"SNAP misreporting on the CPS: Does it affect poverty estimates?"

Julie Parker
Social, Economic, and Housing Statistics Division
US Census Bureau
Presented at 2011 National Association for Welfare Research and Statistics Conference
September 2011
SEHSD Working-Paper Number: 2012-1

I. Introduction

The official U.S. poverty measure has been in place since the 1960's. In 1995, The National Academy of Sciences Report suggested improving this measure. In 2010, an Interagency Technical Working Group provided a guideline for a Supplemental Poverty Measure (SPM) that the U.S. Census Bureau, with the help of the Bureau of Labor Statistics, will develop. The SPM will not replace the official poverty measure but will provide a more intricate measure of poverty.

One aspect of the SPM is that it considers federal government in-kind benefits as near-money because they are designed to reduce poverty. The SPM will include nutritional assistance, subsidized housing, and home energy assistance as family resources. Estimates of these programs will come from the Current Population Survey (CPS). Previous research has shown that benefit receipts of these programs are often underreported.

This paper examines the misreporting of nutritional assistance on the CPS received by Supplement Nutrition Assistance Program (SNAP), formerly known as food stamps.[1] In addition, the paper assesses the difference between the SNAP self-reported amount and the administrative amount in relation to the official poverty measure. This research is conducted using probabilistic record linkage techniques on 2005 Texas, Illinois, and Maryland SNAP administrative data and the CPS 2006 Annual Social and Economic Supplement (CPS ASEC).

II. Literature Review

Data from national surveys are used for a variety of reasons. One common use is to assess the effectiveness of social safety net programs and their take-up rates. If these data are incomplete or misreported, then these estimates could be biased and convey false information that could affect public policy. More specifically, the US Census Bureau, with the help of the Bureau of Labor Statistics, is creating a Supplemental Poverty Measure. This new measure of poverty will include many noncash benefits as near-money. A few noncash benefits come from programs

[1] SNAP is used to describe benefits throughout this paper even though the benefits technically were called food stamps before October 2008.

such as the National School Lunch Program, housing subsidy, and Supplemental Nutritional Assistance Program. These programs are considered near-money, or in-kind benefits, because they are considered a cash equivalent. This distinction is designed to guarantee that recipients will use public assistance in a specified way.

Previous research has shown that program receipt is often underreported on surveys. These studies have included programs such as the Earned Income Tax Credit[2], Medicaid[3], and Supplemental Nutritional Assistance Program[4] (SNAP). This paper will assess how many households misreport SNAP receipt and whether or not the self-reported SNAP amount understates or overstates poverty estimates.

III. Data and Methods

Administrative Data

The Food Stamp Act of 1977 was enacted for the purpose of increasing the food purchasing power of eligible households through the use of coupons to purchase food. The Food and Nutrition Service of the U.S. Department of Agriculture administers SNAP through State and local welfare offices. SNAP is the major national income support program which provides benefits to all low-income and low-resource households regardless of household characteristics (e.g., sex, age, disability, etc.). Eligibility is determined by an asset and income tests. If the family is deemed eligible, their benefit amount is calculated by subtracting 30 percent of net income from the maximum benefit amount.[5] The maximum benefit is based on the cost of a thrifty food plan.[6]

The 2005 SNAP administrative data provide information on program receipt for Illinois, Maryland, and Texas. States collect these data to effectively administrate their program. They include information such as date of receipt, SNAP amount received by month, case number, client number, address, and Social Security Number. The Texas data were provided to the Census Bureau by the Ray Marshall Center for Human Resources, University of Texas at Austin. The source of these data is the Texas Health and Human Services Commission. The Census Bureau obtained Maryland data from the Family Investment Administration, Department of Human Resources of the State of Maryland. These data were provided by the Maryland Department of Human Resources. Illinois data were obtained by Chapin Hall at the University of Chicago. They received the data from the Illinois Department of Human Services.

Survey Data

The 2006 Current Population Survey Annual Social and Economic Supplement (CPS ASEC) is the primary source of information on the labor force characteristics of the U.S. population. The sample is scientifically selected to represent the civilian non-institutionalized population.

---

[2] (Hotz & Scholz, 2002)
[3] (U.S. Census Bureau, 2008)
[4] (Meyer & Goerge, 2011)
[5] The assumption is that a family will spend 30 percent of its net income on food.
[6] (Carlson, Lino, Juan, Hanson, & Basiotis, 2007)

Respondents are interviewed to obtain information about the employment status of each member of the household 15 years of age and older. There is additional data on work experience, income, noncash benefits, and migration. CPS ASEC data are used by government policymakers and legislators as important indicators of our nations' economic situation and for planning and evaluating many government programs. The Supplemental Poverty Measure will derive SNAP estimates from the CPS ASEC.

The questions on participation in SNAP in the ASEC supplement were designed to identify households in which one or more of the current members received food stamps during the previous calendar year. Once a food stamp household was identified, a question was asked to determine the number of current household members covered by food stamps during the previous calendar year. Questions were also asked about the number of months food stamps were received during the previous calendar year and the total face value of all food stamps received during that period.

Matching Discrepancies

When linking data, there are three aspects of the data that need to be taken into consideration to ensure a good match. Those are the linking variable, aligning the data, and the reference period.

Linking Variable

The linking variable is a Protected Identification Key (PIK). This is created by the Census Bureau Person Validation System. This production system validates records by name, address, date of birth, and, if available, social security number (SSN). It then uses an algorithm to randomly generate PIK values for every possible number between 1 and 999,999,999. This number corresponds to a SSN. The Census Bureau converts SSNs to ensure confidentiality of the data[7].

When linking data, it is good practice to verify that the match was done properly by checking that name, address, age, or gender match. Due to new security measures, this opportunity will not be as available as it was in the past. Most researchers' security level will not allow personally identifiable information on the data such as name, address, and date of birth. Personally verifying the match does not intend to understate the validity of the PIK. Creating a PIK is an involved process and the branch that does this does a thorough and good job. Personally verifying the PIK does not imply that the process is inferior.

Not all persons on a survey or administrative record will receive a PIK. Among the reasons for this are that the respondent chose to opt-out of statistical research or the record had either incomplete data or inaccurate indentifying information. The average PIK rate is approximately 90% for surveys and 97% for administrative data. Researchers have shown that non-PIKed persons are not missing at random and will adjust the survey weight. Typically, this reweighting adjustment distributes the weights of cases without a PIK to those with a PIK, but does not compensate for all bias. Some researchers will use strata to reweight. This will account for the bias across the variables used in the reweighting but not in the whole survey and administrative data frame. Another practice is to multiplying the survey weight by the reciprocal of the

[7] For more details, see Appendix A.

probability of receiving a PIK based on demographic data, where the probability is determined by a logistic regression analysis. Reweighting linked data is necessary as the matched subset is a small portion of the full data from which the original survey weights have been calculated from. Usually, the demographics of linked data research are not representative of the full survey sample. This analysis does not adjust the survey weights as the author hopes that will be a broader future research project. Furthermore, the weighted results presented in this paper do not appear different from the non-weighted results.

Aligning Data

It is necessary to check data universe exclusions and ensure unit of analysis agreement when aligning data. The CPS ASEC includes the civilian non-institutionalized US population. This coincides with the state SNAP administrative data. To guarantee unit of analysis agreement for SNAP amounts, the administrative data will need to be summed to a yearly amount as that is how the CPS ASEC reports the SNAP amount. In addition, the administrative SNAP units will need to be aggregated to the CPS ASEC household and then family unit. This topic is discussed in the methodology section.

Reference Period

Depending on the research question, survey respondents should receive the benefit at the time of the interview. Normally a date analysis is necessary where the survey date is before the administrative date. SNAP receipt in the 2006 CPS ASEC refers to receipt in the previous calendar year, 2005. The 2006 CPS ASEC question is "Did (you/anyone in this household) get food stamps at any time during 2005?" Since the reference year is the previous calendar year, a date analysis is unnecessary because the administrative record time frame covers the 2005 calendar year.

Methodology

The Census Bureau has the largest collection of state level SNAP data for 2005. These three states include Texas, Maryland, and Illinois. Before linking the survey and administrative data, the administrative SNAP amount needs to be summed from a monthly amount to a yearly amount. Making this change will take long data files and reshape them into wide form. The procedure is done by identifying SNAP units, transposing the SNAP amount, and then summing these amounts. The administrative data are then matched to the survey data by the PIK.

Once the data have been linked, CPS ASEC observations not in the states of Texas, Maryland, or Illinois are excluded from the subset. The administrative SNAP amount is then summed to the CPS ASEC household. This is necessary because SNAP receipt is report at the household-level on the CPS ASEC. The realignment process is repeated for the CPS ASEC family because poverty is calculated at the family-level. The procedure of realigning the data is similar to above where the household, and then family, unit is determined. The data are then transposed and summed. CPS ASEC households and families were included in the analysis if at least one person could be linked to an administrative record.

A scenario that arose from realigning the family units was cases of multiple families with one administrative case unit. This group comprises 6% of the families in the subset. There seemed

to be two options to reconcile this problem: divide the SNAP administrative amount or change the CPS ASEC family unit. Although these CPS ASEC family units appeared to be non-married families, changing the family units only for these cases could introduce inconsistencies in the data. This option rejects the Census Bureau definition of family, an undesired alternative. Dividing the administrative SNAP amount seemed like a more viable option, but how should the benefit be divided? By the number of people, the number of families, or a combination of the two? The Census Bureau has a method of allocating SNAP amount to families since SNAP is reported at the household-level. This allocation method is the best option. Unfortunately, this part of the analysis was out of scope due to time and resources.

To determine if misreporting affects poverty, SNAP amounts are added to the total money income, the numerator of the equation that determines a family's level of poverty. This comparison will provide a baseline to determine to what extent misreporting is an issue. The official poverty measure is calculated by dividing a family's total money income (TMI) by their official poverty threshold. Each family has a calculated official poverty threshold that is defined as the minimum amount for which that family is considered to be above or below the poverty line. This amount is determined by family size and composition. When this number is divided into the family's TMI, the result is the family's income-to-poverty ratio. There are three comparisons: TMI, TMI + CPS ASEC SNAP amount, TMI + administrative data SNAP amount. These three comparisons are calculated for each family. Each comparison will also be divided by their official poverty threshold to obtain their income-to-poverty ratio.

IV. Findings

Pre-link Counts

As shown in Figure 1, more survey observations, approximately 16%[8] of both the CPS ASEC households and people[9], were lost due to a lack of a PIK compared to the administrative records, 3.5% to 6%. The survey data represents only the households, and the people in these households, that reported receiving SNAP benefits. These numbers do not include false negatives; households that did not report receipt but were found in the administrative records as receiving SNAP benefits. When looking at SNAP benefit amount (Figure 2), a small amount is lost due to a lack of a PIK, $1.4 million for administrative data and $130 million from the CPS ASEC.

Result #1: Misreporting

[8] The data in this report are from the Annual Social and Economic Supplement (ASEC) to the 2005 Current Population Survey (CPS). The estimates in this paper (which may be shown in text, figures, and tables) are based on responses from a sample of the population and may differ from actual values because of sampling variability or other factors. As a result, apparent differences between the estimates for two or more groups may not be statistically significant. All comparative statements have undergone statistical testing and are significant at the 90 percent confidence level unless otherwise noted. Standard errors were calculated using replicate weights. Further information about the source and accuracy of the estimates is available at www.census.gov/hhes/p60_236sa.pdf.

[9] The 16.2% of CPS ASEC households are not significantly difference from the 15.8% of CPS ASEC people.

According to the administrative data, 12% of TX, MD, and IL households received SNAP benefits in the 2006 CPS ASEC. About 50% of these households did not report receipt. This suggests that SNAP receipt is underreported in the 2006 CPS ASEC for TX, MD, and IL. [10] Table 1 illustrates these results and Figure 3 shows an explicit comparison among the reporting categories amongst households and people.

A small percentage of false positives, households that reported receipt but could not be found in the administrative records, were found. It is unknown if these households did or did not receive SNAP benefits as they could have been in an administrative record from another state. According to the CPS ASEC migration variable, this could apply to 28% of false positive households. Another scenario could be that these households were receiving benefits, but from the administrative side, the benefits were delayed. According to practitioners, benefits will be dispersed without an update, or resubmission, to the administrative data recording.

False negatives represent an underreporting of $1.4 billion SNAP dollars, as Figure 4 points out. The net underreporting is reduced slightly by $215 million due to false positives. When considering what was reported in SNAP benefits ($2 billion) and what should have been reported, ($3 billion), it is clear that underreporting has an impact on SNAP estimates. Self-reported households in the CPS ASEC over-reported by $308,000 SNAP benefit dollars. Over-reporting did not cause a statistical difference between the CPS ASEC SNAP and the administrative amount. This suggests that households that correctly reported receipt are also reporting the amount accurately.

Characteristics for the three reporting types are shown in Figure 5. For each category, the majority of people are white, unemployed, and, for adults, have an educational attainment of grade school. Unmarried female households comprise the majority of correctly reported and false positive households, while married families encompass the majority of false negative households[11]. All categories had a majority of families in deep poverty, below one-half the poverty level.

Result #2: Poverty

Figure 6 looks at the effect of SNAP as to whether a family is below or above the poverty line. The graph illustrates a gradual decline among the three comparisons, but this decline is small and there is no statistical difference between the three[12]. That exemplifies that adding SNAP benefits does not move a statistically significant number of families out of poverty. Other studies have shown that SNAP substantially reduces deep poverty and Figure 7 shows the effect of adding SNAP to TMI in greater detail. (Zedlewski & Mon, 2009) Based on TMI, 22% of families are below 50% of poverty. When CPS ASEC SNAP is added, this is reduced to 17% of families,

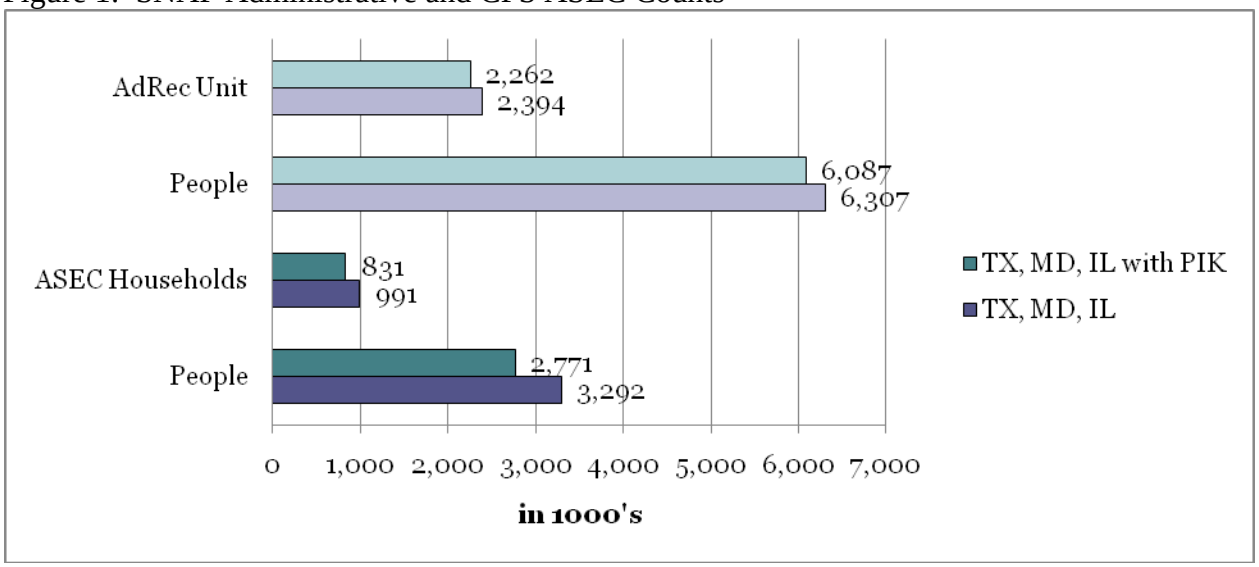
[10] Removing imputed observations resulted in an underreporting of 55%. Although slightly higher than the results that include imputed values, there was no statistical difference.
[11] There is no statistical difference between married families and unmarried female households for false negative households.
[12] Note that the official poverty rate is lower than the rates shown in Figure 6. Keep in mind that the results are for families that either received SNAP benefits or said that they received them, so are more likely to be low-income than the average family.

and if the administrative SNAP amount is used, only 15% of families are below 50% of poverty. When included in determining the family's income-to-poverty ratio, there is no statistical difference between the TMI that includes self-reported CPS ASEC SNAP amount and the TMI that includes the administrative amount. However, there is a statistical difference between the TMI and the two TMI + SNAP for families in deep poverty, <50% poverty. This demonstrates that while adding SNAP benefits to the poverty measurement does not significantly reduce the number of persons below poverty, it does reduce the number of families in deep poverty.

V. Conclusion

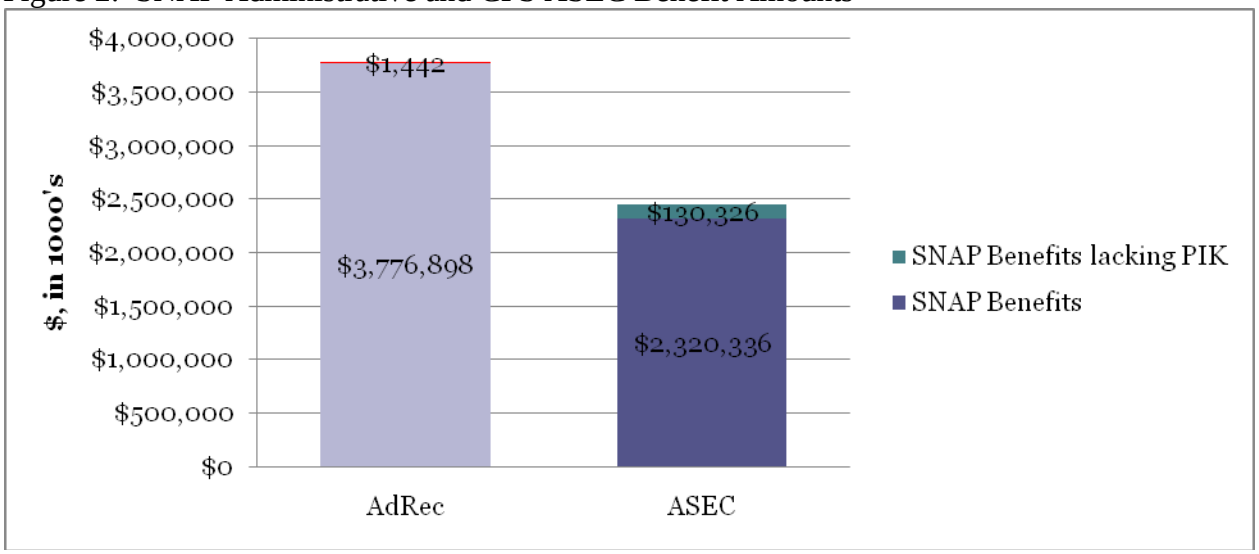
Previous research has shown that benefit receipts of in kind government programs are often underreported. This is evident in this study as SNAP receipt and benefit amount is underreported for TX, MD, and IL in the 2006 CPS ASEC. Providing a baseline to access this misreported revealed that there is a statistical difference between the official poverty measure and this measure with the addition of SNAP benefits for families in deep poverty. It also demonstrated that there is no statistical difference between the self-report CPS ASEC and the administrative SNAP amount.

Figure 1: SNAP Administrative and CPS ASEC Counts



Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 2: SNAP Administrative and CPS ASEC Benefit Amounts



Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

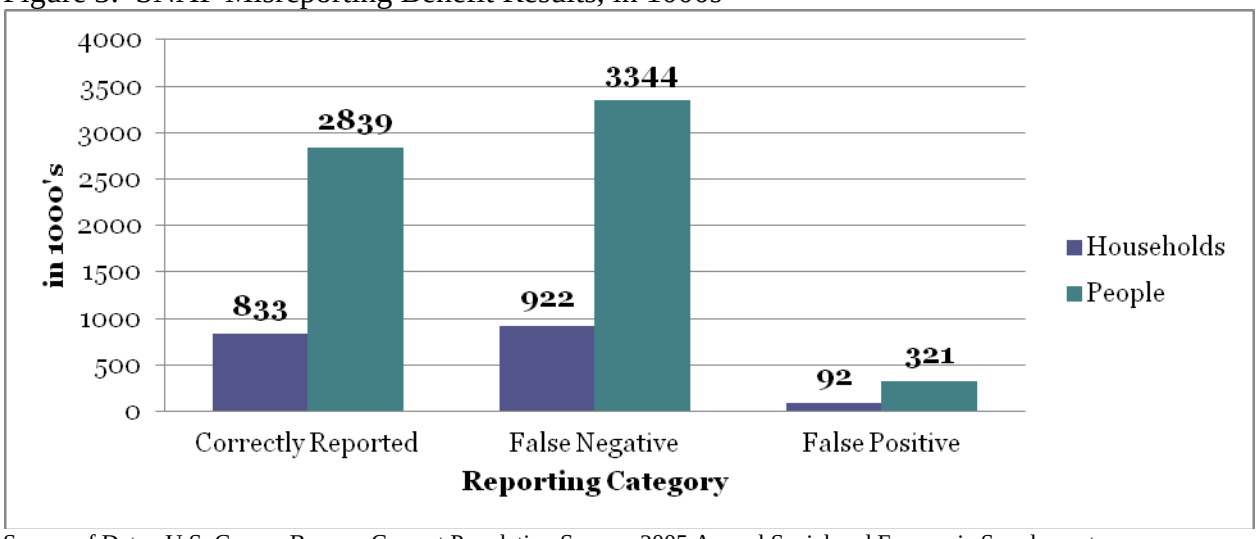
Table 1: Household SNAP Receipt

Administrative Receipt	CPS 2006 Report					
	No SNAP		SNAP		Total	
	Estimate	90% confidence interval (+-)	Estimate	90% confidence interval (+-)	Estimate	90% confidence interval (+-)
2005 TX, MD, IL						
No SNAP	6,397,000		49,000		6,446,000	
	12,526,000	169,662	92,000	19,886	12,618,000	168,504
	87.15		0.64		87.79	
	99.28		0.72		100	
	93.14		9.89		87.79	
SNAP	515,000		469,000		984,000	
	922,000	66,290	833,000	61,079	1,755,000	90,602
	6.42		5.8		12.21	
	52.53		47.47		100	
	6.86		90.11		12.21	
Total	6,912,000		518,000		7,430,000	
	13,448,000	162,021	925,000	63,661	14,373,000	157,871
	93.57		6.43		100	
	93.57		6.43		100	
	100		100		100	

Notes: The entries in each cell from top to bottom are sample count, population estimate, overall %, row %, column %. Estimates are weighted by household weight.

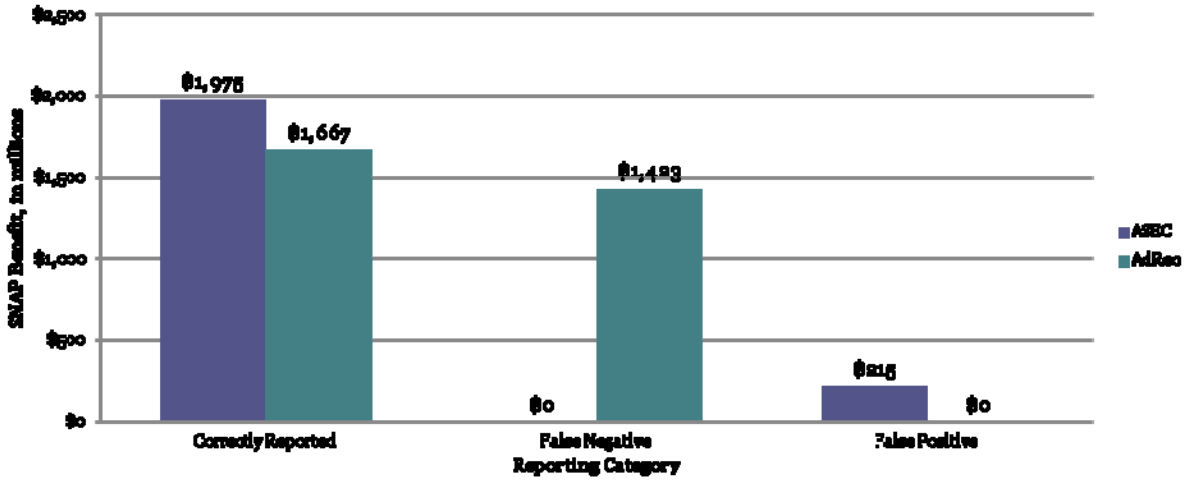
Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 3: SNAP Misreporting Benefit Results, in 1000s



Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 4: SNAP Misreporting Benefit Amounts



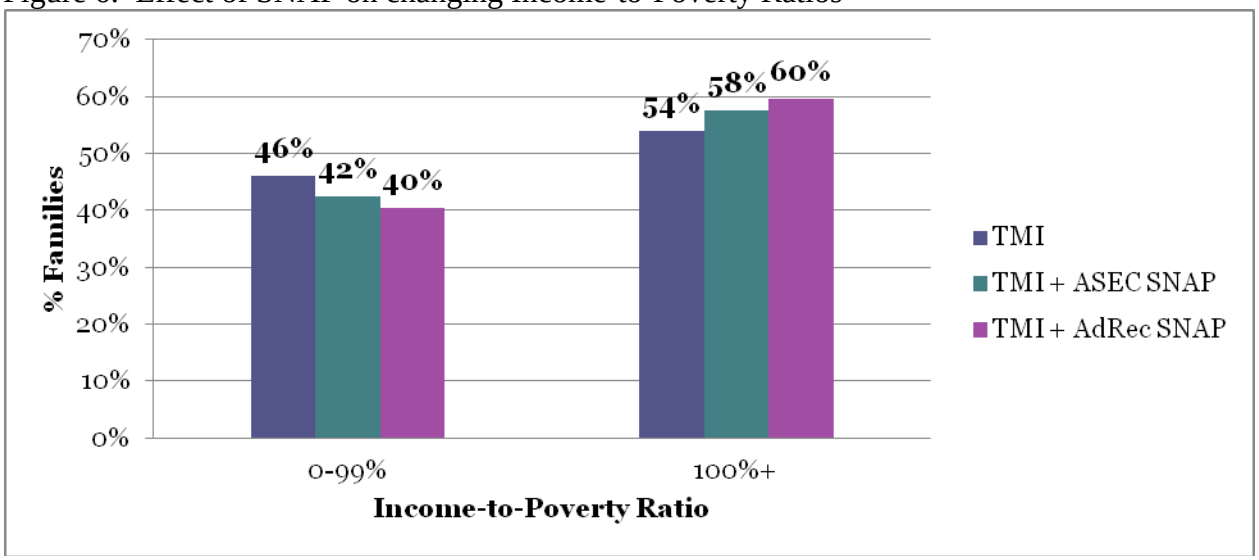
Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 5: Characteristics by Reporting Type

Characteristic	Correctly Reported		False Negatives		False Positives	
	Estimate	90% confidence interval (+-)	Estimate	90% confidence interval (+-)	Estimate	90% confidence interval (+-)
Children (<18)	50%	4%	41%	3%	45%	12%
Adult	43%	4%	52%	3%	50%	11%
Senior	7%	2%	7%	1%	5%	3%
White	64%	4%	62%	3%	72%	11%
Black	33%	3%	34%	3%	25%	11%
Asian	1%	1%	-	-	1%	1%
Hawaiian Pacific Islander	1%	-	1%	1%	2%	3%
Other	2%	1%	3%	1%	-	-
Unemployed	74%	4%	63%	3%	70%	10%
Poverty <50%	30%	4%	15%	2%	36%	12%
50-74%	18%	3%	7%	2%	16%	8%
75-99%	15%	2%	8%	2%	9%	7%
100-124%	13%	2%	10%	2%	12%	7%
125-149%	8%	2%	9%	2%	5%	4%
150-174%	5%	2%	8%	2%	3%	4%
175-199%	2%	1%	9%	2%	4%	4%
200-249%	4%	1%	9%	2%	7%	5%
250-299%	2%	1%	7%	2%	6%	4%
300-349%	1%	1%	6%	2%	-	-
350-399%	1%	1%	4%	2%	2%	3%
400-499%	1%	1%	3%	1%	-	-
500%+	-	-	5%	1%	-	-
Grade School Education	24%	2%	25%	2%	28%	7%
12th Grade, No Diploma	1%	-	2%	1%	2%	2%
High School Diploma or GED	14%	2%	21%	2%	10%	4%
Some College	7%	1%	8%	1%	11%	6%
Associates Degree	2%	1%	2%	1%	1%	1%
Bachelor's Degree or more	1%	-	3%	1%	2%	2%
Child	51%	2%	39%	2%	46%	9%
Married Family Household	26%	4%	42%	4%	23%	9%
Unmarried Male Household	4%	2%	7%	2%	1%	2%
Unmarried Female Household	48%	4%	34%	4%	50%	10%
Single Male	8%	2%	6%	2%	5%	5%
Single Female	13%	2%	12%	2%	21%	10%

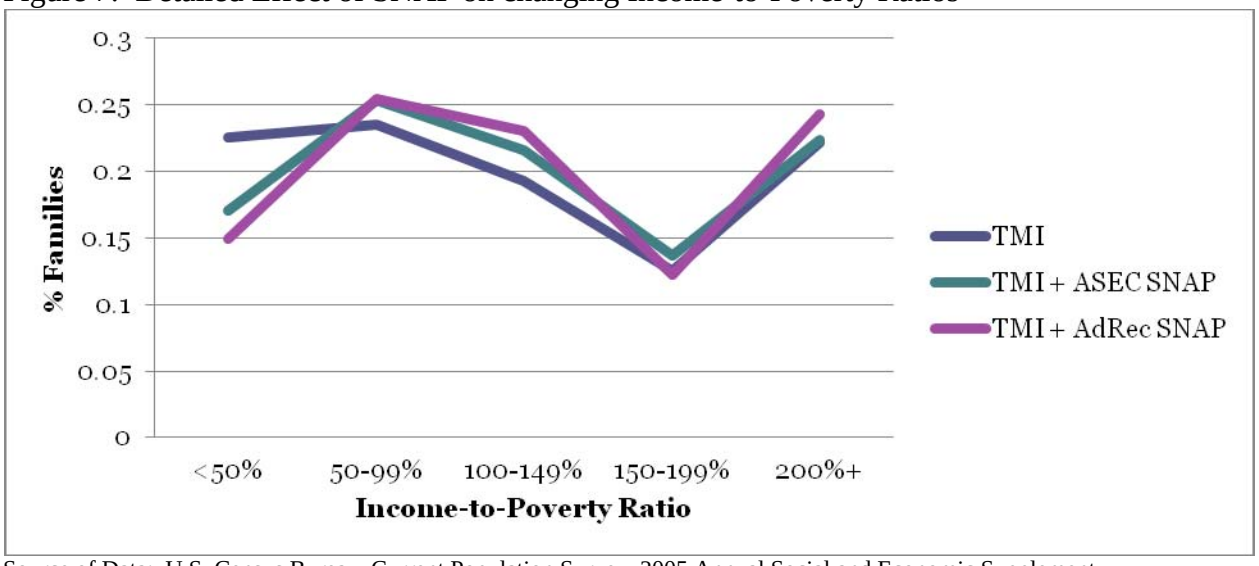
-Represents or rounds to zero.
Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 6: Effect of SNAP on changing Income-to-Poverty Ratios



Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Figure 7: Detailed Effect of SNAP on changing Income-to-Poverty Ratios



Source of Data: U.S. Census Bureau, Current Population Survey, 2005 Annual Social and Economic Supplement.
Link to URL: http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf

Appendix A:

Person ID Validation System (PVS)
The Person Identification Validation System (PVS) provides a fully automated production capability at the Census Bureau for Social Security Number (SSN) validation. Once an SSN is either verified or searched for and assigned, the record is considered validated. The PVS is managed by the Center for Administrative Records Research and Applications Staff. The PVS enables SSN validation for regularly-repeating demographic surveys such as the CPS, SIPP, and ACS as well as for other demographic or administrative files. The PVS also expands data linkage capabilities for merging survey and administrative data sets.

The PVS uses probabilistic matching to verify SSNs contained within an incoming file against those contained within the Census Numident. The processing consists of a verification phase followed by a two-step search phase for assigning SSNs when necessary. For the verification phase, SSNs are matched using several types of demographic data, including names, dates of birth and gender. Specific weights are set to define acceptable matches. Any records not verified through this phase – or without an incoming SSN – are sent forward to the search phase of the system. (Note: In conformance with Census Bureau privacy policy, the PVS does not process any record for which the respondent has refused to provide his or her SSN. Also, due to technical constraints, the PVS does not process records where the respondent withholds his or her first name and surname.)

The search phase of the PVS, also based on probabilistic matching, is comprised of a geokey (address-based) search, followed by a name search. The geokey search consists of logically grouping or "blocking" the data using the geokey, then progressively relaxing the geographic criteria while undertaking multiple passes through a matching routine to achieve agreement on demographic data as cited for the verification phase. Unmatched records remaining after the geokey search fall to the name search, where they undergo a similar demographic matching process but without the use of the geokey.

The final output file of the PVS (created after completion of the verification and the search phases) contains: all records with verified or searched and assigned SSNs; all records where the SSN could not be verified or searched and assigned or where multiple and, therefore, inconclusive SSNs were found; and, all original records withheld from the PVS process due to refusals or wholly blank names. A record is considered ***validated*** when it successfully completes either the verification or the assignment phase (geokey- or name-based search). Only validated records can be used in record linkage.

References

Carlson, A., Lino, M., Juan, W. Y., Hanson, K., & Basiotis, P. P. (2007). *Thrifty Food Plan, 2006.* Washington, DC: Center for Nutrition Policy and Promotion.

Hotz, J. V., & Scholz, J. K. (2002). Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data. In M. eds. Ver Ploeg, R. A. Moffitt, & C. F. Citro, *Studies of Welfare Populations: Data Collection and Research Issues* (pp. 275-313). Washington, DC: National Academy Press.

Meyer, B. D., & Goerge, R. M. (2011). *Errors in Survey Reporting and Imputation and their Effects on Estimates of Food Stamp Program Participation.* Chicago.

U.S. Census Bureau. (2010). *SNACC Research Project to Understand the Medicaid Undercount: the University of Minnesota's State Health Access Center, the Centers for Medicaid and Medicare Services, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, the National Center for Health Statistics, and the U.S. Census Bureau. Phase II Research Results: Examining Discrepancies between MSIS and the CPS ASEC.* Washington, DC.

Zedlewski, S. R., & Mon, E. Y. (2009). *Many Low-Income Working Families Turn to the Supplemental Nutrition Assistance Program for Help.* Washington, DC: The Urban Institute.