Misreporting of SNAP Participation
by
Bruce D. Meyer and Nikolas Mittag

Measurement Error in Survey Data

- Data linkage has demonstrated severe problems of survey accuracy

- For government transfers, most studies link major household surveys to administrative records from one or two states at a time

Key points of this project:

- Examine geographic variation in misreporting to assess whether we can...

  - generalize conclusions of prior studies of one state

  - extrapolate from the validated states to improve national estimates

- Examine how errors in FoodAPS differ from the surveys previously validated to

  - assess to what extent conclusions generalize to other surveys

  - obtain further evidence on how survey design and topic affect reporting

Definitions

- Need a measure of true current receipt corresponding to the survey definition
- FoodAPS reference period is slightly ambiguous, since some people reported SNAP receipt in the follow-up survey
- Using receipt within 30 days of initial interview instead of receipt within 36 days of end of interview week (SNAPNOWADMIN) yields more false positives and slightly fewer false negatives
- Increased false positives suggest including interview week
- Increased false negatives suggest need for further clarifications:
  - Need to clarify why it affects so many people
  - How many households actually started receiving SNAP between interviews? Do they make us overstate false negatives and understate false positives? Would we expect people to report changes in receipt status between interviews?
  - Defining receipt including days between interviews resulted in slightly higher receipt rates

Data Linkage

- Due to availability of administrative data, linkage varies by state

- Four groups of states:
  - Caseload and ALERT data with same IDs
  - Caseload and ALERT data with different IDs
  - No dates in caseload data
  - No caseload data

- Raises several questions we need to examine:
  - For which states can the administrative variable be considered "truth"?
  - Should we use the ALERT data to define "true receipt"? If so, how exactly?
  - When analyzing geographic variation of misreporting, how can we avoid confounding it with variation in linkage quality?
  - If we exclude some states, should we make any adjustments for representativeness?

Data Linkage – Preliminary Evidence

Table 1: Unweighted Error Rates by Definition and Linkage type

Linkage Type	False Positives		False Negatives	
	Caseload only	Caseload and ALERT	Caseload only	Caseload and ALERT
Caseload and ALERT data with same IDs	6.0%	2.4%	2.6%	3.1%
Caseload and ALERT data with different IDs	3.2%	4.6%	2.5%	3.6%
Caseload dates missing	7.6%	7.6%	0.0%	2.0%
ALERT data only	0.0%	14.0%	0.0%	1.6%

Note: This table reports unconditional error rates (percentages of the overall samle count).

Group 1 and 2 only differ in how ALERT was matched.

Need to examine ALERT information for second group to

- Better understand link quality and potential sources of error
- Can we improve the receipt measure?

Extent of Misreporting Program Participation

Table 2: Weighted Misreporting Rates

	Caseload & ALERT		Caseload Only	
	False Positives	False Negatives	False Positives	False Negatives
Caseload and ALERT data with same IDs	**1.5%**	**9.7%**	3.3%	9.0%
Caseload and ALERT data	1.7%	12.1%	2.8%	9.4%
Full Sample	3.3%	11.7%	2.8%	9.4%

Note: Rates are in percent of "true" recipients (false negatives) and non-recipients (false positives).

– Supports using both caseload and ALERT data and excluding groups 3 and 4

– Including second group of states slightly increases both error rates, which likely combines reporting and linkage errors

We will use only the first group in the preliminary analyses here. We hope to be able to include the second group of states in the future.

Extent of Misreporting Program Participation

Suggests that FoodAPS is more accurate than previously validated surveys. Understanding why may help to improve surveys, but differences could be due to several reasons:

- **False negative** rate much lower than in prior studies:
  - Shorter reference period?
  - Specialized survey? Asking the main food shopper?
  - Linkage errors could also understate false negatives

- **False Positive** rate slightly higher:
  - Shorter reference period and telescoping?
  - Specialized survey?
  - Linkage errors and definitions could also overstate false positive rate

Caveats and Definitions

We are still in the process of investigating definitions and samples.

The analyses in the remainder of this presentation are based on preliminary definitions:

1. We only use the most reliable states for now: those for which caseload and ALERT data with unique IDs is available
2. We use SNAP receipt within 36 days of end of interview week as our measure of truth
3. We use the definition of receipt that combines caseload and ALERT data (SNAPNOWADMIN)

The results in the remainder are highly preliminary and mainly an illustration of the types of analyses we will conduct!

Determinants of Errors

Table 3: Marginal Effects of Key Error Determinants

	False Positives		False Negatives	
	MEF	*SE*	*MEF*	*SE*
Number of members 18 or older	0.0172	(0.0187)	0.0061**	(0.0027)
Age >= 50	-0.0562*	(0.0326)	-0.0025	(0.0068)
White Non-Hispanic	0.0596*	(0.0311)	-0.0087	(0.0068)
Unemployed	-0.0998**	(0.0448)	0.0124	(0.0081)
Income Below Poverty Line	-0.1311***	(0.0397)	0.0316***	(0.0079)
Income 100-200% of Poverty Line	-0.0637*	(0.0349)	0.0184***	(0.0070)
Interviews in English	-0.0830*	(0.0478)	-0.0110	(0.0096)
Reported Housing Assistance	-0.1001***	(0.0385)	0.0092	(0.0081)
Reported Welfare, Child Support, Alimony	-0.1253**	(0.0489)	0.0238**	(0.0115)

Note: Estimated from two separate Probit models at the household level using sampling weights.
Demographic characteristics refer to the respondent. The omitted income category is >200% FPL.

The model also includes dummies for household composition, gender, education, employment, disability, citizenship, rural and report of WIC, but the estimated marginal effects are not significant.

Determinants of Errors

- Sizeable effects relative to overall rates, but few significant
- Compared to prior analyses using the CPS, ACS and SIPP in Il, MD and NY, these results suggest:
    - Fewer, but larger effects
    - Income and reports of other programs are key predictors
    - Relation to variables like age and reporting other programs differs
    - No effects of household composition and education
- Will add variables to these models to examine why FoodAPS is more accurate, such as:
    - Recall: last receipt and last use
    - Interview conditions
- Using the larger sample and cleaner definitions should increase precision

Geographic Variation in Misreporting

- Many validation studies use convenience samples, which raises question of their generality
- For transfer programs, often validate one state:
    - Do findings apply to other states?
    - Can we use the validated states to (partly) correct national estimates?
- Both questions hinge on whether misreporting is stable across geography
- FoodAPS validated multiple states, allowing us to examine how much misreporting varies
- A test of equality of misreporting rates across states rejects for both false positives and false negatives (p-values<0.001 in both tests)
- However, extrapolation and most generalizations only require misreporting to be stable conditional on X

Geographic Variation in Misreporting

– To examine whether states conditionally differ in misreporting levels, we add state fixed effects to the error probits

– Coefficients and marginal effects appear small

– Yet, joint tests of coefficient equality reject for false negatives (p-value: 0.03), but not false positives (p-value: 0.27)

– Shows evidence of heterogeneity, but inconclusive and extent not clear.

– We will further examine this by:

  – Decomposing variation into part due to X and due to geography
  – Testing for conditional heterogeneity arising from differences in slopes
  – Assessing how heterogeneity affects extrapolation

Table 4: Bias in Marginal Effects of Probit Models of SNAP Receipt

| | Dependent Variable | | test of |
| | Admin | Survey | equality |
	Receipt	Receipt	(p-value)
Single adult, no children	-0.0844***	-0.0712***	0.46
Single adult, with children	-0.0081	-0.0013	0.59
Many adults, no children	-0.0246*	-0.0153	0.38
Number of members 18 or older	0.0013	0.0084	0.13
Age >= 50	-0.0146	-0.0066	0.42
Male	-0.0110	-0.0074	0.69
Less Than High School	0.0607***	0.0673***	0.49
High School Graduate	0.0771***	0.0734***	0.92
Some college	0.0505***	0.0492***	0.97
White Non-Hispanic	-0.0187	-0.0291**	0.17
Employed	-0.0221*	-0.0242*	0.72
Unemployed	0.0600***	0.0751***	0.11
Income Below Poverty Line	0.2050***	0.2120***	0.11
Income 100-200% of Poverty Line	0.0969***	0.1035***	0.23
Disabled	0.1246***	0.0991***	0.12
Interviews in English	-0.0451*	-0.0333	0.51
Non-U.S. Citizen	-0.0573**	-0.0612***	0.68
Household in Rural Area	0.0008	0.0062	0.54
Reported Housing Assistance	0.0616***	0.0709***	0.26
Reported WIC receipt	0.0689***	0.0606***	0.64
Reported Welfare, Child Support, Alimony	0.0713***	0.0957***	**0.04**

Note: Stars denote significance at the 10,5 and 1% level. Demographics refer to
respondent. The test of equality was performed on coefficients, not marginal effects.

Bias When Receipt Is a Dependent Variable

- Neither individual nor joint tests reject coefficient equality (p-value of joint test: 0.29)

- Using a larger sample and a revised definition should increase precision

- However, differences are also small substantively

- Provides evidence that the more accurate FoodAPS data also yields more accurate estimates

Summary and Very Preliminary Conclusions

- Error rates in FoodAPS appear much lower than in prior studies, particularly false negatives

- Nonetheless, errors are systematically related to key covariates

- Evidence of some geographic heterogeneity, even conditional on covariates

- Evidence that more accurate data indeed improves bias in estimates

Future Work

- Scrutinize definitions and data linkage further

- Further study geographic variation:
  - Measure extent of geographic variation
  - Heterogeneity in slopes?
  - Examine extrapolation in light of geographic variation

- Get an idea WHY FoodAPS is more accurate. Some early thoughts on how we could examine this:
  - Examine how recall matters in shorter FoodAPS reference period
  - Usage data to test what matters for recall and to examine salience
  - Interview conditions: interview variables and interviewer IDs
  - Other measures of survey design that vary?
  - Check whether people report receipt if they use the card after they leave SNAP?