

NBER WORKING PAPER SERIES

HOUSEHOLD SURVEYS IN CRISIS

Bruce D. Meyer
Wallace K.C. Mok
James X. Sullivan

Working Paper 21399
<http://www.nber.org/papers/w21399>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2015

We are grateful for the assistance of many New York Office of Temporary and Disability Assistance (OTDA) and Census Bureau employees including George Falco, Dave Dlugolecki, Graton Gathright, Joey Morales, Amy O'Hara, and Frank Limehouse. The micro-data analysis was conducted at the Chicago Census Research Data Center by researchers with Special Sworn Status and the results were reviewed to prevent the disclosure of confidential information. We would like to thank Pablo Celhay for excellent research assistance. We also thank Dan Black, Constance Citro, Michael Davern, Nikolas Mittag and participants at seminars at the American Enterprise Institute and the Federal Deposit Insurance Corporation for their helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, the New York OTDA or the U.S. Census Bureau.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Bruce D. Meyer, Wallace K.C. Mok, and James X. Sullivan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Household Surveys in Crisis

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

NBER Working Paper No. 21399

July 2015

JEL No. C42,C81,D31,H53,H55,I32,I38

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.

Bruce D. Meyer
Harris School of Public Policy
University of Chicago
1155 E. 60th Street
Chicago, IL 60637
and NBER
bdmeyer@uchicago.edu

James X. Sullivan
Department of Economics
912 Flanner Hall
University of Notre Dame
Notre Dame, IN 46556
James.X.Sullivan.197@nd.edu

Wallace K.C. Mok
Department of Economics
The Chinese University of Hong Kong
Shatin, Hong Kong
wallacemok@cuhk.edu.hk

Large and nationally representative surveys are arguably among the most important innovations in social science research of the last century. As the leadership of the Committee on National Statistics of the National Academy of Sciences wrote: “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). Household surveys are the source of official rates of unemployment, poverty, health insurance coverage, inflation and other statistics that guide policy. They are also a primary source of data for economic research and are used to allocate government funds.

However, the quality of data from household surveys is in decline. Households have become increasingly less likely to answer surveys at all (unit nonresponse), and those that respond are less likely to answer certain questions (item nonresponse). When households do provide answers, they are less likely to be accurate (measurement error).

The survey research and policy community has been pre-occupied with the rising rate of the first of these three main threats to survey quality, unit nonresponse. The nonresponse rate is by far the most cited measure of survey quality. Response rates are subject to White House Office of Management and Budget and journal restrictions. Rising nonresponse has been the subject of two National Academy reports and a journal special issue. The other two threats, item nonresponse and measurement error, have received much less attention. 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.¹

Of course, if nonresponse arises randomly across the population, survey data would still lead to unbiased estimates of distributions. Thus, we also investigate what is known about the extent to which these problems create bias. However, it can be difficult to verify that nonresponse is independent of survey measures of interest. After all, we typically have very limited information on the characteristics those who do not respond. Moreover, a fundamental problem in assessing survey bias due to these problems is the lack of a benchmark measure of the true outcome.

One productive approach to measuring the degree of bias in household surveys, along with addressing potential bias, is comparing survey results with administrative data. In this

¹ In certain cases, additional measurement issues will be important, in particular coverage error and sampling error. See Groves (2004) and Alwin (2007) for an exhaustive list of types of survey errors.

paper, 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. For example, aggregate administrative data for TANF are available in annual reports provided by the U.S. Department of Health and Human Services, Administration for Children and Families, and total SNAP payments are available from the U.S. Department of Agriculture, Food and Nutrition Service. Micro-level administrative data for these same program would typically come from a state welfare agency and take the form of records of monthly benefit payment amounts to individuals along with some information on the each individual's characteristics. Another advantage of focusing on transfers is that, the questions about transfer these programs are often clear and comparable in surveys and administrative sources. We examine the quality of household survey data through comparisons with administrative data from nine large programs that all receive considerable attention from both the research and policy community. For example, we compare the total dollar value of food stamp benefits reported by all respondents in a survey to the total dollar value of food stamp benefits awarded, as recorded in U.S. Department of Agriculture, Food and Nutrition Service administrative data.

Our results show a sharp rise in the (downward) bias in household survey estimates of receipt rates and dollars received for most programs. For example, in recent years more than half of welfare dollars and nearly half of food stamp dollars have been missed in several major surveys. In particular, this measurement error typically takes the form of under-reporting resulting from true program recipients being recorded as nonrecipients. (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.) We argue that although all three threats to survey quality are important, in the case of transfer program reporting and amounts, measurement error, rather than unit nonresponse or item nonresponse, appears to be the threat with the greatest tendency to produce bias.

The under-reporting of transfer income in surveys has profound implications for our understanding of the low income population and the effect of government programs for the poor. We point to evidence from linked administrative and survey data that indicates that this under-reporting leads to an understatement of incomes at the bottom, of the rate of program receipt, and of the poverty reducing effects of government programs, and an overstatement of poverty and of inequality.

The evidence on declining survey quality we present here is not likely to be unique to transfer income. While evidence comparing other survey variables to administrative benchmarks is scarce, there is evidence suggesting that survey biases in self-employment and pension income, education, pension contributions, and some categories of expenditures have also risen.

Our results call for more research into why survey quality has declined. Our preferred explanation is that households are overburdened by surveys, leading to a decline in many measures of survey cooperation and quality. The number and breadth of government surveys rose sharply between 1984 and 2004 (Presser and McCulloch 2011), and the number of private surveys has been rising as well. We discuss the limited evidence concerning some alternative explanations including increasing concerns about privacy, a decline in public spirit, less leisure time, or the stigmatizing effect of giving certain answers to questions. We conclude by noting the need for research on ways to improve the quality of household surveys. In particular, more frequent linking of survey data with administrative microdata provides one potentially fruitful avenue for improving the quality of survey data.

Rising Unit Nonresponse Rates

Unit nonresponse, which occurs 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 US 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. 2007; Brick and Williams 2013). The National Research Council (2013) report provides a thorough summary for US surveys, but the pattern is apparent in surveys in other countries as well (de Leeuw and de Heer 2002).

Indeed, the problem of rising unit nonresponse in major surveys has been a heavily discussed topic in the survey research community. Unit nonresponse was the subject of two National Research Council reports and a special issue of a major journal (National Research Council 2011, 2013, Massey and Tourangeau 2013). The federal government, through its Office of Management and Budget (2006), 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 80 percent. The editorial policy of at least one influential journal, the *Journal of the American*

Medical Association, restricts publication of research using low response rate surveys (Davern 2013).

In Figure 1, we report the unit nonresponse rate for five prominent household surveys during the 1984-2013 period: the Current Population Survey Annual Demographic File/Annual Social and Economic Supplement (CPS), which is the source of the official U.S. poverty rate and income distribution statistics; the Survey of Income and Program Participation (SIPP), which is the best source of information needed to determine eligibility for and receipt of government transfers; the Consumer Expenditure (CE) survey, which is the main source of data on consumption and provides the weights that are put on price changes when calculating inflation as measured by the Consumer Price Index; the National Health Interview Survey (NHIS), which is the primary source for information on the health status of the US population; and the General Social Survey (GSS), which may be the most used dataset across the social sciences for information on social and attitudinal information. Although we do not report their nonresponse rates in Figure 1, for other analyses in this paper we also examine the American Community Survey (ACS), which replaced the Census long form, providing detailed small-area information annually, and the Panel Study of Income Dynamics (PSID), which is the longest running longitudinal survey, which allows tracking specific households over time.

The surveys in Figure 1 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.² The National Research Council (2013) reports a general decline in response rates for a long list of surveys. The decline in response rates seems to be even more pronounced for public opinion surveys (Pew 2012). Interestingly, response rates are often 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.

One of the few notable exceptions to high nonresponse rates for domestic surveys is the American Community Survey. The ACS's low survey nonresponse rate (about 3 percent in recent years) is due in large part to the fact that the survey is mandatory. A Census study showed that a change to a voluntary standard for the ACS led to a rise in nonresponse rates to the mail

² 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. For details, see Appendix Table 1.

version of the survey of more than 20 percentage points (U.S. Census Bureau, 2003). The ACS also contacts potential respondents through multiple modes including mail, telephone, and personal visits. Only about 66 percent of households in most years respond to the initial 2 modes of contact. A random subsample of nonrespondents at that point is selected for a personal home visit, and their responses are given extra weight. Those not selected are given a weight of zero, so they do not contribute to the overall nonresponse rate. The ACS can affect community funding (Reamer 2010), which may also in part account for why the nonresponse rate is so low.

Of the problems with surveys, rising unit nonresponse has gotten the most attention. This emphasis is not surprising given that it is widespread, is often easy to measure, and increases survey costs. However, the rate of unit nonresponse is not particularly informative about the accuracy of statistics from a survey. Unit nonresponse only leads to bias if it is nonrandom, with the exact requirement depending on the statistic in question and the weighting method. However, exploring whether unit nonresponse is random can be difficult, because researchers typically have only limited information on the characteristics of nonresponders. Even if nonresponders look like responders based on a limited set of characteristics—say, age and geography—this does not mean that these groups are similar along other dimensions such as willingness to participate in government programs. Evidence on the extent to which nonresponse leads to bias differs by survey and question. While there are examples of substantial bias, in other cases the resulting bias is small or can be mitigated by appropriate weighting, in which certain demographic variables in the survey are weighted to correspond to the total population (National Research Council 2013, p. 42-43). Even in public opinion surveys with response rates under 10 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 they did across surveys.

Several methods have been proposed for improving unit nonresponse such as sending advance notification of the survey through the mail, increasing the number of times the potential respondent is contacted, improving the training of interviewers, or offering financial incentives for participation, but the evidence suggests only small effects of these efforts (National Research Council 2011). Even when such efforts increase response rates, they do not necessarily lead to a reduction in bias. Indeed, if they mainly encourage the groups that are already overrepresented

in the survey or who are unmotivated to co-operate, they can even make the bias worse (Groves, 2006; Groves 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: improving one measure may come at the expense of making another measure worse. In sum, unit nonresponse is probably not the main threat to the quality of household survey data.

Rising Item Nonresponse

Even if a household agrees to participate in a survey, responses to key questions may not be obtained due to refusal or inability to answer, or failure of the interviewer to record the response. This item nonresponse is distinct from a respondent misreporting that he did not receive a certain type of transfer income, which would be considered measurement error. 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. See Andridge and Little (2010) for more information. Surveys impute responses for all sorts of questions, including those related to demographic characteristics such as age and education, employment, and income. Nonresponse rates are typically low for most questions (Bollinger and Hirsch, 2006), but they 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—as well as the dollars received or the months of benefits received.

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 dollars recorded in two major household surveys that is imputed for six 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

(OASDI) including both retirement and disability benefits, Unemployment Insurance (UI), and Workers' Compensation (WC). These are large national programs that provide benefits to tens of millions of individuals—together they distributed almost \$1 trillion in 2011. We present the imputation shares for the Current Population Survey in Figure 2 and for the Survey of Income and Program Participation in Figure 3. These two surveys focus on income and program receipt, and 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 American Community Survey, the Consumer Expenditure Survey, and the Panel Study of Income Dynamics (Meyer, Mok, and Sullivan 2015).

The imputations rates are quite high, averaging about 25 percent. In 2013, the imputation shares in the Current Population Survey ranged from 24 percent of dollars recorded from Temporary Assistance for Needy Families and the Supplemental Nutrition Assistance Program to 36 percent of Social Security dollars. Overall, the Survey of Income and Program Participation has noticeably higher imputation rates than the CPS.³

Figures 2 and 3 also show an increase in imputation rates over the past two and a half decades. This rise is evident in all programs in both the Current Population Survey and Survey of Income and Program Participation. 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.⁴ The imputation rates for months of receipt (not reported) are similar to those for dollars reported here. In recent years, at least 10 percent of months are imputed in the CPS for all four programs for which we have months. For the SIPP, the imputation shares for months of receipt are sometimes below 10 percent, but are more typically between 10 and 20 percent. The shares have generally risen over time.

³ Imputation procedures in the Survey of Income and Program Participation 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 (2015), Chapter 4 of U.S. Census Bureau (2001), and Pennell (1993) for more information.

⁴ 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 Appendix Table 2, for all six programs and both surveys the coefficient on the time trend is positive. 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.

Transfer income may be imputed when there is missing information either on whether the household receives income from a given program, or on the dollars of such income received. We have also calculated the share of total dollars reported attributable only to those whose reciprocity was imputed. In the CPS and the SIPP this share is typically on the order of 10 percent, but is 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, which includes not only reciprocity imputation but also imputation of dollar amounts when receipt is reported but the dollar amount is not.

While imputation might improve quality assessments based on comparisons to aggregate amounts paid out, there are important limitations associated with imputed values. Studies using linked survey and administrative data show that the rates of false positive and false negative reporting of receipt are almost always much higher among the imputed observations than the non-imputed ones (Meyer, Goerge and Mittag 2014; Celhay, Meyer and Mittag 2015). Also, even in cases where imputations may reduce bias for statistics such as the mean, they may create greater bias in other statistics such as measures of inequality or covariances. For example, Bollinger and Hirsh (2006) show that including imputed values of earnings leads to biased coefficients in regressions of earnings on other attributes.

Measurement Error and Estimates of Bias

Inaccurate responses given there is a response, or measurement error, can contribute to bias (the difference between an estimate and the true value) in common statistics calculated from survey data. One way to test for measurement error is to link survey data on the payments that individuals or households say they have received with administrative microdata on the amounts actually provided to each household. 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). A review of earlier studies can be found in Bound, Brown and Mathiowetz (2001).

An alternative approach to calculating the bias in surveys is to compare aggregate survey data with aggregate administrative data. 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 program.

To provide a more comprehensive look at the magnitude of measurement error in many surveys, across many years, and for several programs, we compare aggregate survey and aggregate administrative data. The aggregate administrative data are available through a variety of reports from government agencies. A complete list of sources for these administrative aggregates is available in Meyer, Mok, and Sullivan (2015). The administrative aggregate data that we use have been audited so we expect bias in these data to be small. There might be some bias that results from different coverage or variable definitions between the administrative aggregates and the survey aggregates, but we make considerable effort to ensure these align closely. Since administrative data sources are heterogeneous, they should not always be taken as accurate. For example, Abowd and Stinson (2013) model administrative and survey measures of earnings, treating both sources as error ridden. Their approach is driven in part by conceptual differences between the survey and administrative measures of earnings that mean that there is not a single true earnings measure. This problem is not present when you examine transfers, because the survey and administrative concepts are generally the same.

Through comparisons to aggregate administrative data, we show that survey measures of whether an individual receives income and how much income is received from major transfer programs are both sharply biased downward, 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—in the following section we argue that the bias is largely due to measurement error.

Here, we will focus on two statistics, the mean receipt of certain transfer programs measured in dollars, and the mean receipt measured in months of receipt. Mean reports of transfer receipt are important statistics. Reports of these means affect distributional calculations

of inequality and poverty, as well as calculations of effects of programs on the income distribution and estimates of what share of those who are eligible for certain programs receive support from the program. Our analyses focus on how under-reporting has changed over time. For a more extensive discussion of how these findings differ across programs and datasets see Meyer, Mok, and Sullivan (2015).

Below we report estimates of the proportional bias in dollar reporting, which we call Dollar Bias, and in month reporting, which we call Month Bias. These biases can be defined as the net reporting rate minus 1, or more specifically

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

and

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

These expressions give us the proportional bias in the mean, and therefore can be thought of as the proportional bias in the total dollars or months or in the per person dollars or months. Also note that the reporting rates in the above definitions are net rates: that is, they reflect both underreporting by true recipients, counterbalanced to some extent 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 Old-Age, Survivors, and Disability Insurance (OASDI) into its retirement (OASI) and disability (SSDI) components.⁵ We also calculate month reporting biases for seven programs. Months of receipt are not available in all cases, including Unemployment Insurance and Workers' Compensation, but they are available for some programs for which we do not observe dollars, including the National School Lunch Program (NSLP) and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). We do this

⁵ 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 (2015) for more details.

for as many individual years as are available for five of the most important datasets for analyzing income and its distribution as well as receipt of transfers: the Current Population Survey, the Survey of Income and Program Participation, the American Community Survey, the Consumer Expenditure Survey, and the Panel Study of Income Dynamics.⁶ If these datasets are understating transfers received in a substantial way—and we will show that they are—this has important 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 the bias understates the measurement problems. To put it another way, by providing values for households that do not report receipt of transfer income, imputations may lead to smaller estimates of bias in our approach even though these imputations introduce considerable measurement error due to the inaccuracy of imputed values.

In Table 1, Panel A presents the average Dollar Bias over the 2000-2012 period for seven programs from five household surveys. In every case, with the single exception of Supplemental Security Income in the Survey of Income and Program Participation, the bias is negative, indicating under-reporting of dollars of transfer income. The upward bias in reporting of SSI appears to be due to confusion among recipients between SSI, which is aimed at low-income people who are blind, disabled, or elderly, and OASI, which is what most people mean by Social Security (Huyhn et al. 2002; Gathright and Crabb 2014). In most cases the bias reported in Table 1 is large. For our main cash welfare programs, Temporary Aid to Needy Families (combined with General Assistance in two cases), four of five surveys have a bias of 50 percent or more, meaning that less than half of the dollars given out are captured in surveys. Even in the SIPP, the survey especially designed to capture transfer program income, more than a third of TANF dollars are missed. For the FSP/SNAP, the bias is at least 30 percent for four of the five surveys. The bias in dollar reporting of Unemployment Insurance and Workers' Compensation is also pronounced: it is at least 32 percent for UI and 54 percent for WC in all surveys. The Social Security Administration programs (the retirement program

⁶ 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.

OASI, the disability insurance program SSDI, and support for the low-income elderly, blind, and disabled through SSI) have much less bias, which may, in part, be due to the fact that receipt of these programs tends to be more regular or permanent.

The average Month Bias for this same period is reported in Panel B of Table 1. These biases are very similar to the corresponding dollar reporting biases in Panel A. In the case of the FSP/SNAP, the similarity is striking, with the bias in the two types of reporting never differing by more than 1.1 percentage points for the three datasets. For both Temporary Assistance to Needy Families and the FSP/SNAP, month reporting comes from a mix of direct questions about each month (in the Survey of Income and Program Participation) and questions about the number of months received (in the Current Population Survey and the Panel Study of Income Dynamics). 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 the 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. In Meyer, Mok and Sullivan (2015) we report several sets of conditions under which equality of the dollar and month bias implies that dollar amounts are reported correctly on average.

For Old-Age and Survivors Insurance and Social Security Disability Insurance 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 Supplemental Security Income, the bias for dollar receipt is actually smaller in absolute value than the bias for monthly receipt for each survey other than the Survey of Income and Program Participation, where

the bias is larger but positive, all of which suggests some over-reporting of dollars conditional on reporting receipt.⁷

The average biases in monthly participation reporting for the National School Lunch Program and for the Special Supplemental Nutrition Program for Women, Infants, and Children are also reported in Panel B of Table 1. Reporting of NSLP months is quite low for both the Panel Study of Income Dynamics and the Current Population Survey, which both have an average bias of about 50 percent. In the Survey of Income and Program Participation, 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 of the reference months of a given wave of the SIPP. WIC is also underreported significantly. The average bias for monthly WIC receipt in the CPS, PSID, and SIPP ranges from 19 to 34 percent.

This large bias in mean receipt of transfer programs has been increasing over time. Table 2 reports estimates from regressions of annual estimates of the proportional bias in dollar reporting on a constant and a time trend for various years from 1967 to 2012 for the five surveys. Most household reports of transfer programs in the Current Population Survey, Panel Study of Income Dynamics, and the Consumer Expenditure Survey show a significant increase in the downward bias—that is, a decline in dollar reporting over time. The downward bias in mean dollars reported of AFDC/TANF in the CPS, for example, increases by about one percentage point each year. The time trends in bias in the Survey of Income and Program Participation and the American Community Survey are less pronounced. The exceptions to the general rise in bias are Supplemental Security Income and Old-Age Survivors Insurance, 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 bias is always positive in recent years.⁸

⁷ For the three Social Security programs--OASI, SSDI, and SSI--the surveys other than the Survey of Income and Program Participation do not report monthly participation, only annual participation. Since our administrative numbers are for monthly participation, we use the relationship between average monthly and annual 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.

⁸ Estimates consistent with those reported in Tables 1 and 2 are available in previous studies for some surveys for a subset of years and programs including: Coder and Scoon-Rogers (1996) for five of our programs for 1984 and 1990 for the Current Population Survey and the Survey of Income and Program Participation; Roemer (2000) for the same

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 Survey of Income and Program Participation data to Social Security Administration data for the Supplemental Security Income and the Old-Age Survivors and Disability Insurance programs. An added benefit of such linking is that one can 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 error in the benefit amount increased by 70 percent for OASDI and by 60 percent for SSI.

The under-reporting of transfer income in surveys has profound implications for our understanding of the low income population and the effect of government programs for the poor. Accounting for under-reporting of receipt, and substantial reporting and imputation error in amounts conditional on correctly reporting receipt, sharply changes what one learns from the survey data. Meyer and Mittag (2015) link data on four transfer programs (SNAP, TANF, General Assistance, and Housing Subsidies) to the New York data from the Current Population Survey over a four year period (2008-2011). 43 percent of SNAP recipients and 63 percent of public assistance recipients are not recorded as receiving benefits. Accounting for the survey errors more than doubles the estimate of the income of those who are reported to have income below half the poverty line. It leads the reported poverty rate to fall by 2.5 percentage points for the entire population and over 11 percentage points for single mothers. It nearly doubles the poverty reducing effect of the four programs overall, and increases it by a factor of over 1.5 for single mothers. The share of single mothers with no earnings or program receipt is cut in half.

Is the declining quality of survey data unique to transfer income? One might argue that potential reasons for declining quality that might be unique to transfer income such as rising stigma or less recognition of the general program names make it a special case. But as we argue below, these reasons do not appear to explain the sharp rise in bias that we find.

The evidence on whether measurement error has grown over time for other outcomes is limited, but this evidence suggests the problem of declining survey quality goes well beyond

five programs for 1990-1996 for the CPS and the SIPP; Wheaton (2007) for four programs between 1993 and 2005 in the CPS and a shorter period in the SIPP; and Duncan and Hill (1989) for the CPS and Panel Study of Income Dynamics for earlier years.

transfer income.⁹ Coder and Scoon-Rogers (1996) and Roemer (2000) find that reporting of self-employment income has worsened, but there is no clear trend for wage and salary income and dividends in the Current Population Survey and the Survey of Income and Program Participation. Comparing earnings aggregates from the Survey of Consumer Finances with those computed using IRS's Statistics of Income, Johnson and Moore (2008) find that respondents over-report earnings, and this over-reporting has worsened over time. They also find a sharp increase in pension income under-reporting. Barrow and Davis (2012), compare reported postsecondary enrollment in the October CPS to Integrated Postsecondary Education Survey administrative data, showing that CPS reporting of type of college attended has gotten worse, though error in reporting of overall enrollment has remained stable. Other studies have shown that measurement error in pension contributions has grown over time (Dushi and Iams, 2010). Also, Bee, Meyer, and Sullivan (2015) shows that while reporting rates for some of the biggest components of consumption have remained stable over time, there have been noticeable declines for some categories such as food away from home, shoes and clothing, alcoholic beverages. Future research on changes in bias in other outcomes would be a valuable extension to this literature.

Decomposing the Overall Bias

The bias estimates we present in Tables 1 and 2 are based on aggregate data, and for that reason they reflect not just measurement error but also coverage error (which arises when the sampling frame does not properly represent the underlying population) and error due to unit and item nonresponse. But for several reasons, we argue that the most important source of the overall bias is measurement error.

Coverage error could explain some of the significant under-reporting we find if the sampling frame for the surveys we examine (typically based on the noninstitutionalized Census population) does not capture the entire population that receives benefits. This argument about underweighting is essentially an argument about individuals being missed in the Census count.¹⁰ Although we do not have undercount

⁹ There are many studies that document substantial bias in levels for other outcomes (see Bound, Brown, and Mathiowetz 2001 for a summary).

¹⁰ We discuss issues related to the institutionalized population that receives transfers in the following section. As a check, for each survey and year, we have confirmed that our weighted population totals are close to Census

data for those who receive transfer income, estimates of the overall Census undercount are small, particularly relative to most of our bias estimates for transfer reports (Hogan 1993; Robinson et al. 1993). Moreover, undercount estimates have declined over time, and the estimates for the 2010 Census suggest an overcount (U.S. Census 2012).

There are also reasons to believe that bias resulting from unit and item nonresponse might be small. While unit nonresponse is surely nonrandom with respect to receipt of transfer income, appropriate weighting may offset much of this bias. Similarly, item nonresponse also appears to be nonrandom, but will not lead to bias in mean reports if imputations are on average accurate.

Empirical studies relying on linked administrative and survey microdata support these arguments. Bee, Gathright, and Meyer (2015), for example, show that for income in the Current Population Survey, unit nonresponse leads to remarkably little bias in the distribution of income. The estimates of bias from studies linking survey and administrative microdata that are most comparable to ours using aggregate data come from Marquis and Moore (1990), which, we should point out, relies on survey data from thirty years ago. Their bias estimates for months of receipt are reported in Column 1 of Table 3. The second column reports our estimated bias based on comparisons of aggregate survey and administrative data for the same year and survey as Marquis and Moore (but not the same months or states). The bias we calculate 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 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 3 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 the Food Stamp Program and for Supplemental Security Income, which are programs that target to the poor—a group that perhaps is most plausibly thought to be underweighted or underrepresented.

population estimates. The sample weights in the Panel Study of Income Dynamics are not appropriate for weighting to the complete population in some years. We adjust them in a manner suggested by the PSID staff, and the Appendix to Meyer, Mok, and Sullivan (2015) provides details.

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.¹¹ In Table 4 we report this decomposition of our estimates of dollar bias for the Food Stamps Program and Public Assistance (combining Temporary Assistance for Needy Families and General Assistance) in three of our surveys in recent years using New York state data for 2007-2012. 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 -0.049 to -0.096 for the FSP and -0.100 to -0.154 for Public Assistance across the three surveys. The item nonresponse bias varies from -0.020 to -0.067 for the FSP and -0.022 to -0.057 for Public Assistance. The bias due to measurement error is substantial for the FSP, ranging from -0.121 to -0.267, and for Public Assistance it is even larger, ranging from -0.529 to -0.584.

Direct evidence of substantial measurement error is not restricted to these two programs. Through linked survey and administrative microdata, Gathright and Crabb (2014) document substantial measurement error in receipt and amounts of Supplemental Security Income and the Old-Age Survivors and Disability Insurance in the Survey of Income and Program Participation, and this measurement error is rising over time.

Methodological Issues when Comparing Aggregate Data

Comparing weighted microdata from surveys to administrative aggregates is an attractive approach for evaluating survey bias because it can be done easily for many years and across many surveys. However, this approach also has some important limitations including possible differences between the survey and administrative

¹¹ We calculate the bias due to the combination of errors in coverage, weighting and unit nonresponse as the ratio of weighted administrative program dollars received by all linked households in the CPS to total administrative dollars paid out minus one. We calculate the bias due to item nonresponse as weighted dollars imputed to those not responding to the benefit question minus the dollars actually received by these households as a share of total dollars paid out. Finally, we calculate the bias due to measurement error as the dollars recorded by non-imputed respondents minus true dollars received as a share of total dollars paid out.

populations, and incomplete information on benefit receipt in some surveys. An additional concern that we will not discuss here is that by looking at net measures of bias, we are missing the extent to which a rise in false negative reports could be counterbalanced by a rise in false positive reports. Most of these problems are not present when linking microdata at the household level.

Survey and administrative data populations do not always align. Our household 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 (for a full description, see Meyer, Mok, and Sullivan 2015). For example, we exclude from the administrative totals payments to those in US territories and those outside the United States. 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 from 0.02 percent (Supplemental Security Income) to about 3 percent (Social Security Disability Insurance). The notable exception is the Food Stamp Program, where dollars paid to US territories constituted about 10 percent of the total prior to 1982.

As another example, to adjust for the fact that the institutionalized can receive some benefits in the Social Security-related programs, we rely on data from the Decennial Censuses (which include the institutionalized) and the 2006 American Community Survey to determine the share of dollars that are likely missed in household surveys that do not cover the institutionalized. 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 in a household survey the next year. We do not adjust for decedents, but assuming that the weights for extrapolating the household survey results to the population are well-chosen, we expect the lack of a specific adjustment for decedents to have little effect on our estimates in most cases.¹²

¹² 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 Social Security Disability Insurance or Supplemental Security Income 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.

Often 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 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 the person to whom the benefit is paid. Awardees and payees may be different people. For example, adults may receive Social Security and Supplemental Security benefits on behalf of their children. Most household surveys provide little information about exactly who is the true awardee of the benefit, although the Survey of Income and Program Participation does provide some partial information about who is the true awardee of Social Security benefits.

Some surveys provide incomplete information on the receipt of benefits. In certain years of the Panel Study of Income Dynamics, for example, we only have information about benefit receipt for the household 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 change slowly over time. Non-head, non-spouse dollars received are typically under 10 percent of family dollars, but exceed 20 percent for Supplemental Security Income in a few years.

Sometimes surveys do not distinguish between different types of benefits received. In some cases we cannot distinguish between different types of Social Security income. In this situation, we apply the Old-Age Survivors and Disability Insurance dollar proportions from published totals to determine participation in these programs. Applying these proportions essentially assumes that an individual can only receive benefits from one of these programs, but not both. In practice, however, individuals can receive benefits from both programs in a year--most commonly those whose disability benefit switches automatically to an old-age benefit when they reach retirement age. This issue leads to a slight bias downward in our Social Security retirement and disability participation estimates.

Reasons for Nonresponse and Errors

Why are nonresponse and measurement error so prevalent? 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 are not interested, do not want to be bothered, or are too busy, while privacy concerns also seem to be important (Brick and Williams 2013, p. 39; 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). Failure to contact has also been offered as a possibility by some who have noted the rise of gated communities and the decline of land-line phones, which could make door-to-door or phone surveys more difficult. However, the rise in nonresponse in household surveys has been primarily driven by refusals by those who are contacted (Brick and Williams 2013), and thus we will not emphasize these potential “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 literature on survey quality has tended to focus on unit nonresponse separately. One reason the three sources of error may be related would arise if some potential respondents are just less cooperative, so that 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 Survey of Income and Program Participation panel report participation in the Food Stamp Program more accurately than those who miss one or more waves. Similarly, Kreuter, Muller and Trappmann (2014) show in a German survey that 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; Marquis and Moore (1990) provide nice examples for the Survey of Income and Program Participation, while Bound, Brown and Mathiowetz (2001) and Groves (2004) provide more general discussions. 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 shorten the time spent on the interview, the stigma of program participation, the sensitivity of income information, or changes in the characteristics of those who receive transfers. Survey and interviewer characteristics such as the interview mode (in person or by phone), respondent type (self or proxy) may also matter for the degree of under-reporting. Notice 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, and time should help in differentiating among the 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: Temporary Assistance to Needy Families, the Food Stamp Program, Unemployment Insurance, and the Special Supplemental Nutrition Program for Women, Infants, and Children. However, other 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.

Another common explanation for under-reporting is that interviewees forget receipt, misremember the timing of receipt, or confuse the names of programs. Such issues should arguably be less common for programs that are received regularly, such as Old-Age and Survivors Insurance, Social Security Disability Insurance, and Supplemental Security Income. And, as shown in Table 1, these three programs typically have smaller bias than the other transfer programs we examine. However, the estimates in Table 1 show that the proportional bias for these programs is still large, particularly for SSDI and SSI, although this could be due to greater stigma for these two programs. Also, all three of these Social Security programs have item nonresponse rates that are no better than for some programs with less regular receipt (see

Figures 2 and 3).

Why has survey quality deteriorated over time? Several studies have considered this question, mostly focusing on unit nonresponse. Among the traditional 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 survey nonresponse as urbanization has slowed and crime has fallen make these less likely explanations for present trends. Tests of the remaining hypotheses are weak, based largely on national time-series analyses with a handful of observations. Several of the hypotheses require measuring societal conditions that can be difficult to capture: the degree of public spirit, concern about confidentiality, and time pressure. The time pressure argument seems inconsistent with the trend toward greater leisure (Aguiar and Hurst 2007) and would suggest that those with higher incomes and less leisure should be less likely to respond to surveys—a pattern that is at best weakly present. We are unaware of strong evidence to support or refute a steady decline in public spirit or a rise in confidentiality concerns as a cause for declines in survey quality. 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.

Changes in survey procedures over time can also provide evidence on the reasons for changes in under-reporting of receipt of government transfers. The reduction or elimination of in-person interviewing seems to have had little effect on reporting rates. For example, reporting rates do not change much after the 1996 reduction of in-person interviewing in the Survey of Income and Program Participation. This result is consistent with the observation by Groves (2004) 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 or not the interviewer explicitly mentions the name of a program (Meyer, Mok, and Sullivan 2015). There is some evidence that adding “bracketed” responses—for example, starting in 2001, when

a specific amount is not provided, the Consumer Expenditure Survey 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, and Sullivan 2015).

Our own reading of the evidence supports the hypothesis that “over-surveyed” respondents are less cooperative resulting in greater nonresponse and measurement error. Presser and McCulloch (2011) document a sharp rise in the number of government surveys administered in the United States over the 1984-2004 period. 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 (Council for Marketing and Opinion Research, 2003). They also note that real expenditures on commercial survey research increased by more than 4 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 crowded out by an annoying press of telemarketers and commercial surveyors. The decline in unit and item response rates may not fully reflect the secular decline in the willingness of households to cooperate, because survey administrators have tried to offset a declining household willingness to be surveyed by altering their methods. For example, Groves and Couper (1998) note cases where the number of attempted contacts with respondents has increased in order to stem the rise in nonresponse.

Taken together, the existing evidence does not provide a complete explanation for why survey quality has deteriorated over time. Households that are over-surveyed seem to contribute to the problem, but other explanations are likely important as well. There is a clear need for further research to fill in the important gaps in this literature.

The Future of Microdata

As the quality of conventional household survey data has declined, the availability of alternative data for research and policy analysis has increased. For empirical research, the role of survey data has declined as that of administrative data has risen; Chetty (2012) reports that the share of non-development microdata based articles in the “top four” general interest economics journals that relied on survey data fell from about 60 to 20 percent between 1980 and 2010, while the share of articles relying on administrative data rose from about 20 to 60 percent. A number of standard sources of administrative data have already been mentioned in this article,

like data from tax records and from transfer programs. In addition, the use of alternative forms of administrative data has been increasing. For example, the work surveyed in Einav and Levin (2014) offers examples like the use of administrative data on student test scores to measure teacher value added or data on earnings to assess the effect of the spread of broadband internet access into different areas.

Administrative data offers a bundle of advantages and disadvantages. 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, which is not possible in cross-sectional household data and can be difficult in longitudinal surveys with substantial attrition. When changes occur in policy or practice, especially when those changes affect only certain populations or geographic areas, administrative data often enable the use of experimental or quasi-experimental research methods. On the other hand, administrative data sets are typically not designed for academic research, and they can be quite heterogeneous in origin, topic, and quality. Researchers can find it difficult to access these data, whether for original research or for replication. Administrative data often offer only a limited set of characteristics of individuals, and these variables are often of low quality, especially for types of information that are collected but are not directly needed by the program for administration or other purposes. Also, administrative data sources often have incomplete coverage and are nonrepresentative, making the data unsuitable for drawing generalizable conclusions or examining population trends.

The limitations of administrative data can potentially be addressed 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 (for example, Burman et al. 2005; Brown et al. 2014; Office of Management and Budget 2014; U.S. House of Representatives 2014). These reports have noted the usefulness of such data for a wide variety of policy analyses. The President's 2016 budget calls for \$10 million for the Census Bureau to "to accelerate the process of acquiring additional key datasets...; expand and improve its infrastructure for processing and linking data; and improve its infrastructure for making data available to outside researchers" (White House 2015). And a recent bi-partisan bill would establish a commission to recommend the structure of a clearinghouse for administrative and survey data (U.S. Senate 2015).

Linking administrative microdata to survey microdata may improve the quality of survey data by providing more accurate information for some variables or by shortening the interview length and reducing the burden on survey respondents no longer asked questions they might be reluctant to answer. Such data linking can also be useful for improving the existing stock of data. For example, Nicholas and Wiseman (2009, 2010) and Meyer and Mittag (2015) show how one can use linked data to correct for under-reporting of transfer income when calculating poverty rates.

A number of examples already exist of linked survey and administrative data. The Health and Retirement Survey is linked to administrative data on Social Security earnings and claims, as well as to Medicaid data. The National Center for Health Statistics is currently linking several of its population-based surveys to administrative data. Many randomized experiments of welfare and training programs linked household survey instruments to Unemployment Insurance earnings records or other administrative datasets (Grogger and Karoly 2005). Ad hoc examples within government have also produced useful research such as the work by Scherpf, Newman and Prell (2014) using administrative data on the Supplemental Nutrition Assistance Program.

Surveys have explored alternative methods to improve survey quality, including multi- or mixed-mode methods to collect information from respondents (Citro 2014). Use of the internet has become an increasingly more common mode. In addition to standard mail, telephone, and face-to-face interview modes, the American Community Survey now allows respondents to respond online. These new methods may, in some cases, reduce costs and have the potential to improve data quality, but whether these approaches effectively reduce bias remains to be seen.

Much of what we know about the conditions of the American public and the information that is used for public policy formation comes from national survey data. The ongoing deterioration of household survey data documented in this paper seems unlikely to end, especially as surveying for commercial purposes and the feeling of being over-surveyed continues to grow. Without changes in data collection and availability, the information infrastructure to formulate and evaluate public policies and to test social science theories will degrade. Efforts to improve national survey data and to reduce nonresponse bias and measurement error have important benefits. Another potentially productive step toward improving the quality of data available for social science research—rather than just seeking to slow the pace of erosion in the quality of that data—is to increase the availability of

administrative datasets and to find additional ways to link them to household survey data and substitute administrative variables for survey questions in a timely fashion.

References

- Abowd, John and Martha Stinson. 2013. Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data. *Review of Economic Statistics*, December, Vol. 95, No. 5, Pages 1451-1467.
- 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.
- Alwin, Duane F. 2007. *Margins of Error*. John Wiley & Sons: Hoboken, NJ.
- Andridge, Rebecca and Roderick Little. "A review of hot deck imputation for survey non-response." *International Statistical Review* 78.1 (2010): 40-64.
- 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.
- Barrow and Davis (2012). "The upside of down: Postsecondary enrollment in the Great Recession." *Economic Perspectives*. 4Q. 117-129.
- Battaglia, Michael P., Mina Khare, Martin R. Frankel, Mary Cay Murray, Paul Buckley, and Saralyn Peritz. 2007. 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, Graton Gathright, and Bruce D. Meyer (2015), "Bias from Unit Non-Response in the Measurement of Income in Household Surveys." University of Chicago working paper.
- Bee, C. Adam, Bruce Meyer, and James Sullivan (2015), "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, 204-240. 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.
- 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, University of Chicago.

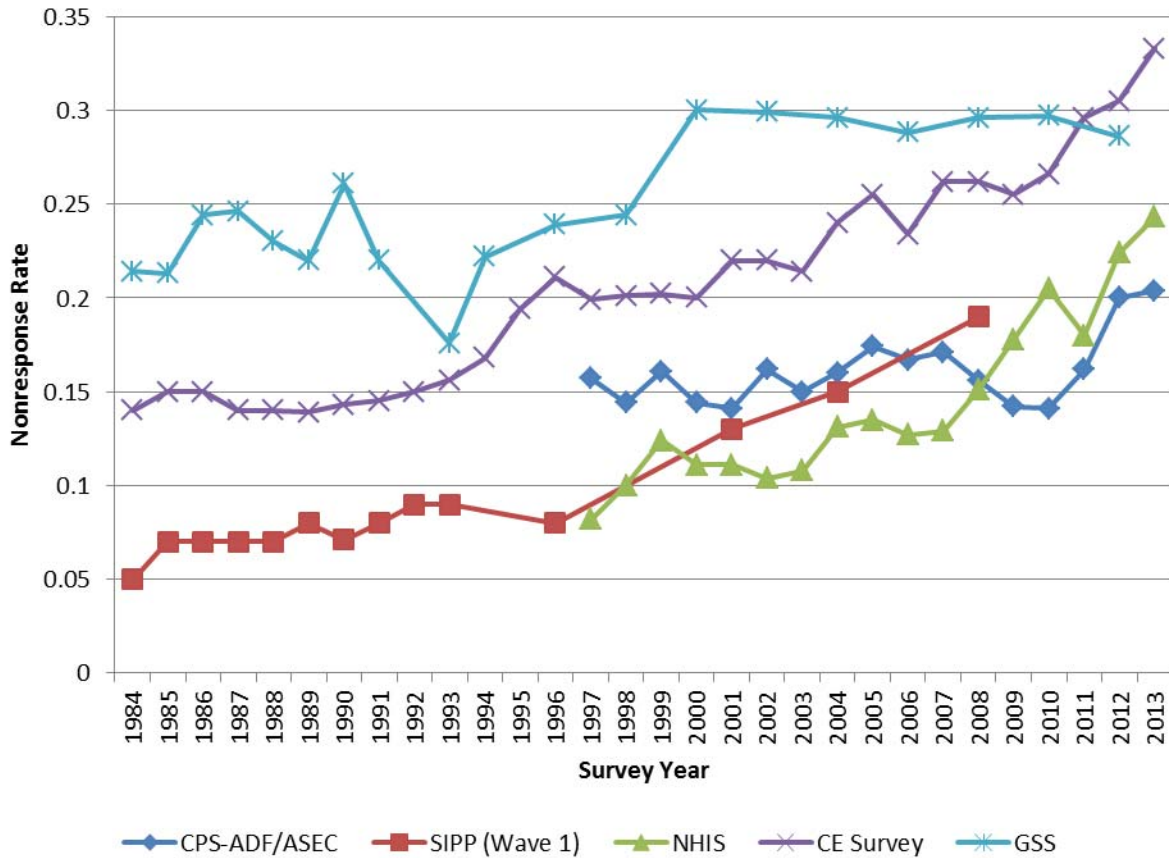
- Chetty, Raj. 2012. "Time Trends in the Use of Administrative Data for Empirical Research." NBER Summer Institute presentation.
- Citro, Constance F. 2014. "From Multiple Modes for Surveys to Multiple Data Sources for Estimates." National Research Council working paper. Presented at the 2014 International Methodology Symposium of Statistics Canada, Ottawa, Canada.
- Council for Marketing and Opinion Research, 2003. "Respondent Cooperation and Industry Image Study."
- 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.
- Davern, Michael. 2013. Nonresponse Rates are a Problematic Indicator of Nonresponse Bias in Survey Research. *Health Services Research* 48 (3): 905-912.
- 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.
- Dushi, Irena and Howard Iams. 2010. "The Impact of Response Error on Participation Rates and Contributions to Defined Contribution Pension Plans." *Social Security Bulletin*, Vol. 70, No. 1, p. 45-60.
- 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.
- Grogger, Jeff, and Lynn A. Karoly. 2005. Welfare Reform: Effects of a Decade of Change. Cambridge, MA: Harvard UP.
- 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.
- Johnson, Barry W., and Kevin Moore. 2008. Differences in Income Estimates Derived from Survey and Tax Data. 2008 SOI Paper Series.
- 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.
- Massey, Douglas S. and Roger Tourangeau, editors. 2013. "The Nonresponse Challenge to Surveys and Statistics." *The ANNALS of the American Academy of Political and Social Science*. 645 (1): 6 – 236.
- 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 and Nicholas Mittag. 2015. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." University of Chicago working paper.
- Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "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.
- Nicholas, Joyce and Michael Wiseman, 2009. "Elderly Poverty and Supplemental Security Income." *Social Security Bulletin* 69(1): 45–73.
- Nicholas, Joyce and Michael Wiseman, 2010. "Elderly Poverty and Supplemental Security Income, 2002-2005." *Social Security Bulletin* 70(2): 1–30.
- Office of Management and Budget. 2006. "Standards and Guidelines for Statistical Surveys." September.
- _____. 2014. "Guidance for Providing and Using Administrative Data for Statistical Purposes," Memorandum for the Heads of Executive 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*. 645(1): 88-111.

- Presser, Stanley and Susan McCulloch. 2011. "The growth of survey research in the United States: Government-sponsored surveys, 1984-2004." *Social Science Research*, 1019-2024.
- Reamer, Andrew D. 2010. "Surveying for Dollars: The Role of the American Community Survey in the Geographic Distribution of Federal Funds." Brookings Institution.
- 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.
- Taeuber, 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. Census Bureau. 2003. "Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Testing the Use of Voluntary Methods." Washington D.C., United States Department of Commerce. Bureau of the Census.
- U.S. Census Bureau. 2012. "Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census." Washington D.C., United States Department of Commerce. Bureau of the Census. http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html.
- 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. Senate (2015). "Murray, Ryan Introduce Bill to Expand Data Use in Evaluating Federal Programs, Tax Expenditures." Accessed at <http://www.murray.senate.gov/public/index.cfm/2015/4/evidence-based-policy-murray-ryan-introduce-bill-to-expand-data-use-in-evaluating-federal-programs-tax-expenditures>.
- 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.
- White House (2015). "The President's Budget Fiscal Year 2016." (accessed at https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/fact_sheets/building-and-using-evidence-to-improve-results.pdf).

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

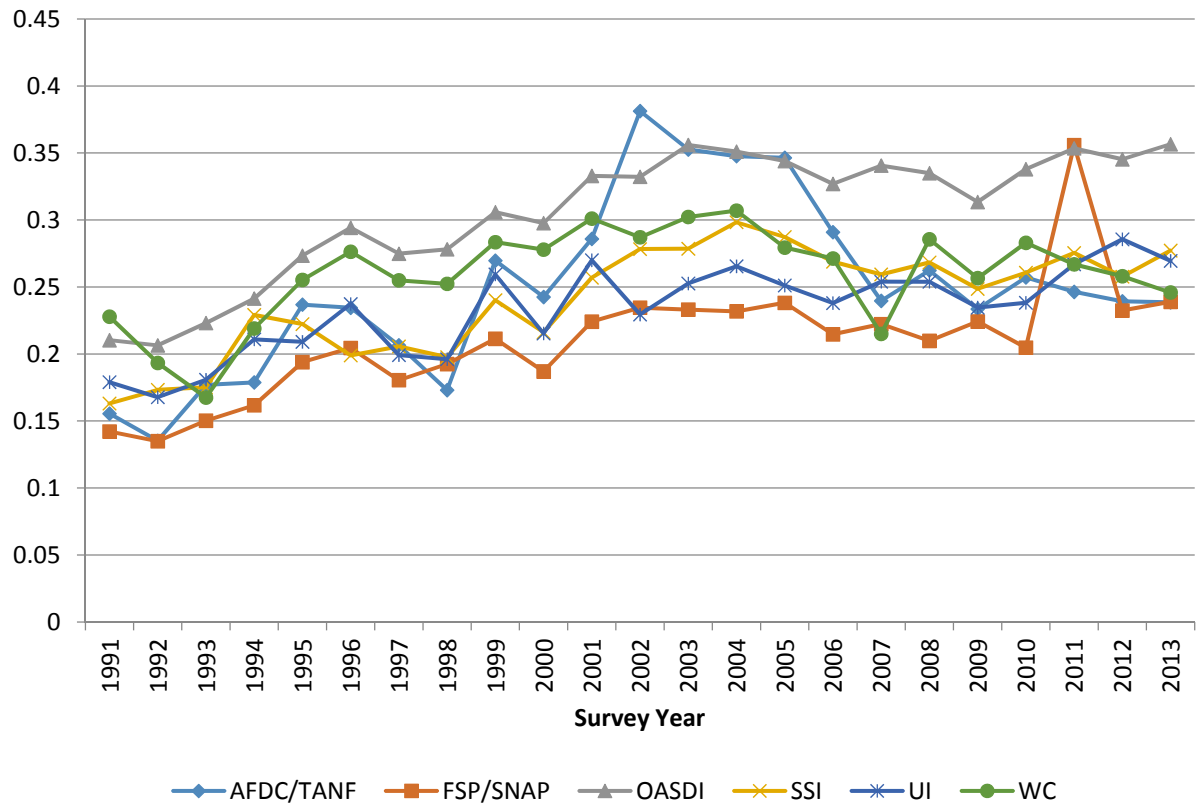


Figure 3
Item Nonresponse Rates in the Survey of Income and Program Participation (SIPP)
by Transfer Program, Calculated as Share of Dollars Reported in Survey that is Imputed,
Excluding Imputation using Previous Wave Information

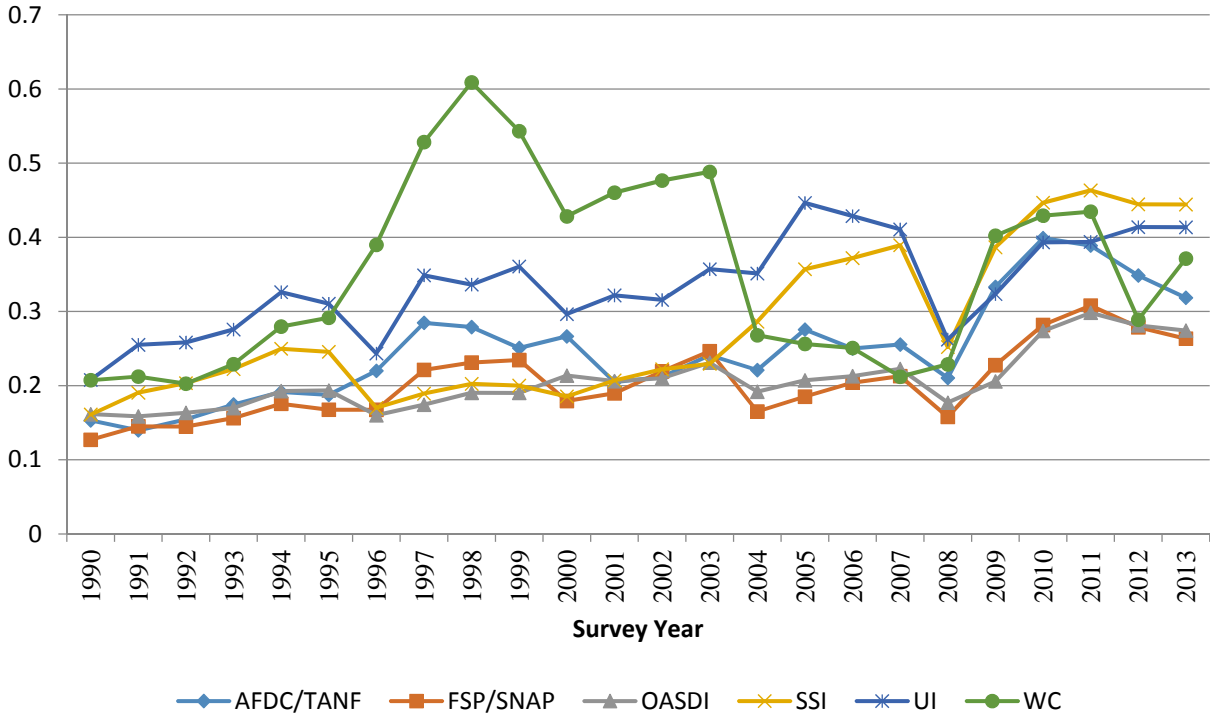


Table 1
Proportional Bias in Survey Estimates of Mean Program Dollars and Months Received,
by Program and Survey, 2000-2012

	AFDC/TANF	FSP/SNAP	OASI	SSDI	SSI	UI	WC	NSLP	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	-0.180
SIPP	-0.232	-0.165	-0.008	0.041	0.023			0.141	-0.197

Notes: Each cell reports the average dollars/months proportional bias for the specified program and survey in the 2000-2012 period. The transfer programs are: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), which is combined with General Assistance in the case of the CE and the ACS; the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP); Social Security, including Old-Age and Survivors Insurance (OASDI), and Social Security Disability Insurance (SSDI); Supplemental Security Income (SSI); Unemployment Insurance (UI); Workers' Compensation (WC); National School Lunch Program (NSLP); and Women, Infants, and Children (WIC). The surveys are: American Community Survey (ACS), Consumer Expenditure (CE) survey, Current Population Survey (CPS), Panel Study of Income Dynamics (PSID), and Survey of Income and Program Participation (SIPP).

Table 2
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, where each observation is a year. The number of years varies across survey and program with as many 45 years for OASI in the CPS (1967-2012, 1969 missing) and as few as 12 for the ACS (2000-2011). 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 3
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 used for the estimates in column 2 are averages of 1983 and 1984 average monthly participation for the entire U.S. We also assume OASDI participation is the sum of OASI and SSDI participation.

Table 4
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	-0.096	na	na	na
	Public Assistance	-0.154	-0.022	-0.529	-0.705
CPS	Food Stamps	-0.049	-0.067	-0.267	-0.382
	Public Assistance	-0.106	-0.057	-0.563	-0.726
SIPP	Food Stamps	-0.056	-0.020	-0.121	-0.197
	Public Assistance	-0.100	-0.043	-0.584	-0.727

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.

Appendix Table 1

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.

Appendix Table 2
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.