Differences in Estimates of Public Assistance Recipiency

Between Surveys and Administrative Records

by

Victoria Lynch and Dean M. Resnick[1],
U.S. Census Bureau

Jane Staveley and Cynthia M. Taeuber,
Jacob France Institute, University of Baltimore

January 2008

---

[1] Lead author contact: victoria.a.lynch@census.gov or dean.michael.resnick@census.gov

Executive Summary

Record-check studies going back several decades show surveys tend to underestimate participation in government safety-net programs.[2] Findings from this record-check study of survey households that received public-assistance income from the state of Maryland support previous findings about the scope of the undercount problem and add to the much smaller set of studies on possible explanations. The study described in this paper found that respondents in households that received public-assistance income had trouble correctly answering the question about public-assistance income on the 2001 American Community Survey/Supplemental Survey (ACS/SS01).

- About half (on a weighted basis) of the households that should have been reported as assistance households were not.
- Many of the households missing from the survey estimate of assistance households are missing because respondents did not report the public-assistance income sent to one or more household members from the state.
- Respondents appear to misreport about public-assistance income when their involvement with the program is relatively complex and correctly answering the question is therefore likely more difficult. This study found that respondents were notably more likely to misreport when they were reporting about someone else in the household and when no household member received assistance in the month the survey was conducted.
- We found no statistical evidence of reporting patterns that might be expected if respondents fear stigmatization should they report public-assistance income for someone in the household.

---

[2] For a comprehensive review of this literature see Lynch, (2006).

Table of Contents

Figure and Tables

I.   Introduction

Public policy-makers and social scientists seek to understand why administrative records yield significantly higher counts of participation in government safety-net programs compared with survey estimates. Record-check studies going back several decades typically show large discrepancies between records from administrative databases and surveys covering program participation (Klerman et al. 2005; Moore and Marquis 1989; Goudreau et al. 1984; Hu 1971).[3] They generally show that surveys underestimate participation and often by as much as 50 percent (Klerman et al. 2005; Taeuber et al 2004; Warburton and Warburton 2004; Kapsalis 1991). The underestimation is problematic because individuals and households misclassified as not participating tend to differ systematically from those correctly classified, meaning analyses using survey samples of program participants tend to produce attenuated measures of association and biased statistics (Klerman et al. 2005; Warburton and Warburton 2004; Bollinger and David, 1997; Moore and Marquis, 1992;). Recent studies also suggest that survey underestimation in programs that provide income assistance has grown worse (Klerman et al. 2005; Roemer 1999; Polivka 1997).

We need to better understand survey errors about participation in programs because surveys remain the most valuable type of data for evaluating program outcomes and equitably funding different programs. Types of data each have pros and cons:

- ✓ *Administrative records alone* – counts of participants can be highly reliable (assuming the data provided are complete and are well understood by the tabulators) but the data may be of limited use in studying dynamics of participation since there is often little information about the characteristics of participants and none about non-participants.

- ✓ *Survey data alone* – survey data have information about non-participants and more detailed information about program participants than administrative records but the quality of responses can be unreliable.

- ✓ *Survey data linked to administrative records* – linked data offer the reliability of administrative data and the richness of survey data. However it can be difficult to create and analyze linked files because of the complexity of linking records and adjusting for differences in the universes covered and the definition of constructs, such as "household". Even more problematic is the fact that very few researchers can use linked files because they require stringent privacy protections, limiting their accessibility to a relatively small group of mostly government researchers who meet the strict legal requirements of U.S.C. Title 13.

Given the limited accessibility of linked data and the resulting need of most researchers to rely on survey data alone, it is important for those with access to linked data to use them as a tool to conduct record-check studies and provide reports about the size and sources of survey error. While we recognize the possibility that administrative records can be incomplete or in some cases in error, by carefully using the administrative record data available to us, we believe we are able to use that data in a way that is useful for understanding survey accuracy and response patterns.

This paper reports the findings from a record-check study of the discrepancy between a survey estimate of public-assistance households from the 2001 American Community Survey/Supplementary Survey (ACS/SS01) and the actual count of public assistance households from administrative records (referred to hereafter as "estimate-count discrepancies"). Although previous studies suggest that estimate-count discrepancies stem from multiple types of survey error[4], we focus on measurement error from false-negative reports, which occur when respondents fail to report participation but administrative records indicate that they should report participation for at least one member of the household. Our scope is limited to those receiving public-assistance income from the State of Maryland and many of our methods are

---

[3] To see a comprehensive review of the literature, refer to Lynch (2006).
[4] In a study on the quality of responses in the ACS/SS01 to a question about Food Stamp recipiency, Taeuber et al. (2004) found evidence of difference in the coverage of housing units and the people in the housing unit between the ACS/SS01 and the administrative records.

drawn from a similar study on participation in the Food Stamp program, which was conducted by researchers at the Census Bureau, the Jacob France Institute of the University of Baltimore, the State of Maryland, and the U.S. Department of Agriculture.[5] One key finding from the Food Stamp study was that survey respondents seemed to respond in terms of their status in the program at the time of response rather than based on their status over the previous year as the question requests. We also draw methods from another similar study[6], which found a discrepancy between administrative records for Baltimore City and the estimates from the 1990 decennial census.[7] Because of the need to understand more about both the size and the causes of the estimate-count discrepancies, the study described in this paper is two-phased. In the first phase, we evaluate the estimate-count discrepancy and identify survey error, and in the second phase, we examine factors associated with why survey respondents give false-negative reports about participation in the Temporary Assistance to Needy Families (TANF) program.

II.   Data

The study uses a research file created from two data sources that were linked using a matching algorithm developed at the Bureau of the Census. The survey data are a Maryland subset of the Census Bureau's American Community Survey/ Supplementary Survey for 2001 (ACS/SS01). The administrative data are from the State of Maryland's Client Automated Resource and Eligibility System (CARES). A description of each of the data sources and the linking algorithm is provided below.

American Community Survey/Supplementary Survey for 2001 (ACS/SS01)
The ACS/SS01 collected detailed demographic, social, economic and housing data using the same methods and the same questionnaire used by the Census Bureau for the American Community Survey (ACS). The ACS collects data from a new sample each month using the current Master Address File and processes the information continuously.[8] The sampling frame is the housing unit population (this excludes persons living in group quarters), and the primary sampling unit is the household, which comprises all residents of a single housing unit. In 2001, the ACS/SS01 had 10,507 completed interviews in Maryland. Data collection is initially attempted by mail with non-responding households being contacted by telephone and then if necessary, by personal visits from field representatives.

The ACS/SS01 instructs survey respondents to complete a set of questions[9] for each person in their household, although not all questions are applicable to each person. Any adult in the household can be the respondent and that person is asked to complete the survey questions for everyone who is living or staying at the unit for more than two months and anyone else resident at the time of the interview who has no other usual place to stay. In this study, we focus on the response to survey question 40f (see Figure 1), which asks about receipt of public assistance or welfare payments. It is sixth in a series of questions about employment and income that apply only to household members who are 15 years and older. As noted in the instructions for the series, the question refers to the receipt of public assistance income in the past 12 months from the time of filling out the survey.

Question 40f refers generically to "any public assistance or welfare payments" so for our Maryland subset the survey question applies to the four cash-assistance programs offered by the state of Maryland and recorded in the CARES database. The largest program is the national TANF program, which is called "Temporary Cash Assistance" (TCA) in Maryland. Other public assistance programs in Maryland are: Transitional Emergency, Medical and Housing Assistance (TEMHA), Emergency Assistance to Families with Children (EAFC), and Refugee Cash Assistance (RCA). A single household may be involved in more than one program, but for the purpose of the survey estimate of public assistance, it is only counted once. In

---

[5] See Taeuber et al. (2004). http://www.ubalt.edu/jfi/jfi/reports/fstampfinrept273004.pdf
[6] See Taeuber et al. (2001).
[7] The administrative count of children who received AFDC in 1989 outnumbered the decennial census count of poor children in 1989.
[8] Information about the accuracy of the sample can be found at
http://www.census.gov/acs/www/Downloads/ACS/Accuracy01_SS01.pdf
[9] The American Community Survey questionnaires are available at:
http://www.census.gov/acs/www/SBasics/SQuest/SQuest1.htm

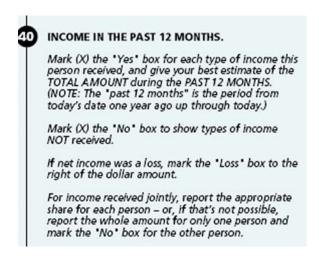Phase 1 of the research, we focus on all four programs. In Phase 2, we focus on just TANF, examining false-negative reporting of TANF-enrolled households only.

Figure 1: ACS/SS01 Instructions for the Income Series and Question 40f on Receipt of Public Assistance by Individuals who are at Least 15 Years Old

Maryland's Client Automated Resource and Eligibility System (CARES)
Maryland public assistance records are maintained on the Maryland Client Automated Resource and Eligibility System (CARES) file. This file also includes data about other assistance programs. All Maryland public assistance recipients are included in the CARES file, and they may live in group quarters such as shelters for the homeless as well as in households. The eligible group for the TCA program is the "assistance unit[10]" which is "a group of eligible individuals living together for whom income assistance has been [jointly] authorized."[11]

For the purposes of this study, the administrative records can reasonably be considered "truth" (i.e. a correct record of actual receipt of public assistance) because CARES determines eligibility and issues benefits for Maryland's public assistance programs and each transaction is recorded in the system. Consequently, we presume that enrollment is consistent with the recordation within administrative records.

---

[10] According to Maryland Department of Human Resources, Family Investment Administration, Temporary Cash Assistance Manual, Chapter VIII, 2000, the following individuals must be included in the assistance unit when they live in the same household: a parent and all the parent's natural or adopted eligible children; the parents and child of a minor parent; minor siblings related by blood, adoption, or marriage; a pregnant woman with no other children if the child, when born and living with the woman, is expected to be eligible; unmarried adults with a child in common or an unborn child.
[11] Annotated Code of Maryland, Article 88A, Title 07 Department of Human Resources, Subtitle 03, Family Investment Administration, Chapter 03 (Family Investment Program), .02B(5).

Linking Algorithm

To create the research file for analysis, we used the matching system developed by the Census Bureau's Data Integration Division (formerly known as the Planning, Research and Evaluation Division or PRED). This system uses probabilistic matching to assign a Protected Identity Key (PIK) to survey and administrative records. The PIK is then used to link records between data sets. The algorithm for the probabilistic match uses names, addresses, and demographic characteristics to output a set of matches, probable matches, possible matches, and non-matches.

We presume failure to find matches is rare because very few survey observations or administrative records were missing the information required for matching and a previous analysis using the same files and process estimated a 95% match rate.[12] In addition, the matching algorithm has been subject to ongoing quality control analyses.[13] Nonetheless, we strengthen the robustness of our findings by conducting sensitivity tests in Phase 1 that account for the possibility of failure to match. As described in the Methods section for Phase 1, match failure would affect results only about the source of error, not the actual size of the problem.

III.   Methods for Phase 1: Evaluating the Size of the Discrepancy between the Survey Estimate and the Count from Administrative Records

The estimate of interest was 30,979[14], which is the ACS/SS01 published point estimate of the number of households receiving public assistance in Maryland. The count of interest was not directly useable for computing an estimate-count discrepancy because CARES and ACS/SS01 were not designed to measure public assistance receipt in the same way. Differences exist in how survey data are structured compared to CARES data. In addition to structural differences, there are also other conceptual and definitional differences between the survey and the administrative records that need to be reconciled before CARES and ACS/SS01 can be legitimately compared. Phase 1 addresses the following questions:

1. What is the estimate-count discrepancy attributable to survey error? In other words, once we adjust the CARES count so it measures the same construct as the ACS/SS01, what is the difference between the CARES count and the published point estimate from ACS/SS01?

2. What is the estimate-count discrepancy attributable to error from respondents giving false-negative reports? In other words, how much smaller is the discrepancy after we use CARES to check the survey side of linked records and "fix" reports about public-assistance income for sample people who were found in CARES but had no public-assistance income reported? What does this figure imply about the size of other types of survey error?

To answer the first question, we computed a count of households from the CARES file that corresponds to the time period covered by the survey. While the CARES data are maintained on a monthly basis, the ACS/SS01 is a rolling survey with data collected throughout the year. As a result, the ACS/SS01 estimate is not an estimate of the number of households that received public assistance in 2001 but rather, it is an estimate of the number of households that have at least one individual who received public assistance sometime during the 12 months prior to the day the respondent reported participation. Because households could be surveyed in any month of 2001, the survey reference period is inconsistent and in aggregate covers 24 months. The best CARES approximation to this survey concept is the monthly average of the number of households recorded in CARES as having received assistance in a given month in 2001 or the 12 months prior to that given month.[15] To reflect the level of analysis of the published estimate, which is

---

[12] Taeuber et al., (2004) found an estimated match rate of 95 percent using the same source files.

[13] Researchers have been using and testing the algorithm in the Data Integration Division at the Bureau of the Census.

[14] The published estimate for the 90-percent confidence interval was 25,818 to 36,140.

[15] Since the administrative records do not indicate when in the survey month the household received public assistance, we cannot determine exactly how many months of CARES records we need to link to a survey record in order to compare the survey response to the records from the 12 months prior to that particular survey date. However we conducted a sensitivity test and concluded that it did not make a difference to our results so we linked each survey record with CARES data for the survey month and 12 prior months.

made at a household level, the CARES data also needed to be restructured so we could tally households from an administrative database of individuals in assistance units. First, we excluded persons under 15 years old from the analysis, because their sources of income were not asked about by the survey.[16] Next, we linked accounts with one or more overlapping individuals or having one or more individuals appearing within the same household in the 2000 decennial census. The latter was done as an extra, precautionary step to prevent the over-counting of households that would occur if household members were in more than one program or enrolled in the same program under separate accounts.[17] The difference between the count resulting from this procedure (62,960) and the published ACS/SS01 point estimate (30,979) constitutes the most basic computation of the estimate-count discrepancy for the sample[18] (see Results from Phase 1).

Although we can only approximate the way the survey measures public-assistance recipiency, the adjusted CARES count is a reasonably accurate measure of the number of public-assistance households in the survey universe because the types of households that may have wrongfully been included in the CARES count are rare. Specifically, we are not able to adjust for the part of the estimate-count discrepancy due to the CARES count inclusion of recipients who could not have been sampled because they were living in group quarters or were homeless. However, we believe the survey estimate missed only about one-half of one percent of the public-assistance population for this reason.[19]

To answer the second question and estimate the proportion of the discrepancy due to false-negative reporting, we use the information on the CARES side of the matched records to check the accuracy of survey reports and correct for Maryland households missing from the published point estimate because the household's respondent failed to report that someone received public assistance from Maryland.[20] This method logically estimates a minimum value because incomplete matching means that there were likely false-negative households not so identified and therefore not used to correct the point estimate.

Since our match probably does not identify all false-negative households, the numerator used to calculate the proportion of the discrepancy comprised by false-negative households is likely smaller than it would be if the match worked perfectly. At the same time, failure to match and identify false-positive households has no impact on the calculation, so the errors only go in one direction and we conclude that our methods tend to produce a minimum value.

Since the results from Question 2 depend on assumptions about the quality of our match, we do a sensitivity analysis by recalculating the results assuming an improbably low match rate. As described above, Question 2 assumes perfect matching and thus assumes we fully correct the published estimate with all the households that are excluded from that estimate because they were incorrectly reported. This

---

[16] It is not necessarily the case that a household with a beneficiary under 15 years old also has a beneficiary who is 15 years old or more. In fact such households fall under the classification "Child Only Case." This can occur when the guardian for a child is not the child's legal parent. In such cases, the benefit is specified as being issued for the minor only. So legally, the guardian is not receiving the benefit.

[17] Only 68 percent of identified CARES individuals could be located in decennial census data.

[18] Due to variability inherent in the sampling processes, alternative samples using the same sampling procedure would generate different estimate-count discrepancies. In order to understand the range of likely discrepancies, it would be necessary to calculate a statistically appropriate confidence interval. Our goal in this paper is more limited: identifying potential components of the discrepancy for the actual ACS/SS01 sample and its point estimate of cash assistance recipient households.

[19] An earlier study (Taeuber et al, 2004) using the same Maryland subset of the ACS/SS01 found fewer than 3 percent of Food Stamp recipients lived in a homeless shelter or other group quarters sometime during the year 2000 and almost all public-assistance recipients also received Food Stamps. Another study (Taeuber et al, 2002) using 1999 data, found that Maryland individuals who lived in a homeless shelter or other group quarters typically lived there for only one or two months of 1999, implying that about 2/12 of them would be excluded from the survey universe in any given month of the ACS rolling sample. Since we can think of no reason why public-assistance recipients would be more likely to be homeless or in other group quarters compared to other types of Food Stamp recipients, we assume that about 0.5 percent (2/12 of the 3 percent) would be outside the survey universe (and thus wrongfully included in the adjusted CARES count).

[20] Note that this correction does not correct for households that received public assistance in a state other than Maryland prior to moving to Maryland and being surveyed there.

underestimates the size of the survey error attributable to false-negative reporting and overestimates the size of the survey error due to other causes. If we try to remedy this shortcoming by instead using our data to estimate a conservative match rate assuming no false positives (households falsely reporting receiving cash assistance), then we can adjust the observed number of false-negative households not included in the published point estimate by a non-match factor calculated from the conservative estimate.[21] This provides an upper bound on the size of the survey error due to false-negative reporting. It also provides a lower bound for error caused by other unidentified factors.

IV. Results from Phase 1

Question 1
Finding: The published ACS/SS01 point estimate is about half[22] of the weighted sum (using the final survey weight) of assistance units *among sampled households* CARES shows actually received cash benefits during the reference period.

After we adjusted CARES to only include public assistance households with at least one person whose income was supposed to be reported (i.e., because they were older than 14), we tallied a count of 62,960. Subtracting the published point estimate of 30,979 yields an estimate-count discrepancy of 31,981.[23] Even if we calculate the estimate-count discrepancy using the 90-percent confidence interval for the published estimate (25,818 to 36,140)[24], the count from the administrative records is much higher than the upper limit of the confidence interval[25].

Question 2
Finding: False-negative reporting accounts for 81 percent[26] of the estimate-count discrepancy between the ACS/SS01 published point estimate and the CARES count of public assistance households in Maryland.

After using CARES to supplement the published point estimate with only the false-negative households that matched to CARES, we calculate a CARES-corrected estimate of 56,775. Even if we make this calculation with the lower limit on the published 90-percent confidence interval (49,592, 63,958), the results indicate the published point estimate does not reflect a sizeable number[27] of households.

Sensitivity Analysis
Finding: It is conceivable that false-negative reporting accounts for virtually all of the discrepancy between the ACS/SS01 published point estimate and the CARES count.[28]

---

[21] The extrapolated estimate assumes the match rate is the same for assistance households that were reported and those that were not and that missing households have the same weights as the matched households.

[22] 30,979 / 62,960 = 49%.

[23] Note that the difference between this modified count and the survey estimate is not an estimate of the number of public-assistance households that were erroneously excluded from the published estimate. To compute the proportion of true-positive cases that would be missing from a survey data set that would be used by researchers to study the dynamics of program participation, one would need to exclude the false-positive households prior to computing the rate. Such a computation is beyond the scope of this project's focus on the estimate-count discrepancy, but one could readily compute it from the figures published here, making the same assumptions about match error (since there is no gold standard for distinguishing false-positive cases from non-matched cases) and the maximum number of true-positive households missed by the survey.

[24] This confidence interval was calculated by corrected recipiency status where indicated by administrative records and then applying the replicate estimates method that is standard for ACS and the Supplemental Survey (see "Instructions for Applying Statistical Testing to ACS Data, U.S. Census Bureau: www.census.gov/acs/www/Products/users_guide/ACS_2005_Statistical_Testing.doc).

[25] 62,960 - 36,140 = 26,820

[26] We made a weighted count of 25,796 households that were classified as not receiving cash assistance when record linkage shows strong evidence that they did in fact receive cash benefits. Taking this count as a percentage of the count-estimate discrepancy gives 25,796 / 31,981 = 81%. See the Methods section for an explanation of why this method logically tends to produce a minimum value (regardless of the point estimate being corrected with administrative data or the survey's sampling variability).

[27] 49,592 - 30,979 = 18,613

[28] This is not to suggest that other sources of error (in particular sample error) are not operative here.

To account for failure to match, we first estimate the match rate as that for households reporting cash assistance under the assumption of no false positives—that is no households reporting cash assistance falsely. If there are no false positives, then the reason not all of the weighted count of 30,979 households reporting cash assistance are found in the match (where only 24,544 were found) is because of match incompleteness[29]. So we estimate a match rate of 24,544 / 30,979 = 79.2 percent. This is a very conservative estimate of the match rate because it assumes that that the unmatched cases (the difference between the published point estimate of 30,979 and the number of matched households, 24,544) all resulted from failure to match and not from false positive reporting. The conservativeness of this estimate is further demonstrated by a similar study[30] that estimated a match rate of 95%. If we extrapolate from this "very conservative" match rate (79.2%), we would estimate that an additional 25,796 / 0.792 = 32,571 households should have been reported in the published point estimate if there had been no false-negative reporting and all true matches were found. Adding this estimate of households that should have been reported to the estimate from households that were reported (the published estimate) yields a corrected point estimate of 63,538. If we consider 79.2% as the minimum possible match rate, then we would expect that only this percentage of matches are being currently identified out of the greater number of those that actually exist. Since our corrected point estimate of 63,538 is based on a weighted count of identified matched cases, it seems reasonable to consider 63,538 an approximate ceiling value for the point estimate that would obtain (using the actually fielded sample) with all true matches found and all false-negatives corrected (as would then be possible). If we use the more plausible match rate (95%) then we would add 25,796 / 0.95 = 27,154 to the published estimate to get a corrected estimate of 58,133. This would mean that most but not all of the error can be explained by false negative reporting and that other sources of survey error (in particular sampling error) may be contributing to the discrepancy. That the ceiling estimate itself (63,538) is not statistically different than the administrative record count (62,960) could suggest respondent misreporting was the only substantial cause of the estimate-count discrepancy. This is conceivable because adjustments for false-negative reporting and incomplete matching may provide an estimate that is as high as (or nearly so or even higher) than the administrative record count.

V.   Methods for Phase 2: Evaluating Errors Causing the Estimate-Count Discrepancy

In the second phase we studied false-negative reporting, that is, why do respondents report "no" to the public-assistance question when CARES shows they should report "yes"? We tested hypotheses about the relationship between household characteristics and the probability that respondents gave false-negative reports about the main type of public assistance, the Temporary Assistance to Needy Families (TANF) program.[31] We addressed the following questions:

- ✓ What household characteristics relate to false-negative reporting?
- ✓ What is the relative risk of false-negative reporting associated with particular household characteristics?

We used the subset of linked records that CARES showed received TANF during the survey reference period. After restructuring the linked records to the household level and dropping households with child-only TANF assistance units, our research file for analysis contained 95 households. We examined the relationship between household characteristics and the probability of false-negative reporting about TANF by conducting a series of chi-square tests of factors we believed would be high-risk factors for false-negative reporting.

---

[29] Actually, if a household received cash assistance from another state but was captured in ACS/SS01 in Maryland they would also contribute to this difference. We expect that this situation happened only rarely but we have not evaluated this data.

[30] See Taeuber et al. (2004).

[31] We did not study the level of error from incorrectly reporting participation (i.e., false-positive cases) because previous studies have shown this to be a relatively smaller source of survey errors.

VI.   Results from Phase 2

The results of cross-tabulating the subset of linked households by accuracy of survey response show that 55 percent of the households were incorrectly reported as not getting public assistance when CARES shows that someone in the household who was 15 years or older received TANF during the survey reference period (Table 1).

Finding: Respondents in the sample of TANF-enrolled households were more likely than not to misreport household participation in public assistance.

Table 1.   Accuracy of Response to the Public-Assistance Question among TANF-Enrolled Households in the ACS/SS01 Sample

	Reported "Yes" to ACS/SS01 Public Assistance Question	Reported "No" to ACS/SS01 Public Assistance Question
Survey households CARES shows had at least one current household member enrolled in TANF in previous 12 months.	43	52

Finding: Results from chi-square tests suggest a significant relationship between TANF misreport (false negative reporting) and the following household characteristics:

- Someone in the survey household was not in TANF (according to CARES).

- The respondent was not one of the household members in TANF (according to CARES).

- No *adult* in the household was in TANF (according to CARES).

- No one in the household was in TANF during the month of the survey (according to CARES).

- Someone in the household had income from employment during the reference period (according to ACS/SS01).

Table 2. Accuracy of Response to the Public-Assistance Question among TANF Households in the ACS/SS01 Sample by Household Characteristics

ACS/SS01 Survey Household Characteristic	Does this Correctly Describe the Household Situation?	Rate of Correct Reporting of Public Assistance for Households in this Situation	Chi-Square	P-Value
Someone in the survey household was not in TANF	Yes (n=56)	34%	7.07	0.0078
	No (n=39)	62%		
The respondent was not one of the survey household members in TANF	Yes (n=40)	15%	25.54	<0.0001
	No (n=55)	67%		
No adult in the survey household was in TANF	Yes (n=17)	18%	6.37	0.0116
	No (n=78)	51%		
No one in the survey household was in TANF during the month of the survey	Yes (n=40)	28%	8.79	0.0030
	No (n=55)	58%		
Someone in the survey household had income from employment during the reference period	Yes (n=73)	40%	3.90	0.0483
	No (n=22)	64%		

Effect of having someone in the survey household not in TANF
Table 2 shows that sample respondents correctly reported TANF participation for only 1 in 3 of the households where the survey roster did not match the list of individuals in the administrative records. This indication of household membership complexity usually occurred because one or more individuals listed in

the survey roster were not in TANF and that could have made decisions about how to respond more difficult. About 60 percent of the TANF households were complex in this way.

Effect of having a respondent who was not one of the household members in TANF
Table 2 shows that only 15% of the respondents who were themselves not in TANF correctly reported participation for the TANF household. It was quite common for sample households to have this risk factor as 42% of the households had respondents who were not in TANF during the survey reference period.

Effect of having no adult in the household in TANF
Table 2 shows that while most of our subset of TANF households included an adult in TANF[32], the child-only households may be a sizeable part of the estimate-count discrepancy since nearly all TANF child-only households were reported incorrectly.

Effect of having no one in the household in TANF during the month of the survey
Table 2 shows that only 28 percent of households were correctly reported if no one was participating in TANF at the time of the survey. This result suggests that part of the problem of false-negative reporting may be attributable to respondents reporting about their household's current TANF participation.

Effect of having someone in the household with income from employment during the reference period
Having income from employment appears related to misreport, and it is notable that the majority of TANF households in the sample had employment income.

Implications from tests finding no relationship
The results of tests of characteristics showing no significant relationship with not reporting (CARES – established) TANF receipt (false negative reporting) are also worth noting because they cast doubt on several explanations that seemed plausible before we conducted this study.

We find no evidence for the hypothesis that false-negative reports can be attributed to confusion about the definition of income. We investigated whether respondents in TANF households misreported because they mistakenly assumed that the series of income questions (which includes the question about public-assistance income) did not apply to household members who have no earned income, such as children or the unemployed. The results showed that no respondents in the sample skipped the income series.

We find no evidence for a statistical relationship between false negative reporting and the following factors that may be associated with a respondent fearing stigmatization should they report that someone in the household needs income assistance.

- ✓ Education—households with a TANF participant with a high-school degree are no more likely to be misreported than those where no one has a degree (and presumably less access to the kinds of jobs that a respondent might feel could financially support a household).

- ✓ Job experience – households with a TANF participant with job experience during the reference period are no more likely to be misreported than those where no one has job experience (and presumably less access to the kinds of jobs that a respondent might feel could financially support a household).

- ✓ Male householder – households with male householders are no more likely to misreport TANF receipt than those with no male householder (and presumably no concern about loss of social status from living with an adult male who has little money, a major determinant of social status, according to studies of social class).

---

[32] We dropped TANF households that had no individuals in TANF older than 14 since the survey instructed the respondent to report public-assistance income only for people older than 14. So these particular child-only cases had at least one child who was at least 15 years of age.

Finding no evidence of factors suggestive of fear of stigmatization may relate to our small sample size; however the findings appear more plausible in light of the additional finding that we see no interviewer effect. False-negative reporting was no higher among households that responded to the public-assistance question in a face-to-face interview than in a paper questionnaire.

VII. Limitations

Results from this study should be considered in the light of this study's limitations. Most notably, our sample is from one state, yet the difficulty of reporting about public assistance likely varies by state. For example, differences in eligibility rules imply that there will be state differences in the relationship between respondents and the public-assistance recipients they report about. In addition, the names of public-assistance programs vary by state so we expect the difficulty of making the connection between the survey question and the name of the state-administered program to also vary by state.

We are also limited in our ability to infer to other survey contexts. For example, an analysis based on a survey with a household-level question about current status in public-assistance programs might find a smaller estimate-count discrepancy and different relationships between household characteristics and survey error. It is also possible that the reference year is a limiting aspect since there is some evidence suggesting false-negative reporting about public-assistance income was higher in the years following 1996, when the federal Aid to Families with Dependent Children program was supplanted by the Temporary Assistance to Needy Families program, devolving administration of the federal welfare program to many different state programs. As persons in assistance households become more familiar with the names and rules of state programs it may become easier for survey respondents to make the cognitive connection between survey items about "public assistance" or "welfare" and the name of the program from which household members receive assistance.

Our ability to make inferences from our results is also limited because we use sample data (instead of data from a census of the entire population of interest) and therefore the findings are subject to sampling error. Participation in public-assistance programs is fairly uncommon and Maryland is a small state, so we had a small sample of matched households for analysis and large confidence intervals.

Our ability to make inferences about the relationship between false-negative reporting and characteristics of the survey household is limited because our bivariate analyses do not control for the effect of other variables. However, the findings from a study that included a regression analysis with the same data suggest that the relationships we find are also found in a multivariate context (Lynch 2006).

Although problems with missing data values can be especially limiting in studies involving record linkage[33] we believe this was rare in our study and moreover, we found no evidence suggesting that missing data values could change the direction of results. Evidence for this belief was provided in the methods sections but bears summarizing again: ACS/SS01 nonresponse was small; cases of item misreport are only missed from analysis when there is also failure to match and the Phase 1 Sensitivity Tests suggests this is rare and only makes our findings too conservative; the administrative database appears to be of high quality.

VIII. Conclusions

Based on the results of Phase 1 and Phase 2, the main reason for the discrepancy between the published point ACS/SS01 estimate and the CARES count of public assistance is false-negative reporting. In addition, false-negative reporting may sometimes be caused by differences between the way TANF defines beneficiaries and the way ACS determines household cash-assistance recipiency status. Specifically, this can occur for survey households that have only children in TANF. For these households, the cash assistance is specified (by the state) as being for the child. Because the survey question does not apply to children under 15 years old, these households will be misclassified as non-recipients of cash assistance if the respondent reports public assistance in terms of who legally receives the TANF cash benefit.

---

[33] Cases can be missed in analysis when they are missing from any one of the data sources or when they are not correctly linked across data sources.

References

Bollinger, C. R., and David, M. H. (1997). "Modeling Discrete Choice with Response Error: Food Stamp Participation". *Journal of the American Statistical Association*, 92, 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 & Econcomic Statistics*, 19, 129-141.

Goudreau, K., Oberheu, H., Vaughan, D. (1984). "An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program" *Journal of Business and Economic Statistics*, 2, 179-186.

Hu, Teh-Wei. (1971), "The Validity of Income and Welfare Information Reported by a Sample of Welfare Families," in Proceedings of the Social Statistics Section, Alexandria, VA: American Statistical Association, 311-313.

Klerman, J., Ringel, J., and Roth, B. (2005). "Under-Reporting of Medicaid and Welfare in the Current Population Survey" (Labor and Population Working Paper, WR-169-3). Santa Moncia, CA: Rand.

Lynch, V. (2006). "Causes of Error in Survey Reports about who in the Household gets Welfare" (Unpublished Paper). College Park, MD: Joint Program in Survey Methodology.

Moore, J. and K. Marquis. (1989). "Using Administrative Record Data to Evaluate the Quality of Survey Estimates," *Survey Methodology*, 15, 129-143.

Moore, J., K. Bogen, and K. Marquis. (1992). "A "Cognitive" Interviewing Approach for the Survey of Income and Program Participation: Development of Procedures and Initial Test Results" in Proceedings of Statistics Canada Symposium

Moore, J., Stinson, L., Welniak, E. (1999). "Income Measurement Error in Surveys" in Sirken, M., Hermann, D.J., Schechter, S., and Tourangeau, R. (Eds.), *Cognition and Survey Research*. New York: Wiley and Sons.

Polivka, A. (1998). "Using Earnings Data from the Current Population Survey After the Redesign," (Unpublished Paper). Washington, DC: Bureau of Labor Statistics.

Roemer, M. (1999) "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990-1996," (Unpublished Paper). Washington, DC: Bureau of the Census, Income and Statistics Branch.

Taeuber, C., Resnick, D., Love, S., Staveley, J., Wilde, P., and Larson, R. (2004). "Differences in Estimates of Food Stamp Program Participation between Surveys and Administrative Records," (Unpublished Paper). Washington, DC/Baltimore, MD: Bureau of the Census, Jacob France Institute, U.S. Department of Agriculture, and the Maryland Department of Human Resources.

Taeuber, C., Staveley, J., and Larson, R. (2004). "Issues in Comparisons of Decennial Census Poverty Estimates with Public Assistance Caseloads in Maryland," (Unpublished Paper). Baltimore, MD: Jacob France Institute and the Maryland Department of Human Resources.

Taeuber, C., Staveley, J., and Larson, R., (2002), "Issues in Comparisons of Food Stamp Recipients: Caseloads from Maryland State Administrative Records and The Census 2000 Supplementary Survey," (Unpublished Paper). Baltimore, MD: The Jacob France Institute and the Maryland Department of Human Resources.

Warburton, R., and Warburton, W. (2004). "Canada Needs Better Data for Evidence-Based Policy: Inconsistencies between Administrative and Survey Data on Welfare Dependence and Education," *Canadian Public Policy*, 30(3), 241-256.

Appendix: Confidentiality Edits

The information below describes the confidentiality edits applied to the American Community Survey (for more information on the accuracy of the ACS statistics, see:
http://www.census.gov/acs/www/Downloads/ACS/Accuracy00.pdf

Confidentiality Edit: To maintain the confidentiality required by law (Title 13, United States Code), the Census Bureau applies a confidentiality edit to the American Community Survey (ACS) data to assure that published data do not disclose information about specific individuals, households, or housing units. As a result, a small amount of uncertainty is introduced into the estimates of ACS characteristics. The sample itself provides adequate protection for most areas for which sample data are published since the resulting data are estimates of the actual characteristics. Small areas require more protection, however. The confidentiality edit is implemented by identifying a subset of individual housing units from the sample data files as having a unique combination of specified person and household characteristics within a block group. The confidentiality edit is controlled so that the basic structure of the data is preserved.

Maryland's Department of Human Resources, Family Investment Administration (FIA), also protects the confidentiality of its clients. FIA is a principal agency of the State of Maryland and is mandated by law to administer the Temporary Cash Assistance (TCA) program, Food Stamp Program, and other public assistance programs that provide income assistance. FIA's responsibilities include collecting and maintaining certain data regarding the participants in the TCA program, the Food Stamp Programs, and other public assistance programs that provide income payments.

The Census Bureau and the FIA signed a Memorandum of Understanding as part of this research to ensure the integrity, security, and confidentiality of information maintained by the FIA and to permit appropriate disclosure and use of such data as permitted by law.

Census Bureau access to the data files is authorized under Title 13, U.S.C., Section 6; Title 13, U.S.C., Section 8(b), and Title 15, U.S.C., Section 1525, provide authority for the Census Bureau to engage in joint statistical projects with public and private entities.

Confidentiality of the FIA data is guaranteed under Title 13, U.S.C., Section 9; 5 U.S.C., Section 552a(3)(b)(4); Title 42 U.S.C., Sections 601-619, 7 C.F.R. Section 272.1(c)(1); 42 C.F.R. Section 431.300; 45 C.F.R. Section 205.50. Only Census Bureau employees and individuals with Census Special Sworn Status (including contract employees), who are working on projects approved by the Census Bureau and the FIA, and have sworn the Census Bureau's oath of confidentiality, will have access to the Title 13-protected data files. The FIA made the specified information available to the Census Bureau pursuant to Maryland's Annotated Code Article 88A, Section 6(a); COMAR 07.01.07.