SOME RESPONSE ERRORS IN SIPP--WITH THOUGHTS ABOUT THEIR EFFECTS AND REMEDIES[1]

Kent H. Marquis and Jeffrey C. Moore
Center for Survey Methods Research, U.S. Census Bureau

I. INTRODUCTION

A first step, when we set out to improve measurement in a household survey, is to understand the nature and extent of the measurement (or response) errors; and then to understand the consequences of the errors. Such knowledge can guide survey designers to better designs and can help them resist pressures for quick, superficial fixes to fundamentally complex measurement problems. Analysts need to know about measurement errors to assess the appropriateness of a data set and to design corrections for their effects.

A record check study is an invaluable way of obtaining the descriptive information about survey errors. Below we report the preliminary results of a record check for the Census Bureau's Survey of Income and Program Participation (SIPP).

For measures of program participation level, we find small to moderate levels of negative response bias and moderate to large levels of response error variance. For measures of change in program participation, we encounter a full range of bias estimates, from small positive to large negative, and very large response error variances. We also find that the sign of the response bias differs according to when the change measure is taken (on or off the seam between interviews), but that the size of the response error variance is largely unaffected by the seam phenomenon. The response biases and response error variances in reports of program benefit amounts show similar patterns.

We demonstrate the consequences of these response errors for estimates of means, correlations, and regression parameters. Many of these estimates are attenuated by substantial amounts, especially estimates using the change measures. In most cases, the degree of distortion is not affected greatly by whether the change measure came from on or off the seam.

In the conclusions, we develop our view that SIPP response errors should be reduced, and that redesigners may need to advance the state of the art to accomplish this goal.

II. METHODS

In this section we discuss SIPP, the record check design, and how we estimate the response errors.

SIPP collects longitudinal data from a national sample to provide federal policy planners and others with information about the economic situation of households and people in the United States. The record check focuses on SIPP reports of participation in government transfer programs and on reports of the benefits received from those programs. SIPP interviews take place every four months with a household, and obtain monthly data about each of the preceding four months. Thus, when we create measures of change between two months, every fourth change measure is assembled from measures of level reported in different interviews.

We use a full record check design (Marquis, 1978), which means that we validate both reports of participation and reports of no participation.

The record check's goal was to get administrative record information for eight programs in each of four states covering the first two SIPP interviews in 1983 and 1984. The states are Florida, New York, Pennsylvania and Wisconsin. The means-tested programs are Aid to Families with Dependent Children (AFDC), Food Stamps (FOOD), and Supplemental Security Income (SSI). The employment-related programs are Unemployment Insurance (UNEMP) and Workers Compensation (WORK). The other programs are Civil Service Retirement (CSRET), Social Security (OASDI) and Veterans Benefits (VETS). We do not include all state/program combinations in this paper because we did not obtain the cooperation of all agencies, and because our processing of some data is still incomplete. The programs included in our analyses, and the group sizes involved, are shown in Appendix 2.

We matched the SIPP and administrative record information using items that uniquely identify individuals, such as social security number, name, address, sex, and birthdate, using the Fellegi-Sunter optimization approach (Fellegi and Sunter, 1969) as it is realized in the Census Bureau's computerized matching procedures (La Plant, 1989).

Following classical measurement theory (e.g., Gulliksen, 1950 or Lord and Novick, 1968) we define a person's survey response (the measure) to be the sum of the person's true value, T, and a response error, e:

$$M = T + e.$$

For binary variables, e is a linear function of the true value and a random error component, u:

$$e = \gamma_0 + \gamma_1 T + u.$$

[γ_0, the intercept, is also the misclassification rate for true participation when participation = 0; γ_1 is a slope parameter and is involved in the expression for the misclassification rate when true participation = 1.]

For estimation, we assume that the record value is truth (R = T). Our estimate of a person's response error, $\hat{e}$, is his measured value minus his record value:

$$\hat{e} = M - R.$$

The mean of the distribution of $\hat{e}$'s over people in the sample is our estimate of response bias. We standardize by dividing by the mean of the record values to yield a percent bias:

$$\text{Percent Response Bias} = [(\Sigma_i \hat{e}_i / N) / (\Sigma_i R_i / N)] \times 100.$$

When we test whether a response bias estimate is significantly different from zero, the inferential statistic assumes simple random sampling. The effect of this assumption, since the SIPP sample design departs from simple random sampling, is to infer statistical significance too often.

We use the variance of the $\hat{e}$ distribution, Var $\hat{e}$, as our uncorrected estimate of the response error variance. We estimate the residual response error variance, Var u, as the residual variance after regressing the measured values on the true values:

$$M = \gamma_0 + (1 + \gamma_1)T + u.$$

To standardize, we divide the estimate by the total measured variance, Var M, to yield the percent residual response error variance:

$$\text{Percent Residual Response Error Variance} = \text{Var } u / \text{Var } M.$$

For each program, we assess errors in the subject matter measures of (1) participation and (2) benefit amounts. For amounts, we restrict attention to people who reported--and whose records verified--participation. The amounts are measured in dollars. For each subject matter variable we look at errors in (1) level and (2) change over time. For benefit amounts we examine the errors in whether a change was reported rather than errors in the size of the change. We also look at the change response error estimates when the subject matter variable is measured on and off the interview seam.

III. DESCRIPTIVE RESULTS

In this section we discuss the basic results of the record check evaluations, first for measures of level and then for measures of change. Within each subsection we examine errors in reports of participation in the programs and in reports of the benefit amounts.

Figure 1 shows the response biases for program participation level for eight programs. The first is OASDI, or Social Security. The standardized bias result, -7%, means that survey respondents reported 7% fewer participation months for sample persons than shown in the records. The asterisk indicates that the estimate is statistically significant (assuming simple random sampling).

All of the bias estimates in Figure 1 are negative, and all are in what we would deem the small to moderate range. As

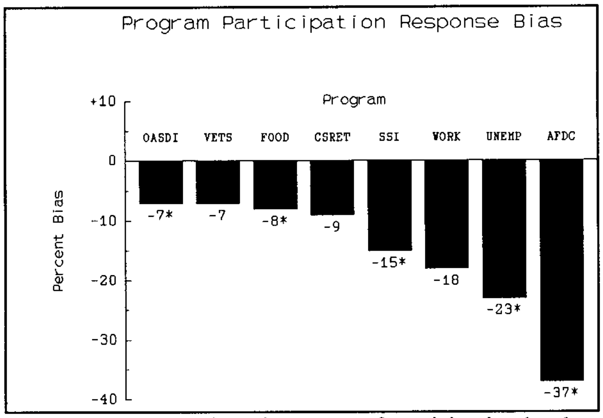

Figure 1: Response biases in reports of participation level are negative and are usually small to moderate in size.

we have discussed elsewhere (Marquis and Moore, 1989; also see Klein and Vaughan, 1980; and Goudreau, Oberheu and Vaughan, 1984) this is not necessarily due to a forgetting or a deliberate withholding bias. For example, much of the minus 37% bias for AFDC can be traced to a majority of Pennsylvania respondents calling their AFDC participation "general welfare." They did not forget or deliberately lie about participation, they just named the program incorrectly. Had it been included in the record check, we undoubtedly would have estimated a large positive bias for reports of "general welfare" participation.

Figure 2 contains our response bias estimates for reports of program benefit amount levels, contingent on both the

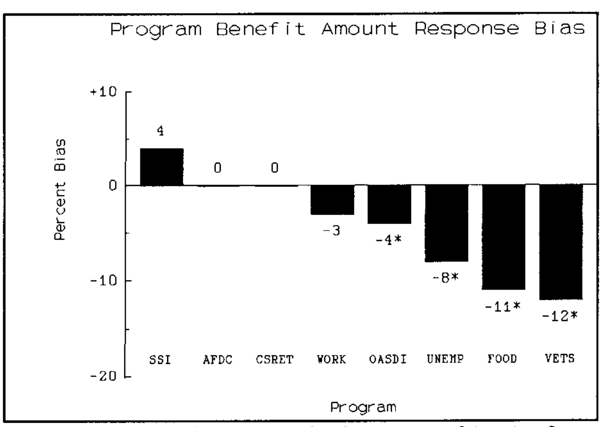

Figure 2: There is little or no bias in reports of levels of program benefit amounts.

survey and record agreeing that the person was participating in the program. A positive bias, as we estimate for the SSI program, suggests that the monthly amounts reported in the survey overstate the true amounts, in this case by 4% on the average. The negative sign, such as the one attached to the estimate for FOOD, suggests that the reports of monthly benefit amounts are lower than the true values, in this case by 11%. The median response bias is between -3% and -4%. In general, the estimates suggest little or no systematic bias for reports of monthly benefit amounts for those who report participation correctly.

Next, we consider the residual response error variances for the level measures. This kind of response error affects estimates of association such as correlations, regression coefficients, and crosstabulations.

In Figure 3 we display the error variance estimates for measures of program participation level. Recall that we

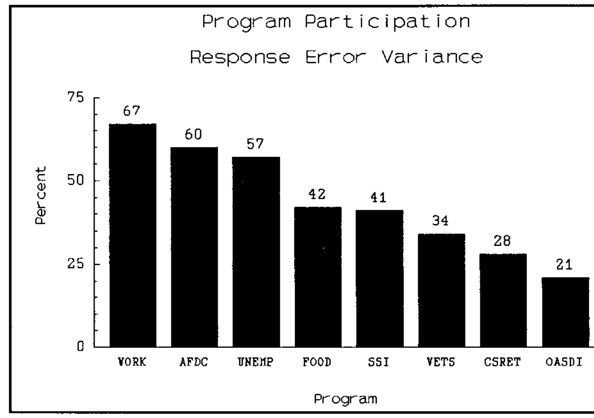

Figure 3: Response error variance is moderate to large in measures of participation level.

standardize by dividing the error variance by the total measured variance, so for WORK (workers compensation), on the left, 67% of the variance SIPP measures among people is "noise" that is unrelated to the true values. At the other extreme, 21% of the measured variance in OASDI participation is noise.

The median residual response error variance is between 41% and 42%. We classify these error variances as moderate to large, and suggest, as a general rule of thumb, that any variable whose measured variation contains more than 33% noise will cause severe problems for a user who is unaware of the problem. More than half of these SIPP measures are in the "problem" range, as are many measures from other household surveys (Marquis, Marquis and Polich, 1986).

Figure 4 suggests that the residual response error variances for benefit amount levels are less extreme, and mainly in the

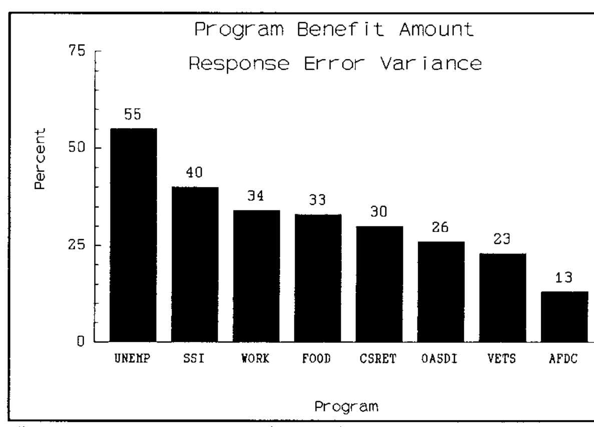

Figure 4: Most measures of benefit amount levels contain moderate percentages of response error variance.

moderate range. Recall that these are estimated only on cases which correctly report positive participation in the program. The median error variance estimate is between 30% and 33%, indicating that about a third of the measured variance is noise.

To summarize up to this point, these SIPP measures of level generally have low to moderate levels of response bias and moderate to high levels of response error variance.

Next we look at errors in the measures of change in program participation and benefit amounts. For amounts, we examine errors in whether a change occurred, rather than the dollar amount of the change.

The participation change response bias estimates, shown in Figure 5, cover a wide range: large and small, positive and

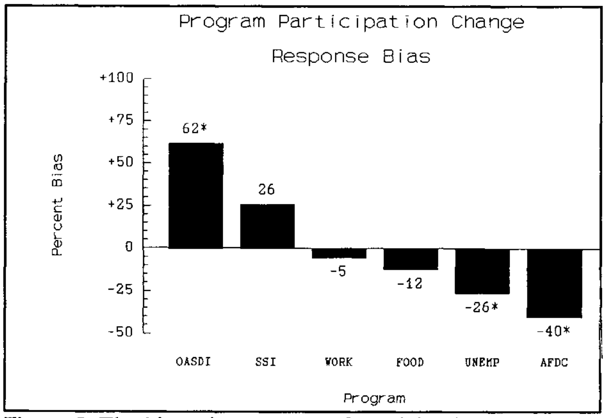

Figure 5: The biases in measures of participation change range from large and positive to large and negative.

negative. The largest estimate shown is for the OASDI program, for which SIPP respondents significantly overstate the frequency of participation changes. (The figure omits two programs--VETS, which has very little true change in participation from month to month, so a small amount of response error got magnified greatly; and CSRET, because there was no true change at all.)

Figure 6 reveals small positive, as well as moderate and large negative, net response biases for whether changes occur

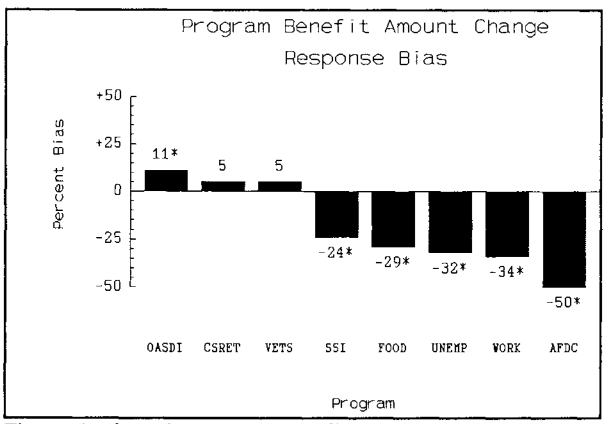

Figure 6: Biases in whether benefit amounts changed range from small and positive to large and negative.

in benefit amounts. The biases for some programs (e.g., retirement) are low and on the positive side. But the estimates for the other programs indicate that respondents underreport amount changes, even though they are correct about participation. People speculate that this may be related to the "seam" problem which we discuss later.

The response error variances for the change measures (Figures 7 and 8) are extraordinarily large. The estimates range from 84% to 100%, with the medians in the 95% to 98% range. These statistics indicate that in month-to-month SIPP measures of program change, most of the variation is not related to true program change or true benefit change.

In SIPP we get our measure of change by noting whether the reports of level in two adjacent months are the same or different. If we use data from two interviews, then we say we are measuring change at the interview "seam;" measurement of change within an interview is "off-seam."

Earlier studies (Moore and Kasprzyk, 1984; Burkhead and Coder, 1985; Hill, 1987) indicate that more changes are measured on- than off-seam. The question is, which estimate

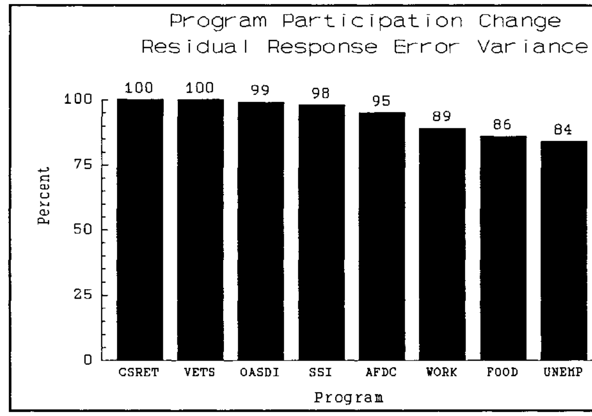

Figure 7: The percent response error variance in measures of participation change is very high.

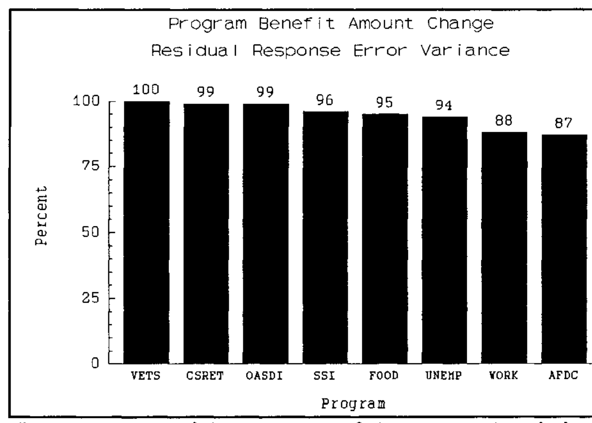

Figure 8: A very high percentage of the measured variation in whether benefit amounts changed is "noise."

is correct? From the record check study we get the answer: "neither." As we discuss below, there are systematic differences between the on- and off-seam response biases, but in neither case is the bias zero, nor does the size of the response error variance depend on the on/off seam classification.

In Figure 9, where we show the response bias estimates for program participation changes, the crosshatched bars

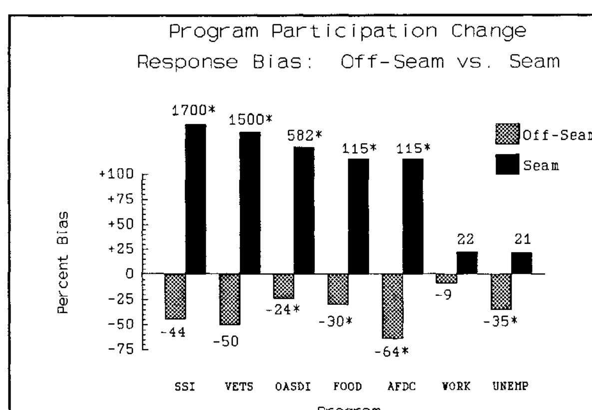

Figure 9: The effect of measuring participation change on versus off the interview seam is to change the sign of the response bias. Usually neither measure is unbiased.

portray the response bias for off-seam measures, which are all negative. The solid bars show the response bias for change measures that span the seam between interviews; they are all positive. Thus, too few changes are measured within the interview and too many are measured between interviews. Neither is unbiased.

Figure 10 shows the on- and off-seam biases for whether changes in amounts of benefits are reported, and the pattern

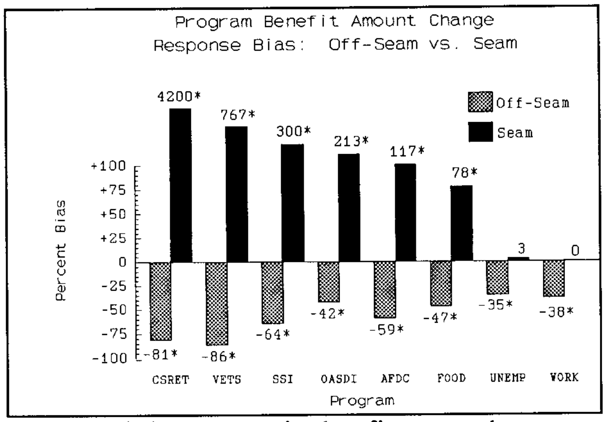

Figure 10: Similarly, measuring benefit amount change on or off the seam changes the sign of the response bias.

of results is the same: too many changes reported on the seam and too few off the seam.

Finally, the residual response error variance estimates for change are generally in the 85% to 100% range, regardless of whether the measure is taken on or off the seam. The lack of a seam effect can be seen in Figures 11 and 12. This result

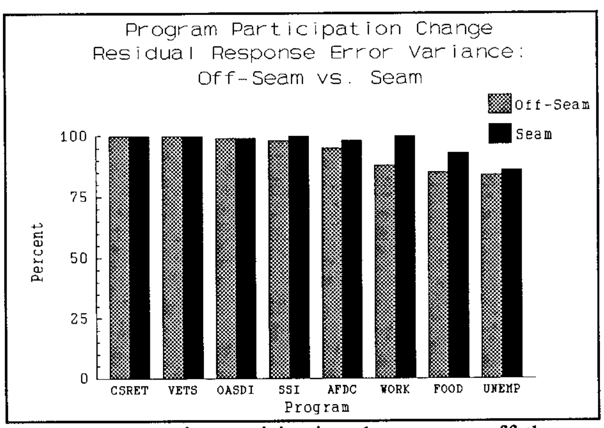

Figure 11: Measuring participation change on or off the seam has no effect on the size of the response error variance.

explains why Young (1989) finds minimal seam effects on the size of correlation and regression coefficients. We address a similar question in the next section.

IV. EFFECTS OF THE RESPONSE ERRORS

It is now appropriate to ask whether the response errors we have identified have any important consequences. Specifically, will the response biases and error variances have major effects on the uses of SIPP data by analysts? And which change measures should analysts use--on-seam or off?

In Appendix 1 we derive the formulas for estimating the effects of the response errors on several types of subject matter estimates--correlations, and regression coefficients

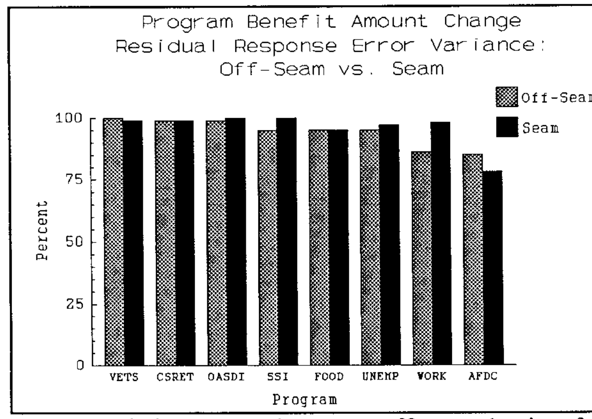

Figure 12: Similarly, there is no seam effect on the size of the response error variance in measures of whether benefit amounts changed.

(slope) and intercepts. Here, we look at the effects of the actual response errors on the subject matter estimates, using the appendix formulas.

We arbitrarily selected the PA FOOD program to illustrate the effects of the measurement errors in reports of participation, starting first, in Figure 13, with response errors in reports of participation level.

The analyst's estimate of the mean rate of food stamp program participation will be about 9% too low for the average month (-9% distortion in Figure 13). The analyst's

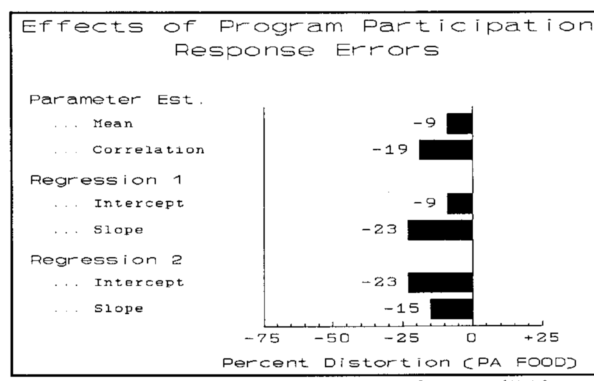

Figure 13: Response errors in measures of level will bias analysts' estimates by small to moderate amounts.

estimate of a correlation between participation and a perfectly measured variable, perhaps years of schooling, will be attenuated by 19%, on average, due to the response errors.

In regression 1 the analyst predicts whether someone participates in the PA FOOD program using a perfectly measured predictor, say employment status. The estimate of the slope of that relationship will be 23% too low, and the intercept will be 9% too low, because of the response errors. The bias in the slope estimate is because the error in the dependent variable is related to true participation and, therefore, to the explanatory variable.

In regression 2 the analyst uses food stamp participation as a predictor of a perfectly measured dependent variable, perhaps poverty level. Here the estimates of intercept and slope would be 23% and 15% too low, respectively.

The effects of response errors in the change measures are much more severe, as Figure 14 demonstrates. Rather than go through the individual estimates in detail, let us point out two important results: first, these change response errors have much bigger effects on the analysts' parameter estimates; and second, for the correlation and slope estimates, the effects of

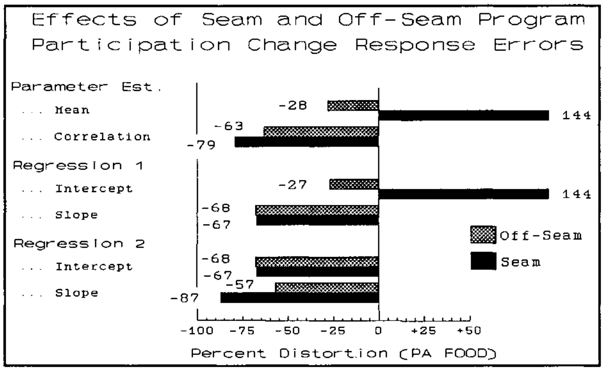

Figure 14: Response errors in change measures can severely distort analysts' estimates.

the response errors are very large regardless of whether the variable is measured on or off the seam.

The implication is that analysts should be aware of the response error structure before using SIPP data to estimate model parameters. In this example, an analyst who is unaware might not go too far astray using SIPP measures of PA food stamps participation for estimating first moments, standard errors, and associations among variables. However, that analyst would seriously underestimate relationships involving the change variables, undoubtedly failing to detect strong associations that truly do exist.

V. CONCLUSIONS

We have described some basic errors in reports of program participation and benefit amounts for a four-state subset of the early Survey of Income and Program Participation. The response biases and residual response error variances are often moderate or high, especially for measures of change in participation status and benefit amounts. The "seam phenomenon" affects the sign of the response bias (negative if the measure is off-seam, positive otherwise), but not the size of the response error variance (which is always extraordinarily high for measures of change). Therefore, eliminating the seam phenomenon will not necessarily remove the response errors that are most responsible for distorting policy model estimates.

These results are preliminary. They may change as we complete processing the data, checking for outliers, etc.

Our examination of effects reveals that the basic response errors (not the seam phenomenon) cause the major distortions in analysts' estimates of model parameters. The most critical analyst problems stem from response errors in the measures of change. These errors can have devastating effects on estimates of relationships among variables.

In our view, response errors in SIPP, although not necessarily worse than those in other similar surveys, should be reduced substantially and as quickly as possible. What are the available cures for the types and sizes of response errors uncovered? Typical remedies for errors in factual items include shortening the recall periods, adopting more stringent self response rules, or making error adjustments based on reinterview estimates. Our initial assessment of these treatments (Marquis and Moore, 1989; Moore and Marquis, 1989) suggests that they will be insufficient.

Shortening the recall period will not help since preliminary results, with one exception, show no evidence of time decay memory effects within the four month SIPP recall period. Naturally occurring variation in self/proxy respondent status also appears to be unrelated to the sizes or signs of the response bias; therefore, more stringent self response rules may not help (unless, as we currently hypothesize, they would reduce the response error variance). Unfortunately, reinterviews or other repeated interview measures will underestimate the response error variance (as evidenced by the substantial correlation of response errors across the four month period between SIPP interviews), and such underestimates cannot adequately serve a statistical error adjustment strategy. Better measures, such as from administrative records, take too long to obtain to be of short term practical value.

If we accept the goal of reducing the response errors in a survey that is already at the current state of the art, then the solutions must be at least unconventional and perhaps untried. While we will expand our analyses to the larger dataset, and evaluate other sets of assumptions relevant to conventional remedies, we suspect that SIPP measurement design improvements will have to accompany advances in the state of the art, using entirely new procedures to reduce response errors to much lower levels.

REFERENCES

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

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

Goudreau, K. Oberheu, H. and Vaughan, D. (1984), "An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program," Journal of Business and Economic Statistics, Vol. 2, pp. 179-186.

Gulliksen, H. (1950), A Theory of Mental Tests, New York: Wiley.

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.

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.

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

Lord, F. and Novick, H. (1968), Statistical Theories of Mental Test Scores, Reading, MA: Addison-Wesley.

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., Marquis, M.S. and Polich, J.M. (1986), "Response Bias and Reliability in Sensitive Topic Surveys," Journal of the American Statistical Association, Vol. 81, pp. 381-389.

Marquis, K. and Moore, J. (1989), "Response Errors in SIPP: Preliminary Results," Proceedings of the Fifth Annual Research Conference, Bureau of the Census, Washington, DC, pp. 515-536.

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), "The Quality of Self and Proxy Data in a Census Bureau Panel Survey: Preliminary

Findings Concerning Response Bias," Paper presented at the American Association for Public Opinion Research meetings, St. Petersburg, FL.

Young, N. (1989), "Wave Seam Effects in SIPP," Proceedings of the Section on Survey Research Methods, American Statistical Association, (forthcoming).

APPENDIX 1

Below we derive the effects of response error on the correlation, the regression slope coefficient and the regression intercept.

Let M = T + e.

For 0,1 variables, e is a linear function of true score, T, and a random variable, u, such that

$$e = \gamma_0 + \gamma_1 T + u.$$

The expected value of u is zero. γ_1 is a parameter representing the degree that error is correlated with true score. For dichotomous variables, this correlation is negative when any response error is present.

Define Z as a perfectly measured variable. Without loss of generality, we also define its mean as zero and its scores as deviations from the mean. Also, CovM,Z = $(1 + \gamma_1)$ CovT,Z.

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

$$r = \text{CovT,Z} / (\text{VarZ VarT})^{1/2}.$$

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

$$r' = \text{CovM,Z} / (\text{VarM VarZ})^{1/2}$$
$$= [(1 + \gamma_1) \text{CovT,Z} / (\text{VarM VarZ})^{1/2}] (\text{VarT} / \text{VarT})^{1/2}$$
$$= [(1 + \gamma_1) \text{CovT,Z} / (\text{VarT VarZ})^{1/2}] (\text{VarT} / \text{VarM})^{1/2}$$
$$= (1 + \gamma_1) (\text{VarT} / \text{VarM})^{1/2} r.$$

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

$$RB(r') = (r - r') / r.$$
$$= [(1 + \gamma_1) (\text{VarT} / \text{VarM})^{1/2} r - r] / r$$
$$= (1 + \gamma_1) (\text{VarT} / \text{VarM})^{1/2} - 1.$$

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

Regression of T on Z. Assume that we want to regress T on Z using the regression model

$$T = \beta_0 + \beta_1 Z + w$$

where β_0 and β_1 are parameters to be estimated and w is a stochastic error term. Define

$$\beta_1 = \text{CovT,Z} / \text{VarZ}.$$

Fit the model using measured participation values

$$M = \beta_0' + \beta_1' Z + w$$
$$\beta_1' = \text{CovM,Z} / \text{VarZ}$$
$$= (1 + \gamma_1) \text{CovT,Z} / \text{VarZ}$$
$$= (1 + \gamma_1) \beta_1.$$

Then the relative bias in the β_1' estimate is

$$RB(\beta_1') = (\beta_1' - \beta_1) / \beta_1 = \gamma_1.$$

Turning to the intercept, define

$$\beta_0 = \overline{T} - \beta_1 \overline{Z}.$$

Fit

$$\beta_0' = \overline{M} - \beta_1' \overline{Z}.$$

If we assume $\overline{Z} = 0$, then the relative bias in β_0' is

$$RB(\beta_0') = (\beta_0' - \beta_0) / \beta_0$$
$$= (\overline{M} - \overline{T}) / \overline{T}.$$

Regression of Z on T

Assume the regression model

$$Z = \beta_0 + \beta_1 T + w$$

where we define

$$\beta_1 = \text{CovZ,T} / \text{VarT}.$$

Fit the model using measured values:

$$Z = \beta_0' + \beta_1' M + w$$
$$\beta_1' = \text{CovZ,M} / \text{VarM}$$
$$= (1 + \gamma_1) \text{CovZ,T} / \text{VarM}$$
$$= [(1 + \gamma_1) \text{CovZ,T} / \text{VarM}] (\text{VarT} / \text{VarT})$$
$$= (1 + \gamma_1) \beta_1 (\text{VarT} / \text{VarM}).$$

The relative bias in β_1' is

$$RB(\beta_1') = [(1 + \gamma_1) (\text{VarT} / \text{VarM}) \beta_1 - \beta_1] / \beta_1$$
$$= (1 + \gamma_1) (\text{VarT} / \text{VarM}) - 1.$$

Define $\beta_0 = \overline{Z} - \beta_1 \overline{T}$.

Fit $\beta_0' = \overline{Z} - \beta_1' \overline{M}$

If $\overline{Z} = 0$, then the relative bias for β_0' is

$$RB(\beta_0') = (\beta_0' - \beta_0) / \beta_0$$
$$= (- \beta_1' \overline{M} + \beta_1 \overline{T}) / (\beta_1 \overline{T})$$
$$= (\beta_1' \overline{M} / \beta_1 \overline{T}) - 1.$$

Since $\beta_1' / \beta_1 = (1 + \gamma_1)(\text{VarT} / \text{VarM})$

and since

$$\overline{M} / \overline{T} = (\overline{T} + \overline{e}) / \overline{T}$$

then the relative bias is

$$RB(\beta_0') = (\beta_1' / - \beta_1) (\overline{M} / \overline{T}) - 1$$
$$= (1 + \gamma_1) (\text{VarT} / \text{VarM}) ((\overline{T} + \overline{e}) / \overline{T}) - 1$$

which we multiply by 100 to express as a percent.

APPENDIX 2: Group Sizes Used in Analyses

PROGRAM	LEVEL Both	CHANGE Both	CHANGE Off Seam	CHANGE On Seam
Participation				
AFDC	10679	10612	10612	10397
CSRET	10679	10612	10612	10397
FOOD	9152	9086	9086	8896
OASDI	10679	10612	10612	10397
SSI	10679	10612	10612	10397
UEMP	7295	7237	7237	7078
VETS	10679	10612	10612	10397
WORK	5159	5117	5117	4995
Benefit Amounts				
AFDC	72	67	67	45
CSRET	67	65	65	58
FOOD	202	187	187	128
OASDI	1508	1498	1497	1370
SSI	101	99	99	85
UEMP	204	153	152	46
VETS	134	134	134	123
WORK	26	21	21	9

NOTE

1. This 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.