

Household Surveys in Crisis

Bruce D. Meyer, Wallace K.C. Mok and James X. Sullivan*

February 4, 2015

Abstract

Household surveys, one of the main innovations in social science research of the last century, are threatened by declining accuracy, due to reduced cooperation of respondents. While many indicators of survey quality have steadily declined in recent decades, the literature has largely emphasized rising nonresponse rates rather than other potentially more important dimensions to the problem. We divide the problem into rising rates of nonresponse, imputation, and measurement error, documenting the rise in each of these threats to survey quality over the past three decades. A fundamental problem in assessing biases due to these problems in surveys is the lack of a benchmark or measure of truth, leading us to focus on the accuracy of the reporting of government transfers. We provide evidence from aggregate measures of transfer reporting as well as linked microdata. We discuss the relative importance of misreporting of program receipt and conditional amounts of benefits received, as well as some of the conjectured reasons for declining cooperation and survey errors. We end by discussing ways to reduce the impact of the problem including the increased use of administrative data and the possibilities for combining administrative and survey data.

*Meyer: Irving B. Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, IL 60637 Email: bdmeyer@uchicago.edu ; Mok: Department of Economics, The Chinese University of Hong Kong, Hong Kong. Email: wallacemok@cuhk.edu.hk; Sullivan: Department of Economics, University of Notre Dame, Notre Dame, IN 46556 Email: sullivan.197@nd.edu. We would like to thank Pablo Celhay for excellent research assistance.

1. Introduction

Household surveys are a critically important source of information to both researchers and policy makers. These large nationally representative surveys are arguably among the most important innovations in social science research of the last century. As the Committee on National Statistics puts it, “It is not an exaggeration to say that large-scale probability surveys were the 20th-century answer to the need for wider, deeper, quicker, better, cheaper, more relevant, and less burdensome official statistics.” (Brown et al. 2014). In recent years, there has been a sharp rise in the number and breadth of government surveys (Presser and McCulloch, 2011). A sizeable fraction of empirical economic research relies on household surveys as indicated by the hundreds of thousands of citations to surveys such as the Current Population Survey and the National Health Interview Survey among others. In addition, these surveys are the source of key information that guides policy decisions including some of the most used national and local statistics such as the rates of unemployment, labor force participation, poverty, and health insurance coverage. They are also the data source for the distribution of income and household spending patterns that determine weights for consumer price indices.¹

However, the usefulness of the data from household surveys is threatened as many measures of their quality decline. Households have become increasingly less likely to answer surveys at all (unit nonresponse), those that respond are less likely to answer sensitive questions (item nonresponse), and when they do provide answers, they are less likely to be accurate (measurement error). All of these issues contribute to bias in common statistics calculated from survey data.

There has been much attention to the problem of rising unit nonresponse in the Federal statistical agencies and the survey research community. The problem has been the subject of two National Research Council reports and a special issue of a major journal (National Research Council 2011, 2013, *Annals, AAPSS* January 2013). The Federal government, through the Office of Management and Budget, has set a target response rate for Federal censuses and surveys, and recommends analysis of nonresponse bias when the unit response rate is less than

¹ Also see “Census surveys provide information that we need” by Robert Groves, *Washington Post*, July 19, 2012.

80 percent.² The emphasis to date on rising unit nonresponse is not surprising given that it is so widespread, is often easy to measure, and increases survey costs.

While the rise of unit nonresponse is important, it provides an incomplete picture of changing survey quality. If unit nonresponse is random, it does not lead to bias. But it is often difficult to verify that the nonresponse is orthogonal to survey measures of interest because we typically have very limited information on the characteristics of nonresponders. There are many examples in the literature of successful approaches to reduce unit nonresponse, but such efforts do not necessarily lead to a reduction in bias, and may even make the bias worse. Several authors have noted that inducing participation from those who are initially reluctant to complete a survey may lead to greater problems with item nonresponse or measurement error.

Rising unit nonresponse is likely a symptom of other concerns—if respondents are increasingly reluctant to participate in surveys, they may also be less likely to respond to questions when they do participate. There is some evidence that rising unit nonresponse impacts other indicators of survey quality such as item nonresponse and measurement error (Bollinger and David, 2001), but much less is known about these other measures of quality.

In this paper we examine changes in the quality of household surveys, focusing on the three main threats to survey quality—unit nonresponse, item nonresponse, and measurement error—and investigate possible explanations for declining quality. The surveys that we analyze are among the most important for social science research and government policy. A fundamental problem in assessing biases due to these problems is the lack of a benchmark or measure of truth. To address this problem, we focus on the accuracy of the reporting of government transfers, because reliable benchmarks for these programs exist from both aggregate and micro-level administrative data. In addition, transfer program definitions are often clear and comparable in surveys and administrative sources. We examine the quality of data from nine large programs that all receive considerable attention from both the research and policy community: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP), Supplemental Security Income (SSI), Social Security retirement (OASI) and disability (SSDI), Unemployment Insurance (UI), Workers' Compensation (WC), the National School Lunch

² See Office of Management and Budget (2006). In most cases there is insufficient information to conduct a reliable nonresponse bias analysis.

Program (NSLP), and the Supplemental Nutrition Program for Women, Infants and Children (WIC). The results we present for transfer programs are consistent with other research that notes low and/or declining quality in survey reports of earnings, expenditures, or demographic characteristics.³

We document a noticeable rise in all three threats to survey quality in many of the most important datasets for social science research and government policy. There has been a pronounced increase in unit nonresponse that is evident in nearly all the major surveys. Item nonresponse is quite frequent and has also increased sharply, as evidenced by a rise in the fraction of responses that are imputed. Over the past two decades, the share of total reported dollars that is imputed for major transfer programs has risen from 20 to almost 34 percent in the Current Population Survey and from about 17 to about 30 percent in the Survey of Income and Program Participation (SIPP). More than 30 percent of Social Security dollars in the SIPP were imputed in 2011.

In order to directly quantify the extent of measurement error in surveys, one needs to compare the responses in the survey to some measure of the “truth”. A common approach, for example, is to link survey data to administrative microdata. We show evidence from such linked data indicating considerable measurement error in reports of receipt of government transfers. But to provide a more comprehensive look at measurement error in many surveys, across many years, and for several programs, rather than calculating measurement error directly we compare aggregate survey and administrative data to calculate bias. We can calculate the bias for key statistics such as the rate of benefit receipt and the mean dollars of transfers received for the population—for example, the share of households who receive Food Stamps, or the mean dollars of welfare received—by comparing weighted survey reports of participation or benefits received to administrative totals for participants or benefits paid out. We calculate the bias for these statistics for nine large transfer programs in five key household surveys. Although these measures of bias include all three threats to survey quality, we show that the bias is largely due to measurement error.

³ There are many papers documenting low response rates and or measurement error for income and earnings questions and there is some evidence that nonresponse rates for earnings have increased over time (Bollinger and Hirsh, 2007) and measurement error for responses has increased (Roemer 2002). Bee, Meyer, and Sullivan (forthcoming) show that underreporting of many categories of expenditures has risen over the past three decades.

Our results document a sharp rise in the (downward) bias for these statistics for most programs in all the surveys we have seen, though the rise differs by program and survey. For example, less than half of AFDC/TANF benefits are reported in the Current Population Survey Annual Demographic File/Annual Social and Economic Survey (CPS), down from about 75 percent twenty years ago. In 2013, only 54 percent of aggregate SNAP benefits were recorded in the CPS, and the SIPP captured only 79 percent. The high and rising share of benefits not reported appears to be largely due to measurement error in responses recorded in the surveys. Rising unit and item nonresponse play smaller roles. Our results suggest that there has been declining survey respondent cooperation which has had deleterious effects on many aspects of survey quality.

These findings have important implications for both researchers and policy makers. That these large government programs are significantly under-reported in surveys means many important analyses are biased. It leads to an overstatement of the dispersion of the income distribution, an understatement of the effect of income transfer programs on this distribution, and an understatement of program takeup—the fraction of those eligible for a program who participate.⁴ For example, the significant under-reporting of receipt of AFDC/TANF will lead analyses to understate the extent to which this program lifts families out of poverty.

Our results call for more research into why survey quality has declined. Some potential explanations include that households are oversurveyed or busier, increasing concerns about privacy, and a decline in public spirit. For survey data on transfer income, the broad patterns of under-reporting that we find do not seem to be consistent with simple explanations such as stigma or saliency, although such explanations may be important for some programs. Our results also emphasize the need for greater efforts to explore ways to improve the quality of household surveys. More frequent linking of survey data with administrative microdata provides one potentially fruitful avenue for improving the quality of survey data.

⁴ Throughout this paper we use under-reporting as a synonym for under-statement or under-recording, since it is likely due to errors by both interviewers and interviewees.

2. Rising Unit Nonresponse Rates

It has been known for some time that unit nonresponse (when a household in a sampling frame is not interviewed at all) has been rising in most surveys. Unit nonresponse rates rose by 3-12 percentage points over the 1990s for six U.S. Census Bureau surveys (Atrostic et al. 2001). In non-Census surveys, the rise in unit nonresponse is also evident, and in some cases even sharper, (Steeh et al. 2001, Curtin, Presser and Singer 2005, Battaglia et al. 2008, Brick and Williams 2013). While National Research Council (2013) provides a thorough summary for the U.S., the pattern is apparent in surveys in other countries as well (de Leeuw and de Heer 2002).

In Figure 1 we report the unit nonresponse rate for five household surveys in the 1984-2013 period: the Current Population Survey Annual Demographic File/Annual Social and Economic Supplement (CPS), the Survey of Income and Program Participation (SIPP), the Consumer Expenditure (CE) survey, the National Health Interview Survey (NHIS), and the General Social Survey (GSS). These surveys are among the most important for social science research and government policy. The CPS is the source of the official U.S. poverty rate and income distribution statistics; the SIPP is the best source of information needed to determine eligibility for and receipt of government transfers; the CE is the source of weights for the CPI and the main source of consumption information in the U.S., the NHIS is the most cited survey of health; and the GSS is the most used dataset with social and attitudinal information (as measured by Google Scholar citations).

All of the surveys show a pronounced increase in unit nonresponse over time, reaching rates in recent years that range from 16 to 33 percent. Between 1997 and 2013 the unit nonresponse rate in the CPS rose from 16 to 20 percent while the rate in the NHIS rose from 8 to 24 percent. Regression estimates of a linear time trend over the available years yields a positive coefficient on year for each of the surveys that is strongly significantly different from zero in four of the five cases and weakly significant in the remaining case (see Appendix Table A1).

National Research Council (2013) reports a general decline in response rates for a long list of surveys. One of the few exceptions is the American Community Survey (ACS), which as the replacement for the Census Long Form, is legally mandated, so is a special case. The decline in response rates seems to be even more pronounced for public opinion surveys (Pew, 2012). However, there is evidence that unit response rates are much higher for surveys in developing

countries. Mishra et al. (2008) report that recent demographic and health surveys from 14 different African countries all had unit response rates above 92 percent.

The rate of survey nonresponse is not particularly informative about the accuracy of statistics from a survey. Nonresponse only leads to bias if it is nonrandom. Verifying that nonresponse is random, however, can be difficult because sampling frames typically include very limited information on the characteristics of nonresponders. Even if nonresponders look like responders based on a limited set of characteristics (such as age and geography), this does not mean that these groups are similar along other dimensions such as program participation. Evidence on the extent to which unit nonresponse leads to bias varies. Some studies suggest that the resulting bias is small or can be mitigated by weighting (National Research Council 2013, p. 42-43). Even in public opinion surveys with response rates under ten percent, researchers have argued that properly weighted responses are largely representative (Pew, 2012). In their survey of bias estimates, Groves and Peytcheva (2008) found that bias magnitudes differed more across statistics (such as mean age or gender) within a survey than across surveys. Abraham et al. (2009) show, for example, that volunteers are more likely to respond to a time use survey than nonvolunteers, and consequently surveys with low unit response rates will overstate estimates of volunteering.

There are many techniques surveys employ to reduce unit nonresponse such as prenotification, incentives, follow-ups, etc. However, even when such efforts increase response rates, they do not necessarily lead to a reduction in bias, and may even make the bias worse (Grove, 2006; Grove and Peytcheva, 2008; Tourangeau et al. 2010; Peytchev, 2013; Kreuter et al., 2014). Inducing participation from those who are initially reluctant to complete a survey may lead to greater problems with item nonresponse or measurement error. There seems to be a tradeoff between different measures of survey accuracy.

3. Rising Item Nonresponse

An additional challenge that household surveys face is that even if a household agrees to participate in a survey, responses to key questions may not be obtained either due to refusal to answer, inability to answer, or failure of the interviewer to record the response. Most surveys, and all of those that we examine, typically impute a response in

these cases of missing data. Many methods are used to impute, though the Census Hot-Deck procedure, where a missing value is imputed from a randomly selected similar record, is probably the most common. Surveys impute responses for all sorts of questions including those related to demographic characteristics such as age and education, employment, and income. Unlike most questions, for which nonresponse rates are typically low (Bollinger and Hirsch, 2006), nonresponse rates can be quite high for questions related to labor and nonlabor income. For transfer programs, surveys may impute reciprocity—whether or not a person received a given type of benefit at all—or dollars or months of benefits received conditional on reported or imputed receipt.

As evidence of the extent of item nonresponse and how it has changed over time, we present imputation rates for survey questions on receipt of transfer income. We calculate the share of recorded dollars that is imputed for six programs: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP), Supplemental Security Income (SSI), Social Security (OASDI) including both retirement and disability benefits, Unemployment Insurance (UI), and Workers' Compensation (WC). These are large, national programs that provide benefits for tens of millions of individuals, and distributed almost one trillion dollars in 2011. We present the imputation shares for the CPS (Figure 2) and SIPP (Figure 3), two surveys that focus on income and program receipt, so they are a good indicator of the state of the art in survey collection over time. Although not reported here, we have also calculated similar imputation rates for the ACS, CE, and the PSID (see Meyer, Mok, and Sullivan, 2009).

On average, the share of dollars imputed is around 25 percent, but imputation has risen over time in both surveys for all programs. In 2013, the imputation shares in the CPS ranged from 24 percent of TANF and FSP dollars to 36 percent of Social Security dollars. Overall, the SIPP has noticeably higher imputation rates than the CPS.⁵

⁵ Imputation procedures in the SIPP take advantage of information collected in previous waves. For example, beginning with the 1996 panel missing data were imputed by using the respondent's data in the previous wave (if available). Starting with wave 2 of the 2004 panel, the SIPP began to use "Dependent Interviewing" in which the interviewers use information from the prior wave to tackle item non-response during the actual interview. For the results in Figure 3 and Table 1 we do not include values imputed from prior wave information in our calculation of total dollars imputed. See Meyer, Mok, and Sullivan (2014), Chapter 4 of U.S. Census Bureau (2001), and Pennell (1993) for more information.

Although the ACS has low unit nonresponse rates, imputation shares (not reported) always exceed ten percent and are fairly similar across programs.

Figures 2 and 3 also show a pronounced increase in imputation rates over the past two and a half decades. For example, between 1991 and 2013 the share of dollars recorded that are imputed in the CPS for AFDC/TANF and FSP/SNAP rose from about 15 percent to nearly 25 percent. This rise is evident in all programs in both the CPS and SIPP. We summarize the trends by regressing the imputation share on a constant and a time trend separately for each program and survey. As shown in Table 1, for all six programs and both surveys the coefficient on the time trend is positive. The estimates suggest, for example, that for AFDC/TANF in the CPS the fraction of dollars imputed is rising by 0.4 percentage points each year. In the case of the CPS the upward trend is statistically significant at the 1-percent level for four of the six programs, while the trend is significant in the SIPP for five of six programs. The imputation rates for months of receipt (not reported) are similar to those for dollars reported here. In recent years, at least ten percent of months are imputed in the CPS for all four programs for which we have months. For the SIPP, month imputation shares are sometimes below ten percent, but are more typically between ten and twenty percent. The shares have generally risen over time.

Transfer income may be imputed when there is missing information on either whether the household receives income from a given program or the dollars of such income received. We have also calculated the share of total dollars reported attributable only to those whose reciprocity is imputed. In the CPS and the SIPP this share is typically on the order of 10 percent, but they are frequently higher. There is substantial variation across programs and over time. For most of the years since 2000, reciprocity imputation exceeds 20 percent for AFDC/TANF. The rise in reciprocity imputation over time is less pronounced than that for overall imputation.

Researchers often take for granted that imputed observations are likely to be of lower quality than other observations, and there is some evidence to support this assumption. Studies using linked survey and administrative data show that the rates of false positive and negative reporting are almost always much higher among the imputed

observations than the non-imputed ones (Meyer, Goerge and Mittag 2014, Celhay, Meyer and Mittag 2015).

4. Measurement Error and Estimates of Bias

Encouraging a household to respond to a survey and answer all of its questions does not ensure the quality of the resulting data, because often the recorded answers are not accurate. These inaccurate responses, or measurement error, can contribute to bias (the difference between an estimate and the true value) in common statistics calculated from survey data. Evidence from survey data linked to administrative microdata indicates considerable measurement error in reports of receipt of government transfers. But to provide a more comprehensive look at measurement error in many surveys, across many years, and for several programs, rather than calculating measurement error directly we compare aggregate survey and administrative data to calculate bias. In this section we establish that mean reports of program receipt and dollars received are sharply biased and this bias has risen over time. Although these measures of bias include all three threats to survey quality (unit nonresponse, item nonresponse, and measurement error), we show that the bias is largely due to measurement error. This conclusion may not be surprising because weighting and imputation have the potential to reduce bias in mean reports resulting from unit or item nonresponse. If unit nonresponse is random or survey weights are able to correct for nonresponse, then the estimates will not be biased. While unit nonresponse is surely nonrandom with respect to receipt of transfer income, the resulting bias after weighting may be small. Similarly, item nonresponse also appears to be nonrandom, but will not lead to bias in mean reports if imputations are on average accurate.

We now examine the bias in a series of statistics, the mean receipt of transfer programs either in dollars or months, by comparing weighted survey data to administrative aggregates. Mean reports of transfer receipt are important statistics that affect distributional calculations of inequality and poverty as well as calculations of effects of programs on the income distribution and estimates of program takeup. We compare weighted survey data to administrative aggregates because this approach can be used for the widest range of transfer programs, the longest time period and many datasets. Our analyses focus on how under-reporting has changed over time.

For a more extensive discussion of how it differs across programs and datasets see Meyer, Mok and Sullivan (2009).

Comparisons to administrative aggregates have been used widely, but results are only available for a few years, for a few transfer programs and for some of the key datasets. Important papers include Duncan and Hill (1989), Coder and Scoon-Rogers (1996), and Roemer (2000), Wheaton (2007). These papers tend to find substantial under-reporting that varies across programs. An alternative approach to comparisons to aggregate data is to compare individual microdata to administrative microdata. Comparisons to administrative microdata on program receipt have been fairly limited in the literature. This approach has often been restricted to a single state, year, program and dataset (Taeuber et al. 2004). Examples of studies that examine more than one program (but still a single dataset) include Moore, Marquis and Bogen (1996), Sears and Rupp (2003) and Huynh et al. (2002).⁶

Below we report estimates of the proportional bias in dollar reporting (Dollar Bias or $Bias_d$) and month reporting (Month Bias or $Bias_m$). These biases can be defined as the net reporting rate minus 1 or

$$Bias_d = \frac{\text{dollars reported in survey, population weighted}}{\text{dollars reported in administrative data}} - 1$$

and

$$Bias_m = \frac{\text{months reported in survey, population weighted}}{\text{months reported in administrative data}} - 1$$

Note that the reporting rates in the above definitions are net rates as they reflect under-reporting by true recipients counterbalanced by over-reporting by recipients and nonrecipients.

We calculate the bias in the mean receipt of transfer dollars for the same programs for which we reported imputation rates above, only now we are able to divide OASDI into its retirement (OASI) and disability (SSDI) components.⁷ We also calculate month reporting biases for seven programs. Months of receipt are not available in all cases but

⁶ A review of studies can be found in Bound, Brown and Mathiowetz (2001).

⁷ In several of the datasets Social Security Disability benefits are in some cases combined with Social Security Retirement and Survivors benefits. To separate these programs, we use data from the Social Security Bulletin (U.S. Social Security Administration, various years) to calculate for each year, age, in school status, and gender cell, the proportions of total social security dollars that are paid to OASI and SSDI recipients. See Meyer, Mok and Sullivan (2009) for more details.

they are available for some programs for which we do not observe dollars. Thus, while we lose UI and WC, we are able to include the National School Lunch Program (NSLP) and the Supplemental Nutrition Program for Women, Infants and Children (WIC). We do this for as many individual years as are available for the CPS, the SIPP, the ACS, the CE Survey and the PSID, which are some of the most widely used surveys and those specifically designed to collect transfer program receipt information.⁸ These datasets are among our most important for analyzing income and its distribution as well as transfer receipt. Thus, the understatement of transfers in these data has major implications for our understanding of the economic circumstances of the population and the effects of government programs. We should emphasize that all of the bias estimates we report include imputed values in the survey totals, so are understated in most cases. A survey may have little estimated bias, in part, because a substantial amount of program dollars or months are imputed.

In Table 2, Panel A we present the average of the proportional bias in dollar reporting ($Bias_d$) over the 2000-2013 period for seven programs from five household surveys. In every case, with the single exception of SSI in the SIPP, the bias is negative, indicating under-reporting of dollars of transfer income.⁹ In most cases the bias is large. For our main cash welfare programs, TANF (combined with General Assistance in two cases), four of five surveys have a bias of 50 percent or more. Even in the SIPP, the survey especially designed to capture transfer program income, more than a third of TANF dollars are missed. For FSP/SNAP the bias is at least 30 percent for four of the five surveys. The bias in dollar reporting of UI and WC is also pronounced. The bias is at least 32 percent for UI and 54 percent for WC in all surveys. The Social Security Administration programs (OASI, SSDI, SSI) have much less bias.

The average proportional bias in monthly participation reporting ($Bias_m$) for this same period is reported in Panel B of Table 2. These biases are very similar to the corresponding dollar reporting biases in Panel A. In the case of FSP/SNAP, the similarity is striking, with the bias in the two types of reporting never differing by more than 1.1

⁸ Our approach of examining biases by calendar year will at times mask differences in reporting rates across SIPP survey panels and over time within panels, especially when data from multiple panels are available for the same calendar year.

⁹ The upward bias in reporting of SSI appears to be due to confusion between SSI and OASI (Huyhn et al., 2002; Gathright and Crabb, 2014).

percentage points for the three datasets. For both TANF and the FSP, month reporting comes from a mix of direct questions about each month (the SIPP) and questions about the number of months received (the CPS and the PSID). In the case of the SIPP, assuming that the reported monthly benefit of those who are true recipients and those who are not is similar, this result suggests that individuals report about the right dollar amount on average, conditional on reporting. Or, put another way, most of bias is due to not reporting at all, rather than reporting too little conditional on reporting. The dollar bias estimates are only slightly larger in absolute value than the month bias estimates, suggesting that there is a small amount of under-reporting of dollars conditional on receipt, nevertheless. In the case of the CPS and the PSID, the evidence suggests that total dollars and months are understated by similar amounts, again suggesting that conditional on reporting receipt, the monthly benefits are reported about right on average.

For OASI and SSDI we see similar biases for monthly receipt and dollar receipt, with the bias for dollar receipt being slightly larger (in absolute value), again suggesting that most of the downward bias results from failure to report receipt rather than underreporting the dollar amount of benefits conditional on reporting receipt. For SSI the bias for dollar receipt is actually smaller in absolute value (or in the SIPP, larger but positive) than the bias for monthly receipt, suggesting some overreporting of dollars conditional on reporting receipt.¹⁰

The average biases in monthly participation reporting for NSLP and WIC are also reported in Panel B of Table 2. Reporting of NSLP months seems to be quite low for both the the PSID and the CPS, which both have an average bias of about 50 percent. In the SIPP, on the other hand, the bias is positive, indicating that more months of participation are reported than we see in the administrative data. This result is likely due in part to our assumptions that all eligible family members (ages 5-18) receive lunches and that they do so for all four months of a given wave. WIC is also underreported significantly. The average bias for monthly WIC receipt in the CPS, PSID, and SIPP ranges from 19 to 34 percent.

¹⁰ For OASI, SSDI, and SSI the surveys besides the SIPP do not report monthly participation, only annual unique participation. Since our administrative numbers are for monthly participation, we use the relationship between average monthly and annual unique participation calculated in the SIPP to adjust the estimates from the other sources. This adjustment step likely induces some error that accounts for the weaker similarity between the bias for monthly and dollar receipt.

Not only is the bias in mean receipt of transfer programs large, it has been increasing noticeably over time. This growing bias is evident in Table 3, which reports estimates from regressions of annual estimates of the proportional bias in dollar reporting on a constant and a time trend. The results indicate that most programs in the PSID, CPS and CE show a significant increase in the downward bias (decline in dollar reporting over time). The downward bias in mean dollars reported of AFDC/TANF in the CPS, for example, increases by 0.96 percentage points each year. The time trends in bias in the SIPP and ACS are less pronounced. The exceptions to the general rise in bias are SSI and OASI, which have rising reporting rates in most cases. However, in the case of SSI in the SIPP, rising reporting leads to greater bias because the baseline bias is positive.

Although our bias estimates reflect not just measurement error but also weighting, coverage error, and unit and item nonresponse, as we argue below, much of this bias is driven by measurement error. The implication that measurement error in survey responses to government programs has grown over time is consistent with findings from Gathright and Crabb (2014) who calculate measurement error directly by linking SIPP data directly to Social Security Administration data for SSI and OASDI. An added benefit of such linking is that one can separately identify false positives and false negatives. Their analysis shows that false positive and false negative rates for reported receipt and the mean absolute deviation of the reported benefit amount from the administrative amount increased between the 1996 and 2008 panels of the SIPP for both SSI and OASDI. During this period the mean absolute deviation of the benefit amount increased by 70 percent for OASDI and by 60 percent for SSI.

While the bias estimates presented here are only for transfer programs, these estimates are likely a good indicator of survey quality more broadly. Other research has shown low and/or declining quality in survey reports of earnings or demographic characteristics.¹¹ Also, as we discuss below, potential reasons for declining quality that might be unique to transfer income, such as rising stigma or less recognition of the general program names, do not appear to explain the sharp rise in bias that we find.

Estimates similar to those reported above are available in previous studies for some surveys for a subset of years and programs. Coder and Scoon-Rogers (1996) provide reporting rates for five of our programs for 1984 and 1990 for the CPS and the SIPP. Roemer (2000)

¹¹ See footnote 3.

estimates reporting rates for the same five programs for 1990-1996 for the CPS and the SIPP. Wheaton (2007) estimates reporting rates for four programs between 1993 and 2005 in the CPS and a shorter period in the SIPP. Duncan and Hill (1989) have also studied the extent of benefit under-reporting in the CPS and PSID for earlier years. Our estimates of bias are generally comparable to those in these earlier studies, and studies that have looked at reporting of transfers over time, such as Roemer (2000), show declines in reporting over time consistent with our estimates.¹² There are some discrepancies between our estimates and those in the literature, but they can mostly be attributed to minor methodological differences.¹³

5. Caveats

Comparing weighted microdata from surveys to administrative aggregates is a particularly attractive approach for evaluating survey bias because it can be done easily for many years and across many surveys for the entire sample (as opposed to only a few states). However, there are some important limitations to this approach including that the survey and administrative populations might differ; that some surveys provide incomplete information on benefit receipt; and that the validity of these comparisons depends on accurate survey weights. In addition, we are estimating net reporting; a rise in false negative reports could be counterbalanced by a rise in false positive reports.

It is important to emphasize that the survey and administrative data populations do not always exactly align. In particular, our survey totals do not include those living outside the 50 states and the District of Columbia, the institutionalized, or decedents. We make a number of adjustments in order to make the administrative and survey data totals comparable.¹⁴ We exclude from the administrative totals payments to those in U.S. territories and those outside the U.S.¹⁵

¹² See Section 5 for a comparison of our results to those from studies of microdata matches.

¹³ For example, our OASDI and SSI numbers differ slightly from Roemer (2000), but this is likely due to differences in accounting for decedents. Roemer also finds a smaller bias for WC reporting, but this difference seems to be due to his exclusion of lump sum payments from the administrative data. For SSI, Duncan and Hill (1989) estimate a smaller bias in PSID reporting than we do. This difference might be due to the difference in the re-weighting algorithm employed, and that we do not account for those who receive benefits but die during the survey year. To account for this latter issue, Duncan and Hill adjust the reporting rate up 5 percent.

¹⁴ A full description of the data sources and methods can be found in Meyer, Mok and Sullivan (2009).

¹⁵ In cases where such information is not available, we subtract estimates of the share of such payments obtained from years when this information is available. For most programs these adjustments are typically small, ranging

To adjust for the fact that the institutionalized can receive some benefits (SSI, SSDI, OASI), we rely on data from the Decennial Censuses (which include the institutionalized) and the 2006 ACS to determine the share of dollars that are likely missed in the surveys. That the surveys do not include decedents is a potential concern because recipients of transfers in one calendar year may subsequently die before being interviewed the next year. We do not adjust for decedents, but we expect this to have little effect on our estimates in most cases.¹⁶

In most cases the reference period for the administrative data (typically a fiscal year) does not exactly align with that for the survey data. We convert fiscal year administrative data to a calendar basis by appropriately weighting the fiscal years. Another noncomparability is that administrative data for transfer income are based on awardees while the survey data typically provide information on who the benefit is paid to. Awardees and payees may be different people. For example, adults may receive social security and SSI benefits on behalf of their children. Unfortunately, most household surveys provide little information about exactly who is the true awardee of the benefit.¹⁷

Another issue is that some surveys provide incomplete information on the receipt of benefits. For example, in certain years of the PSID we only have information about benefit receipt for the head and the spouse. We address this issue by using the share of total benefits received by non-head, non-spouse family members in other years and scaling up the aggregates accordingly. This adjustment assumes that these shares are relatively stable over time.¹⁸

Sometimes surveys do not distinguish between different types of benefits received. For example, in some cases we cannot distinguish between different types of Social Security income. In this situation, we apply the OASI and SSDI dollar proportions

from 0.02% (SSI) to about 3% (SSDI). The notable exception is the Food Stamps Program, where dollars paid to U.S. territories constituted about 10% of the total prior to 1982.

¹⁶ Previous studies have adjusted for decedents by applying age, gender and race specific death rates to the data (Roemer 2000). However, if survey weights have previously been calculated to match survey weighted population totals with universe population estimates by age, gender and race then such an adjustment is unwarranted. A case could be made for adjusting the data if these characteristics are nonstationary (but such an adjustment is likely to be small), or if the adjustments were based on additional individual characteristics which are not used to determine weights but are related to death, such as receipt of SSDI or SSI or other programs, but we do not have this information. Consequently, our estimates of bias for SSDI and SSI are likely to be overstated somewhat, since recipients likely have a higher mortality rate than the average person of their age, gender and race, and consequently are more likely to miss the interview the following year.

¹⁷ The SIPP, however, does provide some information about who is the true awardee of Social Security benefits.

¹⁸ Non-head, non-spouse dollars received are typically under 10 percent of family dollars, but exceed 20 percent for SSI in a few years.

from published totals to determine participation in these programs. A more desirable method would calculate these proportions based on participation rather than dollars. Applying these proportions essentially assumes that an individual can only receive benefits from either SSDI or OASI, but not both, in a particular year. Strictly speaking, individuals can receive benefits from both programs in a year, most commonly those whose SSDI benefit switches automatically to OASI when they reach retirement age. This issue leads to a slight bias downward in our Social Security retirement and disability participation estimates.

We should also note that the validity of these comparisons depends on unbiased survey weights.¹⁹ The weights are based on the Census of Population, so an argument about underweighting is essentially an argument about individuals being missed in the Census count. While several studies have examined overall Census undercounts as well as undercounts for some demographic groups such as by race, gender, age and urbanicity, we have no estimates of the undercount for the populations receiving transfer income.²⁰ Overall estimates of the 1990 undercount are fairly low relative to the bias estimates above, in the range of two percent. Estimates are higher for blacks and renters, but lower for women, especially women of childbearing age. These bias estimates are small relative to most of the bias estimates for transfer reports that we have found.

Comparisons of our bias estimates to those calculated by linking survey and administrative microdata sheds some light on the extent to which inaccurate weighting might affect our bias estimates. Unfortunately, very few of the microdata linking studies are for dollars or average monthly reporting, and those that are often examine a demographic subset of the population. The most comparable estimates of bias from studies linking survey and administrative microdata come from Marquis and Moore (1990). Their bias estimates for months of receipt are reported in Column 1 of Table 4. Column 2 reports our estimated bias based on comparisons of aggregates for the same year (but not the same months or states). The bias in Column 2 is a function of sample weighting, coverage error, unit and item nonresponse, and measurement error, while the bias in Column 1 is only a function of item nonresponse and

¹⁹ As a check, for each survey and year, we have confirmed that our weighted population totals are close to Census population estimates. The PSID sample weights are not appropriate for weighting to the universe in some years. We adjust them in a manner suggested by the PSID staff (see the Appendix to Meyer, Mok and Sullivan 2009 for more details).

²⁰ See Hogan (1993) and Robinson et al. (1993) for 1990 Census undercount estimates.

measurement error. Thus, if the biases in each of these columns are similar, then this suggests that the combination of sample weighting, coverage error, and unit nonresponse is not that important relative to the other sources of bias.

The results in Table 4 suggest that the weights (as well as unit nonresponse and coverage error) are not a substantial source of bias because the bias estimates from the linked microdata are fairly close to our estimates using comparisons to aggregates. Our estimates are particularly close (or higher) for FSP and SSI, programs that target to the poor—a group that might be most plausibly under-weighted or under-represented.

Through linked survey and administrative microdata, one can decompose our bias estimates into three different sources of error: unit nonresponse (combined with coverage error and weighting), item nonresponse, and measurement error. In Table 5 we report this decomposition of our estimates of dollar bias for the FSP and Public Assistance (TANF and General Assistance) in three of our surveys in recent years using New York State data. We find that the bias due to the combination of coverage error, unit nonresponse and weighting is substantial, the bias due to item nonresponse is small, and the bias due to measurement error is always larger than the combination of the other sources of bias combined. The combined coverage, unit nonresponse and weighting bias varies from -0.0X to -0.1X for Food Stamps and -0.1X to -0.1X for Public Assistance across the three surveys. The item nonresponse bias varies from -0.0X to -0.0X for Food Stamps and -0.0X to -0.0X for Public Assistance. The bias due to measurement error is substantial for FSP, ranging from -0.1X to -0.2X, and for Public Assistance it is even larger, ranging from -0.5X to -0.5X.²¹

6. Reasons for nonresponse and errors

In this section we discuss the literatures that address two related questions: why is nonresponse and measurement error so prevalent, and why have these threats to survey quality grown over time. Regarding the high rate of unit nonresponse, disinterest or lack of time appear to be important factors. Based on data recorded by interviewers for two household surveys—the 1978 National Medical Care Expenditure Survey and the 2008 National Health Interview Survey—the most common reasons given for unit nonresponse include that potential respondents

²¹ Digits suppressed until numbers go through disclosure review.

are not interested, don't want to be bothered, or are too busy, and privacy concerns also seem to be important (Brick and Williams 2013, p. 39 and National Research Council 2013). Reasons for unit nonresponse are often divided into three categories: noncontact, refusals, and other reasons (such as language problems or poor health). While the just mentioned explanations are reason for refusal, failure to contact has also been stressed by some who have noted the rise of gated communities and the decline of land-line phones. Since the rise in nonresponse in household surveys has been driven by refusals (Brick and Williams 2013), we will not emphasize these "technological" reasons for noncontact.

One might suspect that the reasons for item nonresponse and measurement error are closely related to those for unit nonresponse, though the survey literature has tended to focus on unit nonresponse separately. One reason the three sources of error may be related is because there are potential respondents who are just less willing participants and their participation is worse in many dimensions. Some research has examined this hypothesis, for example, Bollinger and David (2001) show that those who respond to all waves of a SIPP panel more accurately report FSP participation compared to those who miss one or more waves. Similarly, Kreuter, Muller and Trappmann (2014) show that in a German survey the hard to recruit respondents provided less accurate reports of welfare benefit receipt than those easy to recruit. The reasons for item nonresponse likely differ depending on the nature of the questions. In the case of earnings Groves and Couper (1998) suggest that the most important reason for nonresponse is concerns about confidentiality but that insufficient knowledge is also important.

The reasons for the under-reporting of transfer benefits in household surveys have been catalogued by several authors.²² Interviewees may forget receipt or confuse the names of programs. They may misremember the timing of receipt or who are the true recipients of a program within a family. Errors may be due to a desire to reduce interview burden, the stigma of program participation, or the sensitivity of income information. Survey and interviewer characteristics such as the interview mode (in person or by phone), respondent type (self or proxy) may matter for the degree of under-reporting. Note that all of these explanations may lead to item nonresponse, measurement error conditional on responding, or both.

Information on the extent of under-reporting and how it varies across programs, surveys,

²² Marquis and Moore (1990) provide nice examples for the SIPP, while Bound, Brown and Mathiowetz (2001) and Groves (2004) provide more general discussions.

and time should be informative about the plausibility of different explanations for under-reporting. For example, a standard explanation of under-reporting is the stigma of reporting receipt of “welfare” programs, and the inclination to give “socially desirable” answers (Sudman and Bradburn 1974). This explanation is consistent with the low reporting rates of four of the programs most associated with “welfare” or idleness, AFDC/TANF, the FSP, UI and WIC. However, some of the patterns of reporting by program do not fit with a stigma explanation for under-reporting. Workers’ Compensation has the greatest bias but is presumably not a program that greatly stigmatizes its recipients, given that the program is for those injured while working.

Why has survey quality deteriorated over time? Several studies have considered this question, focusing on unit nonresponse. Among the reasons proposed include increasing urbanization, a decline in public spirit, increasing time pressure, rising crime (this pattern reversed long ago), increasing concerns about privacy and confidentiality, and declining cooperation due to “over-surveyed” households (Groves and Couper 1998; Presser and McCulloch 2011; Brick and Williams 2013). The continuing increase in nonresponse as urbanization has slowed and crime has fallen seem to make these unlikely explanations. The evidence testing the remaining hypotheses is fairly slim, based largely on national time-series analyses with a handful of observations. Several of the hypotheses pre-suppose underlying trends in societal conditions that are themselves hard to measure: the degree of public spirit, concern about confidentiality, and time pressure. The time pressure argument seems inconsistent with the trend in leisure (Aguilar and Hurst, 2007) and would suggest a steep gradient by income that is not apparent. The literature does not report good evidence to support or refute a steady decline in public spirit or rise in confidentiality concern. Some of these hypotheses seem amenable to a geographically disaggregated time-series approach, but little work seems to have been done along those lines. Groves and Couper (1998) show that nonresponse rates differ across demographic groups; cooperation is lower among single person households and households without young children, for example. But more research is needed on whether changes in demographic characteristics such as these can account for declining survey quality.

Our own reading of the evidence suggests that there may be something to the “over-surveyed” hypothesis that Presser and McCulloch (2011) explore. They cite several sources of evidence that show sharply rising rates of surveying of the general public. They document a sharp rise in the number of government surveys administered in the U.S. over the past few

decades. They report that a series of random digit dial telephone surveys found that the share of Americans surveyed in the past year more than quadrupled between 1978 and 2003 (CMOR, 2003). They also note that real expenditures on commercial survey research increased by more than five percent annually for the 16 years ending in 2004. We suspect that talking with an interviewer, which once was a rare chance to tell someone about your life, now is increasingly an annoyance crowded out by the press of telemarketers and commercial surveyors. As we have documented, much of the evidence for deterioration in quality is in the form of measurement error. In this case, explanations such as increasing time pressure, declining public spirit, and rising reluctance due to being over-surveyed seem more plausible. The decline in unit and item response rates may not fully reflect the secular decline in the willingness of households to cooperate, because these rates are endogenous; survey administrators respond to declining response rates by changing survey methods. For example, Groves and Couper (1998) note cases where the number of attempted contacts with respondents has risen in order to stem the rise in nonresponse.

Changes in survey procedures over time can also provide evidence on the reasons for changes in under-reporting. The reduction or elimination of in-person interviewing seems to have little effect on reporting rates. For example, reporting rates do not change much after the 1996 reduction of in-person interviewing in the SIPP. This result is consistent with the Groves (2004) observation that there is no robust evidence of a difference in errors between in-person and phone interviewing. Reporting for transfer programs does not appear to be sensitive to whether the interviewer explicitly mentions the name of a program (Meyer, Mok, Sullivan 2009). There is some evidence that adding bracketed responses (for example, starting in 2001, when a specific amount is not provided, the CE asks interviewees whether the amount falls within certain ranges) leads to increased reporting rates for some programs, but this evidence is not consistent across programs (Meyer, Mok, Sullivan 2009).

7. The Future of Microdata

As the quality of national survey data has declined, the availability of alternative data for research and policy analysis has increased. Economic research increasingly relies on various types of administrative data. Chetty (2012) reports that the share of micro-data based articles in

the “top four” general interest economics journals that relied on survey data fell from about sixty percent to twenty percent between 1980 and 2010. Over that same period, the share of articles relying on administrative data rose from about twenty percent to sixty percent. There has been a rapid increase in the use of alternative forms of administrative data, for example see the work surveyed in Einav and Levin (2014).

Administrative data sets, while quite heterogenous in origin, topic, and quality have many advantages. The datasets often have large sample sizes and low measurement error permitting the estimation of small effects and the testing of subtle hypotheses. The data often allow longitudinal measurement not possible in cross-section household data or difficult in longitudinal data with substantial attrition. They often enable the use of experimental or quasi-experimental methods. However, administrative data often lead to different sets of concerns. Often access to researchers for purposes of replication is difficult. Concerns about limited coverage or nonrepresentativeness of the data often arise, making the data unsuitable for examining population trends.

While administrative data alone can be used to answer many questions, they often suffer from the weaknesses of limited characteristics of individuals covered and incomplete coverage of the population. These limitations can potentially be solved by linking to household survey data. Many recent reports by government agencies, advocacy groups, and politicians have pointed to the advantages of administrative data linked to survey data.²³ These reports have noted the usefulness of such data for a wide variety of policy analyses, and the potential for reducing the burden on survey respondents no longer asked questions they might be reluctant to answer.

There are many examples of linked survey and administrative data. The Health and Retirement Survey is linked to Social Security earnings and claims as well as Medicaid claims. The National Center for Health Statistics is currently linking several of its population-based surveys to these and other administrative data. Many randomized experiments of welfare and training programs link household survey instruments to Unemployment Insurance earnings records or other administrative datasets. Ad hoc examples within government have also produced useful research such as the Scherpf, Newman and Prell (2014) work on SNAP receipt.

²³ See Burman et al. (2005), Brown et al. (2014), Office of Management and Budget (2014), and U.S. House of Representatives (2014).

8. Conclusions

In this paper, we document a decline in many measures of the quality of household surveys. Respondents are less likely to respond at all to a survey, are less likely to answer sensitive questions, and when they do, are less likely to give an accurate response. We have documented this pattern in detail for some of the most used datasets in economics. We have shown that the decline is widespread and pronounced. Maybe most importantly, we find that the decline in respondent cooperation with household surveys has led to increasing bias in an important set of statistics. Focusing on the mean report of dollars or months of government transfer program receipt, we see increasing bias in almost all cases over time. The patterns of bias in reporting that we find do not seem to be consistent with a simple story of stigma or the sensitivity of income reporting. While these reasons are plausible explanations for the low FSP and TANF reporting rates, they cannot explain the very low WC reporting rate. We suspect that other factors, including continuity of receipt, the ease of reporting, the survey structure, and a desire to reduce the length of interviews play a large part in determining the degree of bias. Since one of the most persuasive explanations for the pattern of increasing bias is that the public is over-surveyed, this deterioration seems unlikely to end, as surveying for commercial purposes continues to grow.

Much of what we know about the conditions of the American public and that is used for public policy formation comes from survey data. Thus, without changes in data collection and availability our infrastructure to formulate and evaluate public policies and test social science theories will degrade. There is the potential to reverse this trend through increased availability and linkage of administrative data.

References

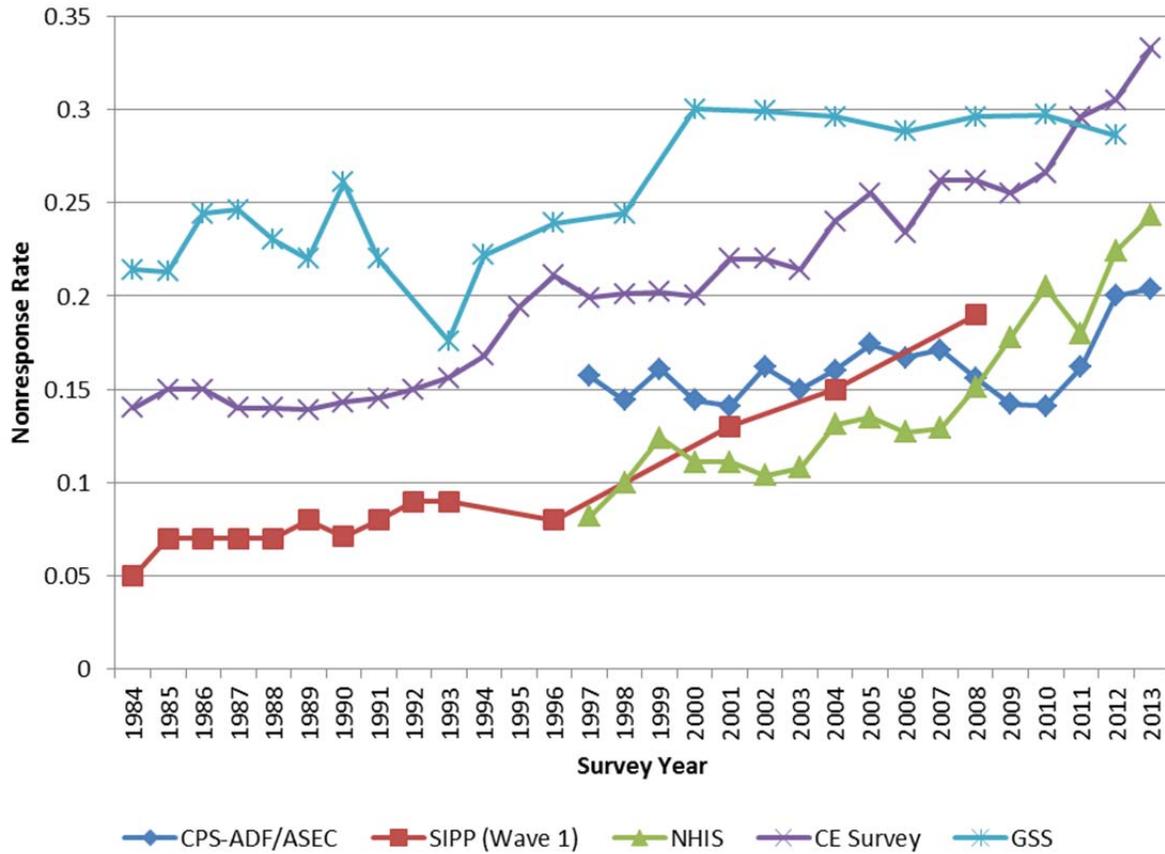
- Aguiar, Mark and Erik Hurst. 2007. Measuring Trends in Leisure: The Allocation of Time Over Five Decades. *Quarterly Journal of Economics*, August, Vol. 122, No. 3, Pages 969-1006.
- Atrostic, B. K, Nancy Bates, Geraldine Burt, and Adrian Silberstein. 2001. "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights," *Journal of Official Statistics*, 17:209-226.
- Battaglia, Michael P., Mina Khare, Martin R. Frankel, Mary Cay Murray, Paul Buckley, and Saralyn Peritz. 2008. Response rates: How have they changed and where are they headed? In *Advances in telephone survey methodology*, eds. James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japac, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 529–60. New York, NY: Wiley.
- Bee, C. Adam, Bruce Meyer, and James Sullivan (forthcoming), "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" in *Improving the Measurement of Consumer Expenditures*, Christopher Carroll, Thomas Crossley, and John Sabelhaus, editors. University of Chicago Press.
- Burman, Len, Dan Feenberg, Austan Goolsbee, Charles Hulten, Bruce Meyer, John Karl Scholz, Joel Slemrod (2005), "Report on the State of Publicly Available Data and Statistics For the Study of Public Economics," working paper.
- Bollinger, Christopher and Martin David (2001), Estimation with Response Error and Nonresponse: Food-Stamp Participation in the SIPP, *Journal of Business and Economic Statistics*, 19:2, 129-141.
- Bollinger and Hirsch 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching, *Journal of Labor Economics*, vol. 24, no. 3.
- _____ (2007), "How Well are Earnings Measured in the Current Population Survey? Bias from Nonresponse and Proxy Respondents," Working Paper.
- Bound, John, Charles Brown, and Nancy Mathiowetz (2001), "Measurement Error in Survey Data," in *Handbook of Econometrics. Volume 5*, ed. by J.J Heckman and E. Leamer. Elsevier: Amsterdam.
- Brick and Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys," *Annals of the American Academy of Political and Social Science*, 645, January.
- Brown, Lawrence, Constance Citro, Carol House, Krisztina Marton, and Christopher Mackie. 2014, "The Past, Present, and Future of Federal Surveys: Observations from the Committee on National Statistics," *JSM 2014 - Social Statistics Section*, p. 75-88.
- Celhay, Pablo, Bruce D. Meyer and Nicholas Mittag. 2015. "Measurement Error in Program Participation." Working Paper.
- Coder, John and Lydia Scoon-Rogers. 1996. "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," Housing and Household Economic Statistics Division, Bureau of the Census.
- Curtin, Richard, Stanley Presser, and Elinor Singer. 2005. Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly* 69 (1): 87–98.
- De Leeuw, Edith, and Wim de Heer. 2002. Trends in Households Survey Nonresponse: A Longitudinal and International Comparison. In *Survey Nonresponse*, ed. by Groves, etc.

- Duncan, Greg J. and Daniel H. Hill. 1989. "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics." *Journal of Business and Economic Statistics*, 441-52.
- Einav, Liran and Jonathan Levin. 2014. "Economics in the age of big data." *Science* 346, 1243089.
- Gathright, Graton and Tyler Crabb. 2014. "Reporting of SSA Program Participation in SIPP." Working Paper, U.S. Census Bureau.
- Groves, Robert M. 2004. *Survey Errors and Survey Costs*. Hoboken, NJ: John Wiley & Sons.
- _____. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly* 70: 646-675.
- Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in household interview surveys*. New York, NY: John Wiley.
- Groves, Robert M. and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias." *Public Opinion Quarterly* 72: 167-189.
- Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association* 88(3): 1047-1060.
- Huynh, Minh, Kalman Rupp, and James Sears. 2002. "The Assessment of Survey of Income and Program Participation (SIPP) Benefit Data using Longitudinal Administrative Records." Social Security Administration.
- Kreuter, Frauke, Gerrit Muller, and Mark Trappmann. 2014. "A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents," *Sociological Methods & Research*. 452-464.
- Marquis, Kent H. and Jeffrey C. Moore. 1990. "Measurement Errors in SIPP Program Reports." In Proceedings of the 1990 Annual Research Conference, 721-745. Washington, DC.: U.S. Bureau of the Census.
- Meyer, Bruce D., Robert Goerge and Nicholas Mittag. 2014. "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation." Unpublished Manuscript.
- Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2009. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." NBER Working Paper 15181. Updated version January 2015.
- Mishra, V. B. Barrere, R. Hong and S. Khan. 2008. "Evaluation of bias in HIV seroprevalence estimates from national household surveys." *Sexually Transmitted Infection* 84: i65-i70.
- Moore, Jeffrey C., Kent H. Marquis, and Karen Bogen. 1996. "The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation." *The Survey of Income and Program Participation*, Working Paper No. 212. Washington D.C.: U.S. Census Bureau.
- National Research Council. (2011). *The Future of Federal Household Surveys: Summary of a Workshop*. K. Marton and J.C. Karberg, rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- _____. (2013). *Nonresponse in Social Science Surveys: A Research Agenda*. Roger Tourangeau and Thomas J. Plewes, Editors. Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

- Office of Management and Budget. 2006. "Standards and Guidelines for Statistical Surveys." _____ . 2014. "Guidance of rProviding and Using Administrative Data for Statistical Purposes," Memorandum for the Heads of Eecutive Departments and Agencies, M-14-06, February.
- Pennell S. G. 1993. "Cross-Sectional Imputation and Longitudinal Editing Procedures in the Survey of Income and Program Participation" U. S. Department of Commerce, Census Bureau.
- Pew Research Center. 2012. "Assessing the Representativeness of Public Opinion Surveys" Washington, D.C.
- Peytchev, Andrew. 2013. Consequences of Survey Nonresponse. *Annals of the American Academy of Political and Social Science*.
- Presser, Stanley and Susan McCullogh. 2011. "The growth of survey research in the United States: Government-sponsored surveys, 1984-2004." *Social Science Research*, 1019-2024.
- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta and Karen A. Woodrow. 1993. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis." *Journal of the American Statistical Association* 88(3): 1061-1071.
- Roemer, Marc I. 2000. "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990-1996." Staff Papers on Income, Housing and Household Economic Statistics Division. Washington D.C.: U.S. Census Bureau.
- _____. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation," LEHD Technical Paper, 2002-22, Washington D.C.: U.S. Census Bureau.
- Scherpf, Erik, Constance Newman and Mark Prell. 2014. "Targeting of Supplemental Nutrition Assistance Program Benefits: Evidence from the ACS and NY SNAP Administrative Records." Working Paper, USDA.
- Sears, James and Kalman Rupp. 2003. "Exploring Social Security Payment History matched with the Survey of Income and Program Participation. Social Security Administration.
- Smith, Tom W.; Marsden, Peter V; Michael Hout; Jibum Kim. 2013. "General Social Surveys, 1972-2012: cumulative codebook," Principal Investigator, Tom W. Smith; Co-Principal Investigators, Peter V. Marsden and Michael Hout. Chicago: National Opinion Research Center.
- Steeh, Charlotte, Nicole Kirgis, Brian Cannon, and Jeff DeWitt. 2001. Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics* 17 (2): 227-47.
- Sudman, Seymour and Norman M. Bradburn. 1974. *Response Effects in Surveys*. Chicago: NORC/Aldine Publishing Company.
- Tauber, Cynthia, Dean M. Resnick, Susan P. Love, Jane Stavely, Parke Wilde, and Richard Larson. 2004. "Differences in Estimates of Food Stamp Program Participation Between Surveys and Administrative Records" working paper, U.S. Census Bureau.
- Tourangeau, Roger, Robert M. Groves, and C.D. Redline. 2010. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly*, 74(3): 413-432.

- U.S. Census Bureau. Various years-a. "Current Population Survey: Annual Social and Economic (ASEC) Survey Codebook," Washington D.C., United States Department of Commerce. Bureau of the Census.
- U.S. Census Bureau. Various years-b. "Survey of Income and Program Participation Codebook ," Washington D.C., United States Department of Commerce. Bureau of the Census.
- U.S. Census Bureau. 2001. "Survey of Income Program Participation Users' Guide." Third Edition. Washington D.C., United States Department of Commerce. Bureau of the Census.
- U.S. Department of Health and Human Services (2014). "National Health Interview Survey – Survey Description," Centers for Disease Control and Prevention, Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, Maryland.
- U.S. Department of Labor (various years), "Consumer Expenditure Interview Survey Public Use Microdata Documentation," U.S. Department of Labor, Bureau of Labor Statistics, Division of Consumer Expenditure Surveys.
- U.S. House of Representatives (2014). "Expanding Opportunity in America, House Budget Committee Majority Staff, July.
- U.S. Social Security Administration. Various Years. "Annual Statistical Supplement to the Social Security Bulletin." U.S. Social Security Administration, Office of Research, Evaluation and Statistics.
- Wheaton, Laura. (2007). "Underreporting of Means-Tested Transfer Programs in the CPS and SIPP," 2007 Proceedings of the American Statistical Association, Social Statistics Section.

Figure 1
Unit Nonresponse Rates of Major Household Surveys



Sources: For CPS, see Appendix G of U.S. Census Bureau (Various years-a). For SIPP, see Source and Accuracy Statement of U.S. Census Bureau (Various years-b). For NHIS, see Table 1 of U.S. Department of Health and Human Services (2014). For CE Survey, see U.S. Department of Labor (various years). For GSS, see Table A.6 of Appendix A – Sampling Design and Weighting in Smith et al. (2013).

Figure 2
Item Nonresponse Rates in the Current Population Survey (CPS) for Transfer Programs,
Calculated as Share of Dollars Reported in Survey that is Imputed

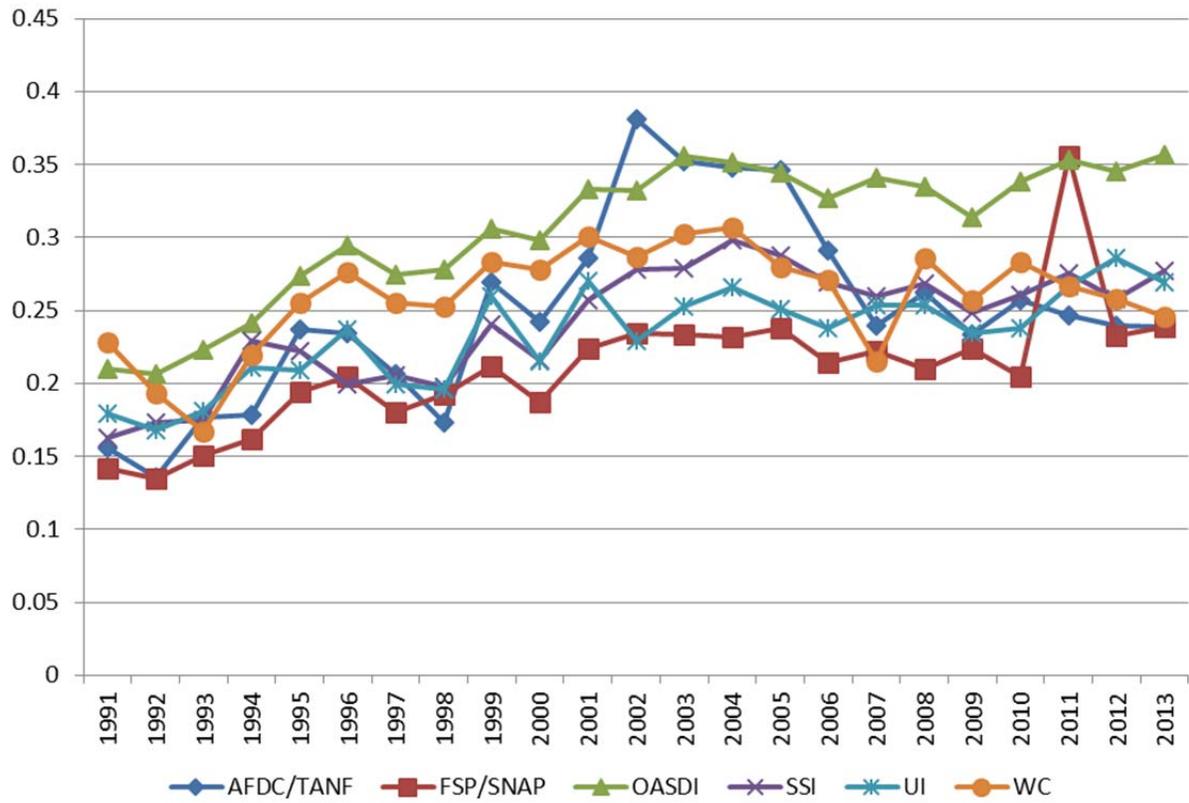


Figure 3
Item Nonresponse Rates in the Survey of Income and Program Participation (SIPP)
by Transfer Program, Calculated as Share of Dollars Reported in Survey that is Imputed,
Excluding Imputation using Previous Wave Information

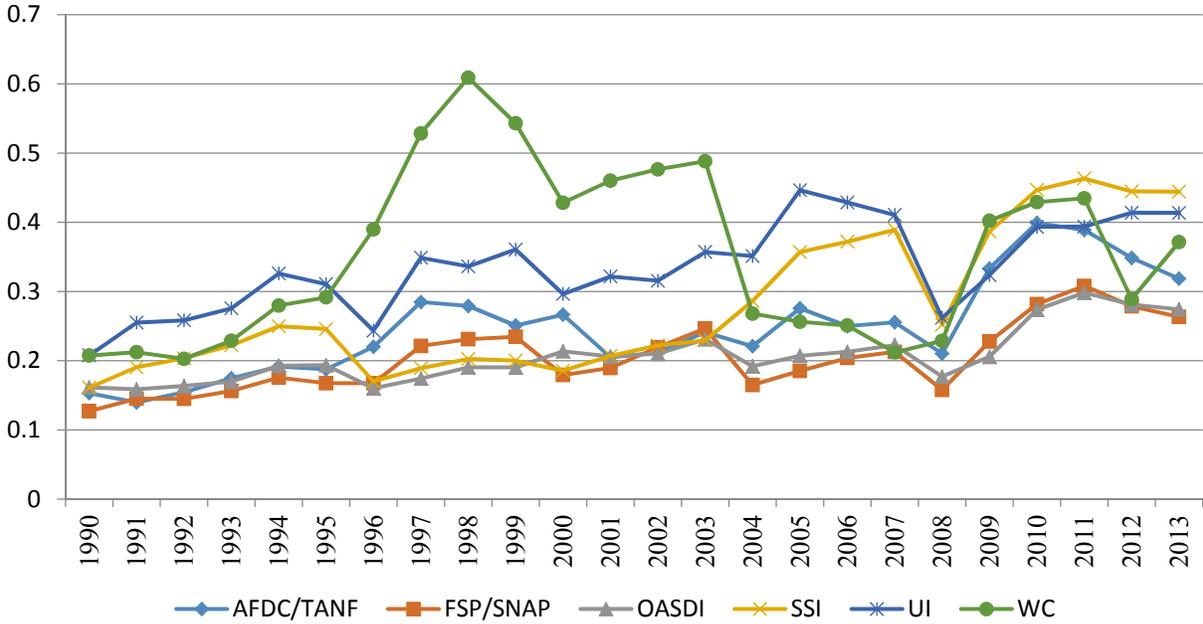


Table 1
Trend in Percentage of Program Dollars Imputed in Survey,
by Program and Survey

	AFDC/TANF	FSP/SNAP	OASDI	SSI	UI	WC
CPS	0.41 (0.35) 23	0.49 (0.08) ^a 23	0.64 (0.13) ^a 23	0.48 (0.12) ^a 23	0.39 (0.06) ^a 23	0.18 (0.17) 23
SIPP	0.79 (0.19) ^a 24	0.53 (0.14) ^a 24	1.25 (0.26) ^a 24	0.48 (0.10) ^a 24	0.69 (0.16) ^a 24	0.40 (0.66) 24

Notes: For each cell, we report the year coefficient from a regression of the percentage reporting rate on a constant and year, with its standard error underneath, followed by the sample size. The regressions correct for first order autocorrelation using the Prais-Winsten procedure. SIPP treats all “Statistical or Logical Imputation using Previous Wave Data” as non-imputation unless the original data are imputed. The superscripts a, b and c, indicate that the coefficient is statistically significantly different from zero at the 1%, 5%, and 10% levels, respectively.

Table 2
Proportional Bias in Survey Estimates of Mean Program Dollars and Months Received,
by Program and Survey, 2000-2013

	AFDC/TANF	FSP/SNAP	OASI	SSDI	SSI	UI	WC	NLSP	WIC
Panel A: Dollars									
ACS	-0.519	-0.458	-0.165	-0.299	-0.046				
CE	-0.767	-0.587	-0.149	-0.214	-0.283	-0.583	-0.618		
CPS	-0.500	-0.417	-0.086	-0.187	-0.162	-0.325	-0.541		
PSID	-0.619	-0.308	-0.086	-0.176	-0.322	-0.360	-0.646		
SIPP	-0.357	-0.170	-0.070	-0.146	0.164	-0.388	-0.651		
Panel B: Months									
ACS			-0.154	-0.261	-0.372				
CPS	-0.453	-0.422	-0.147	-0.154	-0.397			-0.503	-0.341
PSID	-0.574	-0.297	-0.114	-0.121	-0.502			-0.470	-0.192
SIPP	-0.232	-0.165	-0.008	0.041	0.023			0.141	-0.246

Notes: Each cell reports the average dollars/months proportional bias for the specified program and survey in the 2000-2013 period.

Table 3
Trend in Proportional Bias in Mean Dollars Reported in Survey (Including those Imputed),
by Program and Survey

	AFDC/TANF	FSP/SNAP	OASI	SSDI	SSI	UI	WC
ACS	-0.96 (0.87)		0.08 (0.07)	-0.68 (0.11) ^a	3.50 (1.11) ^b		
	12		12	12	12		
CE	-1.87 (0.43) ^a	-1.1 (0.43) ^b	0.07 (0.23)	-0.51 (0.23) ^b	0.05 (0.27)	-0.74 (0.19) ^a	-2.33 (0.38) ^a
	33	33	33	33	33	33	33
CPS	-0.71 (0.20) ^a	-0.59 (0.09) ^a	0.20 (0.02) ^a	-0.61 (0.08) ^a	0.41 (0.12) ^a	-0.39 (0.19) ^c	-0.71 (0.16) ^a
	37	34	45	45	38	26	25
PSID	-1.04 (0.12) ^a	-0.93 (0.27) ^a	0.40 (0.10) ^a	-0.62 (0.23) ^b	-0.04 (0.26)	-0.47 (0.16) ^a	-0.46 (0.12) ^a
	36	38	36	36	34	30	30
SIPP	-0.46 (0.34)	-0.06 (0.15)	0.05 (0.18)	-0.33 (0.49)	1.52 (0.37) ^a	-0.45 (0.22) ^c	-0.50 (0.10) ^a
	29	30	30	30	30	30	29

Notes: For each cell, we report the year coefficient from a regression of the proportional bias in percentages on a constant and year, with its standard error underneath, followed by the sample size. The regressions correct for first order autocorrelation using the Prais-Winsten procedure. The superscripts a, b and c, indicate that the coefficient is statistically significantly different from zero at the 1%, 5%, and 10% levels, respectively.

Table 4
Proportional Bias Estimates from Micro Data and Aggregate Data Compared

Transfer Program	Micro Data Bias Estimate due to Unit Nonresponse and Measurement Error (1)	Aggregate Data Bias Estimate due to All Sources of Error (2)
AFDC	-0.39	-0.21
FSP	-0.13	-0.15
OASDI	0.01	-0.06
SSI	-0.12	-0.14

Note: The microdata are from Marquis and Moore (1990) and use data from the SIPP over June 1983 to May 1984 for months of receipt in Florida, New York (OASDI and SSI only), Pennsylvania and Wisconsin. The aggregate data are averages of 1983 and 1984 from Meyer, Mok and Sullivan (2015) of average monthly participation for the entire U.S. We also assume OASDI participation is the sum of OASI and SSDI participation.

Table 5
Decomposition of Proportional Bias in Dollars Received into its Sources Using Micro Data

Survey	Program	Bias due to Combination of Coverage, Unit Nonresponse and Weighting (1)	Bias due to Item Nonresponse (2)	Bias due to Measurement Error (3)	Total Bias due to All Sources of Error (4)
ACS	Food Stamps	-0.XX	na	na	na
	Public Assistance	-0.XX	-0.XX	-0.XX	-0.XX
CPS	Food Stamps	-0.XX	-0.XX	-0.XX	-0.XX
	Public Assistance	-0.XX	-0.XX	-0.XX	-0.XX
SIPP	Food Stamps	-0.XX	-0.XX	-0.XX	-0.XX
	Public Assistance	-0.XX	-0.XX	-0.XX	-0.XX

Note: Based on New York State data for 2007-2012 from Celhay, Meyer and Mittag (2015). See text for methods. Food stamp dollars received are not reported in these years of the ACS. Digits have been replaced by X until the numbers receive disclosure approval.

Table A1**Trend in Unit Nonresponse Rates of Major Household Surveys**

	CPS	SIPP (Wave 1)	NHIS	CE Survey	GSS
Trend	0.22 (0.12) ^c	0.52 (0.05) ^a	0.90 (0.16) ^a	0.62 (0.06) ^a	0.33 (0.07) ^a
N	17	14	17	30	19
R-squared	0.519	0.934	0.566	0.760	0.791

Notes: For each cell, we report the year coefficient from a regression of the percentage nonresponse rate on a constant and year, with its standard error underneath, followed by the sample size and R-squared. The regressions correct for first order autocorrelation using the Prais-Winsten procedure. The superscripts a, b and c, indicate that the coefficient is statistically significantly different from zero at the 1%, 5%, and 10% levels, respectively.