RESEARCH REPORT SERIES
(Survey Methodology #2010-11)

Using Administrative Record Data to Describe SIPP Response Errors

Jeffrey C. Moore
Kent H. Marquis

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

This paper was originally prepared for presentation at the annual meetings of the American Statistical Association, New Orleans, August 22-25, 1988.

Report made available online: June 9, 2010

USING ADMINISTRATIVE RECORD DATA TO DESCRIBE SIPP RESPONSE ERRORS

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

The SIPP Record Check Study uses administrative record data to describe and model errors in SIPP responses for a variety of income sources and transfer programs. The project uses record linkage techniques to identify SIPP sample persons in four states who are on record as having received payments from any of nine state or Federal programs, and then compares survey-reported dates and amounts of payments with official record values. The paper describes basic considerations in designing the project and presents some early findings describing errors in reported program participation and amounts of payments received. Results of the study are also used to shed light on the "seam bias" problem in SIPP.

Key words: Record check, record linkage, nonsampling error, response bias

Paper prepared for the session "Survey Methodological Research at the U.S. Census Bureau" at the annual meetings of the American Statistical Association, New Orleans, August 22-25, 1988. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are the authors' and do not necessarily reflect those of the Census Bureau.

*Center for Survey Methods Research, U.S. Bureau of the Census, Room 2737 - FB 3, Washington, DC, 20233.

1. INTRODUCTION

This paper addresses issues concerning the use of records to evaluate the quality of survey estimates and describes a specific application to the Survey of Income and Program Participation (SIPP) in the United States.

Matching administrative records to survey observations on a case-by-case basis, which we call a "record check," provides useful information to survey users and designers. A record check enables the analyst to make a full range of measurement error parameter estimates for evaluation purposes. These estimates, in turn, facilitate two basic kinds of activities:

(1) adjusting subject-matter estimates such as means, proportions, correlation coefficients, and multivariate regression coefficients to correct for the measurement errors; and

(2) deriving more efficient survey designs that directly address, for example, the trade-offs between measurement quality and costs.

1.1. Basic Terms

Our focus here will be on using administrative records to assess measurement or response errors, although the record check method can also be extended to evaluate other nonsampling and sampling errors. This is not a technical exposition, but we do need to define some of our basic terms first. We assume that the survey observation from sample element i can be expressed as the sum of the true value and an error, e: $\text{Survey}_i = \text{True}_i + e_i$.

The average bias in a set of N survey observations, which we call the response bias or survey bias, is $\bar{e} = \Sigma e_i / N$ and the response error variance is just Var e. Similarly, the measurement model for administrative record observations is: $\text{Record}_i = \text{True}_i + u_i$, so that the record bias is $\bar{u}$ and record error variance is Var u.

Like all other evaluation techniques, record checks must employ assumptions in evaluating survey measurements. For example, the usual way of estimating the response bias is to assume no record bias ($\bar{u} = 0$) and take the average of the differences between the matched survey and record observed values: Estimated Survey Bias $= \Sigma (S_i - R_i) / N$. Although it is usually impossible to confirm directly the no record bias assumption, one can conduct meaningful sensitivity tests of the effects of possible violations of the assumption on evaluation conclusions.

1.2 Issues in Designing Record Checks

Several issues merit consideration in designing a record check to evaluate survey measurement. We comment on some of the main issues here: incomplete observation designs, matching errors, record errors, true value

differences, and absence of repeated measures or experimental design features.

1.2.1. Incomplete observation designs

Past record checks have often used one-directional or partial designs for data collection--for example, surveying people about owning library cards and checking the records for those who claim to have one, or sampling from a list of people with a diagnosed chronic disease and surveying them to see if they report it in a survey questionnaire. Because these partial designs do not observe the full range of response errors in the correct proportions, they yield biased estimates of such classical measurement error parameters as the response bias and the response error variance. One-directional designs can fail to detect some or all of the true survey bias, can cause up to one-half of the response error variance to be misinterpreted as response bias, and can predetermine the sign of the estimated response bias if the measured variable is binary (Marquis (1978)). Full designs are a necessary (albeit not sufficient) condition for obtaining unbiased estimates of the desired response errors.

1.2.2 Matching errors

The essence of the record check is a one-to-one matching of survey and record observations. This is difficult to do correctly, and matching errors (false matches, false nonmatches) will potentially bias the measurement error estimates of interest. Neter et al. (1965) show that when there are no unmatched cases, the mismatches will bias the estimates of response error variance upward. In terms of the reliability of a dichotomous measure (which is a function of the response error variance), the estimate will be reduced by exactly the match error rate (Marquis et al. (1986)). It is therefore desirable to keep match errors to a minimum and to know something about the errors that remain.

1.2.3. Administrative record errors

A successful record check study requires records which are very good measures of the trait of interest. If the implied assumptions about record measurement bias and record measurement error variance are violated, this can cause the response error estimates to be biased away from zero. For example, bias in the record observations can appear as bias in the survey observations but with the opposite sign. Feather (1972) describes this effect in a record check of physician visits in Saskatchewan, in which an apparently large survey overreporting rate was due to the record's recording a complete treatment procedure rather than the individual visits for the diagnosis. Similarly, the presence of measurement error variance in the record can cause inflated estimates of response error variance in the survey (Marquis (1978)).

1.2.4. True value differences

Problems arise when the survey and record systems use different definitions. This is often the case in "aggregate comparisons" of population parameter estimates made separately by each source. A common difference is in the scope of the populations covered, such as when the survey frame is limited to the civilian, noninstitutionalized population and the administrative records include everybody. Case-by-case matching can minimize the threats posed by differential coverage, but estimates derived from these studies can still be plagued by differences in the concepts or the attributes of the concept.

1.2.5. Absence of experiments and reinterviews

Evaluation record checks can detect errors but are not good at evaluating possible remedies for the errors. To know how well a different survey design might perform, one must usually either test the alternative design options or arrange to estimate parameters of an underlying model from which survey designs can be derived (e.g., a model of forgetting effects). For example, an evaluation record check design can estimate and compare response errors for self and proxy respondents. But without heroic assumptions it cannot suggest how the measurement error parameters would change if the survey's respondent rule were changed (say, to allow only self-response).

Similarly, a record check without a reinterview or another set of independent measures is limited in the number of basic error parameters it can estimate. For example, our initial definitions mentioned three parameters: true value, survey error, and record error. Without a reinterview (or other independent measure) there are only two measures with which to estimate the three unknowns. An additional measure such as a reinterview can help identify the estimates of the parameters in the model.

2. CHARACTERISTICS OF SIPP

2.1. SIPP Overview

The purpose of SIPP is to provide improved information on the economic situation of persons and families in the United States by collecting comprehensive longitudinal data on a wide range of topics including: cash and noncash income; eligibility for and participation in Government transfer programs; assets and liabilities; labor force participation; and household and family dynamics. Core SIPP questions--repeated in each interviewing wave--cover nearly fifty sources of income.

2.2. SIPP Data Collection Design

SIPP started in October 1983 with a sample of approximately 25,000 designated housing units (the "1984 Panel") selected to represent the noninstitutional population of the United States. Starting in February

1985, and continuing in each February thereafter for the life of the survey, new panels are introduced into the sample. Due to budget reductions these subsequent panels have been reduced in size; the sample size for new panels is currently about 15,000 households.

Each sample household is interviewed by personal visit eight times--once every four months for 2-1/2 years. The reference period for each interview is the four months preceding the interview month. At each visit to the household, each person fifteen years of age or older is asked to provide information about himself/herself, but proxy reporting is permitted for household members not available at the time of the visit. Information concerning proxy response situations is recorded and is available for analytical purposes.

To facilitate field operations, each sample panel is divided into four subsamples ("rotation groups") of approximately equal size, one of which is interviewed each month. Thus, one "wave" or cycle of interviewing is conducted over a period of four months for each panel. This design produces steady field and processing workloads, but it also means that each rotation group uses a different four-month reference period.

3. RECORD CHECK DESIGN

The purpose of the record check is to provide an evaluation of some of the data gathered in SIPP. We highlight important features of the design of the record check next, covering the samples, the administrative records, the matching approach, and the analysis.

3.1. Record Check Samples

The records available to the SIPP Record Check Study allow a "full" record check design--one which permits the validation of all observed values in the survey. Design options we did <u>not</u> choose include:

(1) checking records only for people who claimed to be participating in a program; or

(2) drawing a sample only of known recipients and interviewing them to determine how truthfully they report.

As noted above, both of these designs are incomplete and would result in biased estimates of the response error parameters.

The study uses a subset of available data from the 1984 SIPP Panel. First, the sample of people is restricted to households in four target states: Florida, New York, Pennsylvania, and Wisconsin. In the 1984 Panel this translates to approximately 5,000 households. No attempt was made to sample states to be representative of the Nation; they simply had to meet the following criteria:

(1) a computerized, accessible, and complete record system for all target programs;

(2) a large SIPP sample;

(3) reasonable geographic diversity; and

(4) a willingness to share individual-level data for purposes of this research.

Second, the study's sample of calendar time periods includes only the first two waves of the 1984 Panel. Figure 1 illustrates the wave, rotation group, interview month, and reference period structure for the target survey data.

[Figure 1 here]

Third, the SIPP Record Check Study focuses on the quality of recipiency and amount reporting for selected Government transfer programs. The study includes five Federally-administered programs (Federal Civil Service Retirement, Pell Grants, Social Security (OASDI), Supplemental Security Income (SSI), and Veteran's Compensation and Pensions) and four state-administered programs: Aid to Families with Dependent Children (AFDC), Food Stamps, Unemployment Compensation, and Worker's Compensation).

From each participating agency we requested identifying and receipt information for all persons who received income from the target program at any time from May 1983 through June 1984. We obtained these administrative records with the understanding that they would be accorded the same confidentiality protection as data gathered by the Census Bureau under Title 13 of the U.S. Code. Thus, the records may be used only by sworn Census Bureau employees engaged in and for the purposes of this study. Except in the form of non-individually identifiable statistical summary data, the records may not be released or disclosed to any others for any purpose.

3.2. Administrative Record Quality

As noted earlier, errors in the records can cause problems for record check evaluation studies. Although several of the administrative record files obtained for this project contain very minor deficiencies (for example: not listing a middle initial; no sex designation; age, rather than date of birth; etc.), only two appear at all likely to pose major analytical problems because of incomplete coverage of recipients: the New York Worker's Compensation file, and the Veterans' Compensation and Pensions file covering all four states. The New York file excludes an unknown number of cases which were "closed" (i.e., cases which had already been adjudicated and for which payments by a private insurance carrier had already begun) at the time the data base was created several years ago. The veterans file excludes the approximately one percent of all recipients

Wave	Rotation Group	Interview Month	Reference Period Months											
			Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	
1	1	Oct 83	X	X	X	X								
	2	Nov 83		X	X	X	X							
	3	Dec 83			X	X	X	X						
	4	Jan 84				X	X	X	X					
2	1	Feb 84					X	X	X	X				
	2	Mar 84						X	X	X	X			
	3	Apr 84							X	X	X	X		
	4*	May 84								X	X	X	X	

Figure 1: Survey Structure for Data Included in the SIPP Record Check Study

*Technically, rotation group 4 of the 1984 SIPP Panel was not administered a Wave 2 interview. The "missing" interview was transparent to respondents, however, who were simply given their Wave 3 interview at the time they would have received the Wave 2 interview. For present purposes the Wave 3 interview for rotation group 4 is identical to the Wave 2 interview for all other rotation groups, and is included in the Record Check Study in order to have two interviews from all sample cases. All references in the text of this paper to "Wave 2" include the Wave 3 interview for this portion of the panel.

whose benefits were sent to a financial or other institution. There are no known coverage problems with any other files.

3.3. Definitional Differences

A problem which afflicts several of the administrative files is the discrepancy between payout date and receipt of payment. Where the payout date is close to the end of a month it may be difficult to distinguish a forward telescoping error from a legitimate difference between month of payment and month of receipt. Where there are definitional discrepancies, such as this payment date issue, our analyses will attempt to model them explicitly.

4. MATCHING

The goal of matching is to locate and link the true administrative records (if any) to the record of each SIPP sample person. The quality of matching has important effects on some of the most critical response error estimates such as the response error variance. Ideally, variables used to match survey and administrative record observations are measured without error and are able to identify an individual uniquely.

The ideal, of course, is never realized. However, the variables available to this study for matching between SIPP and administrative records should help to minimize match errors. Some, such as social security number (SSN), come close to uniquely identifying an individual even if other information, such as address, is outdated, garbled, or missing. Matching will benefit from special measures the Census Bureau has taken to ensure that SSN information as reported to the SIPP is complete and valid. For all Wave 1 and 2 sample persons, reported SSN's and reports of not having an SSN were verified and, if necessary, corrected, by the Social Security Administration. Sater (1986) estimates that, as a result of this operation, the SIPP file contains a valid SSN for about 95 percent of SIPP sample persons who have one.

The wealth of other data--last name, first name, house number, street name, apartment designation, city, zip code, sex, and date of birth--is sufficient for high quality matching even in the absence of a unique identifier such as SSN. In addition, to aid in evaluating the impact of any remaining match errors, the Census Bureau's matcher produces an ordinal measure of the goodness of the match/nonmatch of each survey observation to its appropriate administrative record counterpart.

4.1. The Census Bureau's Computerized Match Procedures

The Record Check Study uses computerized matching procedures applying Fellegi and Sunter's (1969) theoretical work on record linkage. Computerized matching is the process of examining two computer files and locating pairs of records--one from each file--that agree (not necessarily

exactly) on some combination of variables. The process involves four basic steps:

(1) standardizing the common data fields in the two files which the matcher will examine to determine whether a pair of records is a match or not;

(2) sorting the two files into small subsets of records (or "blocks") which constitute a feasible number of pairs to be examined by the matcher;

(3) determining and quantifying the usefulness of each data field to be considered in the match for identifying true matched pairs; and

(4) implementing the computer algorithms which perform the actual record matching.

4.1.1. Standardization

The SIPP files and most of the administrative record files have been processed through an address standardizer which standardizes the format of various components of an address (e.g., street name, type, and direction; city name; state abbreviation; etc.) and parses each component into a fixed data field. Several programs have been developed for this purpose; we use the ZIPSTAN standardizer developed at the Census Bureau.

In addition to address standardization, many of the files have required modifications to individual data fields to ensure a common format across files for matching. Common examples of problems of this type are sex (which can be represented by either an alpha ("m" or "f") or a numeric ("1" or "2") code); date of birth (which has many variants--e.g., "mm-dd-yy," or "cc-yy-mm-dd," or the Julian format); and name (which may be a single field or which may have separate fields for each component).

4.1.2. Blocking

Blocking--establishing subsets of records for the matcher to examine in searching for matched pairs of records--is a necessary strategy when matching files with large numbers of records (Jaro (1985)). Obviously, the probability of finding all true matches would be highest if, for each record on one file, the entire other file were searched for a match. For large files, however, unrestricted searches for matched records is simply not feasible. Blocking each file into subsets of records makes matching large files feasible, but at the cost of excluding some records from the search, thus increasing the likelihood that some true matches will be missed.

Ideal blocking components must have sufficient variation to ensure the partitioning of the files into many small blocks, and should be effective match discriminators--that is, should nearly always agree in true match record pairs and nearly always disagree in true nonmatch record pairs. The

first criterion (sufficient variation) is easy to achieve; the second is more problematic. Because the success of the match is so sensitive to the blocking scheme, all SIPP/administrative record matches (including those whose results we report here) will be conducted multiple times with independent blocking strategies. This will minimize the likelihood that a true match pair will escape detection as a result of blocking.

The primary blocking strategy for the SIPP Record Check Study employs the first three digits of the United States Postal Service's five-digit zipcode and a four character SOUNDEX code derived from the sample person's/recipient's last name. The three-digit zipcode string is a sub-state geographic indicator which generally is recorded quite accurately according to Census Bureau matching experts. Soundexing is a widely-used algorithm for creating a standard length, standard format code from input character strings of varying lengths. The code is comprised of the first letter of the string (here, the last name) followed by a numeric code which is based on only certain letters in the remainder of the string. One advantage of such encoding for blocking purposes is that it minimizes the effect of common misspelling errors, although it cannot eliminate such errors entirely.

Subsequent blocking arrangements will not be uniform for all matches (because of variations in the availability of some data fields or because of known problems with quality) but will include some combination of sex, month of birth, day of birth, SOUNDEX code for city or street name, or partial SSN.

4.1.3. Data Field Match Weights

Intuitively, the data fields used to match records across the SIPP and administrative files--house number, street name, apartment number, city, ZIP Code, SSN, sex, date of birth, last name, first name, etc.--are not equally useful in determining whether a particular pair is a true match or not. Agreement on sex is not as indicative of a true match as is agreement on SSN, for example. Fellegi and Sunter (1969) discuss weight calculations reflecting different data fields' differing discriminating powers and how these weights feed into optimal match decision rules. The Census Bureau's Record Linkage Research Staff has developed programs using Newton's method for non-linear systems (see Luenberger (1984)) to solve the Fellegi-Sunter equations, and these programs are being used in the SIPP Record Check Study to compute final match weights.

4.1.4. The Computer Matcher

The Census Bureau has developed a computer matcher (CENMATCH) which executes the procedures of Fellegi-Sunter on a user-defined set of data fields on files sorted (blocked) according to user specifications. The user enters the initial match weights for each field, defines the type of agree/disagree comparison for each field (whether the fields must correspond exactly or only approximately in order for the matcher to treat them as agreeing), identifies missing value entries and specifies how they

are to be treated (included or ignored in the calculation of a composite match weight), and sets the composite weight cutoff values for matched pairs and nonmatched pairs. The user generates the appropriate COBOL program codes to conduct a match according to these specifications through GENLINK, the Census Bureau's Record Linkage Program Generator (LaPlant (1987)).

In simple terms, the matcher:

(1) searches each data file for comparable blocks of records--that is, records which agree exactly on the designated blocking components;

(2) counts the number of records in eligible (matched) blocks to ensure that neither file's block size exceeds the preset maximum (matching is not carried out on oversized blocks, but they are flagged for subsequent re-blocking and re-matching);

(3) computes a composite weight (based on the agree/disagree comparisons for all match variables) for all possible pairs of records in the block;

(4) assigns each record in the smaller block to a paired record in the larger block according to a formula which maximizes the total composite weight for all pairs in the block;

(5) applies the Fellegi-Sunter decision procedure to determine whether a pair is a match, a nonmatch, or requires further review; and

(6) produces a "pointer" file map to the skipped records (i.e., records in a block on one file that is not matched with a corresponding block in the other file) and the paired records (matched/review/unmatched) in each file.

5. ANALYSIS

Our goals for the record check study are to estimate selected measurement error parameters for our restricted samples of people, content, and times, and to assess how these errors relate both to each other and to variables that reflect survey design features. We will use the matched data to estimate:

(1) response bias (using the survey-minus-record difference score);

(2) predictors of response bias (using logistic or probit regression techniques or possibly LISREL techniques based upon matrices containing polyserial and tetrachoric coefficients of association (Joreskog and Sorbom (1984));

(3) response error variance (e.g., from regression residuals);

(4) conditions or groups associated with very large and very small response biases and error variances; and

(5) confusion among transfer programs that contribute to the response errors (using covariance structure analysis procedures such as LISREL).

The measurement error issues to be addressed fall into one of two categories: issues which apply to all time periods and issues that require comparing errors across time periods. In the former category are estimates of the amounts of response errors for self and proxy respondents or errors attributable to interviewers. In the latter category are the errors arising from panel surveys with familiar labels such as telescoping, time-in-sample bias, memory decay, rotation group bias, etc.--those implying that measurement errors will differ across time periods when everything else is held constant. To this list we add what Hill (1987) has referred to as the "seam" bias in longitudinal surveys, which we discuss below.

To appreciate the applied questions we wish to address about the different time periods, consider Figure 2, which presents the interview and reference month calendar for one rotation group of SIPP respondents.

[Figure 2 here]

The figure shows two interviews. The first takes place in early October and asks about what happened in June (four months ago), July (three months ago), August (two months ago), and September (last month). The second interview, taken four months later, asks about October, November, December, and January. We refer to the September/October transition as the "seam" because it is between the reference periods covered by two interviews.

To investigate the internal telescoping hypothesis (which asserts that events are not forgotten, just remembered as having happened closer to the present time), we will be testing whether the response bias is negative for the early months of the reference period (June and July in Wave 1 and October and November in Wave 2), and positive for the later months, and that the two biases sum to zero.

Many retrospective surveys (e.g., the National Crime Survey) use a "bounding" procedure to control external forward telescoping--the tendency to report in the survey events which happened outside of the reference period. Bounding is usually accomplished through the use of initial interview whose primary purpose is to serve as an out-of-scope event reference for both interviewer and respondent. We will investigate the bounded interview hypothesis by estimating the extent to which past events influence what is reported for the early time periods after controlling for the true events in those time periods.

To examine the hypothesis about memory decay (that the probability of forgetting an event increases with the passage of time), we will test whether the response bias is more negative for the early months of each reference period than for more recent months.

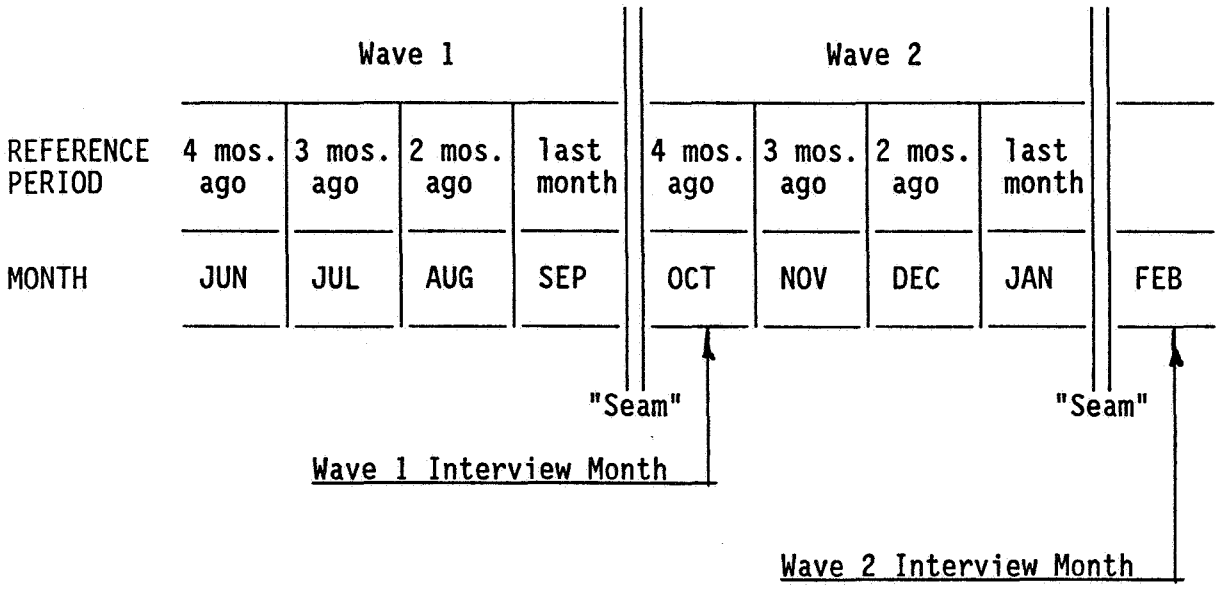

Figure 2: SIPP Survey Time Periods for Rotation Group 1 Showing Reference Months, Calendar Months, Interview Months, and Interview "Seam"

The time-in-sample and rotation group bias hypotheses suggest that response errors will be greater in the second interview than the first, after correcting for any seasonal effects. We plan to examine this and, if we find it to be true, test some of the ideas in the literature about why it may be true. Are the sample elements that survive from the first to the second interview different, as Stasny and Fienberg (1985) suggest, or does the quality of the survivors' reporting deteriorate as the Neter and Waksberg (1966) conditioning hypothesis might predict?

We don't know yet the extent to which SIPP is experiencing these more traditional problems of longitudinal surveys. One problem for which there is evidence, however, concerns the estimation of month-to-month changes in program participation (Burkhead and Coder (1985)). Specifically, more changes in program participation take place at the "seam" between two interviews than between the months covered by any one interview. Moore and Kasprzyk (1984) and Hill (1987) have speculated about what kinds of response, nonresponse, or procedural errors might be producing the seam effect and which set of transition estimates is more accurate. By addressing the problem with administrative data, we hope to come much closer to a definitive explanation about the role of response and nonresponse errors in producing the observed pattern.

Related, possibly, to the seam bias issue is the better-understood phenomenon that measurement error variance tends to inflate estimates of gross change or underestimate stability. Recent literature (e.g., Fuller and Tin (1986)) suggests several possible approaches to the problem. We plan to begin the empirical exploration of the measurement error effects on the transition estimates to learn whether, for example, we can base corrections for the response errors on estimates from reinterviews.

Finally, we have hinted previously at the problems that may arise in getting unbiased estimates of the errors if the administrative records also contain errors. We plan, with the use of reinterview measures (that identify the estimate of Var e) to estimate the record error variance (Var u). However, we have no plans to relax the assumption that the records are unbiased.

6. INITIAL RESULTS

We have just recently reached the stage of putting together matched/merged data files combining SIPP responses and administrative record data, so our analyses are still in their infancy. We present here early results on the quality of reports of participation and amounts received for two income sources--Aid to Families with Dependent Children (AFDC), and Food Stamps--in one state, Wisconsin. (Note that for AFDC and Food Stamps we use the household as the unit of analysis rather than the individual, because these programs generally pay benefits to a group of people.) Because of the limited nature of the data, we report here only simple descriptive statistics; a more formal statistical analysis must await data file construction for more programs in more states.

6.1. Bias in Reported Participation

Table 1 summarizes the match results and presents the mean discrepancy and percent bias estimates for households' participation in AFDC for each month of Wave 1 and Wave 2. For these analyses, a household was considered a SIPP "yes" if any member reported receipt of AFDC, and a record "yes" if any member was indicated as a recipient on the records. (SIPP procedures assume that all Wave 1 households maintain a constant composition through all months of the wave, and thus the number of households in Wave 1 is constant. Composition changes are recorded in Wave 2, so the number of Wave 2 households varies from month to month.)

[Table 1 here]

The AFDC results are quite straightforward. The discrepancies between SIPP and administrative record reports of AFDC participation are small, and none is statistically significant according to a t-test that assumes simple random sampling. (This conclusion would not change if the departures from simple random sampling were taken into account.) There is no consistent directional pattern to the discrepancies. The data offer support for neither consistent underreporting nor memory decay (i.e., decreasing reporting quality with increasing recall length). In short, reports of AFDC participation are not without error, but the average error is minimal.

Table 2 presents results for receipt of Food Stamps. In comparison to the AFDC results summarized in Table 1, SIPP reports of Food Stamps participation appear to be more severely affected by response error. There is a clear suggestion here of substantial and consistent underreporting in SIPP. Seven of the eight discrepancy estimates are negative, and six of those are significantly negative. The data also appear to exhibit a trend supportive of a forgetting decay model--within each wave, underreporting is increasingly severe from the more recent to the more distant months of the reference period--although the fact that response errors are potentially correlated over time prevents us from evaluating this trend with a simple statistical test.

[Table 2 here]

6.2. Bias in Reported Amounts

The second basic datum of SIPP, in addition to reported receipt of various types of income, is reported income amounts. Next, we examine results of the quality of monthly benefit amount reporting for AFDC and Food Stamps among Wisconsin sample persons. There are several ways such an analysis can proceed. We have chosen to restrict the comparison of SIPP and record amounts to households with nonzero amounts in both data sources so as not to confound the estimated quality of amount reporting with misreporting of participation.

Table 1: AFDC Participation According to SIPP and Administrative Records

Table 1a: Match Results

SIPP: / Record:	yes / yes (a)	yes / no (b)	no / yes (c)	no / no (d)	N
Wave-Month:					
1-4 months ago	25	2	1	501	529
1-3 " "	26	2	5	496	529
1-2 " "	25	2	4	498	529
1-1 " "	25	2	4	498	529
2-4 " "	23	1	4	495	523
2-3 " "	25	1	1	503	530
2-2 " "	26	1	1	502	530
2-1 " "	26	1	0	507	534

Table 1b: Estimates

	Participation Rate: SIPP (a+b/N)	Record (a+c/N)	Mean Discrepancy (b-c/N)	Std. Error of Mean	t*	Percent Bias 100x(b-c/a+c)
Wave-Month:						
1-4 months ago	.0510	.0491	.0019	.0033	0.58	3.8%
1-3 " "	.0529	.0586	-.0057	.0050	-1.13	-9.7%
1-2 " "	.0510	.0548	-.0038	.0046	-0.82	-6.9%
1-1 " "	.0510	.0548	-.0038	.0046	-0.82	-6.9%
2-4 " "	.0458	.0516	-.0057	.0043	-1.34	-11.1%
2-3 " "	.0490	.0490	0	.0027	0	0
2-2 " "	.0509	.0509	0	.0027	0	0
2-1 " "	.0505	.0486	.0019	.0019	1.00	3.8%

*This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

Table 2: Food Stamps Participation According to SIPP and Administrative Records

Table 2a: Match Results

SIPP: / Record:	yes / yes (a)	yes / no (b)	no / yes (c)	no / no (d)	N
Wave-Month:					
1-4 months ago	30	1	8	490	529
1-3 " "	32	0	5	492	529
1-2 " "	29	0	4	496	529
1-1 " "	28	0	4	497	529
2-4 " "	24	1	9	489	523
2-3 " "	21	2	10	497	530
2-2 " "	23	1	4	502	530
2-1 " "	25	2	2	505	534

Table 2b: Estimates

	Participation Rate: SIPP (a+b/N)	Record (a+c/N)	Mean Discrepancy (b-c/N)	Std. Error of Mean	t*	Percent Bias 100x(b-c/a+c)
Wave-Month:						
1-4 months ago	.0586	.0718	-.0132	.0056	-2.34	-18.4%
1-3 " "	.0604	.0699	-.0095	.0042	-2.24	-13.5%
1-2 " "	.0548	.0623	-.0076	.0038	-2.01	-12.1%
1-1 " "	.0529	.0604	-.0076	.0038	-2.01	-12.5%
2-4 " "	.0478	.0630	-.0153	.0060	-2.54	-24.2%
2-3 " "	.0433	.0584	-.0151	.0065	-2.32	-25.8%
2-2 " "	.0452	.0509	-.0057	.0042	-1.34	-11.1%
2-1 " "	.0505	.0505	0	.0037	0	0

*This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

Tables 3 and 4 summarize the differences between the monthly benefit amounts as reported in SIPP and according to the administrative records (for households with nonzero amounts in both sources) for AFDC and Food Stamps, respectively. The average discrepancies are generally very small and none is statistically significant. These data suggest that, at least for these sample persons and these two income sources, if the fact of receipt (participation) is reported accurately, then the average bias in benefit amount reporting is also very small.

[Tables 3 and 4 here]

6.3. Bias in Reported Transitions

We have noted above the problems uncovered in SIPP estimates of month-to-month changes in program participation; specifically, the accumulation of apparent changes between pairs of months representing the "seam" between two interview waves. In this section we use the matched SIPP and administrative record data to investigate this issue in more detail. Although most of the attention to the "seam bias" problem has focused on changes in participation status, we also examine response bias for seam/nonseam changes in benefit amounts received.

The unit of analysis for the investigation of transitions is the longitudinal household--a household which continues unchanged from one month to the next. A household continues unchanged over time if both (a) and (b) are true:

(a) The household has the same householder/head and (if applicable) the same spouse in consecutive months, and either the householder or the spouse is a Wave 1 sample person.

(b) The household is of the same type in consecutive months. The five types of households are: (i) married couple household; (ii) other family household, male head; (iii) other family household, female head; (iv) nonfamily household, male head; and (v) nonfamily household, female head.

A participation transition occurs when a household changes its participation status from one month to the next; similarly, an amount transition occurs when a household receives different benefit amounts in consecutive months. Participation transitions are treated equally, regardless of their direction (yes-to-no or no-to-yes).

6.3.1. Participation transitions

Table 5 summarizes month-to-month changes in AFDC participation status according to SIPP and the administrative records. By themselves, the SIPP results in this case do not suggest a seam bias problem--although the effect is in the expected direction, the rate of reported transitions at the seam is not significantly different than the average nonseam transition

Table 3: AFDC Benefit Amounts According to SIPP and Administrative Records for Households With Nonzero Amounts in Both SIPP and Administrative Records

	Mean Benefit Amount: SIPP	Record (a)	Mean Discrepancy (b)	Std. Error of Mean	t*	Percent Bias 100x(b/a)
Wave-Month (N):						
1-4 (25)	$537.08	533.53	3.55	7.00	0.51	0.7%
1-3 (26)	518.62	518.92	-0.31	5.19	-0.06	-0.1%
1-2 (25)	531.40	519.93	11.47	11.42	1.00	2.2%
1-1 (25)	516.24	507.69	8.55	9.86	0.87	1.7%
2-4 (23)	570.13	556.54	13.59	12.94	1.05	2.4%
2-3 (25)	545.00	548.74	-3.74	21.97	-0.17	-0.7%
2-2 (26)	527.58	508.16	19.41	15.57	1.25	3.8%
2-1 (26)	530.73	524.43	6.30	13.23	0.48	1.2%

*This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

Table 4: Food Stamp Benefit Amounts According to SIPP and Administrative Records for Households With Nonzero Amounts in Both SIPP and Administrative Records

Wave-Month (N):	Mean Benefit Amount: SIPP	Record (a)	Mean Discrepancy (b)	Std. Error of Mean	t*	Percent Bias 100x(b/a)
1-4 (30)	$102.00	112.40	-10.40	5.46	-1.90	-9.3%
1-3 (32)	112.97	99.14	13.83	10.27	1.35	14.0%
1-2 (29)	102.48	104.40	-1.92	4.00	-0.48	-1.8%
1-1 (28)	99.39	102.03	-2.63	3.78	-0.70	-2.6%
2-4 (24)	100.92	99.07	1.84	8.54	0.22	1.9%
2-3 (21)	94.19	104.33	-10.14	6.03	-1.68	-9.7%
2-2 (23)	98.48	97.96	0.52	4.32	0.12	0.5%
2-1 (25)	89.84	84.57	5.27	5.58	0.94	6.2%

*This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

rate. The administrative records suggest that the average bias in seam transitions in SIPP is zero, but that nonseam transitions tend to be underreported (although this tendency is statistically significant for only one of the nonseam estimates). The AFDC results, then, suggest that if there is a transition bias at all, the problem is not overreporting at the seam but underreporting of nonseam transitions.

[Table 5 here]

Table 6 presents the results for Food Stamps participation status changes. Here the seam bias effect in SIPP is quite clear. The administrative records suggest a general tendency for nonseam transitions to be underreported on average (although none of the individual estimates is statistically significant), but in this case the SIPP estimate of transitions at the seam clearly exceeds the administrative record figure. Thus, the Food Stamps participation seam bias effect in SIPP is both relative and absolute--seam transitions are overabundant relative to nonseam transitions, and the difference is due to net overreporting at the seam and perhaps net underreporting elsewhere.

[Table 6 here]

6.3.2. Amount transitions

The results we report next concern transitions in reported monthly benefit amounts. Table 7 summarizes month-to-month changes in AFDC amounts (for households which had nonzero amounts in both months) according to SIPP and the administrative records. The SIPP data show a significant seam bias in amount transitions. According to the administrative records there are too many SIPP amount transitions at the seam, and too few in nonseam month pairs (the latter effect is consistent in sign for all nonseam month pairs but statistically significant in only one instance).

[Table 7 here]

The Food Stamps results, summarized in Table 8, present a very similar picture. The seam bias effect in SIPP is very clear--changes in Food Stamp benefit amounts occur significantly more often at the seam than between other pairs of months. And, once again, the administrative records suggest that this effect is a combination of too many SIPP amount transitions at the seam, and too few in nonseam month pairs. In this case, the underreporting of nonseam amount changes is statistically significant for three of the six nonseam month pairs.

[Table 8 here]

Table 5: Month-to-Month AFDC Participation Transitions According to SIPP and Administrative Records

Wave-Month Pair:	Transition Rate: SIPP	Record	N*	Mean Discrepancy	Std. Error of Mean	t**
Wave 1 Nonseam:						
1-4 to 1-3	.0019	.0132	529	-.0113	.0053	-2.13
1-3 to 1-2	.0057	.0076	529	-.0019	.0050	-0.38
1-2 to 1-1	0	.0038	529	-.0038	.0027	-1.41
Seam:						
1-1 to 2-4	.0078***	.0078	513	0	.0055	0
Wave 2 Nonseam:						
2-4 to 2-3	.0039	.0058	514	-.0019	.0043	-0.44
2-3 to 2-2	.0057	.0019	523	.0038	.0038	1.00
2-2 to 2-1	.0019	.0039	518	-.0019	.0033	-0.58

*The unit of analysis is the longitudinal household--a household which continues unchanged from one month to the next. See text for explanation.

**This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

***A paired comparison t-test for the SIPP seam transition rate versus the average nonseam transition rate (for households present in all months of both waves) yields a value of t = 1.26. This test also assumes simple random sampling, and so is subject to the same caveats as previously noted.

Table 6: Month-to-Month Food Stamp Participation Transitions According to SIPP and Administrative Records

Wave-Month Pair:	Transition Rate: SIPP	Record	N*	Mean Discrepancy	Std. Error of Mean	t**
Wave 1 Nonseam:						
1-4 to 1-3	.0057	.0095	529	-.0038	.0053	-0.72
1-3 to 1-2	.0057	.0076	529	-.0019	.0050	-0.38
1-2 to 1-1	.0019	.0019	529	0	.0027	0
Seam:						
1-1 to 2-4	.0195***	.0058	513	.0136	.0070	1.94
Wave 2 Nonseam:						
2-4 to 2-3	.0078	.0039	514	.0039	.0048	0.81
2-3 to 2-2	.0096	.0191	523	-.0096	.0074	-1.30
2-2 to 2-1	.0077	.0135	518	-.0058	.0064	-0.91

*The unit of analysis is the longitudinal household--a household which continues unchanged from one month to the next. See text for explanation.

**This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

***A paired comparison t-test for the SIPP seam transition rate versus the average nonseam transition rate (for households present in all months of both waves) yields a value of t = 2.23. This test also assumes simple random sampling, and so is subject to the same caveats as previously noted.

Table 7: Month-to-Month AFDC Benefit Amount Transitions According to SIPP and Administrative Records for Households With Nonzero Amounts in Both Months

	Transition Rate: SIPP	(N)*	Record	(N)*	Mean Discrepancy	Std. Error of Mean	t**
Wave-Month Pair:							
Wave 1 Nonseam:							
1-4 to 1-3	.0741	(27)	.3600	(25)	-.2859	.1144	-2.50
1-3 to 1-2	.1538	(26)	.2143	(28)	-.0604	.1055	-0.57
1-2 to 1-1	.2222	(27)	.2500	(28)	-.0278	.1145	-0.24
Seam:							
1-1 to 2-4	.3636***	(22)	.1250	(24)	.2386	.1268	1.88
Wave 2 Nonseam:							
2-4 to 2-3	.1250	(24)	.2500	(24)	-.1250	.1127	-1.11
2-3 to 2-2	.1600	(25)	.1923	(26)	-.0323	.1067	-0.30
2-2 to 2-1	.0385	(26)	.1200	(25)	-.0815	.0757	-1.08

*The unit of analysis is the longitudinal household--a household which continues unchanged from one month to the next. See text for explanation.

**This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

***A paired comparison t-test for the SIPP seam transition rate versus the average nonseam transition rate (for households present in all months of both waves) yields a value of t = 2.35. This test also assumes simple random sampling, and so is subject to the same caveats as previously noted.

Table 8: Month-to-Month Food Stamp Benefit Amount Transitions According to SIPP and Administrative Records for Households With Nonzero Amounts in Both Months

	Transition Rate: SIPP	(N)*	Record	(N)*	Mean Discrepancy	Std. Error of Mean	t**
Wave-Month Pair:							
Wave 1 Nonseam:							
1-4 to 1-3	.2667	(30)	.6000	(35)	-.3333	.1233	-2.70
1-3 to 1-2	.2414	(29)	.5152	(33)	-.2738	.1234	-2.22
1-2 to 1-1	.2857	(28)	.3750	(32)	-.0893	.1217	-0.73
Seam:							
1-1 to 2-4	.7500***	(20)	.4138	(29)	.3362	.1434	2.34
Wave 2 Nonseam:							
2-4 to 2-3	.0455	(22)	.4333	(30)	-.3879	.1198	-3.24
2-3 to 2-2	.1905	(21)	.4167	(24)	-.2262	.1374	-1.65
2-2 to 2-1	.3478	(23)	.3478	(23)	0	.1404	0

*The unit of analysis is the longitudinal household--a household which continues unchanged from one month to the next. See text for explanation.

**This estimate of t assumes simple random sampling; the assumption is not true, and the standard errors may be underestimated. Conclusions that differences are statistically significant might change if the sample design effects were taken into account.

***A paired comparison t-test for the SIPP seam transition rate versus the average nonseam transition rate (for households present in all months of both waves) yields a value of t = 4.30. This test also assumes simple random sampling, and so is subject to the same caveats as previously noted.

7. CONCLUSIONS

Administrative record checks are a promising way of evaluating the quality of survey interview responses and of survey estimates. Yet there are a number of pitfalls to be avoided and many areas of uncertainty about basic analytical approaches. We have outlined here some principal considerations in designing and conducting record checks, and we have described the SIPP Record Check Study and how it attempts to avoid the known pitfalls of this technique. We have suggested some of the important survey design issues that this record check will address. And we have presented some initial findings concerning the quality of SIPP's basic estimates--program participation rates, benefit amounts, and their month-to-month change.

The limitations of our initial sample--only two government transfer programs in only one state--preclude drawing general conclusions about SIPP nonsampling errors. However, the results we have described have important implications for how we will address later descriptive modeling and hypothesis testing on the complete sample of people and programs. For example, the size and sign of the participation reporting bias may differ markedly from program to program. Thus, we cannot automatically combine data across programs; we will instead attempt to model the characteristics of programs that are reported well and poorly. If such generalizations are possible, they will suggest where to direct attention on redesigning the SIPP questionnaire, and they may benefit other household economic surveys as well.

Learning how the basic response errors produce the seam effect is also a substantial challenge. With so few cases, we were able to detect a SIPP participation seam bias for only one of the two programs studied--Food Stamps. The administrative record data for this program suggest both a net overreporting of changes at the seam and a net underreporting of true changes that occurred at other times. Modeling that pattern of biases will be difficult, and we suspect that our models will encompass more than the simple average bias parameter mentioned here. Our ultimate goal is to formulate measurement models with several error parameters that adequately describe both the monthly discrepancies (mean and variance) and the month-to-month reporting errors. We expect to include more than one response bias parameter and we are considering how to include error variance parameters representing within and between interviewer error variances. Our current thinking is to model the error variance such that it is highly correlated across months within an interview (wave), and much less so between interviews.

Finally, we need to expand our treatment of errors in reporting benefit amounts and changes in amounts. The results presented here, conditional on correct reporting of participation, suggest that there are only small average errors in the reporting of amounts, but that there may be substantial bias in the reporting of amount changes. This suggests that we may eventually adopt a two-part model of response errors, the first part describing errors in participation reporting and the second describing errors in attributes, conditional on the participation reporting error.

We know that the descriptive modeling phase of the analysis of these data represents a formidable challenge. We are confident, however, that the result of this effort will be a thorough evaluation of SIPP response errors which will be of value to those responsible for assuring the quality of SIPP estimates, to users of SIPP data, and to the survey research community in general.

References:

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

FEATHER, J. (1972).
A Response Record Discrepancy Study. University of Saskatchewan, Saskatoon.

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

FULLER, W. and TIN C. C. (1986).
"Response Error Models for Changes in Multinomial Variables," Proceedings of the Second Annual Research Conference, U.S. Bureau of the Census, Washington, DC. pp. 425-441.

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

JARO, M. (1985).
Current Record Linkage Research. Presentation to the Census Advisory Committee of the American Statistical Association, April 25, 1985.

JORESKOG, K. and D. SORBOM (1984).
LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods. University of Uppsala, Uppsala, Sweden.

LAPLANT, W. (1987).
Maintenance Manual for the Generalized Record Linkage Program Generator (GENLINK) SRD Program Generator System. Statistical Research Division Internal Working Paper, U.S. Bureau of the Census, Washington, DC.

LUENBERGER, D. (1984).
Linear and Nonlinear Programming. Reading, MA: Addison Wesley.

MARQUIS, K. (1978).
Record Check Validity of Survey Responses: A Reassessment of Bias in Reports of Hospitalizations. The Rand Corporation, Santa Monica, CA R-2319-HEW.

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

MOORE, J. and D. KASPRZYK (1984).
Month-to-Month Recipiency Turnover in the ISDP. Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, DC.

NETER, J., S. MAYNES, and R. RAMANATHAN (1965).
"The Effect of Mismatching on the Measurement of Response Errors," Journal of the American Statistical Association, Vol. 60, pp. 1005-1027.

NETER, J. and J. WAKSBERG (1966).
"A Study of Response Errors in Expenditures Data from Household Interviews," Journal of the American Statistical Association, Vol. 59, pp. 18-55.

SATER, D. (1986).
SSN Response Rates and Results of SSN Validation/Improvement Operation. U.S. Bureau of the Census memorandum for R. Herriot, March 11, 1986.

STASNY, E. and S. FIENBERG (1985).
"Some Stochastic Models for Estimating Gross Flows in the Presence of Nonrandom Nonresponse," Proceedings of the Conference on Gross Flows in the Labor Force Statistics, Department of Commerce and Department of Labor, Washington, DC., pp. 25-39.