THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

SIPP RECORD CHECK RESULTS: IMPLICATIONS FOR MEASUREMENT PRINCIPLES AND PRACTICE

No. 126

K. H. Marquis and J. C. Moore
Bureau of the Census

U. S. Department of Commerce BUREAU OF THE CENSUS

A short version of this paper, coauthored with Vicki J. Huggins, was presented at the annual meetings of the American Statistical Association, Anaheim, CA., August 1990, and submitted for publication in the **Proceedings of the Section on Survey Research Methods.** The paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

TABLE OF CONTENTS

SIPP RECORD CHECK RESULTS: IMPLICATIONS FOR MEASUREMENT PRINCIPLES AND PRACTICE

Kent H. Marquis and Jeffrey C. Moore
U. S. Census Bureau

1. INTRODUCTION

Measurement errors in household surveys are inevitable. Yet they are not always well understood and, if one wanted to do something about them, it is not always clear what to do. Our goal in this paper is to contribute to the understanding of response errors in SIPP and begin a consideration of what to do about them.

We will report the results of a record check study for the Survey of Income and Program Participation. The study covers 8 months of reported participation in 8 programs for the SIPP sample in 4 states.

Our results indicate that while misclassification errors in reports of participation level and change are rare, they sometimes result in important amounts of net response bias and in important amounts of attenuation of estimates of relationships.

We show that the effect of the response errors on measures of month-to-month change in program participation depends on which months the change is measured between. The response errors negatively bias change measures between months reported in the same interview and positively bias change measures between months reported in different interviews (which we refer to as measured on "the seam" between interviews). Correlations involving the change measures are always attenuated in the presence of response error, regardless of whether change is measured on or off the seam.

We mention some statistical strategies for learning about the response errors and then introducing corrections. These strategies include edits, raking ratio estimation, record checks, correcting the variance-covariance matrix and instrumental variables techniques. In each case, the survey design is modified to obtain additional information about the measurement errors and the additional information is used to introduce corrections either into the individual data records or to the statistical estimates.

We discuss a collection of behavioral strategies that could prevent many response errors from occurring in the first place. These include shorter reference periods, more memory cueing, dependent interviewing, landmark events, respondent rules, randomized response, using informants, and providing confidentiality reassurances. We also mention a new strategy that we call Task-Focused Interviewing that seems promising in light of what we are learning about the measurement errors.

We conclude by recommending an expanded use of administrative records by SIPP to further understand its error structure, as an aid in experimental research about the causes of errors, as a tool to evaluate statistical and behavioral strategies for controlling measurement errors, and as a key component of a continuous program of total quality management.

2. BACKGROUND AND METHODS

2.1 SIPP

SIPP is a longitudinal panel survey designed to provide improved information about the economic situation of people and families in the United States. For each person fifteen years of age or older, SIPP collects monthly information about earnings, participation in government transfer programs, assets and liabilities, labor force participation, and related topics, for the four months preceding the interview month. Generally, a panel consists of eight such interviews, covering about 2 1/2 years. Proxy reporting is permitted for household members not available for interview at the time of the visit. For a detailed description of the SIPP program, see Nelson, McMillen, and Kasprzyk (1985).

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 follow-up questions to elicit reports of the monthly dollar amounts received from them. In the second and subsequent interviews of the panel, respondents are reminded of which programs were reported for the sample person in the previous interview. For various reasons, final SIPP participation information may differ from the original information (e.g., due to computer editing). We use the final values in the analyses reported in this paper.

2.2 The SIPP Record Check Study Design

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. Below we summarize 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.

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. Full designs are necessary for producing unbiased estimates of the parameters of the response error distribution.

2.2.2 Programs

We obtained program participation records for eight government transfer programs, half administered by the states and half administered by the Federal Government. 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.

2.2.3 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

We conducted the record check study 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:

Wave	Rotation Group	Reference Period Months: 1983 Jun	Jul	Aug	Sep	Oct	Nov	Dec	1984 Jan	Feb	Mar	Apr
1	1	4	3	2	1	(I)						
	2		4	3	2	1	(I)					
	3			4	3	2	1	(I)				
	4				4	3	2	1	(I)			
2	1					4	3	2	1	(I)		
	2						4	3	2	1	(I)	
	3							4	3	2	1	(I)
	4*/								4	3	2	1

KEY: (I) = interview month
Reference Period: 4--3--2--1 = 4 months ago, 3 months ago, 2 months ago, last month.

*/ 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

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. And about 11,000 people lived in the record checked households.

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 us 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.

2.2.5 Matching

We used the 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.

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

---

[3] Our 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.). We identified such SSN refuser matches through our review of all cases in which a child had been matched to 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, we manually relinked the administrative 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).

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 recipients, we relinked any such SSI or OASDI administrative record information to an appropriate adult in the household (e.g., an 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 not completely certain. The numbers of matches for which there was some uncertainty by program are: AFDC=14, CSRET=1, FOOD=13, OASDI=109, SSI=19, UNEM=64, VETS=14, WORK=12.

2.3 Definition of Response Errors

In this paper we estimate errors in reports of program participation, a binary variable where 0 means not participating and 1 denotes participation (in the sense of receiving benefits from the program). 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 2.2.

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

Figure 2.2: Notation for Cross-Classified Reported and True Values.

The letters a, b, c, and d in the table represent frequencies of reported and true characteristics. N is the sample size. The total number of WRONG ANSWERS (or misclassification errors) for a program is b + c. The rate of misclassification is (b + c) / N and the misclassification percent (or percent wrong) is [(b + c) / N] x 100.

The frequency of UNDERREPORT errors is c. The underreporting error rate, which is conditional on a true positive, is c / (a + c), and the percent of underreporting errors is 100 times the rate.

Similarly, the frequency of OVERREPORT errors is b, the rate is b / (b + d), and the percent is 100 times the rate.

We will use the percent wrong in Section 3 (descriptive results) and reserve the underreport and overreport statistics for Section 4 (hypothesis testing results).

2.4 Descriptive and Inferential Statistics

For each program, we usually calculate descriptive statistics (e.g., percent wrong) for each month and report an average over the entire eight months (or other groups of time periods such as wave 1 and wave 2). Unless we say otherwise, the inferential statistics refer to these averages. For the hypothesis tests and other "within person" comparisons, most inferences are based on paired-comparison t-tests that take into account the correlation of the observations for each person over time. We reject the null hypothesis for $p \leq .05$. We discuss other inferential procedures as they are used. For all of our inferential statistics we assume simple random sampling although the SIPP sample design is more complex than this. As a result, our population variance estimates and corresponding p-values are likely to be slightly underestimated for the individual monthly or program-specific analyses. However, we feel that our stated conclusions, based on consistent patterns across programs and time periods, would not change if we were to take the complex sample design into account in our variance estimates.

2.5 Effects of Response Errors

We call the effect of response errors on a parameter estimate a bias. The bias is the difference between the parameter estimated with data containing response errors and the true parameter value. We will examine bias in two kinds of parameter estimates, a mean and a correlation. The bias in the estimated mean, $(a + b) / N$, is $\{[(a + b) / N] - [(a + c) / N]\}$ or $(b - c) / N$. Dividing by $(a + c) / N$ yields the percent bias.

We term the effect of the measurement errors on the bivariate correlation the correlation bias. Our correlation bias estimate is model-based. In the appendix we show the derivation and discuss the sensitivity of the results to the model. For our illustrations, in the text the correlation is between the reported participation status (containing response errors) and a variable that is assumed to be measured without error. In an earlier paper (Marquis and Moore, 1989b) we also derived the expressions for the bias in two forms of the bivariate regression coefficient estimate. However, the correlation result is a reasonably good summary of the two regression results.

3. DESCRIPTIVE RESULTS

In this section we will look at the response error percents for measures of program participation level and change. And we will examine the effects of those errors on statistics that analysts estimate using SIPP data. While the percentage of responses in error is always very small, the errors can have large effects on estimates.

3.1 Misclassification Rates

The misclassification error percentages for monthly reports of program participation or, more simply, the percentages of wrong answers, are very low for each of the eight SIPP programs in the record check study. In Figure 3.1 we average over the eight months of data to look at the percent wrong in reporting participation level. We observe that the error rates range from 0.2 percent (for CSRET) to 2.3 percent (for OASDI). Thus, response errors are rare regardless of which program is involved.

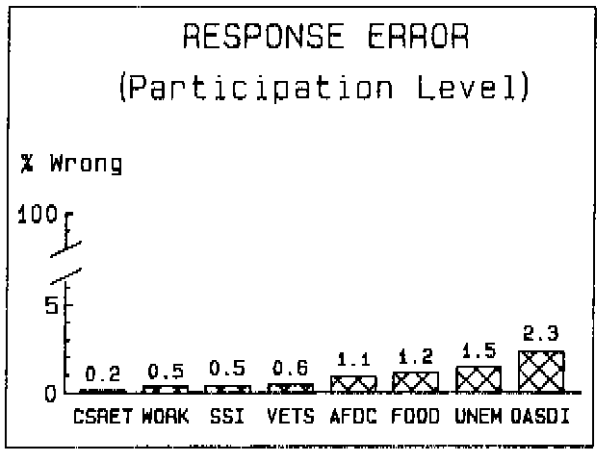

Figure 3.1: Average response error percentages for program participation are very low.

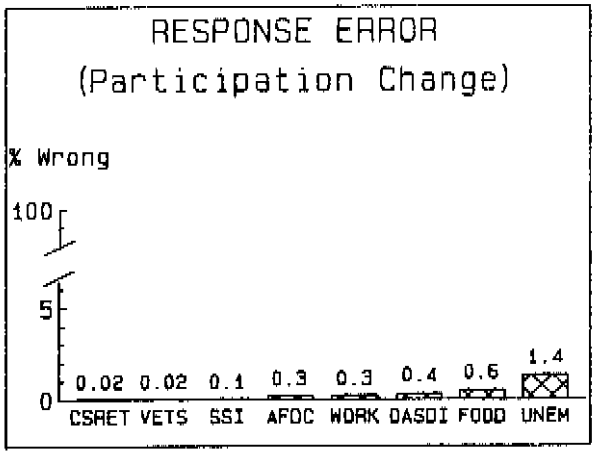

Figure 3.2: Average response error percentages for participation change are also very low.

Next, let us look at the percent response error in measures of program participation change. For any two adjacent months, we say a change has occurred when the program participation status is different (yes in one month and no in the other month, ignoring the direction of change). If the participation status is the same (either both yes or both no), then we say that no change has occurred. In Figure 3.2 we have averaged the percent wrong in change measures over the seven possible pairs of adjacent months and we see even lower error rates. They range from .02 (two-hundredths) percent for CSRET to 1.4 percent for UNEM. So errors in measures of starting or stopping the receipt of program benefits are also rare.

Put another way, almost all respondents report participation status in each of the tested programs accurately almost all of the time.

3.2 Effects of Response Errors on Estimates

Now we ask whether these low response error percents make much difference in the statistical estimates that subject matter analysts might make from SIPP data. If the effects are small, then response error reduction should not be a major concern in the SIPP program. On the other hand, if the effects are large, then it is important to bring the errors under control as quickly and completely as possible.

We will look at the biases induced by response errors in two kinds of estimates: the mean and the correlation. The mean estimate could be something like the proportion of the sample enrolled in the Food Stamps program in the month of June. The correlation estimate could be between education level and participation in the Food Stamp program last month. For both the means and the correlations, we made separate estimates of bias for each of the eight (or seven) time periods and report the average of the monthly biases here.

3.2.1 Effects of Errors on Mean Estimates

Figure 3.3 shows the bias in estimates of the mean level of program participation. The net bias is usually negative for this sample, indicating that the estimated mean is usually lower than the true mean when using the SIPP data containing response errors. Biases for some programs, such as VETS and OASDI are small, only minus three percent and plus one percent. But for other programs, such as the 18 percent underestimate for WORK and the 39 percent underestimate for AFDC, the biases are more serious.

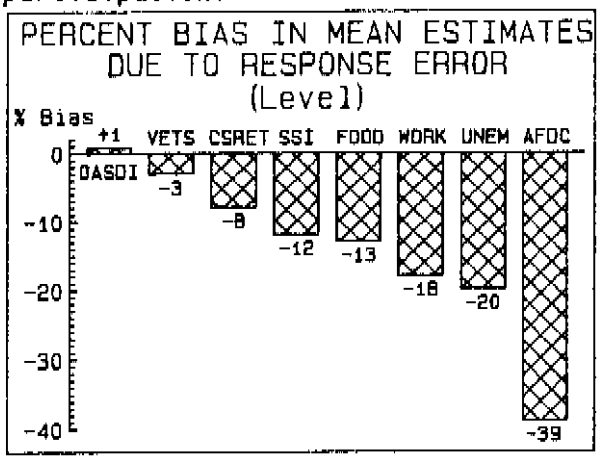

Figure 3.3: Response errors usually bias estimates of program participation levels in a negative direction.

Turning next to the biases in estimates of mean change rates for program participation, we first introduce the concept of the "seam" between interviews, since prior research suggests that the biases may be affected by this timing indicator. Recall that a change refers to whether program participation is the same or different in any two adjacent months. If the two adjacent months are reported in two different interviews, we refer to that time period as "on the seam" between the two interviews, and a change in this period is called an on-seam change. Change measured in any other pair of adjacent months is an off-seam change. This is illustrated in Figure 3.4.

Previous research (Moore and Kasprzyk, 1984; Burkhead and Coder, 1985; Hill, 1987) indicates that many more changes are measured on the seam compared to off the seam. This is also true for this sample as we show in Figure 3.5. Take, for example, the data for the Food Stamps program (FOOD): even though we would expect the rates to be the same, respondents reported change at the rate of 77 per 10,000 people on the seam and at the much lower rate of 32 per

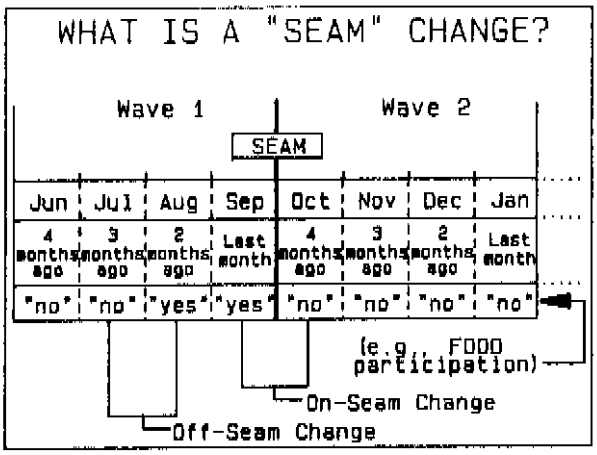

Figure 3.4: A program participation change in "on seam" when it occurs across months covered by different interview; "off seam" changes occur across months within the same interview.

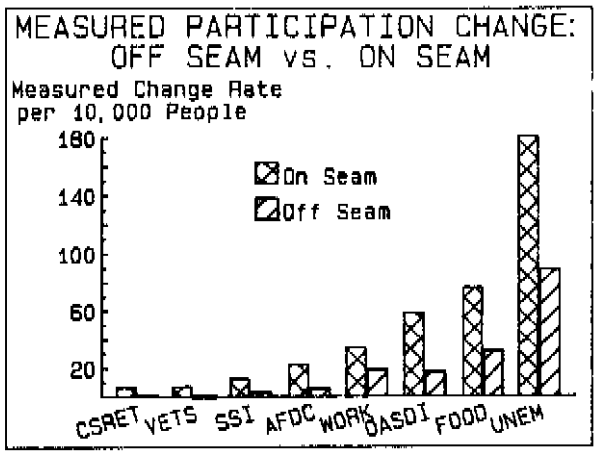

Figure 3.5: Much more change is measured "on seam" than "off seam."

10,000 in the average pair of off-seam months. This pattern is repeated for each of the other programs also. All of the on-off seam differences are statistically significant, assuming simple random sampling. So we turn, now, to the record check data to determine which of these estimates is correct, the on or off seam estimate. The results are surprising since neither estimate is generally correct.

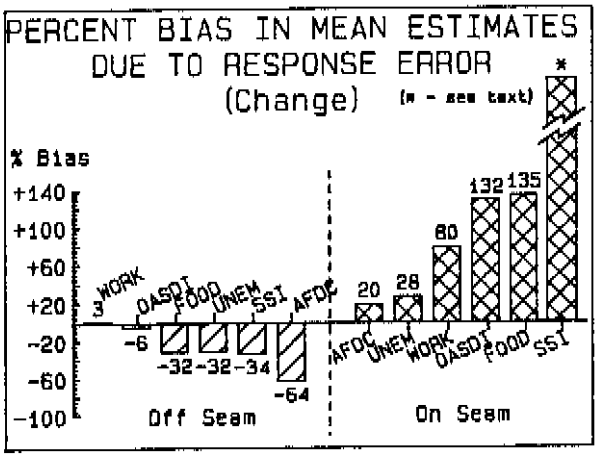

Figure 3.6: The sign of the change bias due to response error depends on whether change is measured on or off the interview seam.

Looking, in Figure 3.6, at the effects on mean change rate on and off the seam, we see that almost all of the off-seam biases are negative and all of the on-seam biases are positive. Thus, too few program participation changes are measured for the off-seam months and too many inferred for the on-seam months. The size of the on-seam bias estimate for SSI is especially uncertain due to a true change rate that, by chance, was abnormally low for the seam time period. Imputing an expected true change rate, based on true change rates for the other month pairs, the new bias estimate would be about 200 percent instead of 900 percent as originally estimated. We have omitted estimates for the two of the eight programs because their true change rate in at least one pair of months was zero (a + c = 0 in the denominator), so the percent bias could not be determined.

(Some may wonder whether the total number of changes is over-, under- or accurately estimated over the two waves. The results--not shown--do not

follow a simple pattern. Total change is overestimated in some instances, underestimated in other instances and, in still other cases, the estimated total comes close to the true total.)

Next, we will look at the effects of the response errors on correlations. These results show very different patterns.

3.2.2 Effects of Errors on Correlation Estimates.

As shown in Figure 3.7, the effect of response error is to attenuate (reduce the size of) the bivariate correlation estimate, causing it to move closer to zero than the true value in the sample. Results indicate small to moderate percentages of bias for the first five programs (20 percent or less) and moderate to large attenuation for the remaining three programs (33 to 51 percent). These effects can cause even the skilled analyst substantial trouble.

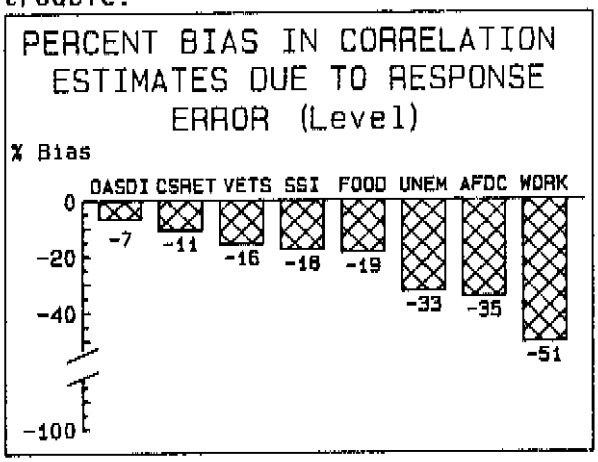

Figure 3.7: For measures of level, biases in estimated correlations due to response errors are trivial for some programs and quite serious for others.

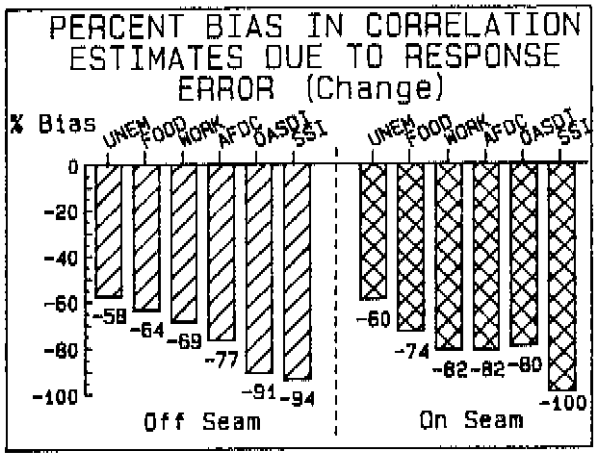

Figure 3.8: For measures of change, correlation biases are consistently very large, regardless of whether they are measured on or off the seam.

We address the bias in estimated correlations for the change measures in Figure 3.8, looking at the effects on and off the seam separately. Note first that the correlations for all programs are severely biased; the least amount of attenuation estimated is 58 percent for Unemployment Insurance (UNEM) when the measure is off the seam. The biases are more negative for the other programs, reaching -100 percent for the Supplemental Security Income (SSI) estimate on the seam. (We have omitted from the figure the two programs that had no true change in at least one pair of months.)

Recall that the on-off seam classification made a big difference in the direction of the biasing effects of error on the estimated mean. Here, however, there is no important effect of the on-off seam classification on the size or sign of the bias in correlation estimates. This is because, while the means of the on and off seam response error distributions have different signs

and sizes, the variances of both error distributions are about the same. This is true for each of the programs. And these variances (not shown) are large enough to make it very difficult to detect the true correlational relationships in the sample. This also explains why Young (1989) finds that the estimates of correlations and regressions are about the same regardless of whether the change measure is made on or off the seam. The reason is, basically, that correlations and regressions are affected mainly by the second moment or variance of the response error distribution and, in this case, the on-seam and off-seam error variances are about the same relative size.

This concludes our description of the errors in reporting program participation. We have shown that while the errors occur at very low rates, they can have very large effects on the kinds of estimates that analysts want to make from SIPP data. Because response errors have these important effects, we need to understand what is causing them and devise strategies for counteracting or removing the causes.

4. PATTERNS OF RESPONSE ERRORS

Next we will examine some of the characteristics of the error distributions, paying particular attention to how errors are distributed over time. We will concentrate on errors in reporting participation level since they appear to underlie the errors in change measures (for a further discussion of the point, see the section on "modeling the seam phenomenon" in Marquis and Moore, 1989a).

4.1 Directional Errors

Up to now we have discussed misclassification errors, which are made up of both underreports and overreports. In this section we will focus on the directional errors, looking at their relative frequencies and how they are distributed over the months of SIPP's 4-month reference period.

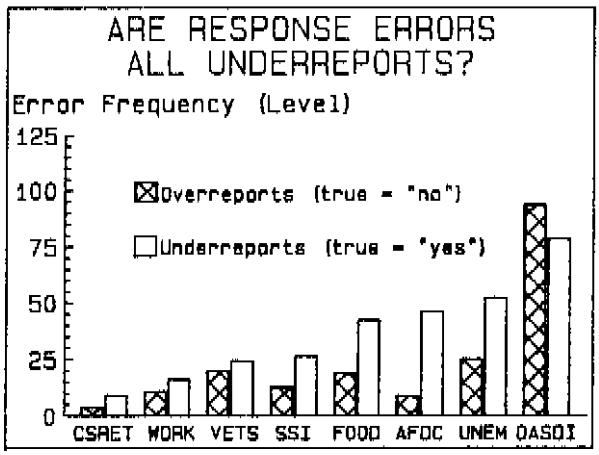

Figure 4.1: Although underreports usually predominate, all programs contain overreports as well.

Many survey designs are based on the assumption that the response errors will be almost entirely underreporting errors. But looking at the average number of monthly underreports and overreports in Figure 4.1, we see that both the overreporting and underreporting frequencies are substantially greater than zero.[4] And while there are usually more underreports than overreports, the overreport frequencies are often substantial as, for example, is the case for social security (OASDI) where there are 94 overreports in each month on the average and 79 underreports. Such relative frequencies alert us to the possibility that the traditional notions about why people make response errors may be too simple.

4.2 Underreporting Over Time

Next, let us look at how error rates are distributed over time. In Figure 4.2 we have plotted the average underreporting rates for participation four months ago and for last month. If memory decay causes the response errors, each line

[4] Based on the standard error of the frequency estimated as $[Np(1-p)]^{1/2}$ where p is the average monthly error probability and N is approximately 7550 for the federally administered programs and 5200 for the state-administered programs.

should slope downward. But most of the lines don't slope downward.[5] In only one case, UNEM, is there a meaningful and statistically significant reduction in the under-report rate for the most recent month.

Let us also point out that while the UNEM slope is consistent with the memory decay prediction, the level of error is not. The average under-reporting percent in the most recent month for the UNEM program is among the highest observed for any of the eight programs. There is nothing in the pure forgetting models that would predict this; in fact, such models generally assume that recent events are recalled with little or no error at all. The results in Figure 4.2 are contrary to this assumption, and for all programs.

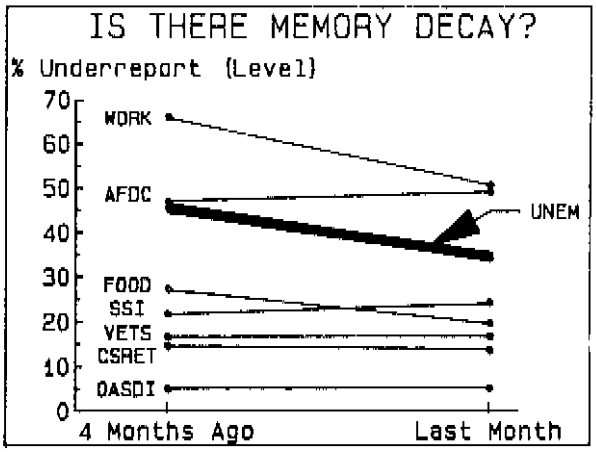

Figure 4.2: Participation under-reports for "4 months ago" versus "last month" show little evidence of memory decay.

4.3 Overreporting Over Time

Next, let us examine the distribution of overreporting errors over time. This is relevant to ideas about internal telescoping. When there is a true change in participation status at the individual level, a respondent who internally telescopes will underreport participation in one month and overreport it in a subsequent month. The implication of this pattern for the whole sample is an increasing overreporting rate (and a decreasing underreporting rate) as one moves from the more distant to the more recent months of the reference period. Since we have already seen (in the memory decay analysis, Section 4.2) that underreport rates seldom show a time effect, we look here at the overreporting percentages, again using the line chart approach.

[5] For each program, the analysis is based on all people who could have underreported (true participation = "yes") either "4 months ago" or "last month" in a wave. Significance testing is for each wave separately, taking account of the within-person correlation of observations over time where appropriate. We report the average underreport percent over waves in Figure 4.2. The t-value for the wave 2 UNEM difference is the only one exceeding 2.00. Numbers of people included in these analyses, by program and wave are: AFDC=111,108 CSRET=69,69 FOOD=215,205 OASDI=1467,1499 SSI=118,121 UNEM=193,203 VETS=149,150 and WORK=42,34.

In Figure 4.3 we compare, for each program, the average overreporting percent for 4 months ago with the percent for last month. The averages are over the two waves.[6] If the forward internal telescoping model fits the data, we should see upward sloping lines.

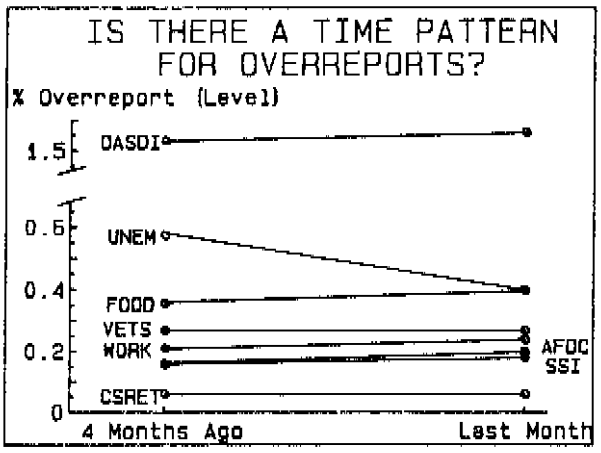

Figure 4.3: Overreports, generally, do not increase over time.

According to the results in Figure 4.3, for no program does the overreport error difference indicate a statistically significant time effect. This result may be due, in part, to low rates of true program participation change for some programs and we cannot rule out the possibility that some individuals may have made internal telescoping errors. However, judging from the temporal pattern of overreport errors, it is clear that internal telescoping is not a major cause of reporting error problems in SIPP program participation.

4.4 Additional Results

Other results mentioned in the previous research report (Marquis and Moore, 1990) include:

- In Pennsylvania, many respondents report AFDC benefits as General Assistance benefits.
- A small number of households confuse Social Security and Supplemental Security benefits.
- Some apparent errors in reporting Food Stamp recipiency are merely mistakes in reporting the correct official recipient in the household.

---

[6] For each program and each wave, we compare the overreport percents based on all people who could have overreported (true participation = "no") either "4 months ago" or "last month." Significance testing is for each wave separately, taking account of the within-person correlation of observations over time as appropriate. For no program was the within-wave difference statistically significant for either of the two waves. Numbers of people included in these analyses, by program and wave, are: AFDC=5129,5127 CSRET=7478,7478 FOOD=5053,5066 OASDI=6114,6093 SSI=7437,7440 UNEM=5136,5140 VETS=7400,7397 and WORK=5195,5190.

- Aside from the above three instances, a comprehensive search failed to reveal other instances of respondent confusion about program names or official recipient designations.

- Since average overreport rates are the same in Wave 1 as Wave 2, external telescoping is not a major determinant of the observed overreporting errors.

- Respondents may learn to deliberately underreport Workers' Compensation and Unemployment Insurance participation because Wave 2 underreporting rates are higher than Wave 1 underreporting rates for those two programs.

- People did not increase their participation in the tested transfer programs in Wave 2 compared to Wave 1, so the household's participation in SIPP did not result in a detectable behavior change.

- Interviewer effects were at the same low levels as found in most major surveys (one or two percent of total variance).

- In general, the directional error levels do not differ by self and proxy status although trends indicate more underreporting by proxy and perhaps more overreporting by self respondents.

In the next section, we begin a discussion about what to do about measurement errors in SIPP and similar surveys that are undoubtedly subject to the same kinds of measurement errors that we have described here.

5. STATISTICAL ERROR CORRECTION AND CONTROL[7]

In this section, we mention statistical strategies that could be used to correct and control measurement errors in future SIPP data. Our purpose is to encourage thinking about the need for error control and the various ways that SIPP and similar surveys might try to minimize the effects of error. These strategies are abstract, they usually require that strong assumptions be satisfied, they cost money to implement and, if implemented, still need to be evaluated themselves. Although we do not recommend adopting any particular single strategy or combination of strategies now, we do recommend that the strategies be developed for application in SIPP and the best ones used as appropriate.

The contribution of the current record check study is its demonstration of the kinds and sizes of measurement errors in reports of program participation. Because of its age and design limitations, we do not recommend using the record check study error estimates to introduce corrections into future SIPP data. The current record check study does indicate the large gains to be had in continuing to collect data that describe measurement (and other) errors and in using the additional information to implement error correction procedures. All of the correction strategies mentioned below obtain additional information that is used to infer something about the errors; then they use the information, along with assumptions, to introduce corrections.

The strategy options we mention introduce corrections either to the individual observations or to the statistical estimates derived from the observations.

5.1 Corrections to Individual Records

We mention three approaches SIPP might consider to introduce corrections for measurement errors at the individual or household record level. In all cases, SIPP has already used at least elements of each approach in its operations or research programs. We refer to the three approaches as edits, raking ratio estimation, and record checks.

5.1.1 Edits

As both ISDP and record check research shows, respondents occasionally make mistakes about the name of the program from which they are receiving benefits. Confusion has been observed between AFDC and general assistance (e.g., Klein and Vaughan, 1980), between means-tested and service-connected veterans programs (Vaughan, Lininger and Klein, 1983), and between Social Security and Supplemental Security (Vaughan, 1978). The edit approach to correcting program name confusion errors entails getting additional data via questionnaire (and/or from past and future waves concerning, for example,

---

[7] This discussion is based, in part, on material originally presented in Part 4 of Marquis, Duan, Marquis and Polich (1981). That presentation also evaluated the correction methods using a simulation methodology. Such evaluation, however, is beyond the scope of the present discussion.

personal characteristics that determine program eligibility) and using logical "edit" rules to verify or reassign participation to the correct program. Coder and Ruggles (1988), for example, have developed and evaluated a procedure to distinguish participation in AFDC from participation in the local general welfare program, and to remove cases that do not belong in either program. Among the information used is the size of the benefit, simultaneous receipt of food stamps, employment status, gender, state-specific eligibility criteria, marital status, and family composition (presence/absence of children). While post-edit distributions of the characteristics of AFDC recipients in SIPP agree generally with patterns in administrative records, the Coder-Ruggles edit reduced the number of SIPP reports of participation in AFDC--yet AFDC is a program that is generally considered to be underreported prior to any editing. The study demonstrates the potential of the editing approach to correcting response errors but it will be necessary to discover additional edits that convert false negative responses about program participation into true positive responses.

5.1.2 Raking Ratio Estimation

Huggins and Fay (1988) describe the use of Internal Revenue Service (IRS) data in connection with raking estimation procedures (e.g., Brackstone and Rao, 1976) to improve the quality of SIPP estimates subject to the effects of measurement and other errors. The technique works by adjusting the sample weights assigned to individual people to force consistency between sample estimates of marginals and corresponding population totals for cross classified variables. The procedure is analogous to the iterative proportional fitting algorithm for contingency tables, which yields maximum-likelihood estimates for hierarchical factorial log-linear models. For a sample of SIPP cases matched to IRS records, Huggins and Fay prepared "population" controls from the IRS data, implemented the estimation for selected SIPP characteristics, and analyzed the effects of the reweighting (which were quite favorable for person-level income). Their paper makes suggestions for further research using the procedure.

5.1.3 Administrative Record Checks

As this study and others have shown, it is possible to match records from SIPP to appropriate administrative records. So, it is possible, in theory, to insert data of higher measurement quality into the individual questionnaire records. Complete administrative records do not become available to the Census Bureau for some time after they are collected, however, and it takes a very long time to implement all the procedures necessary to assure accurate matching. Administrative records of high quality do not exist for all characteristics of interest to SIPP, and we do not yet have experience obtaining and using records from most states and many other federal programs.

If it were possible to implement a comprehensive record check for a <u>sample</u> of the survey cases in a timely fashion, Marquis et al. (1981) discuss several ways of using such data to correct statistical estimates (other than adjusting individual records). We turn to this question of adjusting statistical estimates next.

5.2 Corrections to Statistical Estimates

Earlier we showed that measurement error will produce biased estimates of association such as a correlation. We mention two general ways to introduce corrections into estimates of relationships based on additional information about the measurement errors: variance-covariance matrix correction and instrumental variables. Both techniques appear widely in the literature.

5.2.1 Variance-Covariance Matrix Correction

Perhaps the most widely known and used procedure is to use a reinterview to learn about the variance of the measurement error distribution and introduce this information into a variance-covariance matrix before making a relationship estimate. Fuller and Hidiroglou (1978) present the general theory that has been implemented in SUPER CARP (Hidiroglou, Fuller and Hickman, 1980), computer software for estimation using survey data that contain measurement errors. Fuller (1987) further discusses the theory and an application for the case of labor force status classification (also generally applicable to SIPP program participation).[8]

The model-based procedure assumes the form of the relationship between the observed (or measured) variances on the one hand and the true and error variances on the other. Correcting regression estimates involves additional assumptions about the correlation of measurement errors and true values of the predictor variables. Although most applications of the variance-covariance adjustment approach use a reinterview to obtain estimates of the measurement error variances, Marquis et al. (1981) discuss using other approaches, such as redundant questioning within a single interview (internal consistency) or record checks on a sample of the survey observations. In a panel study such as SIPP, one can also consider overlapping the reference periods of two adjacent interviews to obtain more than one measure of participation in a single time period.

5.2.2 Instrumental Variables

The instrumental variable correction strategy (e.g., Johnson, 1963; Fuller, 1987, Marquis et al. 1981) devotes additional measurement resources to measuring, constructing and using another variable which is assumed to be correlated with the variable of interest but uncorrelated with the measurement error in the variable of interest. The instrumental variable is used in a system of regression equations to produce asymptotically unbiased estimates of the regression parameters of interest, subject, of course, to the validity of the assumptions.

In practice, one cannot use the instrumental variables strategy with dichotomous variables since it is not possible to meet both critical assumptions simultaneously. Nevertheless, the strategy might be useful in

---

[8] For additional ideas about applications to the labor force status classification issue, see Abowd and Zellner (1985), Fuller and Chua (1985) Porterba and Summers (1985) and Lemaitre (1988).

SIPP for analyses involving continuous variables, such as total income or dollar amounts of monthly program benefits.

To recap: Our purpose has been, first, to mention the kinds of approaches one may take in learning about survey errors; and second, incorporating this additional information into statistical correction strategies to improve the data. At this point, these approaches are abstract and clearly dependent on the validity of the underlying assumptions. We hope that SIPP and similar surveys will pursue these correction possibilities in a research program, perhaps using new administrative record data as a source of validating information. The design principle which is new to surveys is to add measurements to production data collection to learn about errors, and then to use the information to produce corrected data and/or estimates. In the long run, rather than correcting errors, however, we desire to learn why respondents make errors and then to change the survey procedures so that the errors don't occur in the first place. In the next section we discuss behavioral strategies for reducing the occurrence of response errors.

6. DESIGN CHANGES TO CONTROL MEASUREMENT ERRORS

A response is a behavior. So response errors are simply one form of behavior. As such, response errors are potentially caused by any of the conditions that normally determine behavior. The survey design problem is to learn which survey conditions are producing the errors so that the conditions can be changed. This section discusses behavioral strategies for minimizing errors rather than statistical strategies. In parallel with the previous section, our goals here are to mention some of the important design remedies for response errors, and to urge further consideration of them--in the form of research and implementation of appropriate strategies. We start by considering design changes aimed at overcoming cognitive difficulties. Then we consider procedures to overcome motivational problems. We conclude by recommending a new line of procedural research designed to teach respondents the problem-solving skills necessary to meet SIPP reporting requirements, and to present respondents with only information and tasks that are consistent with the reporting requirements.

6.1 Strategies to Overcome Cognitive Difficulties

The goal of the first two design features mentioned below is to minimize the effects of poor memory on recall. In discussing them, we also consider what the record check results suggest about the kinds of memory problems that SIPP respondents are experiencing. The next two design approaches we discuss address the misreporting of the time of an event, such as the month of participation in a government transfer program. The final cognitive issue we discuss has to do with how knowledgeable respondents are about their and others' program participation--the self-proxy issue.

6.1.1 Shorter Reference Periods

Normally, surveys seek to use the shortest recall periods possible to minimize the effects of memory decay. The memory assumptions are that recall is very good for recent time periods and increasingly subject to forgetting error as the length of the recall interval increases. The record check results, however, suggest that these assumptions do not apply to SIPP participation reporting, since recall for recent time periods is far from perfect and, in general, does not get worse when the recall interval gets longer. While we feel that SIPP and similar surveys should continue to collect administrative record data to verify the absence of widespread memory decay effects, we do not feel that shortening the reference period is worth substantial further evaluation at this time. People do not correctly report participation status in many instances, but the record check results suggest that the cause is not short term memory problems.

6.1.2 More Memory Cues

Memory retrieval can often be improved by furnishing additional cues to help the search process. For example, it is well known that a recognition task ("Have you seen this before?") is much more likely to result in successful retrieval than unaided recall ("What did you see?"). Survey designers provide additional recall cues by asking specific recognition or short answer

questions rather than asking one global item (e.g., "Tell me all the government transfer programs that members of this household receive benefits from"). When adding memory cues to a questionnaire, one assumes that they will not also increase false positive (overreport) errors.[9]

Marquis, Marshall, and Oskamp (1972), however, have shown that this assumption does not necessarily hold for "low salience" items,[10] which may explain why introducing survey checklists increases the reporting of items of secondary importance. The record check study tells us only that respondents do make overreporting errors--i.e., that some "yes" reports are incorrect. It gives us no information about what would happen to underreporting and overreporting errors if additional cues were introduced into the questionnaire to increase the number of "yes" reports. Experimental research with good validation information will be needed to address this question, perhaps using the signal detection decision model as proposed by Salter and Swets (1984).[11]

The next two design approaches address the misreporting of the time of an event, such as the month of participation in a government transfer program.

6.1.3 Dependent Interviewing

If the interviewer reminds the respondent about what was reported during the last interview of a panel survey, two kinds of telescoping errors may be prevented: duplicate reporting of an event that occurred and was reported in the last interview, and recalling the correct timing of an event change that happened in the current reference period. The reasoning in both cases is that the prior information serves as a time anchor which helps the respondent sort out the dates when things happened, especially when the events are not very recent (and subject to memory decay). The assumptions implicit in the dependent interviewing procedure are (1) that the past information being conveyed again is correct; and (2) that the reminding does not increase the current measurement error beyond what it would have been with a more independent interviewing approach.

While SIPP does use a form of dependent interviewing currently, and is experimenting with an even stronger form, the record check study suggests that the outcomes of such procedures are uncertain. This is true for a number of reasons: memory decay does not appear to be a serious source of reporting

---

[9] This is the familiar "more is better" assumption about amounts reported under two different survey conditions.

[10] Items that have a low probability of spontaneous recall in response to open-ended questions.

[11] The signal detection approach (e.g., Green and Swets, 1966) provides a model, measurement procedures and an estimation technique that separate two aspects of performance: (1) accuracy (the ability to discriminate between a true yes and a true no) and (2) response bias (the effect of other decision factors that people use when translating a lot of information into a binary response).

error; there is little evidence of widespread confusion about dates of participation (although more analyses are planned on this topic); error levels in Wave 1 (without dependent interviewing) are about the same as in Wave 2 (with partly dependent interviewing); and the information from the prior wave, including the most recent month, is far from error-free. Some early analyses from the record check (Marquis and Moore, 1989) suggest that dependent interviewing may increase the correlation of errors across time. So SIPP will gain some potentially important new knowledge about dependent interviewing by continuing its research in this area, especially if the research can reveal whether changes in what is reported represent improved reports.

6.1.4 Landmark Events

When respondents appear to have trouble recalling and reporting the correct months of participation in programs, recent studies (e.g., Loftus and Marburger, 1983) suggest that confusion about timing can be reduced by encouraging respondents to recall important events in their lives and to reconsider how the timing of reported events relates to such landmarks. If further analyses suggest that remembering dates is an important kind of error in SIPP, it would be worth pursuing research on the landmark events procedure, since it may have fewer drawbacks than dependent interviewing.

6.1.5 Respondent Rules

It is possible that some measurement errors arise because the interviewer does not interview the most knowledgeable person(s) in the household. It is assumed that the best information about a person comes from that person directly, and SIPP encourages everyone in the household, present when the interviewer calls, to self-respond. What would happen if SIPP adopted a more stringent rule such as all self-responses?

Marquis and Moore (1990) present a detailed analysis of errors made by self and proxy respondents for each of the 8 record checked programs. Although the data are not from an experimental design, the trends are clear enough to suggest that underreport error levels would be reduced only minimally--nowhere near zero--by an all-self-response rule, while misclassification and overreport error levels might actually increase. Below we hypothesize that none of the household members possess the necessary understanding and skill to report correctly; to do so they need both training and restructured tasks. If this hypothesis is true, it goes part way toward explaining why a simple change in the respondent rule would not greatly affect measurement errors.

While we feel that SIPP should continue to monitor self-proxy reporting errors using administrative records, we do not think that it would be beneficial to conduct experiments with the respondent rules.

6.2 Strategies to Overcome Motivational Problems

Motivational problems are indicated, for example, when respondents deliberately underreport in order to avoid perceived negative consequences. The next three procedures--randomized response, informants and reassurances--are sometimes recommended when the topics of a survey are socially sensitive, such as receiving welfare payments.

6.2.1 Randomized Response

If respondents are not reporting stigmatizing information, a randomized response procedure (Warner, 1965) may help. In its classical form, the respondent is given a randomizing device to determine whether to answer a sensitive question (e.g., about receiving welfare benefits) or an innocuous question (for which the distribution of answers in the population is known). Since the interviewer does not know which question is being answered, it is assumed that the respondent feels protected and answers the selected question truthfully. Using the assumptions, observed proportions, and knowledge about the answer distribution for the innocuous question, one solves for the unknown proportion. Variations are available for reporting continuous variables such as dollar amounts of benefits.

From the record check study we cannot tell whether the underreports of program participation are conscious and deliberate (hence, amenable to correction by randomized response procedures). We do know that many of the measurement errors are overreports, and some are due to cognitive confusion about details of the event. These kinds of errors are unlikely to be helped by a randomized response procedure. And there is always the question of whether people who underreport because they doubt the interviewer's confidentiality pledge will believe that the randomized response procedure gives them privacy protection, instead of being some sort of magic trick that the interviewer is performing.

6.2.2 Informants for Sensitive Topics

Some surveys overcome deliberate underreport errors by collecting sensitive data from another person (a "collateral" or "proxy"). As discussed above, the record check study results suggest that measurement errors will not be reduced importantly by using more proxy responses; indeed, underreporting errors may increase. And aside from administrative sources, we cannot think of other sources of collateral information (e.g., outside of the household) that would be both knowledgeable and willing to report.

6.2.3 Miscellaneous Reassurances

Surveys often place emphasis, both in advance materials and in explanations from the interviewer, on the ways that survey data are used and the ways in which privacy and confidentiality are protected. Sometimes, of course, respondents don't believe those assurances, and this opens up the possibility of additional research on persuasion.

What is clear from the discussion of the motivational design procedures is that the record check gives us very little information about the extent to

which underreporting is deliberate in SIPP and might be reduced if we could identify and remove the barriers to honest reporting. Such research is difficult, but would seem a worthwhile part of an overall program of diagnosing and fixing the causes of measurement errors.

6.3 New Procedures Based on New Research

From the record check study we have been able to rule out many traditional hypotheses about the causes of SIPP response errors. This led us to undertake some additional research to form new hypotheses about the processes that respondents use in answering SIPP questions. We describe the research briefly and then mention a different survey design strategy that we think the new research encourages pursuing.

6.3.1 Preliminary Cognitive Research Results

During the spring of 1989, we trained professional staff members to elicit information about cognitive processes during interviews, using think aloud and explicit questioning procedures. These staff members accompanied SIPP interviewers to nonsample households in the headquarters area. Staff members interrupted the interviews at appropriate places to learn whatever the respondent could reveal about the cognitive answering processes. The data for this research are the staff members' written summaries of the important verbal interactions which occurred during each tape-recorded interview.

One of the main conclusions from our review of the written summaries is that many respondents adopt a simple heuristic or rule of thumb to quickly answer questions about recurring events in the four-month reference period (such as monthly income sources and amounts). Respondents use the simple rule as a substitute for detailed, direct recall and as a substitute for checking their personal records. The respondent might recall an amount received recently, for example, and extrapolate it over the entire 4-month reference period, thus ignoring changes, adjustments, extra payments, etc. that happened during those months. Such "heuristics" were widespread and often subtly encouraged by the interviewer.

While much more cognitive research is needed to reach conclusions with confidence, the preliminary research presents the beginnings of an alternative hypothesis about the causes of most of the response errors. Instead of being caused by memory decay, forgetting, telescoping, deliberate lying, and the like, the measurement errors may be caused by trying to reconstruct a complex past using too simple a rule. If the hypothesis is correct, then research with the approach mentioned next may have a major error-reducing impact.

6.3.2 Task-Focused Interviewing

This is an overall strategy that communicates a limited set of information priorities to the respondent, trains the respondent in how to meet the objectives, uses a questionnaire that focuses clearly and flexibly on the priority information, and constrains interviewer behavior to be consistent with the priorities. This kind of strategy does not yet exist, so a good deal of research would be needed to develop it.

For panel surveys such as SIPP, one of the early goals would be to preempt the use of simple heuristic problem solving strategies and to replace them with training in what sources of information to consult (e.g. personal records), and how to interpret them (the difference between net and gross pay, estimating monthly interest from quarterly statements, keeping track of Food Stamp totals, etc.). Interviewer incentives would be changed to reward obtaining accurate, complete information, while giving much less emphasis to preventing refusals and "getting-through-the-interview-any-way-you-can." Questionnaires would be restructured so that the overall goals of each section were always clear to the respondent, and the interviewer (and respondent) would have discretion to use any of several predetermined questioning, probing, and feedback strategies to meet the objectives.

7. CONCLUSIONS AND IMPLICATIONS

We have described the results of record check research for SIPP which has yielded information about measurement errors in reports of program participation level and change. When we looked at how these measurement errors might affect statistical estimates, we learned that the effects could be considerable, both for estimates of means and for estimates of correlations. This conclusion prompted us to go on to review strategies for mitigating the effects of the measurement errors on estimation. Our considerations included both statistical correction strategies, that attempt to correct existing errors, and design alteration strategies, that attempt to prevent errors from occurring initially.

We do not recommend using current record check results either as a basis for statistical corrections or as a basis for making design change decisions. Instead we suggest obtaining new data and devising new procedures to assure quality. We suggest that the state-of-the-art is already far enough along to make survey product quality a principal and continuing concern in surveys like SIPP.

In general terms, this means constantly monitoring the quality of the data product, making after-the-fact corrections as necessary, and constantly improving the process design to eliminate measurement errors before they occur.

Monitoring quality involves learning as much as possible about the errors, both in a descriptive sense for statistical correction strategies, and in a causal sense for the behavioral design change strategies.

SIPP is in a unique and advantageous position to adopt modern quality assurance procedures because it potentially can use administrative record data to regularly monitor the quality of its priority measurements--program participation and income. For other surveys, high quality administrative record data are not always available.

Administrative record data can also serve an important function in research to develop and evaluate statistical procedures to correct for errors. There are a number of possible correction procedures to be adapted and evaluated. And the evaluation concerns not only the quality of the corrections each strategy produces, but also the quality of its assumptions and the ability of the error measurement procedure that it relies on (e.g., reinterview) to yield correct descriptions of the error characteristics of interest. Administrative record data can potentially serve all of these evaluation objectives.

We also feel that the record check study shows that administrative record data can be very helpful in evaluating selected design features, such as the length of the recall period and the respondent rule. We hope that such evaluations will continue and that design changes will be tested if the data suggest that they are warranted. However, it is clear that new kinds of data, both experimental and descriptive, are necessary, both for a fuller understanding of the causes of SIPP reporting errors, and to develop new ways of collecting household financial data that contain fewer errors.

So our view of the implications of the record check results for measurement principles and practice should now be clear:

1. Measurement errors can cause quality problems for survey data products.

2. Survey designs need to be expanded to include measures of the measurement errors.

3. Survey procedures need to include techniques to correct for measurement errors and to alter the processes that produce the errors.

4. The monitoring, product correcting, and process changing need to be a continuous, high priority part of the entire survey operation.

APPENDIX

Here, we derive the effects of response error on the correlation estimate using a classical measurement model (e.g., Lord and Novick, 1968) and offer comments about an expanded model. Let us begin with the classical model:

$$\text{Let } M = T + e,$$

Where M is the measured response, T is the true value and e is the response error. For 0,1 variables, e is a linear function of true values and a random variable, u, such that:

$$e = \alpha + \beta\ T + u.$$

The expected value of u is zero. β is a parameter representing the degree to which errors are correlated with true values. For dichotomous variables, this correlation is negative when any response error is present.

Define Z as a perfectly measured variable. Without loss of generality, define its mean as zero and its scores as deviations from the mean. We assume $Cov(u,Z) = 0$ so $Cov(M,Z) = (1 + \beta)\ Cov(T,Z)$.

The Pearson product-moment correlation, r, between true participation, T, and a perfectly measured variable whose values are deviation scores, Z, is

$$r' = Cov(T,Z)\ /(Var\ T\ Var\ Z)^{.5}$$

and the correlation, r', using measured participation, M, is

$$\begin{aligned} r &= Cov(M,Z)\ /\ (Var\ M\ Var\ Z)^{.5} \\ &= [(1 + \beta)\ Cov(T,Z)\ /\ (Var\ M\ Var\ Z)^{.5}]\ (VarT\ VarT)^{.5} \\ &= (1 + \beta)\ (Var\ T\ Var\ M)^{.5}\ r. \end{aligned}$$

The bias in the correlation estimate using measured values relative to the correlation using true values, RB(r'), is:

$$\begin{aligned} RB(r') &= (r - r')\ /\ r \\ &= [(1 + \beta)\ (Var\ T\ /\ Var\ M)^{.5}\ r - r]\ /\ r \\ &= (1 + \beta)\ (Var\ T\ /\ Var\ M)^{.5} - 1. \end{aligned}$$

We multiply RB(r') by 100 to express it as a percent.

In this model the response errors depend on the true value in the current time period. A reviewer pointed out that there could be less attenuation in the correlation estimate if response errors also depend on true values in other time periods, and if the value of the Z variable does not change over the extended time period. This might be an appropriate model if the underlying

error mechanism were some form of telescoping, for example. As a result we investigated a model in which:

$$e = \alpha + \beta T + \beta' T' + u,$$

Where β' is a parameter representing the degree of correlation between current period response error and the true value in the previous month (T'). We estimated β' for a sample of 3 time periods for all programs. Typically β' was small. Only about half of the estimates were significantly different from zero (assuming simple random sampling) and all of the non-zero results were positive. For federally administered programs, the estimates of correlation attenuation from the alternative model were between 90 and 100 percent of the estimates from the original model. For state-administered programs, the estimates of correlation attenuation from the alternative model were between 80 and 100 percent of the estimates from the original. Since the alternative model suggests slightly smaller amounts of bias in the correlation estimates, models with additional terms may show even less attenuation.

REFERENCES

ABOWD, J. and ZELLNER, A. (1985), "Application of Adjustment Techniques to U.S. Gross Flow Data," Proceedings of the Conference on Gross Flows in Labor Force Statistics, U.S. Department of Commerce and U.S. Department of Labor, Washington, DC, pp. 45-61.

BRACKSTONE, G. and RAO, J. (1976), "Raking Ratio Estimators," Survey Methodology, Statistics Canada, Vol. 2.

BURKHEAD, D. and CODER, J. (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.

CODER, J. and RUGGLES, P. (1988), "Welfare Recipiency as Observed in the SIPP," SIPP Working Paper No. 8818, U.S. Census Bureau, Washington, DC.

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

FULLER, W. (1987), Measurement Error Models, Wiley, New York.

FULLER, W. and CHUA, T. (1985), "Gross Change Estimation in the Presence of Response Error," Proceedings of the Conference on Gross Flows in Labor Force Statistics, U.S. Department of Commerce and U.S. Department of Labor, Washington, DC, pp. 65-77.

FULLER, W. and HIDIROGLOU, M. (1978), "Regression Estimates After Correcting for Attenuation," Journal of the American Statistical Association, Vol. 73, pp. 99-104.

GREEN, D. and SWETS, J. (1966), Signal Detection Theory and Psychophysics, Wiley, New York.

HIDIROGLOU, M., FULLER, W. and HICKMAN, R. (1980), SUPER CARP, Survey Section, Statistical Laboratory, Iowa State University, Ames.

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.

HUGGINS, V. and FAY, R. (1988), "Use of Administrative Data in SIPP Longitudinal Estimation," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 354-359.

JOHNSON, J. (1963), Econometric Methods, McGraw-Hill, New York.

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.

KLEIN, B. and VAUGHAN, D. (1980), "Validity of AFDC Reporting Among List Frame Recipients," Chapter 11 in Olson, J. (ed.), Reports from the Site Research Test, U.S. Department of Health and Human Services, ASPE/ISDP/SIPP, Washington, DC.

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

LEMAITRE, G. (1988), "The Measurement and Analysis of Gross Flows," Labour and Household Surveys Analysis Division Staff Report, Statistics Canada, Ottawa.

LOFTUS, E. and MARBURGER, W. (1983), "Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events," Memory and Cognition, Vol. 11, pp. 114-120.

LORD F. and NOVICK, M. (1968), Statistical Theories of Mental Test Scores, Addison-Wesley, Reading, Mass.

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., DUAN, N., MARQUIS, S. and POLICH, M. (1981), Response Errors in Sensitive Topic Surveys: Estimates, Effects, and Correction Options, R-2710/2-HHS, the RAND Corp., Santa Monica, CA.

MARQUIS, K., MARSHALL J. and OSKAMP, S. (1972), "Testimony Validity as a Function of Question Form, Atmosphere, and Item Difficulty," Journal of Applied Social Psychology, Vol. 2. pp. 167-186.

MARQUIS, K. and MOORE, J. (1989a), "Response Errors in SIPP: Preliminary Results," Proceedings of the Fifth Annual Research Conference, U. S. Census Bureau, Washington, DC, 1989.

MARQUIS, K. and MOORE, J. (1989b), "Some Response Errors in SIPP--with Thoughts About Their Effects and Remedies," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 381-386.

MARQUIS, K. and MOORE, J. (1990), "Measurement Errors in SIPP Program Reports," Proceedings of the Sixth Annual Research Conference, U.S. Census Bureau, Washington, DC (forthcoming).

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

NELSON, D., McMILLEN, D., and KASPRZYK, D. (1985), "An Overview of the Survey of Income and Program Participation, Update 1," SIPP Working Paper Series, No. 8401, U.S. Census Bureau, Washington, DC.

POTERBA, J. and SUMMERS, L. (1985), "Adjusting the Gross Change Data: Implications for Labor Market Dynamics," Proceedings of the Conference on Gross Flows in Labor Force Statistics, U.S. Department of Commerce and U.S. Department of Labor, Washington, DC, pp. 81-95.

SALTER, W. and SWETS, J. (1984), "Applications of Signal Detection Theory to Survey Research Methodology," BBN Report No. 5785, prepared for the U.S. Census Bureau by Bolt Beranek and Newman, Inc., Cambridge, Mass.

VAUGHAN, D. (1978), "Errors in Reporting Supplemental Security Income Recipiency in A Pilot Household Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 288-293.

VAUGHAN, D., LININGER, C. and KLEIN, R. (1983), "Differentiating Veterans' Pension and Compensation Income in the 1979 ISDP Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 191-196.

YOUNG, N. (1989), "Wave Seam Effects in SIPP," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 393-398.

Table 1

Common Living Benefits Riders Offered by Life Insurance Companies

Company	Policies Rider Can be Attached To	Age Limit on Purchase	Events Causing Payout	Percent of Face Value Paid
American General	Universal & Whole Life	20-65	Terminal Illness	50%
Commonwealth	Whole Life	65	Terminal Illness	50%
Metropolitan	Whole Life	65	Heart Attack, Stroke, Bypass, Kidney Failure, Cancer, Transplants, Alzheimer's, Blindness, Dismemberment	30%
Midland	Universal Life	80	Heart Attack, Stroke, Bypass, Transplants, Renal Failure, Cancer	25%
Old American	Whole Life	80	Terminal Illness	50%
People's Security	Whole Life	65	Terminal Ilness	50%
Sentry Life	Whole Life	75	Terminal Ilness	50%

Table 2

Face Value of Life Insurance by Demographic factors, 1984

	Face Value of Life Insurance						
	None	Less than $5,000	$5,000 to $10,000	$10,000 to $25,000	$25,000 to $50,000	More than $50,000	Mean Face Value for Policy Holders
All elderly persons	39.9%	44.1%	7.5%	5.7%	1.6%	1.0%	$7,389
65-74	34.7%	45.2%	8.8%	7.6%	2.2%	1.5%	$8,931
75 and up	48.3%	42.2%	5.5%	2.9%	0.7%	0.3%	$4,560
Male	30.7%	39.5%	13.3%	11.1%	3.2%	2.3%	$11,519
Female	46.3%	47.3%	3.6%	2.1%	0.5%	0.2%	$3,712
White	39.4%	43.6%	7.9%	6.3%	1.8%	1.1%	$7,734
Non-white	45.0%	48.6%	4.4%	1.3%	0.5%	0.2%	$3,774
Married	34.0%	43.3%	10.2%	8.3%	2.4%	1.8%	$9,745
Not Married	46.6%	45.0%	4.6%	3.0%	0.7%	0.2%	$4,113

Source: Calculations from the 1984 SIPP Panel

Table 3

Face Value of Life Insurance by Health, 1984

	Face Value of Life Insurance						
	None	Less than $5,000	$5,000 to $10,000	$10,000 to $25,000	$25,000 to $50,000	More than $50,000	Mean Face Value for Policy Holders
Persons 65 to 74							
In Poor/Fair Health	39.6%	46.2%	7.1%	5.4%	0.9%	0.7%	$10,673
With Qualifying Condition	41.0%	45.0%	0.6%	5.8%	1.1%	1.2%	$6,880
Persons 75 and up							
In Poor/Fair Health	50.7%	42.6%	4.3%	1.9%	0.3%	0.2%	$5,477
With Qualifying Condition	48.8%	44.7%	3.0%	2.4%	0.6%	0.0%	$4,573

Source: Calculations from the 1984 SIPP Panel

Table 4

Distribution of Income and Wealth
Among Elderly with Life Insurance
Face Value Above $25,000

	Persons with Income
Monthly Income	
Under $1,000	9.0%
$1-2,000	17.1%
Over $2,000	73.9%
	Persons with Net Worth
Net Worth (excluding home equity)	
Under $25,000	16.9%
$25-50,000	14.5%
$50-100,000	25.3%
Over $100,000	43.2%

Source: Calculations from the 1984 SIPP Panel

Table 5

Anticipated Payout From Living Benefits

Estimated Lifetime Nursing Home Costs for SIPP Cohort.....	$455,504,231,792

Estimated Living Benefits Payments

	Percent of Population Receving Amount	
None	97.3%	
$12,500-$25,000	1.6%	
$25,000 or more	1.0%	
Estimated Payout Total.....		$27,971,561,912
Percent of all LTC Costs Paid by Living Benefits.....		6.14%

Source: Wallack (1988) & Calculations from the 1984 SIPP Panel