Income Measurement Error in Surveys: A Review

Jeffrey C. Moore[1], Linda L. Stinson[2], and Edward J. Welniak, Jr.[3]

Because income data are germane to a wide array of important policy issues, income questions are almost ubiquitous in government-sponsored surveys. We review research on the quality of survey measures of income, with a particular focus on U.S. government surveys. We briefly examine two of the more typical quality indicators — "benchmark" comparisons of survey estimates to independent aggregate estimates, and nonresponse — but focus our attention primarily on response error research which compares individual survey respondents' reports to external measures of truth, often obtained from independent record systems. The latter investigation reveals a wide range of error properties across income characteristics (sources, amounts received) and income types, which includes high levels of both random error and bias in some instances. We also examine the recent findings of "cognitive" research into respondents' understanding of the meaning of income questions, their interpretations of the tasks which income questions present, their motivations, etc., and attempt to link what we know about income reporting errors to these cognitive processes.

Key words: Data quality; response error; response bias; random error; income source reports; income amount reports, cognitive research.

1. Introduction and Overview

Of all the statistics gathered in government-sponsored surveys, perhaps none is more ubiquitous than income, or more universally germane to a wide array of important policy issues. This paper reviews research evidence on the quality of survey measures of income, with a particular focus on survey programs of the U.S. government. Its primary intent is to assess what is known about the magnitude and nature of the errors made by individual respondents in providing survey reports of income, as well as what is known about the cognitive bases of those errors.

To assess measurement quality we examine research results of several different types. We start with brief examinations of "benchmark" comparisons of survey estimates to independent estimates, and nonresponse analyses, including both nonresponse to income surveys and to income questions specifically. Primarily, however, we focus on comparisons of individual survey

[1] Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233, U.S.A.

[2] Office of Survey Methods Research, U.S. Bureau of Labor Statistics, Washington, DC 20212, U.S.A.

[3] Housing and Housesold Economic Statistics Division, U.S. Bureau of the Census, Washington, DC 20233, U.S.A.

Acknowledgments: We are grateful to Martin David, Kent Marquis, and Betsy Martin for their useful comments and suggestions. An early draft of this paper was made available as a background document to participants in the "Cognitive Aspects of Survey Methodology II" conference held in Charlottesville, VA, in June 1997. A highly condensed version of the paper will appear in the proceedings monograph from the conference — Sirken, M., D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds.), Cognition and Survey Research, New York, John Wiley and Sons, (forthcoming). The opinions expressed herein are the authors' and do not necessarily represent the official views or positions of either the Bureau of the Census or the Bureau of Labor Statistics.

respondents' responses to external measures of truth, often obtained from independent record systems. We then attempt to link these measurement quality results to the results of recent "cognitive" research into respondents' understanding of the meaning of income questions, their interpretations of the tasks which income questions present, their motivations, etc.

The comparison of nationally-weighted survey estimates to benchmark aggregates in Section 2 shows consistent shortfalls in the survey estimates. The size of these shortfalls relative to their benchmarks varies substantially across different types of income, from quite small (e.g., 5-8% for wage and salary income and for social security payments), to quite large (e.g., 50% or more for interest and dividend income), but the consistency with which the survey estimates fall below the benchmarks is striking. One natural conclusion from this research is: survey respondents underreport income. However, our review of the extent of nonresponse to income surveys and income questions in section 3 suggests that other factors besides underreporting may play a role in the survey underestimates. The response error research summarized in sections 4 (income sources) and 5 (income amounts) suggests that income amount underreport errors do tend to predominate over overreporting errors, but that even more important factors affecting the quality of income survey data are the underreporting of income sources and often extensive random error in income reports. Finally, our review of investigations of the cognitive aspects of providing income data in surveys (section 6) finds many possible contributors to inaccurate reporting, including definitional issues, recall and salience problems, and confusion, as well as the much-suspected issue of data sensitivity.

2. Aggregate Income Estimates

In its role as producer of the nation's "official statistics," the Census Bureau has, over the years, examined the quality of its income estimates through comparisons to independent estimates derived from independent, outside sources — e.g., the National Income and Products Accounts (NIPA), individual income tax data, Social Security Administration records, caseload statistics from agencies that administer various transfer programs, etc. Table 1, derived from Coder and Scoon-Rogers (1995), summarizes the results of recent work comparing survey-based estimates and independent benchmarks for an extensive set of income types. The Census Bureau's two major income surveys, the Survey of Income and Program Participation (SIPP) and the Current Population Survey's (CPS) March Income Supplement, supply the survey estimates. Two conclusions are immediately obvious from Table 1. First, a primary goal of SIPP was to provide more complete income data than CPS. The fact that SIPP's estimates are generally closer to the benchmarks than CPS's suggests that SIPP has had some success in meeting that goal — especially, perhaps, for transfer program income. Second, however, and even more pronounced, is the consistency with which the survey estimates fall short of the benchmarks — across surveys, across time, and across income categories.

It is tempting to conclude from these consistent underestimates an underlying tendency for survey respondents to underreport their income. We urge caution in drawing this conclusion from

these results. As Marquis et al. (1981) and others have pointed out, there are multiple competing explanations for the failure of survey estimates to match benchmark totals, many of which have nothing to do with errors of response to survey questions about income amounts. Even the most carefully constructed aggregate income comparisons are rarely, if ever, straightforward. Data from independent sources are almost never completely comparable to the survey data — due to sampling frame differences, timing differences, definitional differences, etc. — and the adjustments necessary to make them comparable are often inadequate. The flawed adjustments, and the fact that the independent estimates themselves are subject to various errors and omissions, adds some measure of uncertainty to any comparison of survey and benchmark estimates. Thus, while the benchmark comparisons are indicative of a problem, without more information the exact nature of that problem remains uncertain, as does the path toward effective survey design solutions — as does even the importance of finding such solutions (if, for example, the problem lies in the construction of the benchmarks).

3. Income Nonresponse

One important competing explanation for the lack of alignment of survey and benchmark estimates, with definite implications for survey design, is nonresponse. To the extent that nonresponse is both frequent and non-randomly distributed across a survey sample, survey estimates are more prone to inaccuracy, even if respondents' reports are highly accurate. In this section we examine the extent of nonresponse to income questions and find it to be severe across almost all income types.

Surveys conducted under the aegis of the Federal government typically achieve much higher levels of cooperation than non-government surveys (e.g., Heberlein and Baumgartner 1978; Goyder 1987; Bradburn and Sudman 1989). Even with this built-in advantage, however, income questions in U.S. government surveys are particularly prone to high rates of item nonresponse. Table 2 shows results from the 1996 CPS March Income Supplement in the form of imputation rates for most of the income categories whose benchmark results were described in the previous section[4]. As the table amply indicates, even in the best of circumstances about one in five survey "reports" of income is produced not by the intended response process but rather by the process used to fill the holes left by nonresponse, and typically the rate is closer to (or exceeds) one in four. These rates, as large as they are, still do not account for all nonresponse to the income questions in the CPS, since they do not include CPS's typical initial "household nonresponse" rate of around 7-8%. Clearly, the magnitude of the nonresponse to income questions is more than sufficient to wreak considerable mischief with the alignment of survey and benchmark estimates. (See Horowitz and Manski (1995) for a discussion of the impact of

[4] An item's value is "imputed" during processing to fill in missing data resulting primarily from the failure of the interview to produce a valid entry. The vast majority of imputations result from blank, "don't know," or "refused" entries. Imputation values are derived from respondent "donors" who share certain characteristics with the sample person whose information is missing.

household and item nonresponse on survey estimates — in particular, survey estimates of household income in the U.S.)

4. Response Error in Income Source Reports

In this section we begin to examine research evidence on the extent and nature of income *response* (or *measurement*) errors, by which we mean discrepancies between the objective truth concerning a respondent's income and his or her report about that income. It is generally useful to distinguish two facets of response error which can never be attributed with certainty to any individual survey response, but which only emerge with clarity over sufficiently large numbers of responses. These are: first, bias, which affects the size and sign of the discrepancy between the means of the reported and true values for a given set of responses; and second, random error, which refers, roughly, to the distribution of errors around their average value[5]. The two error forms can operate quite independently — low bias and low random error is the goal of every survey, but it is certainly possible for a highly biased response process to display very little random error (e.g., if every response to some income question were to elicit a response that understated the truth by some substantial and comparable amount), just as a completely unbiased process can mask substantial error (e.g., if the overreports and underreports largely cancel each other out). We draw this bias/random error distinction, and highlight each where appropriate in the research we review in this section, because of the likelihood that they represent different response mechanisms, possibly calling for different survey design solutions.

Income reporting in surveys is generally a two-stage process involving first the reporting of income sources, and then the reporting of amounts received from those sources. Response errors can occur at either stage. An entire source of income can be misreported, leading to either the respondent's failure to report true income or his or her "false positive" reporting of income not actually received. Or, the source of income may be reported correctly but the amount received from that source can be misreported. Again, we draw this distinction because of the likelihood that different cognitive mechanisms underlie these different types of error, implying the need for different survey design solutions.

Section 4 examines research evidence concerning response error in respondents' reports of income sources. We separate three major categories of income sources — wage and salary income from employment, participation in government transfer programs designed to provide assistance to the poor, and income-producing assets. We find a general tendency toward underreporting of income sources whose magnitude is highly variable across different income types — underreporting bias seems to affect wage and salary income source reports only very modestly, reports of transfer income sources somewhat more so, and perhaps asset income

[5] More formal, statistical presentations of survey measurement error issues abound — see, for example, Groves (1989). We adopt a much less technical approach here, due in part to the exceptionally wide range of types of indicators of these basic error characteristics in the research literature.

sources most of all. Indicators of random error propensities seem to follow this same trend, although the data are quite scant.

4.1 Wage and salary income sources

This section focuses on the quality of respondents' reports of the receipt of wage or salary income paid by an employer. Wages and salaries comprise the single most important source of personal income in the U.S. and the overwhelming majority of labor income[6].

Checking the accuracy of non-reported wage or salary income sources is simply not feasible, so a direct and thorough examination of response quality for reports of specific work-related income sources is impossible. As a result, research which attempts to address errors in respondents' reports of this income source has fallen back on proxy strategies, the most common of which consists of an assessment of the quality of survey reports of the receipt of <u>any</u> wage/salary income. The most common approach to evaluating measurement errors in wage/salary income source reports involves the use of income tax records. An argument can be made that tax return information is simply a second, independent, survey-like report, and thus more useful for assessing response reliability than response error. Other limitations of tax reports in assessing survey reporting accuracy have to do with: (a) the fact that only incomes above a certain minimum must be reported; (b) the substantial underground work economy that operates outside of the tax system; and (c) the tendency of married couples to file joint returns, making the attachment of job income to particular individuals problematic. However, the legal sanctions associated with the accuracy of income tax reporting, and the official records of wages and salary received that must accompany those reports, make tax reports a reasonable tool for estimating survey response errors, at least among particular segments of the population.

The several studies which have used tax reports as criterion measures have consistently found quite modest negative biases in survey reports of the receipt of any wage/salary income. Miller and Paley (1958) offer perhaps the earliest example of this type of research, in a comparison of a subsample of 1950 census post-enumeration survey income reports (for those reporting "taxable" income) against matched tax forms filed with the IRS. Miller and Paley report that, among the approximately 4,000 matched cases for analysis, the census-derived rate of receipt of wage/salary income falls short of the tax-form derived rate by approximately 2-6% (derived from Table 10, p. 196; the uncertainty in the net bias estimate derives from varying assumptions concerning the behavior of nonrespondents). Miller and Paley do not provide sufficient data to derive the separate overreporting and underreporting components of this "net bias" estimate. Herriot and Spiers (1980) find a similar level and direction of error in a

[6] In 1995, wage/salary work for employers accounted for over three-fourths of all personal income in the United States, and for approximately 95% of all dollars earned via work (versus 5% for self-employment). (Source: U.S. Bureau of the Census, Current Population Reports, P60-193, "Money Income in the United States: 1995 (With Separate Data on Valuation of Noncash Benefits)," U.S. Government Printing Office, Washington, DC, 1996).

comparison of CPS reports of the receipt of wage/salary income matched to tax records. They find a very low (4-6%) rate of inaccurate reporting, leading to a modest net underreporting bias of about -1% among married couples who file joint returns (which increases to about -4% among all tax filers).

Other studies confirm that errors of any kind are quite rare, but suggest an even lower net bias in survey reports of any wage/salary income. An evaluation of married couple, joint filer tax returns matched to 1960 census income reports (U.S. Bureau of the Census 1970) finds very little error and virtually zero net bias in the census reports of receipt of any wage/salary income. About 95% of the approximately 34,000 matched census/tax cases included in the study are in agreement concerning the receipt of wage/salary income. The data show almost equal numbers of underreporters and overreporters, yielding a census rate of wage/salary income receipt only about one-tenth of a percentage point below the rate reported on the tax forms (derived from Table 2, page 9). Coder's (1992) match to tax forms, using a very restricted set of SIPP respondents (e.g., married couples, each with valid social security numbers, matched to married-joint tax returns, with non-zero wage/salary income reported either in SIPP or on the tax return), also finds a very small "net bias" in SIPP's wage/salary income source reports. Again, about 95% of the approximately 5700 cases in Coder's study are in agreement on the receipt of wage/salary income, and the net bias is trivially negative, with the SIPP wage/salary receipt rate about 0.4 percentage points below the tax form rate. A Canadian tax study (Grondin and Michaud 1994) suggests somewhat more underreporting of the receipt of any wage/salary income — in one instance the survey-reported rate was about 8 percentage points below the rate reported on tax returns; in a second the difference was only about 2 percentage points.

Kaluzny's (1978) comparison of the survey responses of participants in a Gary, IN income maintenance experiment to the wage and salary reports of their employers presents a very different picture of the quality of wage/salary source reporting. Kaluzny's results suggest substantial underreporting of wage/salary income sources. The proportion of those with true wage/salary earnings in a calendar quarter who failed to report any earnings in the survey range from a minimum of 10-15% among male household heads, to 35-45% among female heads, to as much as 40-60% among nonheads. (Note that these are rates of underreporting among "true positive" cases, which are only suggestive of the extent to which study participants' income reports are in error, and are not necessarily indicative of the magnitude — or even the sign — of the net reporting bias.) It should be noted that Kaluzny's analysis sample suffers from major limitations (e.g., all participants were poor; several major employment categories are excluded from the record system, representing about 1/3 of all wage earners in the experiment in any quarter; analysis is limited to respondents with both positive state-record earnings and complete interview reports; etc.), and some commentators (e.g., Greenberg and Halsey 1983) have expressed concern about the level of care with which the Gary experiment was administered, leading them to question the validity of the study's results.

The research summarized above concerns the quality of survey reports of wages and salary as an income category. Moore, Marquis, and Bogen (1996) report the only known study

which directly assesses the quality of survey reports of receipt of wage/salary income from a specific employer. Moore et al. tested a revised SIPP interview among a sample of known employees of a single employer. They report that known employees of the target employer failed to mention the employer as a source of income in 4-11% (depending on the interviewing treatment) of the months in which they were known to have received such income. (Note that the design of this study only permitted an assessment of underreporting errors.) Virtually all job holders reported the presence of their job with the relevant employer — the observed underreporting was mostly (totally, in the control treatment of the experiment) a result of failure to report all months of employment in an interview which covered four months.

4.2 Transfer program income sources

Income from government transfer programs also constitutes a substantial proportion of total personal income in the United States, particularly, of course, among the poor[7]. This section focuses on the quality of respondents' reports of their participation in transfer programs. A common assumption is that social desirability pressures will create a general tendency toward underreporting participation in these programs. We are in accord with Marquis et al. (1981) in suspecting that this assumption derives in part from a misinterpretation of the results of flawed research, which tends to focus on the survey behavior of known program participants, and thus is biased toward uncovering only underreport errors. We include such "partial design" studies in our review because they do offer some evidence concerning the frequency with which respondents make errors. In the end, however, after consideration of the results of research which allows the full array of errors to emerge, we find evidence for only rather modest net underreporting bias for most transfer program income sources.

Marquis et al. (1981) review several early studies which assess the accuracy of the survey reports of known program participants. These studies suggest generally modest levels of underreporting error, ranging from only about 2% of known participants failing to report their receipt of General Assistance income (Weiss 1969), to a 7% underreporting rate among a similar population (David 1962), to a 13% failure rate for Supplemental Security Income (SSI) recipients (Vaughan 1978).

Other studies among known program participants suggest that program participation may be somewhat more frequently underreported than the Marquis et al. review indicates. For example, Livingston (1969) compares reports gathered in a 1968 special census of Dane County, WI, regarding receipt of "public assistance" (which could include any of several income types: Aid to Families with Dependent Children (AFDC), old age assistance, general assistance, aid to the blind/disabled, etc.) with local administrative records. He finds that 22% of known recipients

[7] In 1995, government transfer programs paid over 460 billion dollars to individuals and families in the United States, amounting to about 10% of total money income. (Source: U.S. Bureau of the Census, Current Population Reports, P60-193, "Money Income in the United States: 1995 (With Separate Data on Valuation of Noncash Benefits)," U.S. Government Printing Office, Washington, DC, 1996).

failed to report any public assistance income in the census. Similarly, in an examination of the survey reports of approximately 200 respondents selected from families who were known recipients of cash assistance or medical assistance for children, Hu (1971) finds that 27% of the informants failed to report their assistance receipt. Somewhat more recent work by Vaughan and colleagues among small samples of known recipients of AFDC has found underreport levels of 14% (Klein and Vaughan 1980) and 9% (a composite finding which masks the considerable variation among the four states included in the study — Goodreau, Oberheu, and Vaughan 1984).

The above studies offer clear evidence that survey reporting of transfer program participation is far from perfect, but do not support any conclusions concerning respondents' general error tendencies — the overall frequency of errors in program participation reports, for example, and the magnitude and direction of the resulting net response bias. This more complete evidence on the nature of the distribution of response errors requires what Marquis et al. (1981) term "complete design" studies, which permit an assessment of the accuracy of both positive and negative survey reports. Marquis et al. review the results of two such studies. Both rely on small samples of low income families residing in specific locations, so their results are not necessarily generalizable to any larger population. Nevertheless, they certainly "challenge the conventional wisdom that transfer participation is underreported in sample surveys" (Marquis et al. 1981, p. 22). Of the four comparisons generated by the two studies, only one suggests a net underreporting bias, and that is very small — Oberheu and Ono (1975) find that survey reports of AFDC participation "last month" fall short of record data by about 2 percentage points. The other three comparisons actually indicate modest net overreporting of the receipt of General Assistance (by about 5 percentage points — Bancroft 1940), yearly AFDC (8 percentage points), and Food Stamps (6 percentage points) (Oberheu and Ono 1975).

As Marquis et al. are careful to point out, small *net* biases do not necessarily indicate that response errors are infrequent, only that the *difference* between the number of overreports and the number of underreports is small. In fact, the small net biases in the studies they review mask substantial amounts of underlying error. The correlation between survey and record values is reasonably high in the Bancroft (1940) study — about +.8 — but quite low in the various survey-record comparisons reported in Oberheu and Ono (1975) — in the +.3 to +.5 range.

More recently, Grondin and Michaud (1994) compare matched Canadian tax return information concerning the receipt of unemployment benefits to reports in two surveys. Grondin and Michaud find a moderate level of error overall (about 12-19% of the survey reports are in error), with underreports substantially exceeding overreports such that the survey-reported rates of participation fall about 10-13 percentage points below the rate derived from the tax returns. (The report is scant on details. The nature of the sample is not clear, nor are any results presented which would permit the assessment of underreporting and overreporting rates, or the resulting relative net bias in the survey estimate.)

Marquis and Moore (1990) report on the results of perhaps the most comprehensive examination of transfer program participation reporting accuracy — a large-scale, complete

design record check study of survey reports from two waves of the 1984 SIPP Panel. They examine reporting of several major government transfer programs (including AFDC, Food Stamps (FOOD), Unemployment Insurance benefits (UNEM), Workers Compensation benefits (WCOMP), "social security" (OASDI) payments, Supplemental Security Income (SSI), and Veterans' Pensions and Compensation (VETS)) in four states (FL, NY, PA, and WI). Marquis and Moore find that, in general, error rates for transfer program participation reports are extremely low — around 1% or less for all programs except OASDI, for which about 2% of respondents' reports are in error. At the same time, underreporting among true program participants is very common for some programs, exceeding 50% for WCOMP, for example. Other programs in which true participants often failed to report their participation were AFDC (49%, on average across the eight months of the survey), UNEM (39%), and FOOD and SSI (both 23%); underreporting was moderate for VETS (17%) and quite low for OASDI (5%).

Except for OASDI, all programs examined by Marquis and Moore show a bias in the direction of underreporting; that is, there are more true participants who fail to report their participation than there are true non-participants who provide false positive reports. Although the absolute differences between the survey and record participation means for all programs are exceedingly small, the low frequency of true participation in most of these programs magnifies some of the small absolute biases into fairly large biases relative to the true participation mean. Marquis and Moore report trivial relative net biases for some programs (e.g., OASDI=+1%; VETS=-3%), modest levels of bias for others (e.g., SSI=-12%; FOOD=-13%), and relative biases in the survey estimates that are quite high for still others (e.g., WCOMP=-18%; UNEM=-20%; AFDC=-39%).

In subsequent research exploring ways to reduce SIPP measurement errors, Moore, Marquis, and Bogen (1996) describe an experiment to test a new SIPP interviewing approach among a Milwaukee, WI sample of known recipients of four programs. Moore et al. report rates of underreporting that are somewhat lower than those of the earlier record check study, but still substantial — true program participants failed to report 10-12% of their true AFDC participation months (depending on the experimental treatment), 12-17% of their FOOD months, 8-13% of their SSI months, and 41-44% of their UNEM months. Moore et al. also find that most underreporting (from about 60% to 90%, depending on the source and the experimental treatment) is due to participants' failure to report the source at all, as opposed to failure to report all months of participation. Also, all programs yield false positive reports among known non-participants, although, because of the much larger base of true non-participants, overreport rates are substantially lower than underreport rates — e.g., 3-4% for AFDC, 2-3% for FOOD, 3% for SSI, and 1% for UNEM. (Note that this experiment used separate samples of people to produce the underreport and overreport estimates, and so does not permit calculation of a net bias estimate combining the underreports and overreports.)

Yen and Nelson (1996) provide the final example of a type of "complete design" measurement error research, in a comparison of survey reports and record data for a sample of Washington state AFDC recipients (at the time of the initial interview) who were interviewed

annually over a 5-year period. They find 93% of the almost 49,000 person-months available for analysis to have been reported correctly. They also find that for this particular sample, overreports exceed underreports, such that the survey-estimated monthly participation rate exceeds the true rate by a little more than 1 percentage point.

4.3 Asset income sources

The third major type of income source that we examine is income deriving from the ownership of assets or property[8]. With a single exception, all of the research reviewed here is of the partial design variety, involving the comparison of survey reports and financial institution records of asset holdings among samples of known asset holders. The near-unanimous conclusion of this research is that those who own such assets very often fail to report them in response to survey questions. As noted before, however, research of this type is of limited use in describing the general nature of survey response errors. The only legitimate conclusion it allows is that surveys will fail to identify many people who own assets; it can address neither the accuracy of positive survey reports of asset ownership, nor the general tendencies of respondents to under- or over-report asset ownership. We also note, however, that the one example of a complete design study does suggest both high levels of random error and a substantial net underreporting bias in the reporting of asset ownership.

Lansing, Ginsburg, and Braaten (1961) set the tone for much later work in their research on response errors in the survey reports of known owners of savings accounts. In two separate small-scale studies conducted in 1958 and 1959, they report that about one-quarter (24% and 26%) of survey respondents failed to report a savings account's existence. These were not insignificant accounts; the authors deliberately restricted the study populations to owners of large (for that time) accounts, with balances of at least $500 in one case and $1000 in the other, to avoid problems associated with recalling what might be considered trivial income details.

Later in the decade, Ferber and colleagues carried out a number of similar studies on asset (mostly savings account) ownership, reaching similar conclusions about the tendency of asset owners to underreport. In all cases, these researchers find underreporting among true asset owners to be extensive — 19%, 35%, and 22% in three savings accounts studies reported by Ferber (1966); 46% in another savings account study reported by Ferber, Forsythe, Guthrie, and Maynes (1969a); and 30% in a study of stock ownership (Ferber, Forsythe, Guthrie, and Maynes 1969b). Such consistent and large effects undoubtedly led Sudman and Bradburn (1974) to conclude that "for savings accounts ... the major source of response effect is failure to report at all" (p. 56).

---

[8] In 1995, personal property income in the United States — predominantly in the form of interest and dividends — produced over 265 billion dollars in income to individuals and families, amounting to about 6% of total money income. (Source: U.S. Bureau of the Census, Current Population Reports, P60-193, "Money Income in the United States: 1995 (With Separate Data on Valuation of Noncash Benefits)," U.S. Government Printing Office, Washington, DC, 1996).

Maynes (1965), however, reports on a similar study whose results seem something of an outlier. Among approximately 3,300 known savings account owners from a particular financial institution in the Netherlands, only about 5% failed to report ownership of the sample account in a financial survey. In another study conducted outside the United States more recently, Grondin and Michaud (1994) report the results of perhaps the only "complete design" research on the accuracy of asset ownership reporting, in a comparison of individual-level annual income reports from two different surveys with Canadian tax return data. Consistent with the majority of the earlier studies, Grondin and Michaud find high levels of error in asset ownership reports (actually, the reports of any income from interest or dividends) — 43% of survey reports are in error in one survey, 27% in the other — and an overwhelming predominance of underreports relative to overreports. In one instance the survey estimate falls short of the tax record estimate by 39 percentage points; in the other the net difference is "only" 15 percentage points.

5. Response Error in Income Amount Reports

The research summarized in section 4 suggests that one dimension of error in survey reports of income is the reporting of income sources. Clearly, the accuracy of income source reporting can exert a major influence on the accuracy of reported income amounts, since both failures to report income sources and false positive reports of non-existent income sources can produce major errors in the reporting of income amounts received. Our review also shows the extent to which different major income categories are differentially subject to errors at this initial "screening" stage of income reporting.

In this section we switch to the second stage of income reporting, and examine response error for reports of income amounts. Once again we distinguish three major categories of income — wage and salary income (with which we also fold in some studies of total income), income from participation in government transfer programs, and income from assets. In order to assess income amount reporting errors unconfounded by errors in the reporting of income sources, we follow customary practice and generally restrict our focus to studies of amount reporting given correctly reported sources. A review such as this is at the mercy of how authors choose to present their results, and thus we are not always able to avoid the confounding of income source and amount reports. Rather than omit such studies from consideration we choose instead to include them but to point out where the original author's analysis strategy admits the possible contaminating influence of source misreports on amount reporting errors.

In their review of response effects, Sudman and Bradburn (1974) conclude that income data are "pretty well reported, although there are some small understatements of income from sources other than wages and salaries" (p. 56). Radner (1982), drawing on his experience with aggregate comparisons, is quick to draw a much more negative inference about respondents' propensities to report accurately: "It is well-known that estimates of the size distribution of annual family money income obtained from household surveys contain substantial error, **particularly response error**" (p. 19, emphasis added). In this section we attempt to reach our

own conclusion about the accuracy of reported income amounts. In the end the evidence we present seems to side more with Sudman and Bradburn's assessment than with Radner's. Wage and salary income response bias estimates from a wide variety of studies are generally small and without consistent sign, and indicators of unreliability (random error) are quite low. Bias estimates for transfer income amount reporting vary in magnitude but are generally negative, indicating underreporting, and random error is also an important problem. Random error is also marked in asset income reports, although indicators of consistent bias are less clear.

5.1 Wage and salary income (and total family income) amounts

Marquis et al. (1981) review a substantial body of research on response error in reports of wage/salary and total income, including record checks (using a wide variety of types of records) to assess both bias and reliability, and reinterviews. These authors find some evidence suggesting that small and irregular earnings sources might be underreported, but the primary and very obvious conclusion of their review is that there is very little net bias in reports of wage/salary income — "The overall picture that emerges ... is that self-reports of wage and salary data are relatively unbiased" (p. 29). Marquis et al. also find very little random measurement error, as indicated by both high consistency in reinterview studies, and, more compelling evidence, high correlations between individual reports and their matched record values. Although there are fewer studies to draw on, particularly with regard to net bias estimates, a similar conclusion applies to reports of total income as well. In this section we review other studies not included in the Marquis et al. report, primarily studies of more recent vintage. In general we find that the additional data continue to support Marquis et al.'s conclusion of very little bias in survey reports of wage/salary income, and little random error as well.

One of the more common means of assessing income reporting errors is through a comparison of survey reports with tax records. In an early study which somehow escaped the Marquis et al. net, Miller and Paley (1958) compare income reporting for a subsample of 1950 census Post Enumeration Survey cases matched to IRS records. Their comparison of the median census/PES reports of total income for both individuals and for families to the medians from matched tax returns suggests a small net bias which is slightly negative for matched families (the census/PES median is about 3% less than the median tax form amount), but slightly positive for matched unrelated individuals (by about 4%).

A more recent study by Coder (1992) assesses errors in SIPP wage/salary income reports through an exact match comparison of combined calendar year 1990 SIPP reports with tax return information. Using a very restricted set of respondents (e.g., married couples, each with a valid social security number, matched to married-joint tax returns, with no missing or imputed SIPP data, and with non-zero wage/salary income on at least one of the two sources), Coder finds a small negative net bias in the SIPP reports in the form of a mean annual wage/salary income in

SIPP that is about 4% less than the mean computed from the matched tax records[9]. In keeping with the earlier results summarized by Marquis et al., Coder also finds very high reliability in the SIPP reports, which correlate about +.83 with the IRS data.

Two studies conducted outside the United States are consistent with U.S. studies' findings of low levels of net bias in survey respondents' earned income reports as compared to tax records. Andersen and Christoffersen (1982 — cited in Körmendi 1988), in an analysis of 1976 interview data compared to Danish tax assessment records, find a slight positive net bias in respondents' reports, which on average exceed tax record information by 2.5%. In her own subsequent study, Körmendi (1988) compares telephone and personal visit interview reports with national register (tax) information, and also finds "a surprisingly high degree of accuracy" (p. 352) in reports of personal gross and net income and "family" gross and net income. Körmendi's correlational analysis generally shows very high correlations — from +.8 to well over +.9 — between the tax and survey reports, and her analysis comparing the means of the survey reports and tax records suggests very small net biases (even the few significant bias estimates are generally less than ±5%) with no consistent sign across many different comparison groups.

Grondin and Michaud's (1994) Canadian study (cited earlier) presents a picture that differs somewhat from the general findings noted above, although their report offers few details which would permit a careful comparison with other research. Grondin and Michaud find fairly high error frequencies in their comparison of annual income data from two different surveys with tax information. Across both studies, about one-third of the interview reports differed from their matched tax reports by more than ±5%. (As with the Coder (1992) study cited earlier, this figure is biased upward to some extent by the inclusion of an unknown number of cases in which the survey and tax reports did not agree on the presence of any wage/salary income.) The report offers no evidence regarding the sign or magnitude of the resulting net bias in the survey reports.

Other researchers have matched survey reports to Social Security Administration (SSA) earnings records. Bound and Krueger (1991) summarize a study matching earnings reports from the CPS to SSA earnings records, which finds essentially zero net bias in CPS income reports (for those whose incomes did not exceed the SSA's earnings maximum cutoff). Thus, with the possible exception of the Grondin and Michaud work, the evidence from comparisons of survey reports with tax return and other official government data supports Marquis et al.'s (1981) earlier conclusion that wage/salary income (and total income) amounts are reported in surveys with little bias and fairly minimal random error.

---

[9] Note that this figure is probably an upper bound on the extent of response error in SIPP's income *amounts* reports, since it includes an unknown number of cases which did not agree on the presence of any wage/salary income. The reported estimates will overstate the difference in reported amounts to the extent that there were more 0-income SIPP cases matched to tax reports with nonzero income, compared to the reverse. (Coder's results suggest but do not state conclusively that this is the case.)

Employers' records provide perhaps the best source of "true" wage and salary information for the detection of errors in respondents' survey reports. Several of the studies reviewed by Marquis et al. (1981) are of this type; here we examine several additional works. Kaluzny (1978) finds substantial negative bias in the survey responses of Gary income maintenance experiment sample cases matched to income reports provided by their employers. His bias estimates vary widely for different types of respondents, but they are consistently negative (indicating survey underreporting), except for male household heads at the lowest income levels who overreported their quarterly earnings by 15-20%. For other categories of respondents, however, net bias estimates range from a low of -13% (for steadily-working female heads of household) to -79% (for non-steadily working nonheads of household).

Other findings suggest that Kaluzny's results may be something of an outlier. Halsey (1978) reports the results of wage/salary amount validations among similar low income populations in similar income maintenance experiments in Seattle and Denver. Halsey shows statistically significant negative biases, but only on the order of 2-4% — "not large enough to be economically important" (p. (IV)46). He also finds very high correlations (of approximately +.90) between the record and report values, indicating high reliability. In a subsequent paper, Greenberg and Halsey (1983) speculate that the stark differences between these results and those of the earlier Gary experiment are in part attributable to a less careful administration of the latter, leading to more apparent error.

Hoaglin (1978) reports on a study which verified annual and monthly income reported by participants in a housing allowance experiment among low-income renter households in Pittsburgh and Phoenix. The verification was accomplished using consent forms to check all reported income with reported employers; cases included in the verification analyses include only those for which all verifiable income was successfully verified. Hoaglin's presentation of the results makes it impossible to derive precise bias and error estimates. It is clear, however, that the net bias in the survey reports is only very slightly negative. Hoaglin reports mean discrepancies in respondents' annual income reports of -$63 in one site and +$37 in the other (these appear to be within ±2% of the true mean in each case). Errors in monthly reports seem slightly larger (perhaps upwards of -10%) and more consistently negative than the annual reports. Mellow and Sider (1983) use a similar methodology in a validation study of a 1977 CPS income supplement, using waivers to contact reported employers and the pay records of successfully contacted employers to assess survey response errors. They find a nonsignificant tendency for respondents to overreport both their annual job income (by about 5%) and their monthly income (by about 10%); weekly and hourly pay rates, however, tend to be nonsignificantly under<u></u>reported (by about 4-6%).

Duncan and Hill (1985), in a validation study among a small sample of employees of a single large manufacturing company, find essentially zero bias in respondents' annual earnings reports (which were designed to mimic a PSID interview). Duncan and Hill look at reports about two years, and find a net bias of -$55 (-0.2%) in one instance, and -$294 (-1%) in the other. However, these net bias results mask what are rather more substantial, but largely compensating,

absolute errors, which average 7% for one year's annual income reports (that is, the average response differed from the true value by 7%, regardless of the direction of the difference), and 9% for the other year's. (These estimates may be too high — for reasons that are not clear, Bound, Brown, Duncan, and Rodgers (1990), describing the same research, report average absolute discrepancies of only about 3%.)

Rodgers, Brown, and Duncan (1993) report the results of a 1987 followup study of Duncan and Hill's respondents, again using a PSID-type interview among the same employees of a single manufacturing company and validating the respondents' replies with the company's payroll records. They again find relatively little evidence of important net bias in the survey reports, especially for annual income, the mean of which differed from the employer record mean by less than 1% (p. 1213 — no indication of the sign of the bias is provided). Other measures show occasional statistically significant discrepancies, but none seem important (no discrepancy exceeds ±5%). Rogers et al. do find somewhat higher levels of random error than prior investigators, as indicated by the relatively low survey-record correlations (of the logarithms) for earnings last pay period (+.60), and for usual pay period earnings (+.46), although the estimate for annual earnings remains quite high (+.79).

The Moore, Marquis, and Bogen (1996) study cited earlier represents something of a middle ground regarding response errors for wage/salary income amount reports. In an experiment testing a new SIPP interview among a sample of known employees of a single (public) employer, Moore et al. find substantial (but far from complete) agreement between survey reports of monthly income and the employer's payroll records — between 65% and 75% of respondents' monthly reports are accurate within ±5% of the record value.

5.2 Transfer program income amounts

The evidence cited is scant, but Marquis et al.'s (1981) review presents a somewhat inconsistent picture concerning response error in survey reports of income from government transfer programs, although underreporting seems to predominate. Two studies of social security income reporting (Haber 1966; Vaughan and Yuskavage 1976) agree in suggesting modest levels of net bias (of approximately 5 or 6%) — but the sign of the bias is negative in one case and positive in the other. David (1962) finds a net bias of about -18% for "general assistance" amounts reports; Oberheu and Ono (1975) find much more severe underreporting (about -30%) of amounts for a later incarnation of a similar transfer program, AFDC income.

In this section we review additional evidence on the reporting of transfer program income, mostly from studies conducted in the intervening two decades since Marquis et al.'s work. We examine two studies which use tax records as criterion measures; the remainder use the payment records of the administering government agencies to assess errors. This more recent work suggests a fairly consistent negative (underreporting) response bias across almost all transfer programs, but with substantial variation in magnitude across programs. Indicators of unreliability,

however, are uniformly high, indicating the presence of much random error, even for income amount types that show little bias.

Michaud and colleagues have conducted several studies comparing survey reports to Canadian tax records. Grondin and Michaud (1994), in their assessment of annual income data from two different surveys, have only one type of transfer program income to assess — unemployment insurance (UI) benefits. Among those with non-zero amounts in either the survey or the tax record, and excluding the approximately 8% of cases with a missing (DK) survey response, they find evidence of fairly substantial error. In each survey, only slightly more than half of the survey reports — 56% in one case, 55% in the other — are in agreement within ±5% of the tax record amount[10]. (Grondin and Michaud present no evidence regarding the direction of errors for disagreement cases, and thus cannot speak to the net bias in the amount reports.) More evidence on net bias in UI reports is presented in Dibbs, Hale, Loverock, and Michaud (1995). In a test of dependent interviewing procedures for improving the reporting of UI and other benefits, these authors find a modest general tendency to underreport annual benefit amounts, by about 5% on average.

The records of the agencies which administer the various transfer programs represent the most direct and most common means for establishing the nature and level of response error in transfer income amount reports. The several studies that are of this type suggest a general tendency for transfer program income to be at least modestly — and in some instances substantially — underreported. Livingston (1969), for example, compares reports regarding the receipt of "public assistance" income (AFDC, old age assistance (OAA), general assistance, aid to the blind or disabled) as reported in a special census in Dane County, WI, with administrative records. Among those who reported public assistance receipt, and whose receipt is corroborated in agency records, amount underreporting is generally about twice as common as overreporting. For all types of assistance income considered together, the net bias in the census reports is -27% (in a comparison of medians). Underreporting appears to have been concentrated in the very poorest cases, however — those with family incomes of under $1000 reported only 14% of the assistance they received (-86% net bias!); at income levels above $2000 underreporting was trivial (for one category the bias is actually positive in sign). Livingston also considers AFDC and OAA separately and finds both to be marked by rather severe underreporting — OAA recipients on average reported only 80% of their OAA income (i.e., net bias = -20%); for AFDC recipients the level of bias is even more extreme, with only 70% of AFDC income reported in the census.

Halsey's (1978) summary of income reporting verifications carried out in conjunction with the Seattle and Denver income maintenance experiments also suggests a substantial underreporting bias for transfer income amounts, as well as substantial random error. Ignoring experimental treatments (which have almost no effect on the validation results), Halsey analyzes

[10] Note that these figures represent an upper bound on the extent of income amount reporting error. The analysis includes some cases which did not agree on the presence of any UI income, so the reported amount differences are confounded with income source reporting errors.

amount reporting among those for whom either the report or the record was non-zero (and therefore confounds the analysis of amount reporting with source reporting errors to some unknown extent). He finds that reports of transfer program income are mostly negatively biased — for AFDC, for example, the average survey report is approximately 25-30% below the average record amount; for UI benefits the net bias is perhaps as much as -50%. The reliability of the survey reports are also quite low, with correlations between record and report values only in the .40-.60 range.

In contrast, Hoaglin's (1978) results in a housing allowance experiment conducted in Pittsburgh and Phoenix present a very different picture, suggesting that net bias in transfer income reports is essentially zero. The surveys subject to verification include both an initial annual income report, and subsequent monthly reports, of which one, taken approximately 18 months after the initial annual survey, is selected for analysis. Hoaglin's presentation does not permit the derivation of precise estimates, since he only presents median error amount data, without the true (verified) median amounts necessary for the calculation of the magnitude of response bias. Nevertheless, the results clearly indicate that the net bias in transfer income amount reports is negligible, regardless of the source or the reporting period: for both annual and monthly Social Security reports the median error in reported amounts is very slightly negative (but by only about 3% of the median reported amount); however, for monthly and annual SSI and "welfare" (which combines AFDC, general assistance, other welfare) reports, and for monthly UI reports (no annual data are available), the median response error amount is $0.

Goodreau, Oberheu, and Vaughan (1984) compare administrative record data with the survey reports (from the Income Survey Development Program, the precursor to SIPP) of a very small sample of approximately 200 families of known AFDC recipients in four states. In keeping with Hoaglin's results, they find fairly high overall accuracy: averaged across all survey months, about 65% of respondents who reported receipt of any AFDC assistance reported an "accurate" (within $5, or about ±2%) amount. There is a modest underreporting tendency overall, with the total amount of survey-reported assistance reaching only 96% of actually received amount[11], and underreporters outnumbering overreporters by about a 2:1 ratio.

Moore, Marquis, and Bogen (1996) also offer some insights into measurement errors in transfer income reports in their description of the results of a field experiment testing a new approach to SIPP interviewing. Their sample of known program participants who correctly reported their participation reported most program amounts with a fairly high level of accuracy. Averaged across eight months of interviewing, about 70 to 80% of AFDC, FOOD, and SSI recipients reported their income amounts within ±5% of truth. UI reports are substantially less

[11] Note that both this figure and the reporting accuracy results may be somewhat biased in the direction of less apparent response error than truly exists. Goodreau et al.'s analysis includes as AFDC income reporters respondents who reported their AFDC income but mislabeled it as some other type of assistance income. One of the primary means of identifying these "misclassifiers" was through the close correspondence of the true amount paid them, according to the records, with a reported amount in some other income category. This process may well have overlooked misclassifiers who reported highly inaccurate amounts.

accurate, with a maximum accuracy rate of only 61%, and most estimates in the 20 to 30% accuracy range.

5.3 Asset income amounts

With a single exception, studies assessing response errors in reports of income from assets use the records of selected financial institutions to identify measurement errors; the exception uses tax records. The studies also focus on the value of reported assets (generally, savings account balances) rather than the amounts of income they produce — again, with the exception of the study which uses tax records. We summarize these studies in this section, and find that in large measure they offer little evidence of a consistent response bias in reported asset values (i.e., the bias estimates are occasionally substantial, but about as likely to be positive in sign as negative), but substantial levels of random error.

Lansing, Ginsburg, and Braaten (1961) report on the results of several studies of response error in survey reports of (among other things) known owners of savings accounts. Two are of relevance here; each is limited by a very small sample of cases. The first study finds a net bias in reported savings account balances of -14% (i.e., the mean reported balance is almost $500 less than the mean actual balance of $3310), with about twice as many underreported amounts as overreported amounts. The net bias, while fairly substantial in this study, masks a much higher average discrepancy, which approaches 50% ($1571) on average. However, Lansing et al.'s second study finds a much more moderate — and in this case positive — net bias of +2%. (No further details are presented to permit an assessment of the preponderance of the sign of the discrepancies, nor the average absolute discrepancy.) As noted earlier, the sampled accounts in both studies had balances that were quite substantial, in order to avoid problems associated with recalling what might be considered trivial income details. Ferber (1966), like Lansing et al., finds widely diverging bias estimates in three studies, ranging from the substantially large and negative (-20%) in one instance, to the trivially positive (+0.3%) in another, to the moderately positive (+8%) in a third.

Maynes (1965) finds a very moderate level of bias in his record check study of known savings account owners from a particular financial institution in the Netherlands. He reports a net bias of approximately -5% in reported account balances; overall, underreporters outnumber overreporters by about a 2:1 ratio (lack of details make precise estimates impossible, but see Table 2). Maynes also finds a "strong tendency of survey respondents to underreport large balances and to overreport small balances" (p. 379), and he finds more error (but approximately equivalent net bias) for reports about 10-months ago account balances as compared to current balance reports. In a subsequent study, Maynes (1968) finds little difference in the net bias in savings account balance reports according to whether respondents were instructed to consult their records or instructed not to consult records — the bias is +1% in one experimental treatment and -2% in another. However, the amount of random error is clearly affected by the use of records: 85% of record users' account balance reports are within ±1% of the truth, versus 49% among "no records" respondents (91% versus 70% are within ±5%).

Ferber, Forsythe, Guthrie, and Maynes (1969a) also find very little net bias in the reports of savings account balances — less than +0.1% — but substantial random error. Only about 40% of reports fall within ±10% of the true amount. The trivial net bias, however, seems to mask more pronounced biases for different sized accounts; the authors find "a pronounced tendency to over-report the size of small accounts and to under-report the size of large accounts" (p. 441). The same group of researchers (Ferber, Forsythe, Guthrie, and Maynes 1969b) also find a trivial net bias (+0.2%) in the reports of stock holdings; in this case the vast majority of reports, approximately 80%, are "fully accurate" (p. 426), that is, reported with no error.

Grondin and Michaud's (1994) comparison of survey reports with tax records is the only known study of response error in the amounts of income from asset holdings — in this case, interest and dividend income. Unfortunately, they only report the frequency with which the survey reports are in agreement (within ±5%) with the tax data — 37% in one survey, 50% in another[12]. This suggests that such reports carry a substantial random error component, but they offer no clues as to the magnitude or direction of any response bias.

6. Cognitive Factors in Income Reporting

In this final section we review research which addresses the possible cognitive bases of respondents' apparent difficulties in reporting all their income sources, and all their amounts of income received from those sources, fully and accurately. In reviewing this evidence we follow the common practice, suggested by Tourangeau (1984) and others, of distinguishing several distinct stages in the cognitive "work" necessary to answer a survey question, which we label here as understanding, retrieval, and response production.

6.1 Understanding income concepts and terms

Deliberate prevarication — in particular, deliberate underreporting — is probably the most commonly assumed cause of income survey response errors. However, close consideration reveals an abundance of areas of potential difficulty without invoking motivated misreporting at all. Consider for a moment the magnitude of the survey designer's task in designing a survey to collect income data. The most difficult step of this process may, in fact, be the first: defining the construct for survey respondents in clear, simple, and easily understood language. The fact that income has so many varied components partially explains the complexity of the task. Not only are there many different forms of remuneration (e.g., pay-in-kind, non-wage cash payments, cash and noncash program benefits, take-home pay and gross income, fringe benefits, assets, etc.) that may or may not be included within the definition, but there are also varying recipients (e.g.,

[12] As noted earlier, these "amount accuracy" figures are biased upward to some extent by the inclusion of an unknown number of source reporting error cases in which the survey and tax reports did not agree on the presence of any interest/dividend income.

households, families, couples, individuals) and receipt periods (e.g., weekly, biweekly, bimonthly, monthly, quarterly, annual) that must be defined and conveyed.

6.1.1 Types of misunderstanding problems

A reasonable assumption, certainly, is that technical definitions will not be widely known by the general public. An example of a definition not widely understood by the population at large is the phrase "nonwage cash payments." Dippo and Norwood (1994) cite Bureau of Labor Statistics research which shows a wide range of interpretations of "nonwage cash payments," from the inclusion of in-kind payments, such as a car, to the exclusion of even cashable checks. Census Bureau research on some of SIPP's less central income definitions finds more support for the lack of definitional consensus. For example, SIPP's definition of income is intended to include both gains and losses. However, even owners of real estate and other income-generating property, for whom losses (both real and on paper) are commonplace, often fail to include common types of losses (e.g., losses due to the failure to rent property, stock fluctuations, and commodity holdings) in their reports of property income (Cantor, Brandt, and Green 1991).

Other research shows, however, that even the most commonplace terms and phrases elicit highly variable understandings among survey respondents. Stinson (1997) asked respondents what they included when calculating their "total family income." Some respondents reported excluding their own or a spouse's part-time wages, or interest income, because the amounts were small and not a significant contributor to the family's income. Other reasons for omitting income from the family total were that the wages were kept for individual use, or handed over to other parts of the extended family, and thus were not available to the immediate family. Likewise, Bogen (1995), in a cognitive assessment of the SIPP interview, finds that people's idiosyncratic definitions of "income" cause them to omit money earned from sporadic self-employment, their third or fourth job, and "odd jobs," such as occasional baby sitting or maintenance work[13]. Other SIPP research has demonstrated confusion over the distinction between income "earned" during a specified period and income "received" during that time (Marquis 1990; Cantor, Brandt, and Green 1991; Bogen and Robinson 1995). This distinction is critical to SIPP, which is intended to identify short-term ebbs and flows of income receipt, program participation, and poverty status.

6.1.2 Survey design solutions to misunderstanding problems

Fowler (1995) offers two possible solutions to the problems presented by respondents' failure to share survey designers' income definitions. Since the technical definitions of income are not intuitively obvious to respondents, they must be clearly conveyed to them. First, survey designers can develop simple and clear definitions for income-relevant terms and include those definitions within the questions. The obvious drawback to this procedure is that, in some cases, the definitions cannot be simple because the concepts to be defined are not simple. Embedding a

[13] Although they do not speculate about a cognitive mechanism, Marquis et al. (1981) reach a similar conclusion. Based on their observation of a modest underreporting bias for annual earnings from all sources, in contrast to an absence of any such bias in reports of earnings from a specified employer, they hypothesize "that earnings from odd jobs and other irregular sources of earnings are underreported" (p. 27).

complicated definition in an otherwise simple question often renders the entire question so complicated as to risk respondents' comprehension of it. Fowler's second suggestion is to begin with simple but largely undefined questions (e.g., "What is your total family income?"), and follow up with specific probes to cover the aspects of the definition that are frequently misunderstood or omitted. This procedure covers all definitional details incrementally, without requiring that respondents understand everything all at once.

6.2 Retrieval problems

Assuming that the survey designer can formulate income questions that effectively convey his or her intended meanings to respondents, the next challenge is for respondents to retrieve the appropriate income information from memory. The retrieval process can introduce error into survey reports in a wide variety of ways. In this section we review a number of studies which demonstrate some of the common ways in which faulty retrieval has been shown to affect the quality of income reports.

6.2.1 Types of retrieval problems

Lack of knowledge is an immediate and common problem; respondents often simply do not carry the necessary information in memory. Nonresponse to income questions is always comprised of a significant proportion of "don't know" responses as well as refusals. In a SIPP investigation of respondents' income reporting strategies, those who did not use records, particularly when reporting asset income, often appeared to observers to almost blindly guess at both asset types and amounts, because the information concerning their assets was simply not stored in memory (Cantor, Brandt, and Green 1991).

Even when knowledge about income is present in memory, the Moore, Marquis, and Bogen (1996) results demonstrate that the low salience of some income types might cause some income sources to be overlooked in a cursory recall effort. Their test of experimental SIPP interview procedures found that an income "free recall" reporting task, without specific probes for specific types of income, was sufficient for eliciting over 90% reports of job income, pensions, and major transfer income sources such as OASDI, SSI, and AFDC — the apparently highly salient, easy-to-retrieve income sources. Other important income sources, such as Food Stamps and asset income sources were much more rarely reported in free recall. In fact, less than 40% of all Food Stamps reports eventually produced in the first interview administration were elicited during "free recall" (this increased to 64% in the second interview, still substantially below job income and other transfer programs). Assets are apparently even less salient — only 22% of interest-earning savings accounts were reported in response to free recall procedures, 23% of interest-earning checking accounts, 34% of stock dividends, 35% of money market account interest, etc. Most of these income sources required specific "recognition" probes before they emerged in respondents' reports, even in a second interview wave.

One demonstrated correlate of transfer income reporting problems which is probably related to salience issues is brief "spells" of receipt, which often leads to recall failure. For

example, Livingston (1969) compares census reports of receipt of "public assistance" with administrative record information and finds that underreporting of receipt is associated with both a small total benefit amount and fewer months of receipt. Hu's (1971) subsequent examination of the survey reports of approximately 200 "welfare family" respondents for whom detailed data were available from local records finds that failure to report is associated with younger household heads and higher reported household income, and, therefore, less time receiving welfare. Similarly, Goodreau, Oberheu, and Vaughan (1984) find that known AFDC recipients are three times more likely to underreport their AFDC participation if they were "part-period" participants versus participants throughout the entire survey reference period.

Error-prone "reconstruction" strategies: When the required income information is not directly retrievable from memory, respondents adopt any of a number of strategies to "reconstruct" that information, some of which may be more successful than others. Cantor, Brandt, and Green (1991) report that very few SIPP respondents who were asked to report their monthly earnings attempt a direct retrieval. Most use one of three main reconstruction strategies: (1) recalling paycheck amounts (and occasionally actually looking at paycheck records) and then calculating the number of paychecks received within the relevant time period; (2) trying to estimate the number of hours worked during the reference period and then multiplying by the hourly wage rate; and (3) dividing an annual salary by twelve. Marquis (1990) describes the same types of reconstruction strategies in the absence of knowledge and records, as well as some additional ones:

- identifying an "average" pay amount and multiplying that by a "typical" pattern of hours worked;
- recalling a paycheck amount, rounding to hundreds, and translating into a monthly estimate;
- identifying an "average" pay amount and then estimating whether more or less than that average amount was received in a given month;
- recalling exact amounts for recent months and then recalling circumstances that would raise or lower monthly amounts for more distant months;
- recalling the annual gross salary and then dividing by twelve for a monthly estimate;
- recalling take-home pay from the last paycheck and then multiplying it by the number of pay periods per month;
- recalling hourly amounts, multiplying by the number of hours per week and the number of weeks per month;
- recalling weekly amounts and multiplying by four.

Bogen (1995) provides additional examples of income amount reporting strategies that result in inexact estimates. Instances of this imprecision include:

- reporting weekly pay, but failing to include a fifth payday in a month with five weeks;
- severe rounding of reported dollar amounts; and
- reporting "average" amounts.

Another frequent pitfall for income reporting accuracy is confusion among income category labels. This can be especially troubling for a detailed accounting of asset income, and many researchers — most notably, Vaughan and his colleagues (Vaughan 1978; Klein and Vaughan 1980; Vaughan, Lininger, and Klein 1983; Goodreau, Oberheu, and Vaughan 1984) — have noted respondents' tendencies to confuse transfer program names. Marquis and Moore (1990), for example, find that a significant proportion of AFDC response errors are actually due to misclassification of AFDC income as "general welfare" income. Other Census Bureau investigations have found that respondents have great difficulty recognizing and distinguishing the official names for various sources of income, such as "Medicaid," "Supplemental Security Income," and "Worker's Compensation" (Cantor, Brandt, and Green 1991; Bogen 1994).

Some evidence suggests that recall period length contributes to faulty recall. In a study comparing respondents' self-reports of income collected as "current income" in one year and as "last year's income" collected through re-interview one year later, Withey (1954) found that only 61% of the cases reported the same income in their current and retrospective reports. On the other hand, Marquis and Moore (1990) compare reports about "last month" with reports about "four months ago" and find evidence of increased forgetting" (underreporting) with the passage of time for only one of the eight programs in their SIPP record check study. Nor do Marquis and Moore's data show consistent increases in error rates for the more distant reference month — for four of the programs there are numerically more errors in the "four months ago" reports than in the "last month" reports, but for three other programs this trend is reversed (and one program shows inconsistent results; see Appendix Table 1 in Marquis and Moore 1990).

6.2.2 Survey design solutions to retrieval problems

Survey methodologists have directed a great deal of attention toward attempts to solve or reduce income survey recall/retrieval problems. For example, one common method for avoiding the inevitable "don't know's" and guesses that result from a respondent's lack of knowledge is to find a memory system which does contain the desired information (i.e., use a more knowledgeable respondent). Moore's (1988) review of self/proxy effects on survey response quality finds that self-response conditions do tend to reduce nonresponse to income questions, although the benefits of self-response on response completeness may be washed out by increases in household and person nonresponse.

A second solution to recall/retrieval problems is to bypass memory retrieval altogether through the use of records. A number of investigations have tried this approach, with some notable successes in reducing response error. For example, in the Maynes (1968) study cited earlier, respondents who were instructed to consult their records reported their savings account balances with much greater accuracy than those instructed not to consult records. Similarly, Grondin and Michaud (1994) find major reductions in income amount reporting errors among respondents who used a pre-mailed notebook, or referred to tax forms during the interview, compared to those who answered the interview questions unaided. About 85% of "assisted" wage/salary income reports were accurate (within ±5% of the true (tax report) value), versus 36% of amounts reported without notebook or tax forms; for UI the comparable figures were about

75% versus 16%; and, for interest/dividend income, about 70% of those who used some form of record were accurate versus 10% of those who did not. Grondin and Michaud also note that these memory aids reduce survey "don't know's" to practically zero. Finally, Moore, Marquis, and Bogen's (1996) experimental SIPP interview procedures achieved notable success in getting respondents to use their income records, which produced significantly improved income amount reports among record users, although both survey nonresponse and interviewing costs may have been adversely affected by the record use procedures.

Another important potential drawback of encouraging record use to assist income reporting is interviewers' lack of enthusiasm for such procedures. Observers have noted that interviewers do not readily encourage respondents to use records, and on occasion overtly discourage record use (Marquis 1990). (Moore, Marquis, and Bogen (1996), however, find that interviewers did quite well administering experimental SIPP procedures involving major emphasis on record use.) Nor are records always helpful. Marquis (1990) reports one instance in which retrieved pay stubs were in sufficient disorder as to render them of almost no use in collecting accurate information; in another case, a respondent was unable to decipher the record information which, it turned out, also failed to cover the appropriate time period. Most respondents in the Marquis study reported that they would have been willing to use their income records had the interviewer requested them. A few, however, cited reasons why they might be unlikely to use records, including:

- difficulty finding the records;
- the "convenience" of simply recalling a close approximation of the desired income amount without the bother of a search for and interpretation of records; and
- the lack of availability of records for some types of income.

Other proposed solutions have to do with modifications to the <u>timing and focus</u> of the income inquiry. For example, Sudman and Bradburn (1982) suggest that the recall of total family income may be more accurate if the specified time period is "the last calendar year," especially if data collection occurs when people are preparing their income tax returns in the spring. Another strategy to address the challenge of accurate recall seeks to avoid the problems caused by fluctuations in hours and wages, in turn caused by such events as holidays, vacations, illness, and overtime, by requiring respondents to remember and report only the most recent pay period (Rodgers, Brown, and Duncan 1993). A related approach would reduce the emphasis upon precise recall and shift toward the integration of estimates into a report of "usual pay." However, the results of research in the Panel Study of Income Dynamics (PSID) by Rodgers, Brown, and Duncan (1993) call into question the proposed solutions. Rodgers et al. report a +.79 correlation between company-recorded yearly earnings and respondent reports — substantially higher than the correlation between records and survey reports of income received in the last pay period (+.49) or the respondent's "usual" pay (+.46).

6.3 Sensitivity

While both misunderstanding of income concepts and memory fallibility play roles in the inaccurate reporting of income, these two factors alone do not completely explain reporting problems. It seems clear that in some cases respondents' sensitivity about discussing their income may also lead to "motivated mis-remembering." This should not be surprising since the taboo against speaking about money is so strong in our society that to even raise the topic is considered rude. A therapist in San Francisco specializing in the "psychology of money" estimates that 90% of her clients report never discussing money in their homes (Gallagher 1992). In a study on attitudes toward money conducted on a non-representative, Washington, DC area convenience sample, van Melis-Wright and Stone (1993) found that the two most frequently endorsed statements were "I think it is impolite to ask others about their financial situation," and "Surveys asking about my finances should be completely anonymous." Clearly, social conventions of politeness and the desire for anonymity may contribute some of the sensitivity respondents feel about reporting their income, and may also create hesitancy on the part of interviewers to even ask the questions.

The difficulty experienced by respondents and interviewers alike when discussing income seems to be borne out by a behavior coding study conducted by the Bureau of the Census (Miller, Bogen, and Moore 1997), which indicated that SIPP interviewers often make major changes to income questions. These changes include omitting definitions, reference periods, and in some cases, entire questions. The mode of the interview may also contribute to the difficulty respondents and interviewers experience when talking about income (Brehm 1993). There is some indication that telephone respondents are more likely to feel uncomfortable discussing income than are respondents to face-to-face interviews (Groves 1989). Jordan, Marcus, and Reeder (1979) find more missing values for family income in telephone surveys than in personal interviews. This may reflect respondents' greater discomfort about discussing income over the telephone than in person — or, on the other hand, the telephone may simply lower some of the social barriers against expressing such discomfort overtly. Laumann, Ganon, Michael and Michaels (1994) report the recent experience of the National Opinion Research Center's landmark survey of sexual practices in the United States. The study was designed so that the most sensitive questions from this exceedingly sensitive interview could be self-administered and not depend upon questioning by interviewers. Among the questions included in the self-administered privacy section was a question about personal family income, since for many of their respondents the income question was the most personal and sensitive question asked.

Singer, Von Thurn, and Miller (1995) have shown that assurances of confidentiality reduce nonresponse and/or increase the accuracy of responses to income questions and other sensitive survey topics. In addition, recent technological advances raise some possibilities for gathering income information under more anonymous conditions, without the direct involvement of an interviewer, which may substantially reduce sensitivity pressures and thus yield more complete reporting. Research is starting to show the benefits of anonymous reporting — for example, via audio computer-assisted self-interviewing (A-CASI) — on the quality of sensitive behavior reports (Turner, Lessler, and Devore 1992; Tourangeau and Smith 1996), although as yet there has been no clear demonstration that income reports would similarly benefit.

7. Summary and Conclusions

Our review has demonstrated many problems with the measurement of income in surveys, some of which are visible to the naked eye, and others whose existence can only be confirmed under the microscope. In the former category are the consistent and often large shortfalls in nationally-weighted survey estimates as compared to independent benchmarks, and the considerable level of nonresponse; in the latter are the errors in individual respondents' survey reports of both income sources and income amounts that, collectively, signal the presence of both bias and random error effects. The propensities for problems vary substantially across different types of income. For wage and salary income, by far the largest component of personal income in the U.S., indicators of both response bias and random error are quite low. For other types of income, however, the situation is notably worse.

Our review of investigations into the cognitive aspects of providing income data in surveys suggests, first of all, that the field is a long way from having final and definitive information on how respondents understand, think about, and form answers to income questions. The review does serve to highlight the many possible contributors to inaccurate reporting, including lack of knowledge, misunderstanding and other definitional issues, recall problems and confusion, as well as the much-suspected issue of sensitivity. Taken all together, it is apparent that asking respondents to report their income is taxing in many ways, although no single cognitive issue seems predominant. The positive side of this situation is that many avenues are available for making inroads on income measurement error problems. The more daunting specter that it raises, however, is that so many problems must be solved in order to significantly improve measurement quality, some of which — respondents' lack of knowledge of some aspects of their own income situation, for example — seem largely immune to "cognitive" solutions.

Table 1: Ratio of SIPP and CPS March Income Supplement Aggregate Income Estimates to Independent Aggregate Income Estimates for 1984 and 1990 (source: adapted from Coder and Scoon-Rogers, 1996)

Source of Income	1984			1990		
	Indep. Aggregate Estimate (billions)	SIPP (%)	CPS (%)	Indep. Aggregate Estimate (billions)	SIPP (%)	CPS (%)
EMPLOYMENT INCOME:						
Wages and salaries	$1,820.1	91.4	97.3	$2,695.6	91.8	97.0
Self-employment	192.6	103.1	70.2	341.4	78.4	66.8
ASSET INCOME:						
Interest	244.8	48.3	56.7	282.8	53.3	61.1
Dividends	59.3	65.9	51.8	126.3	46.1	31.3
Rents and royalties	19.4	211.3	95.4	44.1	102.9	87.8
GOVERNMENT TRANSFER INCOME:						
Social Security	160.5	96.2	91.9	225.5	98.3	93.0
Railroad Retirement	5.6	96.4	71.4	6.9	95.7	66.7
Supplemental Security Income	9.9	88.9	84.8	13.6	94.9	89.0
Aid to Families w/Dep. Children	13.9	83.5	78.4	19.7	70.1	71.6
Other cash welfare	2.0	135.0	120.0	2.9	86.2	86.2
Unemployment Insurance	16.3	76.1	74.8	17.7	84.2	80.2
Workers' Compensation	14.1	56.7	48.2	14.6	86.3	94.5
Vets' Pensions and Compensation	13.9	82.0	59.7	13.8	84.1	77.5
RETIREMENT INCOME:						
Private pensions	65.2	63.8	57.2	70.2	107.1	110.8
Federal employee pensions	20.3	98.0	84.7	30.4	73.4	82.6
Military retirement	15.6	105.1	98.1	20.4	92.2	89.2
State and local employee pensions	21.9	88.1	71.7	36.1	75.1	80.1
MISCELLANEOUS INCOME:						
Alimony	2.7	100.0	81.5	2.5	116.0	124.0

Table 2. Imputation Rates and Amounts for Selected Sources of Income from the 1996 CPS March Income Supplement (source: unpublished data from the U.S. Bureau of the Census)

Source of Income	Total Number of Persons with Income (000's)	Number of Persons with Income Amount Reported	Number of Persons with Imputed Income Amount	Percent of Persons with Imputed Amounts
EMPLOYMENT INCOME:				
Wage and salary income	134,135	98,930	35,205	26.2
Nonfarm self-employment income	11,618	8,622	2,996	25.8
ASSET INCOME:				
Interest	107,871	60,354	47,518	44.1
Dividends	29,697	15,253	14,444	48.6
GOVERNMENT TRANSFER INCOME:				
Social Security	37,530	26,490	11,039	29.4
Supplemental Security Income	4,808	3,855	943	19.6
Public assistance	4,943	3,855	1,088	22.0
Workers' Compensation	2,064	1,498	566	27.4
Vets' Pensions and Compensation	2,549	1,875	674	26.4
RETIREMENT INCOME:				
Pension income	14,350	10,013	4,337	30.2
MISCELLANEOUS INCOME:				
Alimony	463	329	134	28.9
Child Support	5,190	4,180	1,010	19.5

8. References

Andersen, B. and Christoffersen, M. (1982). Om Sporgeskemaer. Socialforsknings-instituttets studie 46, Copenhagen, Denmark. *[In Danish; cited in Körmendi (1988)]*

Bancroft, G. (1940). Consistency of Information from Records and Interviews. Journal of the American Statistical Association, 35, 377-381.

Bogen, K. (1994). Documentation of SIPP CAPI Cognitive Interviews. Unpublished U.S. Bureau of the Census memorandum, June 13, 1994.

Bogen, K. (1995). Results of the Third Round of SIPP CAPI Cognitive Interviews. Unpublished U.S. Bureau of the Census memorandum, April 28, 1995.

Bogen, K. and Robinson, J. (1995). Results from the Fourth Round of SIPP Redesign Cognitive Interviews. Unpublished U.S. Bureau of the Census memorandum, October 17, 1995.

Bound, J., Brown, C., Duncan, G., and Rodgers, W. (1990). Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Validation Study Evidence. In Hartog, J., G. Ridder, and J. Theeuwes (eds.). Panel Data and Labor Market Studies, Amsterdam: North-Holland, 1990, 1-19.

Bound, J. and Krueger, A. (1991). The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? Journal of Labor Economics, 9, 1-24.

Bradburn, N. and Sudman, S. (1989). Polls and Surveys. San Francisco: Jossey-Bass.

Brehm, J. (1993). Who is Missing? In The Phantom Respondents: Opinion Surveys and Political Representation. Ann Arbor: University of Michigan Press, 23-38.

Cantor, D., Brandt, S., and Green, J. (1991). Results of First Wave of SIPP Interviews. Unpublished Westat report to the U.S. Census Bureau (memorandum to Chet Bowie). February 21, 1991.

Coder, J. (1992). Using Administrative Record Information to Evaluate the Quality of the Income Data Collected in the Survey of Income and Program Participation. Proceedings of Statistics Canada Symposium 92: Design and Analysis of Longitudinal Surveys, Ottawa: Statistics Canada, 295-306.

Coder, J. and Scoon-Rogers, L. (1995). Evaluating the Quality of Income Data Collected in the Annual Supplement to the March Current Population Survey and the Survey of Income and Program Participation. SIPP Working Paper Series, No. 215, U.S. Bureau of the Census, Washington, DC.

David, M. (1962). The Validity of Income Reported by a Sample of Families Who Received Welfare Assistance During 1959. Journal of the American Statistical Association, 57, 680-685.

Dibbs, R., Hale, A., Loverock, R., and Michaud, S. (1995). Some Effects of Computer-Assisted Interviewing on the Data Quality of the Survey of Labour and Income Dynamics. Statistics Canada: SLID Research Paper Series No. 95-07.

Dippo, C. and Norwood, J. (1994). A Review of Research at the Bureau of Labor Statistics. In Questions About Questions (J. Tanur, ed.). New York: Russell Sage Foundation.

Duncan, G. and Hill, D. (1985). An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data. Journal of Labor Economics, 3, 508-532.

Ferber, R. (1966). The Reliability of Consumer Reports of Financial Assets and Debts. Urbana, IL: Bureau of Economic and Business Research, University of Illinois.

Ferber, R., Forsythe, J., Guthrie, H., and Maynes, E. (1969a). Validation of a National Survey of Financial Characteristics: Savings Accounts. The Review of Economics and Statistics, LI, 436-444.

Ferber, R., Forsythe, J., Guthrie, H., and Maynes, E. (1969b). Validation of Consumer Financial Characteristics: Common Stock. Journal of the American Statistical Association, 64, 415-432.

Fowler, F. (1995). Improving Survey Questions. Applied Social Research Methods Series, Vol. 38. Thousand Oaks, Ca.: Sage.

Gallagher, N. (1992). Feeling the Squeeze. Networker, 16, 16-23, 26-29.

Goodreau, 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, 2, 179-186.

Goyder, J. (1987). The Silent Minority: Nonrespondents on Sample Surveys. Boulder, CO: Westview.

Greenberg, D. and Halsey, H. (1983). Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment. Journal of Labor Economics, 1, 380-407.

Grondin, C. and Michaud, S. (1994). Data Quality of Income Data Using Computer-Assisted Interview: The Experience of the Canadian Survey of Labour and Income Dynamics.

Proceedings of the Survey Research Methods Section, American Statistical Association, 830-835.

Groves, R (1989). Survey Errors and Survey Costs. New York: John Wiley and Sons.

Haber, L. (1966). Evaluating Response Error in the Reporting of the Income of the Aged: Benefit Income. Proceedings of the Social Statistics Section, American Statistical Association, 412-419.

Halsey, H. (1978). Validating Income Data: Lessons from the Seattle and Denver Income Maintenance Experiment. Proceedings of the Survey of Income and Program Participation Workshop — Survey Research Issues in Income Measurement: Field Techniques, Questionnaire Design and Income Validation, Washington, DC: U.S. Department of Health, Education and Welfare, January 16-17, 1978, (IV)21-(IV)51.

Heberlein, T. and Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. American Sociological Review, 43, 447-462.

Herriot, R. and Spiers, E. (1980). Measuring the Impact on Income Statistics of Reporting Differences Between the Current Population Survey and Administrative Sources. Studies from Interagency Data Linkages, Report 11, U.S. Department of Health, Education, and Welfare, Social Security Administration, 39-49.

Hoaglin, D. (1978). Household Income and Income Reporting Error in the Housing Allowance Demand Experiment. Proceedings of the Survey of Income and Program Participation Workshop — Survey Research Issues in Income Measurement: Field Techniques, Questionnaire Design and Income Validation, Washington, DC: U.S. Department of Health, Education and Welfare, January 16-17, 1978, (IV)52-(IV)95.

Horowitz, J. and Manski, C. (1995). Identification and Robustness with Contaminated and Corrupted Data. Econometrica, 63, 281-302.

Hu, T. (1971). The Validity of Income and Welfare Information Reported by a Sample of Welfare Families. Proceedings of the Social Statistics Section, American Statistical Association, 311-313.

Jordan, L, Marcus, A., and Reeder, L. (1979). Response Styles in Telephone and Household Interviewing: A Field Experiment from the Los Angeles Health Survey. In NCHSR Proceedings of the Health Survey Methods Third Biennial Conference, 116-123.

Kaluzny, R. (1978). Validation of Respondent Earnings Data in the Gary Income Maintenance Experiment. Proceedings of the Survey of Income and Program Participation Workshop — Survey Research Issues in Income Measurement: Field Techniques, Questionnaire

Design and Income Validation, Washington, DC: U.S. Department of Health, Education and Welfare, January 16-17, 1978, (IV)1-(IV)20.

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. Washington, DC: US Department of Health and Human Services/ASPE.

Körmendi, E. (1988). The Quality of Income Information in Telephone and Face to Face Surveys. Chapter 21 in Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls, W., and Waksberg, J. (eds.). Telephone Survey Methodology. New York: John Wiley & Sons, 341-356.

Lansing, J., Ginsburg, G., and Braaten, K. (1961). An Investigation of Response Error. Urbana, IL: Bureau of Economic and Business Research, University of Illinois.

Laumann, E., Ganon, J., Michael, R., and Michaels, S. (1994). The Social Organization of Sexuality. Sexual Practices in the United States. Chicago: The University of Chicago Press.

Livingston, R. (1969). Evaluation of the Reporting of Public Assistance Income in the Special Census of Dane County, Wisconsin: May 15, 1968. Proceedings of the Ninth Workshop on Public Welfare Research and Statistics, 59-72.

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

Marquis, K., Duan, N., Marquis, M., and Polich, J. (1981). Response Errors in Sensitive Topic Surveys: Estimates, Effects, and Correction Options. Santa Monica, CA: Rand Corporation.

Marquis, K. and Moore, J. (1990). Measurement Errors in SIPP Program Reports. Proceedings of the Sixth Annual Research Conference, Washington, DC: U.S. Bureau of the Census, March, 1990, 721-745.

Maynes, E. (1965). The Anatomy of Response Errors: Consumer Saving. Journal of Marketing Research, II, 378-387.

Maynes, E. (1968). Minimizing Response Errors in Financial Data: The Possibilities. Journal of the American Statistical Association, 63, 214-227.

Mellow, W. and Sider, H. (1983). Accuracy of Response in Labor Market Surveys: Evidence and Implications. Journal of Labor Economics, 1, 331-344.

Miller, E., Bogen, K., and Moore, J. (1997). Documentation of the SIPP Redesign Pretest and Content Test Wave 1 Behavior Coding Project. Unpublished U.S. Bureau of the Census report, February 1997.

Miller, H. and Paley, L. (1958). Income Reported in the 1950 Census and on Income Tax Returns. An Appraisal of the 1950 Census Income Data, Princeton, NJ: Princeton University Press, 179-201.

Moore, J. (1988). Self/Proxy Response Status and Survey Response Quality. Journal of Official Statistics, 4, 155-172.

Moore, J., Marquis, K., and Bogen, K. (1996). The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation. Unpublished U.S. Bureau of the Census report, January 11, 1996.

Oberheu, H. and Ono, M. (1975). Findings from a Pilot Study of Current and Potential Public Assistance Recipients Included in the Current Population Survey. Proceedings of the Social Statistics Section, American Statistical Association, 576-579.

Radner, D. (1982). Distribution of Family Income: Improved Estimates. Social Security Bulletin, 45, 13-21.

Rodgers, W., Brown, C., and Duncan, G. (1993). Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages. Journal of the American Statistical Association, 88, 1208-1218.

Singer, E., Von Thurn, D., and Miller, E. (1995). Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature. Public Opinion Quarterly, 59, 66-77.

Stinson, L. (1997). The Subjective Assessment of Income and Expenses: Cognitive Test Results. Unpublished U.S. Bureau of Labor Statistics report, January 1997.

Sudman, S. and Bradburn, N. (1974). Response Effects in Surveys. Chicago: Aldine.

Sudman, S. and Bradburn, N. (1982). Asking Questions. San Francisco: Jossey-Bass.

Tourangeau, R. (1984). Cognitive Science and Survey Methods. In Jabine, T., M. Straf, J. Tanur, and R. Tourangeau (eds.). Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington, DC: National Academy Press, 73-100.

Tourangeau, R. and T. Smith (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. Public Opinion Quarterly, 60, 275-304.

Turner, C., Lessler, J., and Devore, J. (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In Turner, C., Lessler, J., and Gfroerer, J. (eds.) (1992). Survey Measurement of Drug Use: Methodological Studies. Rockville, MD: National Institutes on Drug Abuse.

U.S. Bureau of the Census (1970). Evaluation and Research Program of the U.S. Census of Population and Housing, 1960: Record Check Study of Accuracy of Income Reporting. Series ER60, 8, Washington, DC.

van Melis-Wright, M. and Stone, D. (1993). Psychological Variables Associated with Respondents' Sensitivity to Income Questions— A Preliminary Analysis. American Statistical Association Proceedings of the Section on Survey Research Methods, II, 1124-1129.

Vaughan, D. (1978). Errors in Reporting Supplemental Security Income in a Pilot Household Survey. Proceedings of the Survey Research Methods Section, American Statistical Association, 288-293.

Vaughan, D., Lininger, C., and Klein, R. (1983). Differentiating Veterans' Pensions and Compensation in the 1979 ISDP Panel. Proceedings of the Survey Research Methods Section, American Statistical Association, 191-196.

Vaughan, D. and Yuskavage, R. (1976). Investigating Discrepancies Between Social Security Administration and Current Population Survey for 1972. Proceedings of the Social Statistics Section, American Statistical Association, 824-829.

Weiss, C. (1969). Validity of Welfare Mothers' Interview Responses. Public Opinion Quarterly, 32, 622-633.

Withey, S. (1954). Reliability of Recall of Income. Public Opinion Quarterly, 18, 197-204.

Yen, W. and Nelson, H. (1996). Testing the Validity of Public Assistance Surveys with Administrative Records: A Validation Study of Welfare Survey Data. Paper presented at the 1996 Annual Conference of the American Association for Public Opinion Research, Salt Lake City, UT, May 16-19, 1996.