Misreporting of Food Stamp Participation in a Household Survey: Results from a Single-State Pilot Study of the National Health and Nutrition Examination Survey[§]

John A. Kirlin[1], Lisa B. Mirel[2], Daniel Schroeder[3], and Michael Wiseman[4]

[1]Economic Research Service, U.S. Department of Agriculture, jkirlin@ers.usda.gov
[2]National Center for Health Statistics, LMirel@cdc.gov
[3]Ray Marshall Center for the Study of Human Resources, schroed@raymarshallcenter.org
[4]The George Washington University, WisemanM@gwu.edu

Introduction

Data from household surveys play an essential role in monitoring and assessing the consequences of food assistance programs for the well-being of Americans. Compared to other sources—most notably administrative data—surveys like the National Health and Nutrition Examination Survey (NHANES) provide more information on individual and household characteristics such as food security, nutrition, and body weight, and these are characteristics likely to be affected by participation in food assistance programs. Thus, by using survey information on both program participation and household or individual characteristics (and statistical methods for handling non-response), surveys can provide powerful evidence on the impacts of food assistance programs.

The utility of survey data for evaluating food program outcomes is diminished by respondent misreporting of participation in these programs. At least in principle, misreporting issues can be addressed by linking administrative records to household survey data. The potential benefits of such linkage are widely recognized, especially in social policy research. In practice, however, efforts at linkage encounter many problems, but experience is growing, and with gains in experience, more of the potential benefits are being realized.

The Texas NHANES Pilot (TNP) is part of a cooperative effort by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention of the U.S. Department of Health and Human Services, the Economic Research Service (ERS) of the U.S. Department of Agriculture, and the Ray Marshall Center for the Study of Human Resources (RMC) at the University of Texas, Austin to link administrative records from the Food Stamp Program (FSP)[1] in Texas to survey records from NHANES. The Ray Marshall Center determined if NHANES participants had records in the Texas food stamp program database under contracts. We conducted this study in one state as a pilot to learn more of what would be required to develop and implement match protocols appropriate to the NHANES should more general effort ever be contemplated. The immediate goals of this venture were to: (a) assess the feasibility of matching records from these disparate sources; (b) evaluate the accuracy of the match process; (c) use the matched data to estimate the prevalence and correlates of misreporting of FSP participation for this one-state subset of NHANES; and (d) assess the likely impacts of matching error on our estimates of misreporting.

[§]This paper is a short summary of our work and findings. A more complete working paper is available from the authors. The analysis and opinions expressed are the authors' and may not represent the views of their respective organizations.

[1] On October 1, 2008, what was formerly called the Food Stamp Program became the Supplemental Nutrition Assistance Program (SNAP). As virtually all of our data predate this transition, we use the old name and acronym throughout this paper.

Ultimately, we sought to obtain information to improve analyses of food assistance programs and policies through more informed use of valuable information gathered during household surveys.

This paper presents results to date from the TNP. Part 1 briefly describes the match process used to link data from food stamp administrative files in Texas to NHANES participants for interviews conducted in Texas for survey cycles 2005 through 2008. Part 2 summarizes the results of that data linkage in terms of how often food stamp participation status as reported in the survey did or did not correspond to food stamp participation status as indicated in the administrative records. Part 3 provides a general discussion of our results. In our larger working paper (available from the authors) we provide more specifics about the linkage algorithm, misreporting using weighted estimates and logistic regressions models, creating a face validity scoring algorithm for sensitivity analyses of the matched results, and lessons learned through this linkage project. This paper serves as a short summary of our work and findings.

It is important to remember that, while the NHANES is designed to produce nationally representative estimates of numerous measures of the health and well-being of the non-institutionalized U.S. population, the results presented herein are based on the sample from one state and cannot be generalized to the U.S. population. The results of this pilot should be interpreted with caution and in light of our objectives.

Part 1: The Match Process

The administrative data for this pilot were derived from the Texas System of Application, Verification, Eligibility, Referral, and Reporting system (SAVERR) and the Texas Integrated Eligibility Redesign System (TIERS). SAVERR is a legacy system used to process applications for FSP, Temporary Assistance for Needy Families (TANF), and Medicaid with data going back to 1978. Implementation of TIERS began in 2003 and has continued through to the present. Operating under contract to the Texas Health and Human Services Commission, the Ray Marshall Center regularly receives a concatenated file of SAVERR and TIERS records covering the entire state FSP caseload.

Match procedures were divided into two matching phases and a "finalization" phase, with the first phase yielding a large pool of potential pair-wise matches between NHANES participants and administrative records. The second phase used more conservative match criteria to yield a reduced sample of potential linked records. Additional steps were then taken to produce a final dataset of linked records.

The overall match process was informed by the seminal work by Fellegi and Sunter (1969) and further discussion by Winkler (1993). The TNP's match process may be fairly characterized as a hybrid approach between purely probabilistic and deterministic matching that allows users to introduce their own estimates of the discriminatory power of a matching variable by assigning how many points that variable may potentially contribute to the overall match score. The components of the TNP match process are still being evaluated.

NHANES participants were eligible for linkage if they were part of the Texas NHANES sample, if they did not refuse to provide a social security number (SSN) at the NHANES interview and they provided sufficient personally identifiable information (PII) for linkage. Eligibility for linkage for a NHANES participant was independent of the questionnaire responses to the food stamp questions and independent of a participant's SNAP program eligibility.

NHANES 2005-2008 included 2,755 SPs from Texas. Of this total, 455 were not eligible for data matching, and an additional 40 records did not include information on FSP status. Another five records were successfully matched to FSP administrative files, but the administrative data indicated that these people were in FSP cases newly certified for receipt at some time during their sample month. Since the NHANES food stamp questions refer to receipt, it was not clear how the status of newly certified recipients should have been reported in the family interview and, as a

result, how the NHANES response should be compared to the administrative data. We excluded all five. This left us with an analysis sample of 2,255 persons.

Part 2: Misreporting of Food Stamp Program Participation

We focused on differences between survey responses and administrative data related to FSP status of the households of sample participants (SPs) at the time of the NHANES interviews and in the recent past ("recent" refers to connection with the FSP at any time during the 12-month period leading up to the survey month and status during the survey month itself). We employed data from NHANES 2005-2008.

For our initial examination of survey misreporting of food stamp participation, we assumed that our match of these data with administrative records was successful, and that all persons in the "matchable" (i.e., eligible for match) sample who were in fact either current or recent FSP recipients had been so identified. Those SPs not successfully linked were assumed to be non-recipients. Our definitions of current and recent status identified various types of FSP status misreporting. Assuming status as revealed in the administrative data is "true," the survey reports may differ from truth in several ways. Table 1 illustrates types of both under- and over-reporting. For example, SPs in cell 3 were under-reporters: The administrative data indicated they were current recipients, yet no receipt was reported for anyone in their household in NHANES. In contrast, SPs in cell 4 were possibly over-reporters; they claimed (or were claimed by the person completing the NHANES family interview) to be current recipients, which they were not, according to the administrative database. They were, however, recent recipients.

Table 1: Types of Misreporting, NHANES FSP Participation [1]

Types of Misreporting, NHANES FSP Participation			
	FSP participation status as reported in NHANES (subcategories are mutually exclusive)		
FSP participation status per administrative records (mutually exclusive)	**Current**	**Recent [2], but not Current**	**No FSP**
Current	1	**2 (UR)**	**3 (UR)**
Recent, but not Current	**4 (OR)**	5	**6 (UR)**
No FSP	**7 (OR)**	**8 (OR)**	9
FSP *under-reporters (UR)* are represented by gold cells (2, 3, and 6).			
FSP *over-reporters (OR)* are represented by red cells (4, 7, and 8).			
Correctly reported status represented by grey cells (1, 5, and 9).			

[1] Numbers in cells indicate cell position not sample size.

[2] Recent refers to connection with the FSP at any time during the 12-month period leading up to the survey month and status during the survey month itself.

Table 2 illustrates the results of the types of misreporting based on the Texas NHANES Pilot linked data. Eighty-three percent (1,865/2,255) had correctly reported their status. About 11% (241/2,255) were classified as under-reporters, and about 7 % (149/2,255) were classified as over-reporters.

Table 2: Joint distribution of Reported and Actual Food Stamp Program Participation Status in Texas NHANES Pilot: 2005-2008, unweighted counts.

	Joint Distribution of Reported and Actual Food Stamp Program Participation Status in the Texas NHANES Pilot: 2005-2008 (unweighted counts)			
FSP participation status per administrative records (mutually exclusive)	**FSP participation status as reported in NHANES (subcategories are mutually exclusive)**			
	Current n	**Recent[1], but not Current** n	**No FSP** n	**Total** n
Current	279	60	107	446
Recent, but not Current	*16*	27	74	117
No FSP	*80*	*53*	1,559	1,692
Total	375	140	1,740	2,255

FSP *under-reporters* (UR) are represented by gold cells (cell number 2, 3, and 6).

FSP *over-reporters* (OR) are represented by red cells (cell numbers 4, 7, and 8).

Correctly reported status represented by grey cells (cell numbers 1, 5, and 9).

Source: Tabulations by authors from special NHANES extract.

[1] Recent refers to connection with the FSP at any time during the 12-month period leading up to the survey month and status during the survey month itself.

Preliminary modeling of how SP and household characteristics relate to the likelihood of food stamp misreporting suggests that the SP's race/ethnicity and citizenship and the household reference person's gender and education level may be correlated with misreporting, but further refinements to model specification and structure are needed.

Part 3: Discussion

Due to the special nature of the analysis sample, the empirical results are not generalizable to the U.S. population, but the results do provide suggestions for future research related to linking FSP data with survey data. They also may provide some guidance for modification of future questionnaire design in the NHANES.

In the course of this research effort, we and other project members at ERS, NCHS, and RMC have learned a great deal about procedures for identifying NHANES participants in administrative files and the incidence of misreporting for this subgroup. Given the importance of understanding the connection between instruments of nutrition policy such as food stamps and health, continued work appears justified. We think there would be benefit by expanding the effort to include subsequent NHANES cycles for Texas (as planned) and to replicate the effort for another state. Replication in another state would serve two important purposes. One is to compare results from another state to the Texas data about FSP/SNAP misreporting. The other is to provide more experience with the replicability of our linkage procedures in other administrative contexts and to compare results following our protocol with what has and is being done by others. However, replication in another state may be difficult if there is not a centralized database of food stamp records, if their systems are not designed for linkage, and if there are cost implications. In addition,

the TNP linkage protocol is still being evaluated and improved. Nevertheless, more pilot efforts could provide useful information to develop linking strategies for surveys and state based programs.

References

Fellegi, I. P., and A. B. Sunter (1969), "A Theory for Record Linkage," Journal of the American Statistical Association, 64, pp. 1183-1210

Winkler, William E. "Matching and Record Linkage." U.S. Bureau of the Census, Working Paper. URL: http://www.census.gov/srd/papers/pdf/rr93-8.pdf; accessed on 12/14/2011.