

NBER WORKING PAPER SERIES

COMBINING ADMINISTRATIVE AND SURVEY DATA TO IMPROVE INCOME
MEASUREMENT

Bruce D. Meyer
Nikolas Mittag

Working Paper 25738
<http://www.nber.org/papers/w25738>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2019

Any opinions and conclusions expressed here are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the National Bureau of Economic Research. We would like to thank the Alfred P. Sloan, Russell Sage and Charles Koch Foundations for their support. Mittag acknowledges support from the Czech Academy of Sciences (RVO 67985998), the Czech Science Foundation (16-07603Y) and Charles University (UNCE/HUM/035).

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Bruce D. Meyer and Nikolas Mittag. 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.

Combining Administrative and Survey Data to Improve Income Measurement
Bruce D. Meyer and Nikolas Mittag
NBER Working Paper No. 25738
April 2019
JEL No. C81,D31,I32,I38

ABSTRACT

We describe methods of combining administrative and survey data to improve the measurement of income. We begin by decomposing the total survey error in the mean of survey reports of dollars received from a government transfer program. We decompose this error into three parts, generalized coverage error (which combines coverage and unit non-response error and any error from weighting), item non-response or imputation error, and measurement error. We then discuss these three sources of error in turn and how linked administrative and survey data can assess and reduce each of these sources. We then illustrate the potential of linked data by showing how using linked administrative variables improves the measurement of income and poverty in the Current Population Survey, focusing on the substitution of administrative for survey data for three government transfer programs. Finally, we discuss how one can examine the accuracy of the underlying links used in the combined data.

Bruce D. Meyer
Harris School of Public Policy
University of Chicago
1307 E 60th Street
Chicago, IL 60637
and NBER
bdmeyer@uchicago.edu

Nikolas Mittag
Economics Institute of the
Academy of Sciences of
the Czech Republic (CERGE-EI)
Politických vězňů 7
Praha
Czech Republic
and Charles University in Prague
nikolas.mittag@cerge-ei.cz

1. Introduction

Large, nationally representative surveys are among the key sources of information for both academics and policy makers. Household surveys are the basis of official rates of poverty, unemployment and inflation and many other official statistics in the U.S. Survey estimates not only provide information that guides policy choices, but are also directly used to allocate federal funds. Policy makers often rely on detailed household characteristics from surveys for purposes such as improving the targeting of policies. Such information is also essential to academic researchers studying household decisions and the effects of policies. This situation makes the collection of large, detailed and accurate data indispensable. However, recent research has documented an alarming and often growing extent of survey error in key measures, such as income, education and health, in important household surveys. See Bound, Brown and Mathiowetz (2001) for a review of earlier studies and Meyer, Mok and Sullivan (2015) for the recent literature on income and transfer programs.

This chapter reviews how combining administrative and survey data can improve survey accuracy. Building on an overview of the growing applied literature that links administrative records to survey data, we argue that linking administrative records to household surveys can be used to measure the importance of each error source and suggest ways to reduce them. Recent projects have demonstrated these potential benefits of data linkage by linking administrative microdata to U.S. household surveys, such as the Current Population Survey (CPS), the American Community Survey (ACS) and the Survey of Income and Program Participation (SIPP). Databases that contain large parts or the entire survey population of interest are increasingly becoming available from administrative records, private companies, or public sources such as the internet. Advances in computation and methods to link data sources have not only greatly reduced the cost of data linkage, but also increased the accuracy of the linked data. These advances make it possible to implement linkage broadly and frequently enough to continuously evaluate and improve surveys. Thereby, linking administrative records to survey data offers many promising ways

to alleviate the problem of survey error and provides us with the accurate and detailed data necessary to design good policies. Focusing on data quality and its policy applications, this chapter is closely related to two sets of other chapters. Data quality issues are central to chapters 1, 4, 5 and 7. The usefulness of linked data for policy is emphasized in several chapters, but especially chapter 14. We specifically emphasize recent research on government income transfer programs and discuss the advantages of administrative data and record linkage for data users in academia and policy. Thereby, we complement chapters 1 and 2 of this book, which put more emphasis on the perspective of data producers and statisticians.

Linking administrative records to surveys at the household or individual level allows us to add variables to the survey. This process can add information on questions that were not asked, or eliminate the need for some existing questions, reducing respondent burden. Here, we are concerned with measuring and improving the accuracy of survey responses, so we focus on cases where data linkage is used to add a second measure of a variable that is included in the survey data, but measured with error. If the administrative data contain accurate measures that are comparable to the survey definitions and the data sources can be linked with little error, then the variable from the linked administrative records can be considered a measure of truth. We focus on such cases, which heavily rely on high quality administrative records and linkages as discussed in section 3 of this book, as well as on researchers to ensure and assess the accuracy of the linked data as described in section 2 of this volume. See Celhay, Meyer and Mittag (2018a), Courtemanche, Denteh and Tchernis (forthcoming) and Meyer and Mittag (forthcoming) for evidence on data accuracy for specific linked data sets. Many of the methods we review can still be used as long as errors in the linked administrative variable are sufficiently rare that it provides a reasonable approximation to truth. In section 7, we discuss this issue further and refer to methods for cases in which both survey and administrative measures are subject to error.

We focus on the case of U.S. government transfers, where the linked administrative measure can often provide a measure of true program receipt including the amount received. We do not mean to argue that administrative records or transfer program records are more accurate in general. Niehaus and Sukhtankar (2013) provide an impressive example to the contrary. However, by virtue of being paid from government resources, transfer payments usually leave an accurate paper trail to ensure accountability. The administrative records are usually scrutinized to avoid fraud both on behalf of the government agency administering the program and the individuals who receive the transfers. The records often contain official identifiers, which facilitate linkage. Therefore, linking program records can provide us with an accurate measure of a survey concept. Past research on government transfer programs, such as Medicaid (Davern et al., 2008), the Supplemental Nutrition Assistance Program (SNAP, formerly The Food Stamp Program, see Bollinger and David, 1997, Celhay, Meyer and Mittag 2018a, Harris, 2014, Kirlin and Wiseman 2014, Marquis and Moore 1990, Meyer, Mittag and Goerge 2018, Taeuber et al. 2004), cash welfare (Celhay, Meyer and Mittag 2018a, Lynch et al. 2007) and social security and pensions (Nicholas and Wiseman 2009, 2010, Gathright and Crabb 2014, Bee and Mitchell 2017a,b), has used data linkage to demonstrate substantial survey error that has important effects on survey estimates. Both policy makers and academics heavily rely on the data in which these studies document sizeable error. Transfer programs are key tools of the government to combat poverty and inequality, so understanding their consequences is important to academics examining the distribution of income, work incentives, and the well-being of recipients.

We use the case of transfer receipt to show how data linkage can improve the accuracy of survey data. We first show how linked data can be used to estimate total survey error and decompose it into key components, such as generalized coverage error (the combination of coverage error, unit non-response error and errors from weighting), item non-response error and measurement error. Knowing the magnitude of different errors should be the first step in a plan to reduce them. We then discuss each of these three error sources in sections 3-5. For each source, we summarize how linked data can be used to

examine the nature and consequences of errors, review the evidence on the extent of error and discuss how data linkage can contribute to improving the problem. We provide an illustration of how data linkage can improve measures of poverty and the income distribution in section 6. In section 7, we discuss likely effects of errors in the linked variable and review ways to assess and improve data accuracy.

2. Measuring and Decomposing Total Survey Error

Recent research has documented survey error from non-sampling sources such as non-response and measurement error that is often more problematic than sampling error. Nonetheless, survey producers routinely provide users with information on sampling error, but data users usually do not have enough information about non-sampling error to assess its impact on their estimates or account for it. To understand and address the problem as well as to allow researchers to assess the reliability of their estimates, we need a framework to catalogue and measure non-sampling error. Decomposing survey error according to its sources can provide a joint assessment of multiple sources of error in the same survey, allowing easier and more meaningful comparisons.

Understanding the importance of each error source is crucial to reduce overall survey error in a cost-effective manner. For example, survey producers often spend considerable resources to increase response rates, even though the impact on final estimates and non-response bias seems to be small (Groves, 2006; Groves and Peytcheva, 2008). Reducing non-response may also amplify measurement error (Sakshaug, Yan and Tourangeau, 2010; Tourangeau, Groves and Redline, 2010), which we find to be a much larger source of bias in mean transfer dollars received than all other error components (Meyer and Mittag, 2019b). This observation raises the question whether the same overall error reduction could be achieved at a lower cost by shifting some resources toward the reduction of measurement error. To answer such questions, we need to understand the magnitude of each error source, as well as how the measures taken and funds spent affect the error from each source. This requires methods to measure survey error and its components on a routine basis.

To illustrate our points, we rely on a framework and examples from some of our recent work. In Meyer and Mittag (2019b), we argue that administrative records that cover the entire population of interest and data linkage enable us to estimate total survey error (TSE) and its components (Groves and Lyberg, 2010). In particular, we define measures of total non-sampling survey error in estimates of the sum or mean of a variable as well as components of this error due to survey generalized coverage, item non-response and measurement error. Accurate records on the entire population linked to the survey data can provide the measures of truth required to estimate these measures of error. Thereby, data linkage provides an inexpensive and easily implementable way to analyze and summarize survey error.

The unlinked population data enable us to calculate the target amount that the survey should yield. Administrative records usually include individuals that the survey does not intend to cover and records that cannot be linked to the survey data due to linkage problems, so estimating the target amount likely requires some adjustment. For example, in Meyer and Mittag (2019b), the statistic of interest is average dollars received from a welfare program. The surveys that we use do not include individuals in group quarters and the homeless. We adjust for this intentional difference by subtracting payments to the two groups from the total amount paid according to the unlinked administrative data. We adjust for linkage problems such as linkage failure in a similar way. Both adjustments are likely typical in that they contain components that can be calculated from the administrative data (such as payments to the homeless) and components that need to be estimated from other surveys (such as payments to individuals in group quarters). We estimate the TSE in average dollars of transfer payments, ε_{TSE} , as the difference between the estimate of average dollars received according to the survey reports in the linked data and average dollars paid according to the unlinked administrative records adjusted for intentional coverage differences and linkage problems.

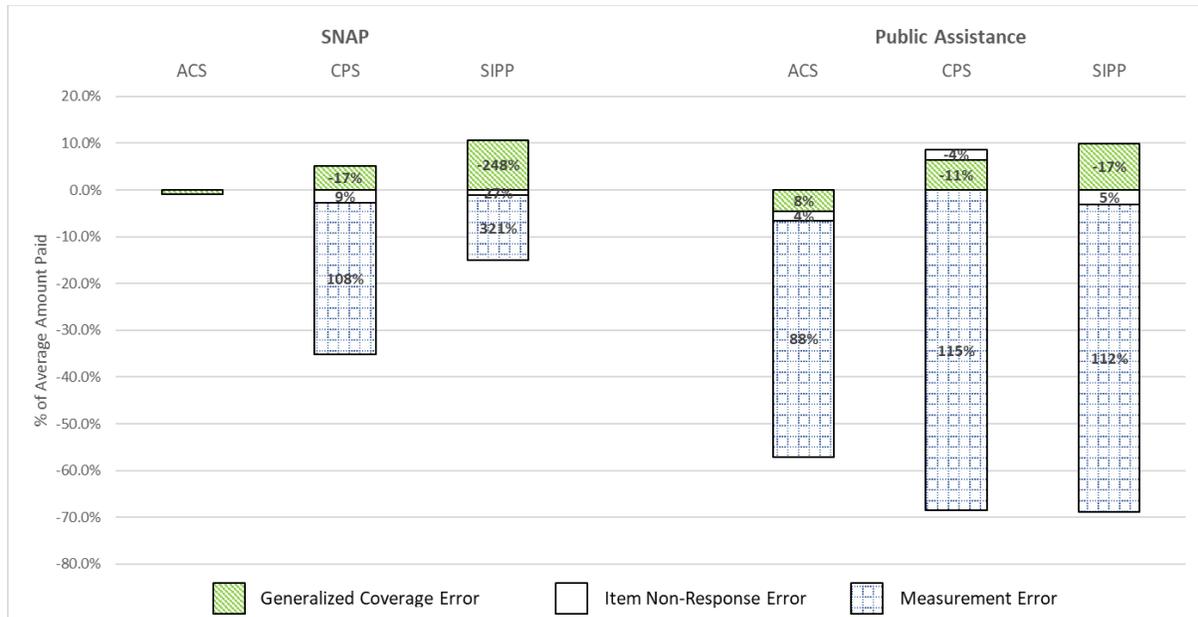
We then decompose TSE into generalized coverage error, item non-response error and measurement error:

$$\varepsilon_{TSE} = \varepsilon_{GCE} + \varepsilon_{INR} + \varepsilon_{ME}$$

Generalized coverage error, ε_{GCE} , is the difference between average dollars actually received by those in the survey population and the target amount of the survey in our case. More generally it is the difference between the statistic of interest, such as the mean of a variable, in the population the survey sample actually captures and the same statistic in the population that the survey intends to represent. We can estimate generalized coverage error as the difference between the weighted average of the administrative variable in the linked data and the target amount for the survey.

Item non-response error, ε_{INR} , is the contribution to TSE from the difference between the statistic of interest according to imputed and accurate values of non-respondents. We estimate item non-response error from the linked data as the weighted sum of the difference between the imputed survey responses and the accurate values from the administrative data for item non-respondents divided by the population

Figure 1: TSE Components in Percent of Average Dollars Paid and as Shares of TSE in New York State



Source: Figure 1 in Meyer and Mittag (2018a).
 Note: The size of the bars indicates the error amount as a percentage of average dollars received per household by the relevant population (which varies across surveys). The numbers in the bars provide the error due to each component as a percentage of total survey error for the respective survey and program. All estimates are based on the linked NY sample using survey weights adjusted for PIK probability. Our samples include survey year 2008-2012 (ACS), 2008-2013 (CPS) and wave 10-12 of the 2004 SIPP combined with wave 1-14 of the 2008 SIPP panel. The ACS does not ask for SNAP amounts, so we can only estimate coverage error. The SIPP is not claimed to be representative of NY state, so the estimates of generalized coverage error for the SIPP should be interpreted with caution.

size. Finally, measurement error, ε_{ME} , is the contribution to TSE from the difference between the statistic of interest according to survey reports and true values among respondents. Just as item non-response error, we can estimate it from the linked data as the weighted sum of the difference between survey reports and the accurate administrative values among respondents divided by the population size. We further decompose item non-response and measurement error into two parts due to misclassification, false positives and false negatives, and a part due to errors in amounts among those correctly classified. This division is useful for the many important variables that combine a continuous part with a mass point at zero, such as income or expenditure and their components. The decomposition can be extended to other TSE components such as processing error and other measures of TSE, such as mean squared error.

In Meyer and Mittag (2019b) we estimate and decompose survey error in average transfer dollars received from SNAP and Public Assistance (cash welfare) in the NY sample of the ACS, CPS and the SIPP using administrative records covering 2007 to 2012. TSE is substantial both in absolute and relative terms in all cases except for SNAP in the SIPP. All three surveys understate transfer dollars received, with larger overall error for Public Assistance than for SNAP. TSE ranges from missing 4.3 percent of dollars received (SNAP in the SIPP) to missing almost two out of three dollars of Public Assistance in the CPS. Figure 1 presents the results of our TSE decomposition. The size of the vertical bars indicates survey error as a fraction of the target amount. The numbers in each error band provide the error due to each component as a percentage of TSE for the respective survey and program. TSE is negative in all cases, but some error components are positive and thus reduce overall error. The sign and the magnitude of the error components vary across surveys and programs, which makes the composition of TSE differ substantially across the columns of Figure 1. The only component that leads surveys to understate dollars received in all cases is measurement error. For transfer dollars received, measurement error is by far the largest source of error. Further decomposing it into binary and continuous parts shows that it is mainly due to underreporting of binary receipt. Our estimates of generalized coverage and item non-response error are

much smaller and often bias estimates in opposing directions. The TSE due to these two components combined is less than 10 percent of the target amount in all cases.

3. Generalized Coverage Error

Generalized coverage error is a combination of several types of error. The sampling frame may not include all units in the population of interest (frame error) and some sampled units may not respond to the survey at all (unit non-response error). In addition, survey processing, such as adjusting the sample weights to correct for unit non-response, may introduce additional error. We focus on studies that link microdata from surveys to administrative records to assess and improve the problem of survey representation. If the administrative records include the entire population of interest and data linkage is accurate or one can adjust for missed links, then the linked data are representative of the population of interest. Then, data linkage enables us to observe the same variable for the population and the survey sample. Therefore, we can directly compare survey estimates to their population values without any concerns for comparability of what the variables in the two data sources measure. Thereby, data linkage provides us with a simple, but powerful tool to analyze survey representativeness and coverage in the aggregate. Note that this does not require the variable in the administrative data to be an accurate measure. If the variable from the administrative data is measured with error, we can still estimate the population size of linked units and compare it to the actual population size in the administrative data. The same idea can be applied to subpopulations of interest, such as single parent households or the elderly as long as they can be identified in the administrative records.

A few recent studies have gone even further by linking the administrative records to the sampling frame or all sampled units rather than survey participants only. Linking directly to the sampling frame is more complicated, because usually little information beyond addresses is available for units that were not sampled or did not respond. However, this type of data linkage allows us to assess the problem of survey representativeness in much more detail. First, it makes it possible to isolate the different sources of error

that lead to generalized coverage error. For example, frame error is the difference between the statistic of interest according to the administrative variable linked to the survey frame and according to the unlinked population data. Unit non-response error can be estimated from the records linked to unit non-respondents. The effect of the weight adjustment is the difference in the statistic of interest calculated using the adjusted and the original sample weights. Second, it allows researchers to study these sources of error at the household or individual level. For example, Bee, Gathright and Meyer (2017) link the 2011 CPS Annual Social and Economic Supplement (CPS ASEC) frame to IRS Form 1040 for tax year 2010 using address information from both sources. They use the resulting linked data to compare characteristics of respondents and non-respondents obtained from the tax records. These characteristics include income, self-employment status, marital status, presence and number of children, and the receipt of pensions and certain government benefits.

The preliminary results from a few recent studies that have used data linkage to examine survey representation and coverage are encouraging both in terms of the usefulness of this approach and in terms of survey accuracy. To our knowledge, Meyer and Mittag (2019b) is the only study of this kind that examines transfer programs. As the results in Figure 1 show, they find that the coverage of program recipients in all three surveys is high, at least compared to the magnitude of other sources of error. Bee, Gathright and Meyer (2017) focus on unit non-response and assess the consequences for measures of income in the CPS ASEC. They are unable to reject that the distribution of household income among non-respondents is the same as that of respondents, though the groups differ on other characteristics such as marital status and number of dependents. Quite remarkably, an analysis of the SIPP by Mattingly et al. (2016) using the same approach of linking tax records to respondent and non-respondent addresses finds nearly the same result of no significant differences between respondents and nonrespondents at a wide range of percentiles of income, but does not examine other characteristics of the two groups. Brummet et al. (2017) use a similar approach to examine income in the Consumer Expenditure Survey.

Both linking administrative records to the sample frame, and to respondents only, can directly point to methods to improve survey representation. The problem of survey coverage is notoriously difficult to study, so that even descriptions of who is missing from the survey can help to address the problem. The methods described above yield a detailed description of the geographic distribution of units that are missing from the survey. If the administrative data include individual or household characteristics, then they also provide estimates of the demographic characteristics of those missing from the survey. This information can be used to improve the survey at the design stage, by improving the sampling frame or increased efforts to include groups that are known not to respond. It can also be used ex post by adjusting the sample weights to make information from the two data sources match.

4. Item Non-response and Imputation Error

Item non-response is the problem that some survey respondents may refuse to answer some questions or sections of the survey, so that some information is missing for these observations. Survey agencies often fill in these values using imputation methods such as the hot deck, see e.g. Little and Rubin (2002) for an overview. Applied researchers usually deal with the problem of item non-response either by using only respondents with complete interviews or by including the imputed values of item non-respondents. For consistency, both strategies assume that item non-response is conditionally random (MAR). Data linkage can provide an accurate measure of the response for both respondents and non-respondents in a survey. Thereby, it allows us to test MAR, examine how respondents differ from item non-respondents as well as how imputed values differ from accurate responses and to directly estimate the bias from each strategy.

Item non-response is often found to be related to other observable characteristics. However, the crucial question for the consistency of estimates is whether respondents differ in their true response conditional on the observed characteristics in the model of interest. See Heitjan and Rubin (1991) and Heitjan (1994) for discussions. The availability of an accurate measure for the entire sample in linked data

makes it possible to examine how item non-respondents and respondents differ in terms of the true response. For instance, Bollinger, Hirsch, Hokayem and Ziliak (forthcoming) and Chenevert, Klee and Wilkins (2016) compare linked tax records of respondents and non-respondents to the question on earned income. Both reject the assumption that item non-response is conditionally random.

In Celhay, Meyer, and Mittag (2018a), we use administrative records on SNAP and Public Assistance linked to the ACS, CPS and SIPP to examine non-response to the question on program receipt. We conduct two tests of MAR: First, we test whether the probability of program receipt among item non-respondents differs from the overall population after conditioning on covariates typically included in models of program receipt. Second, we test whether the conditional distribution of program receipt differs between non-respondents and the entire population. Both tests provide evidence that non-respondents differ systematically from the entire population even conditional on many covariates. Thus, excluding item non-respondents will bias not only unconditional population statistics, but also estimates of models that condition on these covariates. More specifically, the results in Celhay, Meyer and Mittag (2018a) show that item non-respondents are more likely to be program recipients than the overall population both unconditionally and conditional on key covariates. As such, estimated receipt rates are likely to be biased when excluding item non-respondents. We also find that item non-respondents and respondents differ in how variables, such as income, relate to program receipt. We show that coefficient estimates associated with these variables are biased in models of program receipt that do not include item non-respondents.

Imputation allows researchers to include item non-respondents in such analyses. This solution avoids the sample selection bias from excluding item non-respondents, but may introduce new bias due to imputation error. Unfortunately, little is known about the accuracy of imputations and the merits of different imputation procedures, see e.g. Andridge and Little (2010) and Little and Rubin (2002). Recent studies provide evidence that imputation can induce substantial error in survey data (Meyer, Mittag, and Goerge, 2018; Celhay, Meyer, and Mittag 2018a) and in estimates derived from them (Lillard, Smith, and

Welch, 1986; Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006). Celhay, Meyer and Mittag (2018a) use linked data to examine how imputations of transfer receipt differ from the accurate administrative variable. They reject the hypothesis that imputed values correctly reproduce the joint distribution of program receipt and covariates commonly used in studies of transfer programs. In line with the results in Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006), the differences are particularly severe for variables that were not used in the imputation procedure.

This result raises the question whether data users should use or exclude the imputed observations. There is considerable disagreement on this issue, see Angrist and Krueger (2001) for a discussion. Our results from Celhay, Meyer, and Mittag (2018a) above imply that neither including imputations nor excluding item non-respondents yields consistent estimates. We use the linked data to directly estimate and compare the bias from each strategy for models of transfer receipt. We find the bias to be smaller when excluding item non-respondents than when using their imputed values. This finding suggests that it may be better to use only respondents than to include imputed values. However, the magnitude of the bias terms also depends on the model of interest and the imputation procedure, so it is not clear whether this finding generalizes. Nonetheless, the results in Celhay, Meyer, and Mittag (2018a) can provide some guidance, because one would expect the bias from non-response to be determined by the differences between respondents and non-respondents and the bias from imputation error to depend on how the imputed values differ from the true values.

Linked data can help to rectify the problem of item non-response both directly and indirectly. Linked data can provide us with the information on the extent of error among the imputed observations that is needed to make informed decisions, such as whether to use imputed values or not. The accurate measure of the response that linked data can provide also makes it possible to evaluate the accuracy of different imputation procedures. Studying the merits and disadvantages of different imputation methods can help to refine existing procedures and choose the optimal method in specific cases. The results on match bias

(Hirsch and Schumacher, 2004; Bollinger and Hirsch 2006) and the mounting evidence that item non-response is not conditionally random also indicates that it is important to incorporate information beyond key survey variables in the imputation procedure. Data linkage allows survey producers to incorporate information from the accurate administrative records in the imputation procedure. See Davern, Meyer and Mittag (forthcoming), Mittag (2019) or Blackwell, Honaker and King (2017) for further discussion. Linked data may even make it possible to replace the missing values with those from the administrative data, see section 6 for an illustration.

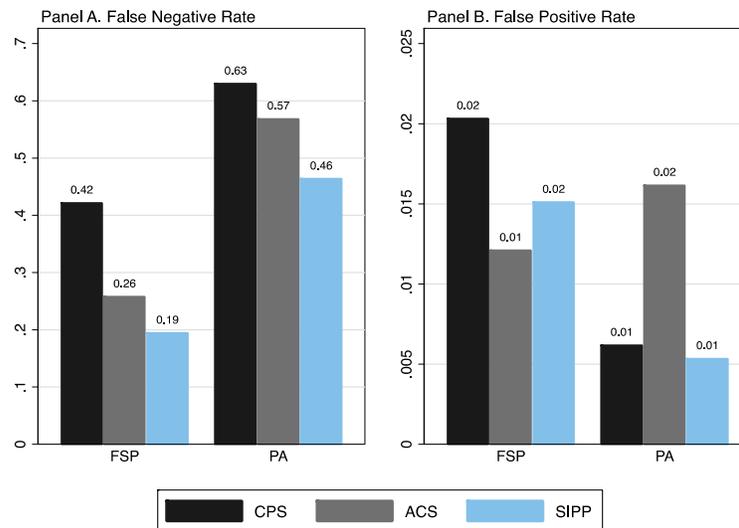
5. Measurement Error

Measurement error in surveys is defined as the difference between the value of a variable of interest obtained from survey responses and the true value of the same variable. Such survey response errors have been documented for many different variables, such as income (Abowd and Stinson, 2013; Bound and Krueger 1991; Bollinger 1998; Dahl, DeLeire and Schwabish 2011), education (Black, Sanders, and Taylor 2003) or drug use by teenagers (Johnson and Fendrich 2005). The main challenge of studying measurement error in surveys is the availability of a variable's true value. Unlike studies that compare aggregate numbers from survey data and official records (Meyer, Mok and Sullivan, 2015), linking survey data to administrative records allows studying measurement error at the individual level, significantly expanding the scope of analysis. Contrary to aggregate comparisons, studies that use linked data at the individual level can separate the measurement error component from other sources of survey error such as frame, coverage or survey non-response error (Meyer, Mok, and Sullivan 2015; Meyer and Mittag 2019b). An accurate measure of the response at the individual level also makes it possible to explore the nature and consequences of these errors in greater detail, for example by examining their distribution and relation to other variables. Finally, data linkage enables us to study the extent of both over- and underreporting in addition to net reporting. For the example of binary variables such as transfer receipt, these data make it possible to observe program recipients that do not report participation (false

negatives) and non-recipients that report participation (false positive). Using linked data, researchers can estimate how much of the differences between surveys and official records are due to underreporting of government transfers and how much is offset by overreporting. This simple distinction is relevant since the two types of error may cancel out in aggregate comparisons of survey receipt, clouding the magnitude of measurement errors in surveys. In addition, the sum of the two error rates rather than the net error rate determines the bias in common models--see Hausman, Abrevaya and Scott-Morton (1998) for an example.

As pointed out in the introduction, studies of various welfare programs have found shockingly high error rates in the report of program receipt in different surveys and programs. For example, sixty percent of welfare recipients in the CPS and the same share of pension recipients in the ACS fail to report receipt (Meyer and Mittag 2019a; O’Hara et al. 2016). Celhay, Meyer, and Mittag (2018a) link multiple surveys

Figure 2: Error Rates in Reported Receipt of SNAP and Public Assistance in the NY CPS, ACS, and SIPP



Source: Celhay, Meyer, and Mittag (2018a).

Notes: Rates of misreporting program receipt among recipients (Panel A) and non-recipients (Panel B) of Food Stamps (FSP or SNAP) and Public Assistance (PA) according to the linked administrative program receipt variable. All estimates are based on the linked NY sample using survey weights adjusted for PIK probability. Our samples include survey year 2008-2012 (ACS), 2008-2013 (CPS) and wave 10-12 of the 2004 SIPP combined with wave 1-14 of the 2008 SIPP panel.

to administrative records from NY State covering 2007 to 2012. They report that the probability of a false negative response of SNAP receipt varies from 19% in the SIPP to 42% in the CPS, while the false positive

rate varies from just above 0.5% for cash welfare in the SIPP to just above 2% for SNAP in the ACS. Figure 2 summarizes these results.

These results are problematic for estimates of aggregate statistics of receipt, such as take-up rates or total dollars received. These biases can be large compared to other well-studied sources of survey error such as imputation or non-response (Celhay, Meyer and Mittag, 2018a). Even worse, the studies mentioned above show that response errors are not independent of other respondent characteristics or the true value of the variable, so that they likely bias both causal and descriptive estimates obtained from survey data (e.g. Bollinger and David 1997; Bound, Brown, and Mathiowetz 2001; Meyer, Mittag and Goerge, 2018; Meyer and Mittag, 2017, 2019b). For example, Meyer, Mittag and Goerge (2018) find that survey error leads to biased estimates of the determinants of program receipt. Using linked data their results indicate that the survey data understate participation by single parents, non-whites, and the elderly and how participation declines as incomes rise.

Using linked data also enables researchers to study the nature of measurement error by exploring the determinants of reporting errors in surveys. To understand and address the problem of measurement error it is important to examine how respondent characteristics, behavior and survey design affect response accuracy. Inaccurate responses may arise from cognitive problems such as recall or respondent cooperation. Survey features related to measurement error include the recall period, choice of survey mode, the role of interviewers, and data post processing. Reviews of the main theories of misreporting can be found in Sudman and Bradburn (1974), Sirken (1999) and Bound, Brown, and Mathiowetz (2001). In Celhay, Meyer, and Mittag (2018b), we use administrative records linked to the ACS, CPS and SIPP to examine several hypotheses related to respondents' cognition, their cooperation and survey design.

The literature on respondents' characteristics usually discusses how cognitive aspects of respondents and their situation affect survey response (e.g., Tourangeau 1984). An important source of error arises from memory or recall error, when respondents are asked about past events (Groves et al. 2009, pp. 231).

In Celhay, Meyer, and Mittag (2018b) and Meyer and Mittag (forthcoming), we show that the length of time since last program receipt substantially increases measurement error in reported program receipt. Such recall error is among the most important sources of measurement error we find, but this relationship may depend on the question at stake and survey design. Celhay, Meyer, and Mittag (2018b) also provide evidence of survey error arising from confusion regarding the dates of program receipt. This confusion causes some respondents to report program receipt that happened either before or after the reference period (telescoping). We find that households that are more dependent on government transfers are better reporters on average, which supports the hypothesis that salience improves reporting.

The quality of survey data is also affected by respondents' motivation or attitudes towards surveys. Linked data enable us to test whether respondents who are more cooperative are less likely to misreport in surveys. For example, Bollinger and David (2001) use SIPP data linked to administrative records to show that reporting errors predict later attrition. Extending their analysis in Celhay, Meyer and Mittag (2018b), we find that respondents who have higher imputation rates in other sections of a survey are more likely to misreport participation in questions of program receipt. Another frequently discussed reason for misreporting is the unwillingness to provide socially undesirable answers. In the case of government transfers, Celhay, Mittag, and Meyer (2017d) provide evidence that stigma associated with participation in welfare programs can explain why people fail to report participation to interviewers. Finally, many survey design features are hypothesized or known to be related to reporting errors, although the empirical evidence is often mixed (Groves, 2004). Linked data can be used to examine how survey design features affect response accuracy. However, this requires exogenous (conditionally random) variation in survey design features.

Linked data can help to improve the problem of measurement error in various ways. The insights gained from linked data can improve the way we use surveys and how we interpret survey estimates. Linked data allow us to describe the extent and predictors of errors and provide estimates of the bias for

specific cases. This can help survey users assess the accuracy of their estimates and the likely biases, particularly in the absence of theoretical results. If formulas for (asymptotic) bias from measurement error are available, estimating the bias is helpful to gauge the applicability of the formulas to finite samples and whether the magnitudes and directions of the bias conform to what the patterns of survey error suggest (Meyer and Mittag, 2017). For example, given the high error rates in reports of transfer receipt found in linked data, survey users should be skeptical of survey estimates of receipt rates. Yet, results from linked data on the bias in models of program receipt suggest that strongly significant estimated associations in a given direction are unlikely to be reversed even when the dependent variable suffers from high and systematic misclassification (Meyer and Mittag 2017, Celhay, Meyer and Mittag, 2018a).

Understanding the nature of survey errors not only helps to assess the bias, but also puts us in a better position to correct it. Several bias corrections and consistent estimators have been proposed in the literature. However, they require additional assumptions on the errors, such as conditional independence. Linked data can inform us whether these assumptions are plausible and hence which correction methods are likely to improve estimates. For example, Meyer and Mittag (2017) find that corrections for misclassification of the dependent variable in binary choice models that require misclassification to be unrelated to covariates can easily make estimates worse if this assumption does not hold. On the other hand, corrections that incorporate information on these correlations from linked data can substantially improve estimates. The insights on the nature of errors and the bias discussed above are particularly useful if results from linked data are available for similar variables in similar surveys. By providing insights on the reasons for misreporting, linked data can also help to gauge the reliability of survey data more generally. Studies such as Celhay, Meyer and Mittag (2018b) point to conditions under which survey errors are particularly problematic that likely hold more generally.

Linked data can also provide us with a better understanding of how survey design affects measurement error. Linking surveys to administrative records on a more routine basis would shed more

light on the effects of (changes in) survey design on response accuracy. A better understanding of the causes and predictors of misreporting that can be gained from linkage studies would also allow survey producers to provide users with more information on the likely accuracy of a given response. For example, we find the share of an individual's responses that required imputation predicts survey errors in the individual's other responses (Celhay, Meyer and Mittag, 2018b). Investigating which paradata predicts response errors and publishing the results could provide data users with better means to assess the likely prevalence of measurement error in the samples they use. An emerging literature also examines how such information can be used to improve the accuracy of estimates derived from error-ridden data (e.g. Da Silva, Skinner, and Kim 2016). Finally, linking more accurate measures of existing variables in surveys can also be used to improve survey accuracy directly. Survey producers could replace the misreported variable by a more accurate administrative variable or combine survey responses and administrative data in synthetic variables if direct substitution is infeasible or undesirable. See Davern, Meyer and Mittag (forthcoming) and Mittag (2019) for further discussion. We review an example of direct substitution in the next section.

6. Illustration: Using Data Linkage to Better Measure Income and Poverty

The sections above describe ways in which data linkage can help us to understand and measure the errors in survey data. Data linkage can also improve survey quality more directly. If an accurate measure of a survey variable can be linked to the data, one can replace the error-prone survey variable by the more accurate linked variable. We use our work that corrects analyses of poverty and the income distribution to illustrate that such direct substitution is feasible and can alter important conclusions obtained from the contaminated survey data. See Bee and Mitchell (2017a) for similar analyses that use direct substitution.

In Meyer and Mittag (2019a), we combine the New York State (NY) sample of the 2008-2013 CPS ASEC—the source of many inequality statistics and the official poverty rate—with administrative records for SNAP, Public Assistance and housing assistance to better measure payments made, and the impact of

the programs on the income distribution and poverty. The linked data confirm that misreporting, particularly underreporting of program receipt, is severe. More importantly, we show that replacing the error-ridden survey variable by the accurate linked variable in several prototypical analyses of poverty and transfer programs, sharply alters the picture of well-being at the bottom of the income distribution and the effects of transfer programs. Using the administrative variables, poverty and inequality are lower, program effects are larger and fewer individuals have been missed by the combination of programs. Incomes below the poverty line are substantially understated in the CPS ASEC. This is particularly severe for incomes below half the poverty line where more than \$1,400 per person from the programs we examine alone are missing in the survey. These missing transfer dollars exceed cash income reported by this group. Although underreporting as a share of income becomes smaller as income rises, substantial dollars are missed even toward the middle of the income distribution.

Underreporting of transfer receipt also makes government anti-poverty policies appear much less effective: The four programs moved a much larger fraction of people out of poverty than the CPS suggests. Including transfers from these programs in the income definition reduces the poverty rate by 2.8 percent according to the CPS reports. This reduction rises to 5.3 percent when using the more accurate administrative measures. The difference is particularly pronounced for housing assistance, where correcting for underreporting triples the poverty reduction. Both the understatement of household income and the poverty reducing effect in the survey are even more pronounced for some subpopulations that are at particular risk of deprivation. The understatement is particularly large for single mothers: correcting for survey errors increases their overall poverty reduction due to the four programs by 11 percentage points, amplifying the poverty reducing effect of public assistance more than 6-fold and that of housing assistance more than 10-fold. In addition, we find that the fraction of non-working single mothers missed by government transfers is much lower than previously reported. This underlines that the coverage of the safety net is better than the survey suggests.

These results show that survey errors, mainly the misreporting of government transfer receipt and amounts, but also non-response and inaccurate imputation, lead to a greatly distorted view of the situation of those with the fewest resources and the effects of transfer programs. However, the results also underline that enhancing survey data with administrative records can provide improved answers to questions of relevance for both policy makers and academics. These analyses were conducted on the most recent vintage of the data at the time of initial writing, so such enhanced data products could be made available on a timely enough basis to inform policy.

7. Accuracy of Links and the Administrative Data

The analyses above consider the linked administrative variables to be accurate versions of the survey concepts. In many cases one will not be able to exactly match the survey concepts or the data linkage will not be error-free. In such less than idealized circumstances, many of the benefits of record linkage we discuss above still apply. See Courtemanche, Denteh and Tchernis (forthcoming) and Meyer and Mittag (forthcoming) for discussions of the merits and perils of examining the accuracy of a specific survey using linked data in less-than-ideal circumstances.

For the analyses of survey representativeness and coverage, one mainly needs the administrative records to be complete and reliably linked to the survey, as we discuss in section 3 above. Similarly, analyses of unit non-response remain meaningful as long as data linkage provides a measure of the survey response that does not differ systematically between respondents and non-respondents. If response status does not predict errors in the linked variable, then statistics such as differences in means or tests whether the distribution of the linked variable differs by response status remain valid. This condition may not hold, for example if the links are (partly) made based on information provided only by respondents. By the same logic, many analyses of item non-response remain valid as long as item non-response status does not predict errors in linkage.

Studying measurement error at the household or individual level requires a measure of “truth” at the same level, so that linkage errors cause bias in such analyses. See Meyer, Mittag, and Goerge (2018) for a discussion of the likely consequences for estimated misreporting rates. If the rates of linkage error are negligible compared to the error rates in the survey responses, comparing the administrative and the survey variable may still yield valuable insights about survey error. If the error rates in the linked variable are too high to be ignored, one can still analyze many of these problems by using methods that rely on two error-ridden variables. These methods are beyond the scope of this chapter, see Abowd and Stinson (2013), Kapteyn and Ypma (2007), Oberski et al. (2017), or Meijer, Rohwedder and Wansbeek (2012) for applications. What can researchers do to assess the extent to which their data deviates from this gold standard and how can they address problems they identify? A growing methodological literature develops methods to examine and address inaccuracies in the administrative data and linkage errors. Chapters 4 to 7 of this book discuss methods to assess the accuracy of administrative records and summarize the conditions under which administrative records are likely accurate. We only provide key references and focus on strategies commonly used in the literature on income and transfer programs.

A common way to examine the overall accuracy of the unlinked administrative records is to compare them to control totals that the administrative records should match in aggregate. For government programs, it is usually possible to obtain aggregate statistics, such as the total number of recipients or the total amount paid out, from independent sources. For example, the total amount paid out by a government program is often published by the agency that administer the program, such as the Department of Health and Human Services or the Department of Agriculture in the U.S. Meyer, Mok and Sullivan (2015) provide information on the sources of aggregate payments for many U.S. transfer programs, and discuss how these numbers can be made comparable to survey estimates.

Validating the survey response also requires matching the concept of the survey question. For the case of transfer reporting, a key issue is matching the reference period of the survey. This matching is

possible if the reference period in the survey is clearly defined and the administrative data contain detailed information on the timing of receipt. See Meyer and Mittag (forthcoming) for a discussion. This issue is less straightforward when validating more complex concepts such as income. One problem is that the two data sources may define income in different ways. Tax records, for example, often contain net income after some tax deductions such as pension contributions, while surveys often do not exclude tax deductions and may ask for gross income. A second issue is that the tax records may include only taxable sources of income (but not nontaxable or in-kind income) and miss informal sources that may be captured in the survey. See Abowd and Stinson (2013) and Bollinger et al. (2015) for discussions of income sources that tax records may not capture.

Finally, obtaining an accurate measure from the administrative data requires data linkage to be error free. If data linkage is deterministic based on official identifiers, as is often the case with administrative records in Europe, such errors can only arise from mis-assigned official identifiers. However, most record linkages in the U.S. use probabilistic record linkage. Chapter 8 provides an overview; see NORC (2011) and Wagner and Layne (2014) for descriptions of the Person Identification Validation System of the U.S. Census Bureau, which was used to link the data in many of the studies discussed so far. An advantage of the Person Identification Validation System is that it links both the administrative and the survey records to a third data source from which common identifiers are obtained. Linking both data sources to such a population register lets the researcher observe whether a record was not linked to the other source, because it was not included in the data (e.g. because the household did not receive the program), or because the record was unlinkable. In the former case, the unit has an identifier from the population register, but this identifier was not present in the other data source. This information makes it likely that the record would have been linked if it were in the other data source, particularly if linkage failure in the other data source is rare. For example, administrative records on transfer receipt often contain linkable identifiers for almost all records. Thus, not finding a survey unit with a linkable identifier in the

administrative data makes it very likely that the unit did not receive the transfer. However, if a survey or administrative record is not linkable, no identifier from the population register is attached to it. This situation makes it possible to examine the extent of linkage failure in both data sources and can often be used to improve or solve the problem of linkage error.

Survey units that cannot be linked present a missing data problem: If they are not missing completely at random, the linked sample is not representative of the survey population. Adjusting for this problem is only necessary if one needs a representative sample of the population and linkage failure predicts the outcome of interest. Otherwise, one may be able to use the linked sample without further problems. If the remaining records are correctly linked, one can still examine survey error in the linked data. However, if one is interested in population estimates or concerned that survey errors are related to linkage failure, one can often use the detailed information in the survey to examine or even address this problem. One can estimate binary choice models where the dependent variable is an indicator for whether or not the survey unit is linkable, to examine whether linkage failures are random and if not, how they vary with characteristics of the unit. This estimation is often informative about likely biases. If linkage failure is random conditional on the variables in the survey, one can use methods for missing data to address it. A common approach is inverse probability weighting (Horvitz and Thompson, 1952; Wooldridge, 2007). However, the assumption that linkage failure is conditionally random is not testable, because the administrative variable is not observed for the survey units that cannot be linked. One way to obtain evidence on whether the procedure yields consistent estimates is to examine whether the linked data reproduce the distribution of other survey variables, most importantly, of survey reports of the variable of interest.

There are usually far fewer unlinkable administrative records than survey units and in some of the applications discussed above, this problem is negligible. For administrative records that could not be linked, we still observe the value of the administrative variable of interest. Therefore, we can examine the

extent and importance of this problem, for example by calculating the share of individuals or payments that are missing from the linked data due to linkage failure in the administrative data. It is also often possible to adjust aggregate numbers, for example as in Meyer and Mittag (2019b).

8. Conclusions

Recent research that links administrative and survey microdata has demonstrated the potential of data linkage to remedy the problem of survey inaccuracy. As the work we review here underlines, data linkage can provide us with evidence on the extent of survey error and its key sources. Understanding the extent of error is important for survey users to gauge the reliability of their estimates. Measuring the error and its components is crucial for survey producers to cost-effectively increase accuracy by spending resources on the error sources that are important and on measures to address them that have been shown to reduce the error. Linked data can also be used to study specific sources of error, such as the key components of generalized coverage error, item non-response error and measurement error that we survey here. By adding an accurate measure of the variable of interest to the survey data, data linkage puts researchers in the unique position to study errors at the level of the unit of observation. As the literature on survey errors in transfer receipt we review above shows, this addition provides unprecedented detail on the nature of errors, which often points to potential remedies to the identified problems.

However, data linkage has yet to reach its full potential. The methods and benefits outlined in this chapter extend beyond studies of government transfers and other income components, because they are applicable whenever an accurate measure can be obtained from another data source. Many extensions are possible. The analyses could be extended to study further error components, such as processing error. The existing studies mainly document survey error in means or totals, but the same or similar methods could be used to study other statistics such as variances or possibly regression coefficients. Most studies in this review are case studies that used a short time period from a few surveys, mainly because most

studies linked a convenience sample based on data availability. The full benefits of such analyses will accrue with more routine application. With data linkage becoming more common, methods to assess the accuracy of linked data and to understand and address the problems that arise will likely improve further. Linking data on the entire U.S. population or at least multiple states will help to examine geographic heterogeneity in survey error and provide larger sample sizes. As the results in Celhay, Meyer and Mittag (2018a) underline, many studies would likely yield more informative results if they were able to link larger samples. Finally, routinely linking surveys to administrative data would allow survey producers to monitor how survey errors change over time and whether the measures taken to reduce them yield the desired results. As the study of poverty and the income distribution in section 6 shows, linking the data on a timely basis would also allow government agencies and policy makers to rely on more accurate information.

References

- Abowd, J.M., and Stinson, M.H. 2013. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics*, 95(5), 1451–1467.
- Abowd, J.M., Stinson, M. and Benedetto, G. 2006. "Final report to the social security administration on the SIPP/SSA/IRS public use file project." U.S. Census Bureau, Longitudinal Employer-Household Dynamics Program.
- Andridge, R.R. and Little, R.J.A. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review*, 78(1): 40–64.
- Angrist, J.D., and Krueger, A.B. 2001. "Empirical Strategies in Labor Economics", in O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Vol 3A. Elsevier: Amsterdam.
- Bee, C.A., and Mitchell, J. 2017a. "The Hidden Resources of Women Working Longer: Evidence from Linked Survey-Administrative Data." In C. Goldin and L.F. Katz (Eds.), *Women Working Longer: Increased Employment at Older Ages*. Chicago: University of Chicago Press.
- Bee, C.A. and Mitchell, J. 2017b. "Do Older Americans Have More Income Than We Think?" U.S. Census Bureau SESH Working Paper #2017-39.
- Black, D.A., Sanders, S., and Taylor, L. 2003. "Measurement of Higher Education in the Census and Current Population Survey." *Journal of the American Statistical Association*, 98(463), 545–554.
- Blackwell, M., Honaker, J. and King, G. 2017. "A Unified Approach to Measurement Error and Missing Data: Overview and Applications." *Sociological Methods & Research*, 46(3), 303-341.
- Bollinger, C.R. 1998. "Measurement error in the Current Population Survey: a nonparametric look." *Journal of Labor Economics*, 16(3), 576–94.

- Bollinger, C.R. 1998. "Measurement error in the Current Population Survey: a nonparametric look." *Journal of Labor Economics*, 16(3), 576–94.
- Bollinger, C.R. and David, M.H. 1997. "Modeling Discrete Choice with Response Error: Food Stamp Participation." *Journal of the American Statistical Association*, 92 (439) pp. 827-835.
- Bollinger, C.R. and David, M.H. 2001. "Estimation with Response Error and Nonresponse: Food-Stamp Participation in the SIPP", *Journal of Business and Economic Statistics*, 19:2, 129-141.
- Bollinger, C.R., and Hirsch, B.T. 2006. "Match Bias Due to Earnings Imputation: The Case of Imperfect Matching." *Journal of Labor Economics* 24 (3).
- Bollinger, C.R., Hirsch, B.T., Hokayem, C.M., and Ziliak, J.P. forthcoming. "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch." *Journal of Political Economy*.
- Bound, J., and Krueger, A.B. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9(1), 1-24.
- Bound, J., Brown, C. and Mathiowetz, N. 2001. "Measurement error in survey data." In *Handbook of Econometrics*. Vol. 5, eds. J.J. Heckman and E. Leamer, Chapter 59, 3705 – 3843. Amsterdam: Elsevier.
- Brummet, Q., Flanagan-Doyle, D., Mitchell, J. and Voorheis, J. 2017. "Investigating the Use of Administrative Records in the Consumer Expenditure Survey." Center for Administrative Records Research and Applications, U.S. Census Bureau.
- Celhay, P., Meyer, B.D. and Mittag, N. 2018a. "Errors in Reporting and Imputation of Government Benefits and Their Implications" Unpublished Manuscript.
- Celhay, P., Meyer, B.D. and Mittag, N. 2018b. "What Leads to Measurement Error? Evidence from Reports of Program Participation in Three Surveys." Unpublished Manuscript.
- Celhay, P., Meyer, B.D. and Mittag, N. 2018c. "Stigma in Welfare Programs." Unpublished Manuscript.
- Chenevert, R.L., Klee, M.A. and Wilkin, K.R. 2016. "Do Imputed Earnings Earn Their Keep? Evaluating SIPP Earnings and Nonresponse with Administrative Records." U.S. Census Bureau SEHSD-Working Paper 2016-18.
- Courtemanche, C., Denteh, A., and Tchernis, R. forthcoming. "Estimating the Associations between SNAP and Food Insecurity, Obesity, and Food Purchases with Imperfect Administrative Measures of Participation." *Southern Economic Journal*.
- Da Silva, D.N., Skinner, C. and Kim, J.K. 2016. "Using Binary Paradata to Correct for Measurement Error in Survey Data Analysis." *Journal of the American Statistical Association* 111 (514): 526–37.
- Dahl, M., DeLeire, T., and Schwabish, J.A. 2011. "Estimates of Year-to-Year Volatility in Earnings and in Household Incomes from Administrative, Survey, and Matched Data." *Journal of Human Resources*, 46(4), 750–774.
- Davern, M., Call, K.T., Ziegenfuss, J., Davidson, G., Beebe, T.J., and Blewett, L. 2008. "Validating Health Insurance Coverage Survey Estimates: A Comparison of Self-Reported Coverage and Administrative Data Records." *Public Opinion Quarterly*, 72(2), 241-259.
- Davern, M., Meyer, B.D., and Mittag, N. forthcoming. "Creating Improved Survey Data Products Using Linked Administrative-Survey Data." *Journal of Survey Statistics and Methodology*.

- Gathright, G.M.R., and Crabb, T.A. 2014. "Reporting of SSA Program Participation in SIPP." Working Paper, U.S. Census Bureau.
- Groves, R.M. 2004. *"Survey Errors and Survey Costs."* New York: John Wiley & Sons.
- Groves, R.M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (4): 646-75.
- Groves, R.M., Peytcheva, E. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly*. 72: 167-189.
- Groves, R.M., Fowler, F.J. Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R.. 2009. *"Survey Methodology."* Wiley Series in Survey Methods. New York: John Wiley and Sons.
- Groves, R.M., and Lyberg, L. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly*, 74(5): 849–79.
- Hausman, J.A., Abrevaya, J. and Scott-Morton, F.M., 1998. "Misclassification of the dependent variable in a discrete-response setting." *Journal of Econometrics*, 87(2), pp.239-269.
- Heitjan, D.F. 1994. "Ignorability in General Incomplete-Data Models." *Biometrika*, 81, 701-708.
- Heitjan, D.F., and Rubin, D.B. 1991, "Ignorability and Coarse Data." *Annals of Statistics*, 19, 2244-2253.
- Hirsch, B.T., and Schumacher, E. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics*, 22(3): 689–722.
- Horvitz, D.G., Thompson, D.J. 1952. "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association*. 47, 663–685.
- Johnson, T., and Fendrich, M. 2005. "Modeling Sources of Self-Report Bias in a Survey of Drug Use Epidemiology." *Annals of Epidemiology* 15 (5): 381–89.
- Kapteyn, A., and Ypma, J.Y. 2007. "Measurement error and misclassification: A comparison of survey and administrative data." *Journal of Labor Economics*, 25(3), 513-551.
- Kirlin, J.A., and Wiseman, M. 2014. "Getting it Right, or at Least Better: Improving Identification of Food Stamp Participants in the National Health and Nutrition Examination Survey." Working Paper.
- Lillard, L., Smith, J.P., and Welch, F. 1986. "What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation." *The Journal of Political Economy*, 489–506.
- Little, R.J.A. and Rubin, D.B. 2002. *"Statistical analysis with missing data."* 2nd ed., New York: John Wiley.
- Lynch, V., Resnick, D.M., Stavely, J. and Taeuber, C.M. 2007. "Differences in Estimates of Public Assistance Reciprocity Between Surveys and Administrative Records." U.S. Census Bureau Working paper.
- Mattingly, T., Choi, J., Finney, T., Hoop, R., Hornick, D., Nieman, D., Rothhaas, C., Westra, A. and White, M. 2016. "Results of a Nonresponse Bias Analysis Using Survey of Income and Program Participation (SIPP) Addresses Matched to Internal Revenue Service (IRS) Data." Memorandum, U.S. Census Bureau.
- Marquis, K. H., and Moore, J.C. 1990. "Measurement Errors in SIPP Program Reports." In *Proceedings of the 1990 Annual Research Conference*. 721–745. Washington, D.C.: U.S. Bureau of the Census.
- Meijer, E., Rohwedder, S., and Wansbeek, T. 2012. "Measurement error in earnings data: using a mixture model approach to combine survey and register data." *Journal of Business & Economic Statistics*, 30(2).

- Meyer, B.D., Mittag, N. and Goerge, R. 2018. "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation." *NBER Working Paper 25143*.
- Meyer, B.D., Mok, W.K.C. and Sullivan, J.X. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives*, 29(4): 199–226.
- Meyer, B.D. and Mittag, N. 2017. "Misclassification in Binary Choice Models." *Journal of Econometrics*, 200(2), 295-311.
- Meyer, B.D. and Mittag, N. 2019a. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." *American Economic Journal: Applied Economics* 11(2).
- Meyer, B.D. and Mittag, N. 2019b. "An Empirical Total Survey Error Decomposition Using Data Combination." IZA Discussion Paper 12151.
- Meyer, B.D. and Mittag, N. forthcoming. "Misreporting of Government Transfers: How Important are Survey Design and Geography?" *Southern Economic Journal*.
- Mittag, N. 2019. "Correcting for Misreporting of Government Benefits." *AEJ: Economic Policy*, 11(2).
- Nicholas, J. and Wiseman, M. 2009 "Elderly Poverty and Supplemental Security Income." *Social Security Bulletin*, Vol. 69(1).
- Nicholas, J. and Wiseman, M. 2010 "Elderly Poverty and Supplemental Security Income, 2002-2005." *Social Security Bulletin*, Vol. 70(2).
- Niehaus, P. and Sukhtankar, S. 2013. "Corruption dynamics: The golden goose effect." *American Economic Journal: Economic Policy*, 5(4): 230-269.
- NORC. 2011. "Assessment of the US Census Bureau's Person Identification Validation System." NORC at the University of Chicago Final Report presented to the US Census Bureau.
- Oberski, D.L., Kirchner, A. Eckman, S. and Kreuter, F. 2017. „Evaluating the quality of survey and administrative data with generalized multitrait-multimethod models." *Journal of the American Statistical Association*, 112(520): 1477-1489.
- O'Hara, A., Bee, C.A. and Mitchell, J. 2016. "Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records" 2015, American Community Survey Research and Evaluation Report Memorandum Series #ACS16-RER-6, U.S. Census Bureau.
- Sakshaug, J.W., Yan, T. and Tourangeau, R., 2010. "Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items." *Public Opinion Quarterly*, 74(5), pp.907-933.
- Sirken, M. 1999. "Cognition and Survey Research." Vol. 322. Wiley-Interscience.
- Sudman, S., and Bradburn, N.M. 1973. "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association* 68 (344): 805–15.
- Taeuber, C., Resnick, D.M., Love, S.P., Stavely, J. Wilde, P. and Larson, R. 2004. "Differences in Estimates of Food Stamp Program Participation Between Surveys and Administrative Records" Working Paper, U.S. Census Bureau.
- Tourangeau, R. 1984. "Cognitive Science and Survey Methods." *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, 73–100.

- Tourangeau, R., Groves, R.M. and Redline, C.D. 2010. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly*, 74(3):413–32
- Wagner, D., and Layne, M. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." U.S. Census Bureau.
- Wooldridge, J.M. 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." *Journal of Econometrics*, 141(2): 1281–1301.