected to show no less error than the original proxy reports, and more stringent rules would not result in better data.

In our analyses we control for one major source of confounding by modelling two different types of response errors separately. However, there may be other factors which also affect both response quality and self/proxy status which we have not taken into account. The non-experimental nature of the research design necessarily limits the conclusions we can draw to statements about how consistent the findings are with hypotheses attributing causality to self/proxy response status.

3. METHODS

This section describes in some detail the assumptions and procedures which were used to produce the results which follow. We first define our response error assumptions, with examples of the derivation of our error estimates, and then outline the analytical procedures used to examine self/proxy response error differences in SIPP.

3.1 Definition of Response Errors

We treat both reports of monthly program participation and month-to-month participation change reports as binary variables, where 0 means not participating (or no change in participation status between adjacent months) and 1 denotes participation (or change). The response error scores are derived by comparing responses from SIPP to the true values from administrative records. We discuss several kinds of response error, all defined from the 2 x 2 table in Figure 3.1.

REPORTED PARTICIPATION	TRUE PARTICIPATION YES = 1	TRUE PARTICIPATION NO = 0	
YES = 1	a	b	
NO = 0	c	d	
	a + c	b + d	N

Figure 3.1: Notation for Cross-Classified Reported and True Values.

Cell entries a, b, c, and d represent frequencies of reported and true characteristics (in this case for program participation, although the logic is identical for participation change) for any SIPP reference month. N is the sample size. The total number of wrong answers (or MISCLASSIFICATION errors) for a program is b + c. The rate of misclassification errors is (b + c) / N. The frequency of UNDERREPORT errors is c. The underreporting error rate, which is conditional on a true positive, is c / (a + c). Similarly, the frequency of OVERREPORT errors is b, and the rate, conditional on a true negative, is b / (b + d).

3.2 Analytical Approaches

We use two basic approaches to investigate possible response error differences between self- and proxy respondents. First, for each program, we calculate self and proxy monthly participation error estimates for the three types of errors defined above for each month, and average these estimates over the four months of each wave. Similarly for participation change, we calculate the error estimates for each pair of adjacent months, focusing on the wave 1 and wave 2 averages for these estimates for month-pairs within each wave, as well as the estimate which straddles the interview seam[5]. For both analyses it is necessary to treat each wave separately because self/proxy status can change from one wave to the next.

Our second approach is to investigate two additional variables which might mediate self/proxy response differences, but which are hidden in the initial global analyses. In essence, this approach looks for conditions under which one might expect proxy data to suffer especially severely, to see whether

---

[5] Note that the within-wave averages for the directional errors are unweighted--that is, the within-wave average underreporting rate for a respondent who was "true yes" in only one month of the reference period is given equal weight as the rate for a respondent who was "true yes" in all four months.

under these extreme conditions the "expected" self/proxy response error effects might manifest themselves. We use as additional potential explanatory variables two characteristics which have been suggested as possible causes of reduced proxy data quality--increased difficulty of recall due to a longer recall interval, and a high reporting "load" due to the presence of many persons in the household for whom to respond.

We test the effects of recall length by comparing the average recall length effect (which we define as the "4 months ago" error rate minus the "last month" error rate) for self-respondents with the effect for proxies. To investigate the effects of reporting "load" we use an analysis of variance approach, with an F-test of the interaction between self/proxy response status and small/large load (1- or 2-person versus 3-or-more-person household) on the various types of response errors. Again, because self/proxy status can change from one wave to the next, both the recall length and load analyses treat each wave separately[6]. In addition, both analyses are restricted to monthly participation reports because of the problem of sparse cells for participation change.

We test self/proxy differences for statistical significance using an approximate t-test procedure that adjusts for unequal variances and uses the Satterthwaite approximation for the degrees of freedom. The test assumes that self- and proxy reporters are independent groups, which. of course. they are not, and thus fails to take into account the within-person correlation of errors. For some analyses we treat each reference month's report as an independent observation, again failing to take into account the within-person (over time) correlation of errors. The result of these assumptions is a conservative bias in the t-test, increasing the chances of failing to detect a true self/proxy difference. However, bias in the opposite direction results from the fact that for these and all of our inferential statistics we assume simple random sampling. This assumption is not true, and could lead us to reject the null hypothesis with a higher probability than the nominal .05 level. These uncertainties, and the low power of some of our tests due to small numbers of cases, necessitate caution in the interpretation of the statistical inferences. In fact, the consistent patterns in the data across programs and time periods may be a more appropriate primary indicator of the true underlying effects than the statistical test results.

4. RESULTS

4.1 Self and Proxy Response Errors in Monthly Program Participation Reports

Table 4.1 presents self and proxy response error rate estimates for monthly program participation reports, averaged across each wave, for all three types

---

[6] SIPP procedures are blind to any household composition changes which might have occurred in the reference period for the first interview. In subsequent waves, however, household size can change monthly. For wave 2 we defined the household size category based on the largest household size in any month of the reference period.

of response error: total misclassifications, underreports, and overreports. The parenthetical entries in the (n) rows indicate the numbers of cases in each respondent status class for each type of error. Because these are averages over four months, a person will appear in the analysis of both overreports and underreports if the administrative record indicates both a "yes" and "no" participation in the program for the period (so the sum of the underreport and overreport n's may exceed the n in the misclassification column).

As shown in Table 4.1, when the dependent variable is misclassification error, the self-proxy difference is statistically significant in both waves for only the OASDI program (a double underline under the difference value indicates statistical significance at or beyond the .05 level). Interestingly, however, the pattern of the differences clearly suggests that self-respondents make more errors, overall, than proxies. This is the direction of all three of the significant differences, and six of the eight programs show consistently (both waves) more misclassification error in self-response data than in proxy data. This is a nice example of the potential for confounding that exists when an experimental design is absent--the probability of error is higher when the true value is "yes" (as can be seen from the fact that underreport error rates far exceed overreport rates), and the true value is more likely to be "yes" for self-respondents (as indicated by the greater ratio of self-respondents to proxies in the underreport (true "yes") column than in the overreport (true "no") column).

Thus, we controlled for the effects of true participation differences between self- and proxy respondents by estimating underreport and overreport errors separately. The underreporting results in Table 4.1 show consistently and significantly higher rates of underreporting for proxy respondents in only one of the eight programs, AFDC. However, all five of the significant differences are in this direction, and for six of the programs the consistent direction of the effect is for proxy responses to contain more underreport errors. The overreport results resemble the misclassification results. Only OASDI shows consistent, statistically significant differences in both waves, although the pattern of results--indicating that self-respondents make more overreporting errors than proxy respondents--is mirrored in four other programs.

The monthly participation data, then, yield somewhat conflicting evidence regarding quality differences between self- and proxy reports. The overreport results suggest that when there is no true participation, self-respondents may be more prone to false positive reports than proxies. On the other hand, when there is true participation to report, proxies may fail to do so more often than self-respondents. However, especially in the latter case, a focus on self-proxy differences is in one sense highly misleading. Underreporting error rates in SIPP are extremely high, especially for the more volatile state-level programs, regardless of self/proxy response status. Even if proxy reporting were completely eliminated it is clear that underreporting error levels would remain unacceptably high.

Next we turn our attention to errors in the measurement of month-to-month participation change.

TABLE 4.1: RESPONSE ERROR RATES FOR MONTHLY PROGRAM PARTICIPATION BY TYPE OF ERROR, SELF/PROXY RESPONDENT STATUS, PROGRAM, AND WAVE

Part 1: State Programs

		Misclassification			Underreport			Overreport		
		Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference
PROGRAM	WAVE									
AFDC	1	.0119	.0091	+.0028	.468	.773	-.305	.0023	.0018	+.0005
	(ave. n)	(3541)	(1671)		(94)	(22)		(3471)	(1659)	
	2	.0126	.0069	+.0057	.482	.754	-.272	.0019	.0017	+.0002
	(ave. n)	(3400)	(1812)		(91)	(19)		(3325)	(1802)	
FOOD	1	.0140	.0093	+.0047	.253	.361	-.108	.0055	.0039	+.0016
	(ave. n)	(3541)	(1671)		(184)	(36)		(3407)	(1648)	
	2	.0126	.0090	+.0036	.241	.483	-.242	.0063	.0039	+.0024
	(ave. n)	(3400)	(1812)		(173)	(35)		(3277)	(1793)	
UNEM	1	.0126	.0183	-.0057	.413	.545	-.132	.0056	.0094	-.0038
	(ave. n)	(3539)	(1669)		(151)	(68)		(3490)	(1657)	
	2	.0156	.0155	+.0001	.434	.515	-.081	.0094	.0081	+.0013
	(ave. n)	(3397)	(1811)		(153)	(74)		(3358)	(1795)	
WORK	1	.0046	.0070	-.0024	.445	.733	-.288	.0024	.0036	-.0012
	(ave. n)	(3540)	(1672)		(32)	(15)		(3527)	(1668)	
	2	.0053	.0052	+.0001	.571	.750	-.179	.0023	.0014	+.0009
	(ave. n)	(3399)	(1813)		(28)	(12)		(3384)	(1806)	

Part 2: Federal Programs

CSRET	1	.0027	.0011	+.0016	.159	.167	-.008	.0008	.0007	+.0001
	(ave. n)	(5139)	(2408)		(63)	(6)		(5076)	(2402)	
	2	.0019	.0011	+.0008	.117	.286	-.169	.00041	.00038	+.00003
	(ave. n)	(4894)	(2653)		(62)	(7)		(4832)	(2646)	
OASDI	1	.0250	.0171	+.0079	.056	.099	-.043	.0180	.0094	+.0086
	(ave. n)	(5140)	(2410)		(1230)	(238)		(3935)	(2179)	
	2	.0266	.0169	+.0097	.056	.084	-.028	.0211	.0102	+.0109
	(ave. n)	(4895)	(2655)		(1186)	(313)		(3744)	(2349)	
SSI	1	.0057	.0046	+.0011	.261	.241	+.020	.0015	.0024	-.0009
	(ave. n)	(5139)	(2409)		(90)	(29)		(5054)	(2383)	
	2	.0057	.0041	+.0016	.269	.200	+.069	.0018	.0023	-.0005
	(ave. n)	(4894)	(2654)		(91)	(30)		(4812)	(2628)	
VETS	1	.0066	.0047	+.0019	.178	.173	+.005	.0032	.0017	+.0015
	(ave. n)	(5139)	(2408)		(107)	(42)		(5034)	(2366)	
	2	.0061	.0057	+.0004	.162	.178	-.016	.00271	.00268	+.00003
	(ave. n)	(4894)	(2653)		(105)	(45)		(4789)	(2608)	

4.2 Self and Proxy Response Errors in Month-to-Month Participation Change Reports

Table 4.2 presents the measurement error estimates for month-to-month program participation change. At the interview seam the "Self" category means self-response in both interviews; "Proxy" refers to all other circumstances (i.e., proxy involvement in either or both waves). The parenthetical entries in the (n) rows indicate the numbers of cases in each respondent status class for each type of error. Again, a person can appear in both the analysis of over-reports and underreports if the administrative record indicates both "change" and "no change" for the period, so the sum of the underreport and overreport n's may exceed the n in the misclassification analysis.

As shown in Table 4.2, no program shows a consistent (i.e., significant in both waves) self/proxy difference in participation change misclassification errors, nor is there much of a trend to the direction of the differences. For three programs the within-wave differences are consistently positive, four have mixed signs, and one shows consistently negative estimates. At the interview seam the difference is positive for seven of the eight programs, indicating more errors for self-respondents, but none of the effects is significant.

Elsewhere (Marquis and Moore, 1990), we have shown that the "seam bias" in SIPP change reports--the tendency for measured change to be much higher at the seam than across months within a single interview--results from a combination of underreporting of within-wave change and overreporting of change at the seam[7]. Thus, the directional errors are central to the investigation of self/proxy differences. The data in Table 4.2 offer only very modest evidence that the seam bias effect is at all exacerbated when proxy respondents are involved.

First, consider the within-wave underreporting component of the seam bias effect. None of the six analyzable programs (those with some true change for both self- and proxy respondents to underreport) shows a consistently significant within-wave difference. There is only one significant effect, although the fact that eight of the twelve within-wave underreport differences are negative may indicate a general tendency for greater proxy error. Here again, however, a focus on differences diverts attention from the only truly noteworthy aspect of these results--and that is SIPP's inability to elicit correct reports of true participation change within a survey wave, regardless of respondent status. In the <u>best</u> of circumstances (wave 1 self-respondents' WORK reports), 52.6% of the true changes were undetected by the survey. With a baseline level of error this high, neither eliminating proxy reporting nor

---

[7] It is perhaps a fine point, but it is somewhat misleading to refer to "misreports" of participation change, since SIPP respondents do not actually report change; they report monthly participation from which change can be inferred. Although this distinction may be quite blurred within an interview, it is certainly appropriate for changes measured at the seam, where the reports for each month of the adjacent month-pair are gathered four months apart.

TABLE 4.2: RESPONSE ERROR RATES FOR PROGRAM PARTICIPATION CHANGE BY TYPE OF ERROR, SELF/PROXY RESPONDENT STATUS, AND PROGRAM: WITHIN WAVE 1, WITHIN WAVE 2, AND AT THE WAVE 1 / WAVE 2 INTERVIEW "SEAM"

Part 1: State Programs

PROGRAM	WAVE	Misclassification			Underreport			Overreport		
		Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference
AFDC	1	.0024	.0022	+.0003	.833	.900	-.067	0	.0006	-.0006
	(n)	(3541)	(1671)		(24)	(10)		(3541)	(1671)	
	2	.0025	.0018	+.0006	.938	.889	+.049	.0012	.0003	+.0009
	(n)	(3400)	(1812)		(16)	(9)		(3400)	(1812)	
	"seam"	.0042	.0026	+.0016	.875	.500	+.375	.0017	.0022	-.0004
	(n)	(2889)	(2323)		(8)	(2)		(2881)	(2321)	
FOOD	1	.0050	.0034	+.0016	.600	.705	-.105	.0024	.0012	+.0012
	(n)	(3541)	(1671)		(50)	(13)		(3541)	(1670)	
	2	.0066	.0037	+.0029	.770	.750	+.020	.0027	.0017	+.0010
	(n)	(3400)	(1812)		(50)	(16)		(3399)	(1812)	
	"seam"	.0087	.0077	+.0009	.455	.833	-.379	.0069	.0056	+.0013
	(n)	(2889)	(2323)		(11)	(6)		(2878)	(2317)	
UNEM	1	.0109	.0132	-.0023	.618	.661	-.043	.0049	.0052	-.0003
	(n)	(3539)	(1669)		(102)	(56)		(3539)	(1666)	
	2	.0143	.0144	-.0000	.646	.793	-.147	.0087	.0055	+.0032
	(n)	(3397)	(1811)		(114)	(58)		(3395)	(1810)	
	"seam"	.0194	.0194	+.0000	.511	.593	-.082	.0113	.0126	-.0014
	(n)	(2887)	(2321)		(47)	(27)		(2840)	(2294)	
WORK	1	.0025	.0044	-.0019	.526	.818	-.292	.0016	.0023	-.0006
	(n)	(3540)	(1672)		(19)	(11)		(3540)	(1672)	
	2	.0021	.0020	+.0000	.692	.900	-.208	.0014	.0013	+.0001
	(n)	(3399)	(1813)		(13)	(5)		(3399)	(1813)	
	"seam"	.0048	.0043	+.0005	.800	.800	0	.0035	.0026	+.0009
	(n)	(2888)	(2324)		(5)	(5)		(2883)	(2319)	

Part 2: Federal Programs

CSRET	1	0	.0001	-.0001	x	x	x	0	.0001	-.0001
	(n)	(5139)	(2408)		(0)	(0)		(5139)	(2408)	
	2	.0001	0	+.0001	x	x	x	.0001	0	+.0001
	(n)	(4894)	(2653)		(0)	(0)		(4894)	(2653)	
	"seam"	.0005	.0009	-.0004	x	x	x	.0005	.0009	-.0004
	(n)	(4168)	(3379)		(0)	(0)		(4168)	(3379)	
OASDI	1	.0034	.0028	+.0007	.920	.955	-.035	.0021	.0014	+.0008
	(n)	(5140)	(2410)		(25)	(11)		(5140)	(2410)	
	2	.0040	.0016	+.0024	.914	.909	+.005	.0021	.0004	+.0017
	(n)	(4895)	(2655)		(35)	(11)		(4895)	(2655)	
	"seam"	.0074	.0059	+.0015	.688	.667	+.021	.0048	.0053	-.0005
	(n)	(4168)	(3382)		(16)	(3)		(4152)	(3379)	
SSI	1	.0005	.0007	-.0002	1.000	1.000	0	.0002	.0001	+.0001
	(n)	(5139)	(2409)		(5)	(3)		(5139)	(2409)	
	2	.0009	.0005	+.0004	.889	1.000	-.111	.0003	0	+.0003
	(n)	(4894)	(2654)		(9)	(4)		(4894)	(2654)	
	"seam"	.0017	.0012	+.0005	x	1.000	x	.0017	.0009	+.0008
	(n)	(4168)	(3380)		(0)	(1)		(4168)	(3379)	
VETS	1	.0001	.0001	-.0000	1.000	x	x	0	.0001	-.0001
	(n)	(5139)	(2408)		(2)	(0)		(5139)	(2408)	
	2	0	0	0	x	x	x	0	0	0
	(n)	(4894)	(2653)		(0)	(0)		(4894)	(2653)	
	"seam"	.0012	.0006	+.0006	1.000	x	x	.0010	.0006	+.0004
	(n)	(4168)	(3379)		(1)	(0)		(4167)	(3379)	

simply eliminating the difference between self and proxy error rates would produce worthwhile improvements in SIPP estimates.

The other component of the seam bias effect is the overestimate of participation change at the interview seam. Table 4.2 shows no evidence that

the involvement of proxies in either of the seam months increases the tendency to overreport change. None of the seam overreport estimates is significant, and the direction of the differences is evenly split, with half indicating more overreport errors for self-respondents and half indicating more overreports for proxies.

With regard to SIPP, both the monthly participation results and these data for participation change should put to rest the notion that proxy reporting is an important cause of SIPP's response quality problems. More generally, both sets of results offer only weak support for the common sense notion that proxy reports are beset with greater response error than self-reports. Our final analyses attempt to push this notion one step farther, with an investigation of additional variables which might be expected to mediate self/proxy response error differences--recall difficulty due to a longer recall interval and reporting "load" effects.

4.3 Self/Proxy Response Error Differences as a Function of Recall Interval

Although we have not found the expected large and consistent effects in general, it may be that self/proxy response error differences would be revealed under more difficult recall conditions, such as a long recall interval as compared to a short recall interval. Table 4.3 summarizes the effects of recall interval duration on self- and proxy response errors for reports of monthly program participation. Entries in the "Self" and "Proxy" columns indicate the difference between the "4 months ago" error rate and the "last month" rate; thus, a positive difference indicates more error in the more distant month's report. The "Difference" column indicates the difference between the self-response recall interval effect and the proxy effect. The parenthetical entries in the (n) rows indicate the numbers of cases in each respondent status class for each type of error. Only respondents who had the same true participation status at both ends of the reference period are included in the underreport or overreport estimates (because otherwise the "4 months ago"-"last month" difference cannot be determined for each individual), so the sum of the directional error n's is generally less than the n in the misclassification analysis[8].

---

[8] In the directional analyses, the restriction to respondents with the same true participation status in both target months permits the calculation of a difference score for each respondent, which is necessary for statistical testing. A reviewer suggested that restricting the sample in this manner may tend to leave only the uninteresting cases for analysis--i.e., "4 months ago" recall for those whose participation is constant across all months of the wave may be no more difficult than recall for "last month." Therefore, I examined the overall mean recall interval effect for self-respondents (the "4 months ago" average minus the "last month" average), and the effect for proxies, and their difference, for each program and wave. Although the level of the observed effects in some instances changed substantially, the patterns of the new results yield no evidence that would alter the conclusions drawn from the restricted data set (data not shown).

TABLE 4.3: RESPONSE ERROR RATE TIME EFFECTS (" 4 MONTHS AGO" - "LAST MONTH") FOR MONTHLY PROGRAM PARTICIPATION BY TYPE OF ERROR, SELF/PROXY RESPONDENT STATUS, PROGRAM, AND WAVE

Part 1: State Programs

		Misclassification			Underreport			Overreport		
PROGRAM	WAVE	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference
AFDC	1	-.0017	-.0018	+.0001	0	0	0	0	0	0
	(n)	(3541)	(1671)		(71)	(12)		(3452)	(1649)	
	2	.0003	-.0011	+.0014	0	0	0	-.0009	0	-.0009
	(n)	(3400)	(1812)		(75)	(10)		(3310)	(1794)	
FOOD	1	.0020	.0042	-.0022	.022	.087	-.065	.0006	.0012	-.0006
	(n)	(3541)	(1671)		(136)	(23)		(3360)	(1637)	
	2	.0021	.0022	-.0001	-.008	.053	-.061	-.0012	-.0006	-.0006
	(n)	(3400)	(1812)		(127)	(19)		(3230)	(1777)	
UNEM	1	.0045	.0096	-.0051	-.125	.000	-.125	.0009	.0006	+.0003
	(n)	(3539)	(1669)		(56)	(16)		(3405)	(1610)	
	2	.0035	.0077	-.0042	-.063	.000	-.063	.0012	.0000	+.0012
	(n)	(3397)	(1811)		(48)	(20)		(3260)	(1745)	
WORK	1	.0008	.0000	+.0008	.077	0	+.077	.0000	-.0012	+.0012
	(n)	(3540)	(1672)		(13)	(4)		(3511)	(1659)	
	2	.0015	.0006	+.0009	.067	.143	-.076	-.0003	-.0006	+.0003
	(n)	(3399)	(1813)		(15)	(7)		(3375)	(1803)	

Part 2: Federal Programs

CSRET	1	0	.0004	-.0004	0	0	0	0	.0004	-.0004
	(n)	(5139)	(2408)		(63)	(6)		(5076)	(2402)	
	2	.0004	0	+.0004	.032	0	+.032	0	0	0
	(n)	(4894)	(2653)		(62)	(7)		(4832)	(2646)	
OASDI	1	-.0002	-.0017	+.0015	.010	.000	+.010	-.0010	.0000	-.0010
	(n)	(5140)	(2410)		(1205)	(231)		(3910)	(2173)	
	2	-.0006	-.0011	+.0005	.006	.000	+.006	-.0003	0	-.0003
	(n)	(4895)	(2655)		(1151)	(306)		(3709)	(2342)	
SSI	1	.0000	-.0004	+.0004	.012	0	+.012	0	.0004	-.0004
	(n)	(5139)	(2409)		(85)	(26)		(5049)	(2381)	
	2	-.0010	-.0008	-.0002	.012	0	+.012	0	0	0
	(n)	(4894)	(2654)		(82)	(26)		(4803)	(2624)	
VETS	1	-.0004	.0004	-.0008	0	.024	-.024	0	0	0
	(n)	(5139)	(2408)		(105)	(42)		(5032)	(2366)	
	2	0	0	0	0	0	0	0	0	0
	(n)	(4894)	(2653)		(105)	(45)		(4789)	(2608)	

The recall interval hypothesis predicts that the direction of the recall interval effect will be generally positive, indicating greater error with a longer interval, and, more importantly for present purposes, that the effect for proxies will be more pronounced--more positive--than the self-respondent effect, so that the self-proxy difference will be negative. The data in Table 4.3 offer neither strong nor consistent support for either component of the hypothesis. At the most immediate level, the statistical analysis of self-proxy differences certainly provides no indication that recall interval affects self- and proxy respondents differentially. None of the differences, for either the total misclassifications or either of the directional errors, even approaches the standard .05 level for statistical significance. In addition, apart from the lack of statistically significant differences, there are no clear trends in the direction of the effects.

Consider first the misclassification results. The state-administered programs show a general tendency for a "typical" recall interval main effect--for three of the four programs there is consistently more error in the "4 months ago" reports than in the "last month" reports for both self-respondents and proxies--but there is certainly no evidence of this trend among the Federal programs. And, for neither set of programs is there evidence that proxy data are more likely to be affected by recall interval. In fact, the self-proxy differences are evenly split between positive (greater effect on self-respondents) and negative (greater effect on proxies) effects.

A cursory review of the underreporting results suggests some support for the hypothesis, at least among the state-level programs. Most of the effects in the state-level "Proxy" column are positive, the "Self" effects are evenly split between the positive and the negative, and the self-proxy differences are mostly negative. In fact, however, there is little similarity to the patterns of results across the four programs, only one of which--FOOD--shows the hypothesized pattern of consistently greater and positive recall interval effects for proxies: AFDC shows absolutely no recall interval effect for either respondent group; UNEM shows a consistent reverse recall interval effect for self-respondents and no effect for proxies; and WORK shows almost perfectly inconsistent differences from one wave to the next. The most consistent pattern in the underreport results for the Federal programs suggests that recall interval affects <u>only</u> self-respondents, and proxies not at all.

The overreport trends also provide scant support for the notion that a long recall interval will cause special difficulties for proxy reporting. In the state-level programs, neither self- nor proxy respondents show a consistent recall interval effect, and the self-proxy differences are an even mix of positive and negative effects. Overreporting is mostly unaffected by recall interval in the Federal program data, regardless of respondent status.

4.4 Self/Proxy Response Error Differences as a Function of Reporting Load

Another potentially important difference between self- and proxy respondents is the additional data proxies generally must provide. SIPP procedures direct interviewers to conduct self-response interviews with each reporter before the reporter goes on to provide proxy data for other sample persons. Therefore,

not only do self-respondents usually report only for themselves, their self-reports are (almost) always their first reports. Proxies, on the other hand, always provide more than one report, and are (almost) always reporting after having already provided at least one report. This additional "reporting load" may produce greater boredom or fatigue among proxies, and thus may adversely affect the quality of their survey reports, as some researchers have suggested (e.g., Cartwright, 1957), leading especially to errors of omission. The notion here is that the incipient self-proxy underreport error differences we see in Table 4.1 may represent the effects of reporting load, and that such effects may become even more evident in the reports of especially burdened proxy reporters.

Table 4.4 offers a rough examination of this reporting load notion, operationalized here as the number of people in the household for whom the reporter potentially must report. It shows the effects of household size on self- and proxy response errors for reports of monthly program participation. Entries in the "Self" and "Proxy" columns represent the "household size" effect--the difference between the error rate for respondents in large households (three or more persons) versus those in small households (one or two persons). A positive difference indicates more error for large households than for small households. Again, the "Self-Proxy Difference" column indicates the difference between the self-response reporting load effect and the proxy effect. The parenthetical entries in the (n) rows indicate the numbers of cases in each respondent status and household size class for each type of error. Because these are averages over four months, a person will appear in both the analysis of overreports and underreports if the administrative record indicates both a "yes" and "no" participation in the program for the period (so the sum of the underreport and overreport n's may exceed the n in the misclassification analysis).

The simple notion here is that the effects of reporting for large households will be more severe (more positive) for proxies than for self-respondents, so that the self-proxy difference will be negative. The reporting load hypothesis predicts an interaction, in other words, which we test with an analysis of variance F-test on the interaction between self/proxy response status and small/large household size.

The test is admittedly rough, and the cell frequencies are often quite small--so small, in fact, that for some programs there are simply too few cases for the underreporting analysis. Nevertheless, where there are sufficient data, the results in Table 4.4 offer virtually no support for the reporting load hypothesis. None of the misclassification interactions is significant, and there is no consistent trend to the direction of the effects. For the directional errors, no program shows consistent significant effects in both waves. Of the three scattered significant interactions, two are in the "wrong" direction, indicating a greater household size effect for self-respondents than for proxies. Nor is there a trend to the direction of the differences, which are about as likely to suggest that household size affects self-respondents more than proxies as they are the reverse. Small n's notwithstanding, there is certainly no evidence here that the reports of especially burdened proxies suffer additional data quality problems.

TABLE 4.4: RESPONSE ERROR RATE REPORTING LOAD EFFECTS (LARGE (3+ PERSON) HOUSEHOLD - SMALL (1 OR 2 PERSON) HOUSEHOLD) FOR MONTHLY PROGRAM PARTICIPATION BY TYPE OF ERROR, SELF/PROXY RESPONDENT STATUS, PROGRAM, AND WAVE

Part 1: State Programs

		Misclassification			Underreport			Overreport		
PROGRAM	WAVE	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference	Self	Proxy	Self-Proxy Difference
AFDC	1	.0167	.0051	+.0116	[insufficient data]			.0025	.0023	+.0002
(large hh n)		(1727)	(1295)					(1668)	(1285)	
(small hh n)		(1814)	(376)					(1803)	(374)	
	2	.0152	.0067	+.0085	[insufficient data]			.0027	.0022	+.0005
(large hh n)		(1673)	(1390)					(1609)	(1380)	
(small hh n)		(1727)	(422)					(1716)	(422)	
FOOD	1	.0114	-.0009	+.0123	.030	-.037	+.067	.0051	.0007	+.0044
(large hh n)		(1727)	(1295)		(111)	(27)		(1647)	(1279)	
(small hh n)		(1814)	(376)		(73)	(9)		(1760)	(369)	
	2	.0084	.0017	+.0067	-.021	.219	-.240	.0070	-.0027	+.0097
(large hh n)		(1673)	(1390)		(108)	(24)		(1597)	(1377)	
(small hh n)		(1727)	(422)		(65)	(11)		(1680)	(416)	
UNEM	1	.0036	.0072	-.0036	.067	-.163	+.230	.0005	.0075	-.0070
(large hh n)		(1725)	(1294)		(81)	(57)		(1692)	(1285)	
(small hh n)		(1814)	(375)		(70)	(11)		(1798)	(372)	
	2	.0011	.0078	-.0067	-.093	-.130	+.037	-.0015	.0106	-.0121
(large hh n)		(1670)	(1389)		(90)	(61)		(1645)	(1377)	
(small hh n)		(1727)	(422)		(63)	(13)		(1713)	(418)	
WORK	1	-.0019	.0005	-.0024	[insufficient data]			-.0021	.0038	-.0059
(large hh n)		(1726)	(1296)					(1720)	(1293)	
(small hh n)		(1814)	(376)					(1807)	(375)	
	2	-.0007	.0045	-.0052	[insufficient data]			-.0017	.0010	-.0027
(large hh n)		(1672)	(1391)					(1664)	(1386)	
(small hh n)		(1727)	(422)					(1720)	(420)	

Part 2: Federal Programs

CSRET	1	-.0038	-.0010	-.003	[insufficient data]			-.0008	.0009	-.0017
(large hh n)		(2518)	(1878)					(2507)	(1875)	
(small hh n)		(2621)	(530)					(2569)	(527)	
	2	-.0029	-.0028	-.0001	[insufficient data]			-.0008	.0005	-.0013
(large hh n)		(2417)	(2047)					(2406)	(2044)	
(small hh n)		(2477)	(606)					(2426)	(602)	
OASDI	1	-.0193	-.0168	-.0025	.067	.110	-.043	-.0259	-.0165	-.0094
(large hh n)		(2519)	(1880)		(225)	(95)		(2301)	(1788)	
(small hh n)		(2621)	(530)		(1005)	(143)		(1634)	(391)	
	2	-.0138	-.0187	+.0049	.105	.121	-.016	-.0273	-.0284	+.0011
(large hh n)		(2418)	(2049)		(217)	(121)		(2213)	(1932)	
(small hh n)		(2477)	(606)		(969)	(192)		(1531)	(417)	
SSI	1	-.0026	-.0074	+.0048	.258	-.157	+.415	-.0001	-.0018	+.0017
(large hh n)		(2518)	(1878)		(17)	(17)		(2504)	(1863)	
(small hh n)		(2621)	(531)		(73)	(12)		(2550)	(520)	
	2	-.0031	-.0042	+.0011	.192	.082	+.110	-.0002	-.0036	+.0034
(large hh n)		(2417)	(2047)		(19)	(17)		(2401)	(2032)	
(small hh n)		(2477)	(607)		(72)	(13)		(2411)	(596)	
VETS	1	-.0042	-.0043	+.0001	.068	-.021	+.089	-.0023	-.0028	+.0005
(large hh n)		(2518)	(1878)		(31)	(30)		(2488)	(1848)	
(small hh n)		(2621)	(530)		(76)	(12)		(2546)	(518)	
	2	-.0048	-.0012	-.0036	.023	.115	-.092	-.0021	-.0031	+.0010
(large hh n)		(2417)	(2047)		(28)	(34)		(2389)	(2013)	
(small hh n)		(2477)	(606)		(77)	(11)		(2400)	(595)	

5. SUMMARY AND CONCLUSIONS

Contrary to common assumptions, but in keeping with the results of some recent investigations, the results of this research offer only weak and inconsistent support for the notion that survey proxy reports about government transfer

program participation are generally more error-prone than the reports of self-respondents. We find this to be true for two basic estimates using data from SIPP--monthly reports of participation and month-to-month participation change. Even including in our analyses variables which might reasonably be expected to mediate "proxy bias"--to identify extreme conditions under which the presumed response error differences would surely be brought to light--failed to produce the expected effects.

Why do we not find important and consistent self/proxy response error differences, even when we stack the deck against proxies? Why is it at all difficult to confirm what appears to be such an obvious hypothesis that it hardly needs testing--that survey data about person A gathered from person A are less likely to be in error than survey data about A gathered from person B?

We have no sure explanations, but offer the following line of thought for consideration. The problem, perhaps, is too narrow a focus on knowledge retrieval as the cognitive basis for survey response. It may be reasonable to assume that proxies, in general, are less likely to have information available in memory about the person for whom they are to respond than the person him- or herself, and that the information they do have may be less detailed, less organized, less integrated with other information in memory, and less often accessed or practiced[9]. However, to the extent that respondents do not respond to survey items by retrieving information from memory, any such differences are unimportant as determinants of response error.

Some recently conducted, small-scale, qualitative research investigating the cognitive aspects of SIPP response behavior (Marquis, 1990) suggests that this may be the case with SIPP--that SIPP interview procedures often do not succeed in tapping into respondents' knowledge in any deep and systematic way. What seems to happen instead is that respondents reject an active, effortful search of memory for specific knowledge (which is, on the surface, the SIPP task) in favor of an easier approach which relies on only superficial knowledge fragments, highly fallible rules, and perhaps some minor adjustments.

Thus, any differences in knowledge between self- and proxy respondents are not important, because not much real knowledge is used. Neither self- nor proxy respondents are really retrieving information from memory, they are estimating. This response strategy produces high levels of response error--particularly for underreports--for both self-respondents and proxies, and thereby minimizes response error differences between them.

---

[9] We also suspect, however, that there are many situations of interest to survey researchers in which a proxy's knowledge may certainly be <u>different</u> than that of the person being responded for, without necessarily being reduced in richness of detail, integration with other memories, etc. For example, while memories of another's spells of illness or incapacity may lack details about physical symptoms, activity limitations, etc., there may be many other effects with a <u>greater</u> impact on the proxy--increased home and family maintenance burdens, greater financial responsibilities, the need to make and monitor medical care arrangements, etc.

Although the data we have are directly relevant only to SIPP, perhaps this knowledge-retrieval-avoidance process applies more often than is realized to other retrospective surveys as well, thus accounting for the general difficulty of finding convincing and consistent evidence of self/proxy response error effects. If so, then the task not only for SIPP, but for the survey research enterprise much more widely, is daunting. We must find new ways of approaching the interview situation which head off error-prone shortcuts and engage respondents--and interviewers--in the mutual task of producing accurate survey responses.

REFERENCES

BERK, M., C. HORGAN, and S. MEYERS (1982), "The Reporting of Stigmatizing Health Conditions: A Comparison of Proxy and Self-Reporting." Proceedings of the Section on Social Statistics, American Statistical Association, pp 506-509.

BOWIE, C. (1989), Self/proxy response figures for the 1984, 1985, and 1986 SIPP Panels, personal communication.

BURKHEAD, D. (1985a), "SIPP Waves 1 and 2: Month-to-Month Changes of Work Status and Income Recipiency Status." U.S. Bureau of the Census memorandum to J. Coder, February 21, 1985.

------ (1985b), "SIPP Waves 1, 2, and 3 (partial): Month-to-Month Changes of Work Status and Income Recipiency Status." U.S. Bureau of the Census memorandum to J. Coder, March 12, 1985.

------ (1985c), "SIPP Waves 1-3: Changes in Occupation and Industry Codes Between Waves." U.S. Bureau of the Census memorandum to J. Coder, May 7, 1985.

BURKHEAD, D. and J. CODER (1985), "Gross Changes in Income Recipiency from the Survey of Income and Program Participation." Proceedings of the Social Statistics Section, American Statistical Association, pp 351-356.

CARTWRIGHT, A. (1957), "The Effect of Obtaining Information from Different Informants on a Family Morbidity Inquiry." Applied Statistics, Vol. 6, pp 18-25.

CODER, J. (1986), "Monthly Transitions from the SIPP Longitudinal Research File." U.S. Bureau of the Census memorandum to P. Schneider, May 20, 1986.

CZAJKA, J. (1983), "Subannual Income Estimation." Paper 7 in M. David (ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.

ENTERLINE, P. and K. CAPT (1959), "A Validation of Information Provided by Household Respondents in Health Surveys." American Journal of Public Health, Vol. 49, pp 205-212.

FELLEGI, I. and A. SUNTER (1969), "A Theory for Record Linkage." Journal of the American Statistical Association, Vol. 64, pp 1183-1210.

HAASE, K. and R. WILSON (1972), "The Study Design of an Experiment to Measure the Effects of Using Proxy Respondents in the National Health Interview Survey." Proceedings of the Social Statistics Section, American Statistical Association, pp 289-293.

HILL, D. (1987), "Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp 210-215.

HORVITZ, D. (1952), "Sampling and Field Procedures of the Pittsburgh Morbidity Survey." Public Health Reports, Vol. 67, pp 1003-1012.

JARO, M. (1989), "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida." Journal of the American Statistical Association, Vol. 84, pp 414-420.

JUDKINS, D. (1986), "Problems with Gross Flows from SIPP." Paper presented at the U.S. Bureau of the Census' SIPP Supervisors' Conference, September, 1986.

KASPRZYK, D. and R. HERRIOT (1986), "Some Aspects of the Survey of Income and Program Participation." U.S. Bureau of the Census, SIPP Working Paper Series, No. 8601.

KILSS, B. and W. ALVEY (1976), "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972." Proceedings of the Social Statistics Section, American Statistical Association, pp 471-476.

LAPLANT, W. (1989), "Users' Manual for the Generalized Record Linkage Program Generator." Statistical Research Division, U.S. Bureau of the Census.

MARQUIS, K. (1978), "Inferring Health Interview Response Bias from Imperfect Record Checks." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp 265-270.

MARQUIS, K. (1990), "Report of the SIPP Cognitive Interviewing Project." Unpublished report, U.S. Bureau of the Census, August, 1990.

MARQUIS, K. and J. MOORE (1989), "Response Errors in the Survey of Income and Program Participation: Preliminary Results." Paper presented at the U.S. Bureau of the Census' Fifth Annual Research Conference, March 19-22, 1989, pp 515-536.

MARQUIS, K. and J. MOORE (1990), "Measurement Errors in SIPP Program Reports." Paper presented at the U.S. Bureau of the Census' Sixth Annual Research Conference, March 18-21, 1990, forthcoming.

MATHIOWETZ, N. and R. GROVES (1985), "The Effects of Respondent Rules on Health Survey Reports." American Journal of Public Health, Vol. 75, pp 639-644.

MOORE, J. (1988), "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." Journal of Official Statistics, Vol. 4, pp 155-172.

MOORE, J. and D. KASPRZYK (1984), "Month-to-Month Recipiency Turnover in the ISDP." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp 210-215.

MOORE, J. and K. MARQUIS (1989), "Using Administrative Record Data to Evaluate the Quality of Survey Estimates." Survey Methodology, Vol. 15, pp 129-143.

NELSON, D., D. McMILLEN, and D. KASPRZYK (1985), "An Overview of the Survey of Income and Program Participation." SIPP Working Paper Series, No. 8401, U.S. Bureau of the Census.

RYSCAVAGE, P. and A. FELDMAN-HARKINS (1986), "Work Experience Data from the SIPP." Paper presented at the meetings of the Allied Social Science Association, New Orleans, LA, December, 1986.

SINGH, R., L. WEIDMAN, and G. SHAPIRO (1988), "Quality of the SIPP Estimates." Paper presented at the SIPP Conference on Individuals and Families in Transition: Understanding Change through Longitudinal Data, Annapolis, MD, March 16-18, 1988.

SUDMAN, S. and R. ANDERSEN (1977), "Health Survey Research Instruments." Advances in Health Survey Research Methods, National Center for Health Services Research, pp 7-12.

SUDMAN, S. and N. BRADBURN (1974), Response Effects in Surveys. Chicago, IL: Aldine.

WEIDMAN, L. (1986), "Investigations of Gross Changes in Income Recipiency from the Survey of Income and Program Participation." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp 231-236.