

APPENDIX B
FIELD TEST RESULTS

This page has been left blank for double-sided copying.

MEMORANDUM**TO:** FoodAPS TWG Members**FROM:** Mathematica Policy Research**DATE:** 7/11/2011

Rev. 9/29/2011

SUBJECT: FoodAPS Field Test -- Sampling Design and Response Rates

Mathematica Policy Research is currently conducting the field test of the National Household Food Acquisition and Purchase Survey (FoodAPS), referred to in the field as the National Food Study. The results of this field test will be used to inform the design of the full-scale survey scheduled for a six-month field period from March 2012 through September 2012. This memo provides a summary of the field test sampling design, sampling procedures, weighting, and response rates. We also provide an overview of operational challenges that may have affected response rates, and recommended changes in procedures for the full-scale survey. Some of the first section of this memorandum was provided to you in the "Overview of the Field Test Design" dated May 2.

SAMPLING DESIGN

The design for the full-scale survey calls for a multistage design with 50 primary sampling units (PSUs) selected at the first stage, and 8 secondary sampling units (SSUs) selected within each PSU at the second stage for a total of 400 SSUs. Addresses will be selected at the third stage of sampling within each SSU.

For the field test, our target sample size was 400 completed cases, divided among the groups of interest:

- 200 SNAP households;
- 120 non-SNAP households with income between poverty and 185 percent of poverty (low income); and
- 80 non-SNAP households with income less than poverty (very low income).

These sample sizes were chosen to allow adequate precision for estimates of response rates, household burden, and data quality, both overall and for the SNAP and non-SNAP groups. For estimates of response rates, we expected 95 percent confidence intervals of between ± 5.7 and ± 6.2 percentage points for the sample as a whole (depending on the extent of the design effect) and of no more than ± 8.1 percentage points for each of the groups of SNAP and non-SNAP households. This design would also enable us to evaluate sampling and data collection procedures for the two groups. The sample of 80 very low-income non-SNAP households was

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 2

not intended to` provide precise estimates for this group but can inform us of large differences between this group and others.

First Stage of Selection

The field test was conducted in two purposively selected PSUs (counties) in New Jersey. The TWG had recommended purposive selection of PSUs that provide a mix of observed food acquisitions so as to test our survey protocols and gauge the quality of the data obtained from FoodAPS. We measured the expected diversity of food acquisitions by the racial/ethnic mix of the population and by the retail food environment (distribution of SNAP redemptions by type of retailer). Atlantic and Essex counties were selected for the field test. These counties have the characteristics shown in Table 1.

TABLE 1. FIELD TEST PSUS

County	SNAP retailers		Estimated SNAP/non-SNAP distribution of households			Population distribution by race / ethnicity			
	# SNAP retailers	%SNAP redemptions at supermarkets	% SNAP HHs	% non-SNAP <100% FPL	% non-SNAP <185% FPL	White	Hisp	Black	Asian
Atlantic	219	78.5%	8.1%	2.8%	15.7%	65.6%	14.3%	17.8%	6.9%
Essex	763	57.1%	11.5%	3.8%	14.1%	42.6%	18.7%	42.0%	4.6%

Second Stage of Selection

Eight SSUs (Census Block Groups) were selected within each PSU, with a goal of obtaining low-income areas with a variety of retail food environments. SSUs were selected using probability proportional to size (PPS) sampling. The measure of size (MOS) was a composite of three measures: the number of SNAP households in the SSU (as calculated from SNAP administrative data), estimates of the numbers of low-income non-SNAP households in the SSU, and estimates of the numbers of very low-income non-SNAP households in the SSU. The composite measure, of numbers within each SSU of households in each of the sampling strata, reflects the relative overall sampling rate of households within the SSU. The composite MOS is intended to enable us to obtain samples of households that have nearly equal probabilities of selection within each study population group.

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 3

Potential inaccuracies in the data used to create the MOS¹ and the fact that the SNAP frame contained a lower than expected proportion of SNAP households (discussed below) could have led to a higher than hoped for increase in sampling error (design effect) due to unequal weighting.

We set the sampling rates for the field test assuming that the SNAP frame would yield almost all of the SNAP households. The sampling rates for the two frames would have been a bit different if we had better anticipated the respective yields of the two frames.

Some of these issues will be alleviated in the full study. We plan to use five-year ACS files² to create the MOS. We will also try to make sure the SNAP frame used for sampling within SSUs is more current. If needed we can adjust the sampling rates within frames to reflect the actual yields. Either or both should reduce the design effects of weighting within Quota groups, and the first approach should reduce field costs somewhat since we will be able to more efficiently identify SNAP households. However, there are “costs” to these approaches: ensuring the currency of the SNAP data may entail getting multiple SNAP frames over the course of the field period; additional fine-tuning of sampling rates during the field period will increase the costs of sampling, and will not reduce field costs.

Third Stage of Selection

Sampling of addresses at the third stage of selection followed the same procedures planned for the full-scale survey. Within sampled SSUs, we sampled addresses for screening from two or three sources (sampling frames):

- **SNAP frame.** Address list of SNAP participant households as of November 2010, obtained from the State SNAP Agency. These addresses comprised the frame for selecting households expected to be receiving SNAP.
- **Non-SNAP frame.** An Address-Based Sampling (ABS) frame, a commercial list of addresses compiled from the United States Postal Service Delivery Sequence File. Addresses in the SNAP frame were eliminated from the non-SNAP frame.

¹ Because current Block Group level data were not available from the American Community Survey (ACS) or the Decennial Census, we used estimates provided by our sample vendor, Marketing Systems Group. Their estimates are derived from Claritas, now Nielsen Claritas.

² The ACS Block Group level files are tabular rather than microdata, so constructing more accurate MOS may be more challenging than we anticipated.

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 4

- **Listing frame.** Some multi-unit buildings appeared as a single record in the ABS frame; such records contained an indication of the number of units, but no unit numbers.³ If these buildings were “hit” in the sampling process, they were field listed to obtain unit numbers and units were sampled from the listings.

Addresses in the SNAP and non-SNAP frame were selected directly. Buildings in the Listing frame were selected with PPS and after listing, addresses were selected randomly from each listed building.

Initial samples of 400 SNAP and 4,000 non-SNAP addresses were selected for screening.⁴ We used two methods for sampling addresses from the non-SNAP frame. In 12 of the 16 SSUs we selected an equal probability sample of non-SNAP addresses. In the other 4 SSUs (2 per PSU), we oversampled addresses on the non-SNAP frame that are adjacent to addresses on the SNAP frame. The theory underlying this test is that nonparticipants who are eligible for SNAP, or close to the SNAP’s income cutoff, are more likely to live in close proximity to SNAP households. We will evaluate this procedure to see if it can reduce data collection costs, and also estimate the extent to which it may increase the design effect.

The addresses in the initial samples were randomly sorted into 20 replicate subsamples and assigned to 3 release waves. Assignment to replicates provided flexibility. After the target number of interviews for any of the three groups (SNAP, low-income non-SNAP, very low-income non-SNAP) was attained, that group could be made ineligible for interviewing in subsequent releases. Further, if the SNAP target was met before all replicates had been released, subsequent releases would not include any addresses from the SNAP frame.

As a check on the completeness of the sampling frame, we identified an adjacent address for each sampled address. Field interviewers were instructed to confirm that the adjacent address was in fact adjacent to the sampled address and to report any building or unit that was between

³ These buildings are called “drop points” on the ABS frame.

⁴ Mathematica expected to need 300 SNAP addresses, with plans to keep a random sub-sample of 100 in reserve. For SNAP households, we assumed 95 percent of addresses contacted would be eligible. (The other 5 percent would either be invalid addresses, non-household housing units or would, at the time of the contact, be occupied by a household that is not eligible for the survey.) In addition, we assumed that 87 percent of the addresses provided will result in a contact. For other households, we expected the screening eligibility rate for the eligible non-SNAP group to be 25 percent in the SSUs selected. We expected to make contact with 90 percent of households and obtain a screener completion rate of 75 percent and complete data collection among 81 percent of eligibles. We expected 80 percent of the addresses on the ABS frame to be deliverable household addresses. If these assumptions were correct, we would need a sample of 1828 addresses after unduplicating with SNAP records, and a total of 2504 before unduplicating. To allow for inaccuracies in our assumptions we selected an initial sample of 4000 addresses.

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 5

the sampled address and adjacent address. Adjacent addresses were assigned according to the type of sampled address:

- Single unit building – adjacent address was identified by sorting street addresses that are on the same side of the street.⁵ The adjacent address was the address before the sampled address in numerical order.
- Multiple unit buildings – If the sampled unit was not the highest numbered unit in the building, field interviews confirmed the number of units in the building. If the sampled unit was the highest numbered unit, field interviews confirmed the adjacent address which was identified the same way as for single unit buildings.

We will report on the results of the checking procedures in a follow-up memo.

Sample Release

The sample was released in stages as shown below.

TABLE 2. TIMING OF SAMPLE RELEASES

Release	Replicates	Number of sampled addresses	Mailing date for advance letters	Date sent to field	Comments
1	1-9	1,223	January 14	January 31	12 of 16 SSUs
1	1-9	423	January 23	January 31	3 SSU
1	1-9	151	February 8	February 14	1 SSU
2	10-15	1,104	February 8	February 14	SNAP & ABS frames released. On March 13, we pulled back all non-SNAP cases leaving 118 SNAP addresses from release 2.
1	From listing	55	March 23	March 30	
3	16-20	102	April 9	April 14	SNAP only
Total in field	1-18	2,017			

As noted above, each sampled address was randomly assigned to a survey protocol and incentive level. Field interviewers confirmed the presence of a housing unit at each sampled address and administered a screener to determine the household's eligibility for the survey. Eligibility was determined by membership in a quota group:

⁵ In practice, all street addresses were assigned the postal service delivery point code and sorted by that code.

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 6

1. Quota group A – Non-SNAP, household income \leq 100% of poverty
2. Quota group B – Non-SNAP, household income between 100-185% of poverty
3. Quota group C – Non-SNAP, household income \geq 185% of poverty
4. Quota group D – SNAP participant household

Households in Quota group C were not eligible for the field test. Households screened into quota group B from releases 2 and 3 were also not eligible for the field test. (Quota group B was closed to cases in release 2 when the SNAP frame of release 2 was re-released to the field.)

SNAP participant households were identified in both the SNAP and ABS frame. Among all households completing a screener, 46 percent of those reporting SNAP participation were from the ABS frame. Among completed interviews (defined as households completing a data collection week), 48 percent of SNAP households were from the ABS frame.

Table 3 shows the final status, at the time that we ended field operations, of the 2017 sampled addresses released to the field. A small percentage (2.5) of released addresses were outside the sampling area.⁶ We determined the dwelling unit status for 97.5 percent of addresses, of which 84.3 percent were occupied dwelling units; 83.1 percent of occupied dwelling units were contacted for screening (12 percent of cases were retired after a maximum number of attempts⁷). Of those contacted and eligible for screening, 72.2 percent completed the screener and 27.8 percent refused.

Among households completing the screener, 76.1 percent were eligible for the study and 62.9 percent completed the initial household interview. Overall, half of refusals occurred during screening, with the other half occurring during Household Interview #1 or training. Among households who refused during the screener, most (90 percent) agreed to complete additional questions (“Refusal Form”) which provide demographic information for non-response analysis.

⁶ 50 of the addresses were outside the sampled SSUs and resolving their status did not require field contact.

⁷ Starting on April 11, we retired cases after a maximum number of attempts (18 attempts for replicates 1-6 and 10-20; 12 attempts for replicates 6-9). The maximum number was based on a review of the distribution of completed cases which showed that 99 percent of completes had 18 or fewer attempts.

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 7

TABLE 3. FINAL STATUS OF RELEASED SAMPLE

(Each line shown in bold is the denominator for calculating the unweighted percentages in the next section.)

	Total		Atlantic County		Essex County	
	Number	Percent	Number	Percent	Number	Percent
Distribution of Released Sample by Screener Status						
Sample Released	2017	100.00	1085	100.00	932	100.00
Outside sampling area	50	2.5	48	4.4	2	0.2
 Within sampling area	1967	97.5	1037	95.6	930	99.8
DWELLING UNIT DETERMINATION RATE (DDR)						
Dwelling unit not determined ^a	58	2.9	49	4.7	9	1.0
Dwelling units determined	1909	97.1	988	95.3	921	99.0
DWELLING UNITS						
Dwelling unit = No (vacant)	299	15.7	156	15.8	143	15.5
Dwelling unit = Yes	1610	84.3	832	84.2	778	84.5
SCREENING ELIGIBILITY DETERMINATION RATE (EDR)						
Not eligible for screening ^b	272	16.9	123	14.8	149	19.2
Eligible for screening	1338	83.1	709	85.2	629	80.8
SCREENER COOPERATION RATE (SCR)						
Refusal to complete screener	371	27.8	199	28.2	172	27.3
Screener complete	963	72.2	506	71.8	457	72.7
ELIGIBILITY FOR STUDY						
Ineligible for study (Income >185% FPL)	230	23.9	128	25.3	102	22.3
Eligible for study	733	76.1	378	74.7	355	77.7
HH #1 COMPLETION RATE (HH1CR)						
Completed HH #1	461	62.9	214	56.6	247	69.6
DISTRIBUTION OF ELIGIBLE HOUSEHOLDS THAT DID NOT COMPLETE HH1 #1						
1. Refused HH#11 at screening						
Completed refusal form	89	12.0	73	19.3	15	4.2
No refusal form	10	1.4	4	1.1	6	1.7
2. Refused at HH#1 or training	94	13.0	60	15.9	35	9.9
3. Appointment for HH#1 pending and not completed	76	10.4	26	6.9	50	14.1

^a The dwelling unit could not be determined for some addresses in locked buildings and gated communities.

^b Not eligible for screening if (a) case expired due to maximum attempts (N=199); language was not English or Spanish (N=52); respondent had physical impairment (N=5); respondent unavailable during field period (N=16); group quarters (N=4).

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 8

WEIGHTING

For analysis of the field test data, it is appropriate that each PSU be given equal weight at least at the screener level. We note that if the 2 field test PSUs had been selected with PPS, as the PSUs for the main study have been, the sums of the analysis weights for the households in each would be approximately equal (the composite MOS and non-response adjustments would introduce some differences). Within PSUs the field test analysis weights adjust for differences in probability of selection and propensities to respond.

The field test screener weights have 3 components:

1. The first component of a household's weight is the inverse of its probability of selection. We calculated the probabilities of selection for each household as the product of the selection probabilities of the SSU to which it belongs and the address at which it resides. The probability of selection for the address had only one component if the building was not listed. For buildings that were listed the address probability of selection is the product of the probability of the building being selected and the probability of the listed address unit being selected within the building.
2. The next component is a non-response adjustment calculated separately within PSU for each frame: SNAP addresses and the ABS frame. Before constructing the weights we examined the distribution of responses by SSU to determine if SSUs or their characteristics should be used in adjusting for non-response, but decided not to. Insufficient data about non-responding units precluded using other characteristics as weighting factors.
3. The screener weights were scaled so that the sums of the weights are the same for each PSU. The non response adjusted weights were examined to determine if trimming is appropriate and it was decided not to trim them.

Two additional weights were constructed: HH1 weights are for respondents to Household Interview #1 (462 households that started the data collection week); HH3 weights are for households that completed Household Interview #3 (411 households that started and completed the data collection week). For the HH1 and HH3 data the weights are the screener weights adjusted for further sampling⁸ and for response to these particular data collection efforts. The non-response adjustment cells were the same as used for the screener.

⁸ Depending on the quota group, a completed case might not be selected to continue. Quota Group B households were included in the Household Interview 1 sample only if they were part of Release 1.

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 9

RESPONSE RATES

Response rates were calculated in a way that is equivalent to AAPOR Response Rate 4 (AAPOR 2011). The screener response rate (SRR) is the product of

- The dwelling unit determination rate (DRR) or the percent of attempted addresses where it was determined whether the unit was a dwelling unit.
- The screening eligibility determination rate (EDR) or the percent of dwelling units where we determined whether or not household lived there.
- The screener cooperation rate (SCR) or the percent of known households that completed the screener.

Thus:

$$\begin{aligned} \text{SRR} &= \text{DRR} * \text{EDR} * \text{SCR} \\ &= (97.05\%)(83.11\%)(72.19\%) \\ &= 58.2\% \end{aligned}$$

For Household Interviews 1 and 3 (HH1 and HH3), the response rates (RRH1 and RRH3) were calculated as the product of the SRR and the completion rates for HH1 and HH3 (CRH1 AND CRH3).

$$\begin{aligned} \text{RRH1} &= 58.2\% (62.9\%) = 36.6\% \\ \text{RRH3} &= 58.2\% (56.3\%) = 32.7\% \end{aligned}$$

RRH3 can be considered the response rate for the entire survey.

Response Rates Compared to Expectations

Our proposed sampling and data collection budget was based on four assumptions about the behavior of the sample. These assumptions are shown in Table 4 along with field test results.

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 10

TABLE 4. SAMPLING ASSUMPTIONS AND PRELIMINARY INDICATORS
 FROM THE FIELD TEST

	Proposal		Field Test	
	SNAP	ABS	SNAP	ABS
Percent of addresses ineligible for survey (invalid address; non-housing unit; group home)	5.0%	10.0%	12.3%	16.5%
Screener completion rate (completed screeners as a percent of eligible addresses)	87.0%	67.5%	76.9%	70.89%
Percent of screened households eligible for survey	95.0%	39.0%	88.0%	72.3%
Percent of eligible households agreeing to participate	90.0%	90.0%	74.8%	72.3%

These finding from the field test have implications for the full-scale survey:

1. **Percent of ineligible addresses** is higher than expected in both sampling frames – implies that we must release and work a larger sample
2. **Screener completion rate** is lower than expected for the SNAP frame – implies that we must release and work a larger sample from the SNAP frame
3. **Eligibility rate among screened households** is higher than expected in the ABS frame – implies a smaller release of households from the ABS frame
4. **Participation rate among eligible households** is lower than expected in both frames⁹ – implies that we must release and screen more households

The implications of these findings for sample release are offsetting because the eligibility rate in the ABS frame is higher than expected. However, there are two factors that make extrapolating from our experience a bit difficult:

1. The high ABS eligibility rate is because both field test PSUs were low income urban areas, whereas the full study will have a broader mix of areas (the 39 percent estimate was based on national data).
2. The overall eligibility of the ABS frame was increased by the inaccuracies (or aging) of the SNAP frame. (A large proportion of the SNAP households came from the ABS.) If the SNAP frame is more current (through change #4), the eligibility rate for ABS will be lower.

⁹ The low and high incentives provided a small difference in participation rates (74 versus 78 percent).

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 11

Therefore the main implication from these findings is the lower than expected screener completion rate for SNAP and lower than expected participation rate in both groups. These findings imply a larger than anticipated release of sample for the full-scale survey. If we assume that rate #3 (participation of eligible households) is 80% and not 90%, then we must release and screen **16% more addresses** than anticipated.

VARIANCE ESTIMATION

The field test sample was designed to provide adequate precision to evaluate data collection procedures and data quality and to provide adequate precision and adequate power to detect differences between the experimental treatments on cooperation rates.

Expected design effects for the full-scale survey are shown in Table 5, and estimated design effects from the field test are shown in Table 6. For the full-scale survey, we estimated design effects (Deffs) to range from 1.38 to 2.38 for measures with low intracluster correlations (ICCs). These estimates are based on Deffs reported in Cohen et al. 1999¹⁰ and further analysis of the same data. We expect values of the ICCs to be between 0.01 and 0.05, and Table 5 shows the values of Deff at the ICC values of both 0.01 (first two rows) and 0.05 (last two rows). The value of Deff_w of 1.07 was derived from the same study.

TABLE 5. EXPECTED DESIGN EFFECTS FOR THE FULL-SCALE NATIONAL FOOD STUDY, FOR THE WHOLE SAMPLE AND SUBGROUPS

Group	Completed Households	PSUs	b-1	ICC	Deff_c	Deff_w	Deff	Effective n
All	3,000	50	59	0.01	1.59	1.5	2.38	1,257
SNAP	1,500	50	29	0.01	1.29	1.07	1.38	1,086
All	3,000	50	59	0.05	3.95	1.5	5.92	506
SNAP	1,500	50	29	0.05	2.45	1.07	2.62	572.

Notes: $Deff = deff_c * deff_w$
 deff_c is the design effect due to clustering
 deff_w is the design effect due to unequal weights
 $deff_c = 1 + ICC(b-1)$
 ICC is the intracluster correlation
 b is the number of cases per PSU

¹⁰ Cohen, Barbara, James Ohls, Margaret Andrews, Michael Ponza, Lorenzo Moreno, Amy Zambrowski, and Rhoda Cohen. "Food Stamp Participants' Food Security and Nutrient Availability: Final Report." Princeton, NJ: Mathematica Policy Research, July 1999.

MEMO TO: FoodAPS TWG Members
 FROM: Mathematica Policy Research
 DATE: 7/11/2011
 PAGE: 12

Estimated design effects and ICC from the field test are means of DEFFs and ICC for 19 outcome measures (Attach A).¹¹ The design effects are computed treating the PSUs as strata and the SSUs as PSUs. If these estimates of the ICC are indicative of the SSU level clustering for the main study, the overall design effect of clustering will be somewhat higher, but not inconsistent with the higher level in Table 5. For the main study, the number of households per SSU will be smaller than in the field test so the impact of the SSU level clustering will be less. Since we expect the level of clustering (ICC) at the PSU level to be much less than at the SSU level, the total design effects of clustering for the main study should be less than the higher level presented in Table 5.

The design effect of weighting was higher than anticipated for subgroups, but this is probably due to two factors: sampling frame issues (discussed above) and the strategy of over sampling households next to SNAP addresses.

TABLE 6. DESIGN EFFECTS¹² FROM THE FIELD TEST

Quota Group	Average Completed Households	b-1	ICC	Deff-c	Deff-w	Deff	Effective n
All	390.4	23.4	.035	1.82	1.48	2.69	144.9
A. Non-SNAP, income <100% FPL	64.0	4.0	.020	1.05	1.32	1.39	46.0
B. Non-SNAP, income 100-185% FPL	134.4	8.4	.034	1.31	1.47	1.93	69.6
D. SNAP	193.6	11.1	.019	1.21	1.31	1.59	121.8

Notes: Deff = deff_c * deff_w
 deff_c is the design effect due to clustering
 deff_w is the design effect due to unequal weights
 $deff_c = 1 + ICC(b-1)$
 ICC is the intracluster correlation
 b is the average number of cases per PSU (for the field test, these are the SSUs)

¹¹ These include five measures normalized per adult male equivalent (number of food-at-home (FAH) and food-away-from-home (FAFH) acquisitions, FAH and FAFH expenditures, total food expenditures); number of school meals per school-age children; percent of FAH acquisitions that were (a) free, (b) at a supercenter, (c) respondent reported a saved receipt, (d) scanned data matched a receipt; percent of FAFH acquisitions that were (a) free, (b) at a top 30 fast food or full service restaurant, (c) respondent reported a saved receipt; number of stores where household shopped last month; whether primary store is a supercenter; whether household reported no change in food shopping in the data collection week; household food insecurity; awareness of the MyPyramid; whether respondent uses nutrient facts panel when shopping.

¹² Deff-w was estimated as 1 plus the relvariance of the weight variable. Deff was computed by SUDAAN, the software used to analyze the field test data. Deff-c was estimated as $Deff-c = Deff / Deff-w$. The ICC was derived as $ICC = (Deff-c - 1) / (b - 1)$. Nominal n is (average completed households / Deff).

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 13

NON-RESPONSE BIAS ANALYSIS

We will report on non-response bias analysis in a subsequent memo.

OVERVIEW OF OPERATIONAL CHALLENGES

The FoodAPS Field Test was an ambitious undertaking. We learned that households were willing to participate in the survey (information about the quality of collected data is presented in a separate memo). However, the collection of food acquisition data over a 7-day period had not previously been tested and this field test revealed some challenges which led to a longer than expected field period and possible impact on response.

Field activities were initially scheduled to run eight weeks from January 23 through March 17. The field period lasted 15 weeks from January 31 through May 22. The beginning of field operation was postponed to begin with the SNAP benefit distribution schedule, and a one-week break in field activities was scheduled to conduct preliminary analyses.¹³ The length of the field period was affected by the following:

1. Delay in obtaining printed field materials – OMB clearance was obtained Dec 27, 2020. Printing the large number of materials required to field two different protocols took longer than expected. During the first week of the field period, interviewers were able to work only half of the sample assigned to the single book protocol.
2. New Jersey experienced a very harsh winter – Winter weather slowed field activities even when interviewers were in the field. Interviewers experienced navigation problems, limited parking, and respondents' reluctance to converse in their doorway.
3. Greater need for training and monitoring for the screening effort – Our planning for the field test focused primarily on procedures for collecting and processing food acquisition data. We designed an efficient screening instrument, but the three-day field interviewer training provided little room for training interviewers on efficiently organizing and implementing the screening effort.¹⁴

¹³ The one-week break was taken from March 6-13 while we conducted preliminary analyses requested by OMB.

¹⁴ The three-day field interviewer training focused on training households to use the food reporting instruments under two separate survey protocols (day 1), administering the screener and CAPI interviews (day 2), practicing mock interviews and reporting time and expenses (day 3).

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 14

Two corrections were made while in the field:

- In mid-March we trained “screener-only” interviewers. These interviewers worked in teams with other interviewers and only administered the screener. After screening an eligible household, they handed the case off to an interviewer who was trained and equipped to administer the household interviews and train the household to track food acquisitions. This strategy allowed us to increase field staff without the full training, however, it necessarily imposed a lag between screening and the start of the data collection week for the household and may have contributed to the number of refusals that occurred after screening.
 - In April we provided specific guidelines to interviewers regarding days of the week and time slots that must be filled when making attempts on each case before a case could be retired. At the same time we set a maximum number of attempts after which the case was retired.
4. Computer systems were not well suited for managing the large screening effort – Mathematica’s standard sample management system proved to be a cumbersome tool for logging the large number of attempts that interviewers made on days when they spent all of their time knocking on doors and screening households. As a result, the ratio of administrative time to field time was much higher than expected, resulting in less field time.

Our mid-field correction included deployment of “screener-only” interviewers to increase field time; and implementation of an alternate logging system whereby field interviewers reported their activity to the survey operations center by telephone each night.

5. Too few scanners in the field – During the field test it became apparent that the project budget had planned for a number of scanners to provide for average demand (average number of households active in a week), rather than peak demand. One hundred scanners were in the field, allocated across about 35 interviewers. If an interviewer identified an eligible household and did not have an available scanner, they made an appointment to begin data collection at a future date.

During our debriefings with field interviewers, we learned that: (1) Some interviewers did not initially understand that they could continue screening when they had no scanner available. This led to a delay in screening activities. (2) Some interviewers hoarded their scanners rather than share them with team members

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 15

because they firmly believed that they would “lose a household” if they had to make an appointment. (3) Only a minority of interviewers thought they had a high success rate when appointments were made.¹⁵

RECOMMENDED CHANGES FOR THE FULL-SCALE SURVEY

The field test revealed several opportunities for improving field operations. These are:

1. **Acquire additional SNAP data** – One of the findings from the field test is that the SNAP caseload is fluid, thus reducing the efficiency of the SNAP sampling frame. Half of all SNAP households were identified from the ABS sampling frame. Presumably SNAP households from the ABS frame enrolled in SNAP after the time that we obtained the SNAP caseload data for sampling (SNAP participation will be confirmed with administrative data at the end of the field period).

One option for increasing the efficiency of the sampling design is to refresh the SNAP frame. This would be achieved by randomly splitting the ABS master list of addresses into two parts to work in the first and second halves of the field period. SNAP data obtained in January 2012 will be matched to the first half of the ABS frame, providing a SNAP frame to sample cases for the first half of the field period (March-May). SNAP data obtained in April will be matched to the second half of the ABS frame and provide sample for the second half of the period (June-August).

Use of additional SNAP data will involve:

- Coordinating with 27 States for an additional data extract
 - Geocoding additional SNAP extract from each state and matching it with the ABS frame
 - Managing the split sample and the added complexity of developing sampling weights
2. **Revise the field interviewer training schedule** – Training for the full-scale survey will not include training on two survey protocols and two incentive levels. Therefore, the schedule may be revised to include more structured training on screening activities, including finding sampled addresses, mapping out efficient routes, and making attempts within a structured schedule of time slots and days of the week.

¹⁵ Information about the lag between screening and the start of the data collection week is presented in a second memo.

MEMO TO: FoodAPS TWG Members
FROM: Mathematica Policy Research
DATE: 7/11/2011
PAGE: 16

- 3. Implement a new Sample Management System** – Mathematica suggests implementation of a web-based system to manage the screening effort for the full-scale survey. A web-based system will be more responsive and intuitive for field interviewers, compared with the legacy system used for the field test. A web-based system will also provide real-time reports with drill-down capability for supervisory staff who monitor the screening effort.
- 4. Acquire additional scanners** – The number of scanners acquired for the field test was based on an average number of households participating per week. It became clear in the field test that scanners should be fielded to serve peak rather than average demand.
- 5. Provide replicate weights** – ERS has asked about the feasibility of providing replicate weights for FoodAPS. While such weights were not budgeted, they may be useful if ERS wishes to produce a Public (or Restricted) Use File, Mathematica has the experience and capabilities to create replicates and replicate weights for such a file.

REFERENCE

Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys.
<http://www.aapor.org/Content/aapor/AdvocacyandInitiatives/StandardsandEthics/StandardDefinitions/StandardDefinitions2011.pdf>

ATTACHMENT A

**DESIGN EFFECTS USED TO ESTIMATED
INTRA-CLUSTER CORRELATION**

TABLE A.1. DESIGN EFFECTS USED TO ESTIMATED INTRA-CLUSTER CORRELATION

	Survey Strata			
	Overall	SNAP	Very-Low Income, Non-SNAP	Low Income, Non-SNAP
	Design effect			
Mean # of FAH acquisitions	1.69	2.27	0.33	1.33
Mean # of FAFH acquisitions	0.74	1.62	0.92	0.35
Mean weekly FAH spending (\$)	3.34	1.16	1.43	3.35
Mean weekly FAFH spending (\$)	2.34	1.09	1.59	1.47
Mean total Weekly food spending (\$)	4.97	1.26	1.84	4.76
Mean # of school meals ¹	0.73	1.06	0.91	0.70
Mean # of stores where household buys food	2.62	2.77	1.18	2.01
Percent w/primary food store a supercenter	4.62	2.14	2.05	2.25
Percent not changing acquisition behavior during data collection week	2.22	2.55	0.53	2.64
Percent food insecure	3.13	1.80	1.47	1.72
Percent aware of MyPyramid	2.52	1.00	1.43	2.07
Percent using nutrient facts panel when shopping	2.76	0.79	2.46	1.57
Percent of FAH acquisitions obtained for free	1.60	2.17	3.20	0.51
Percent of FAH with saved receipt	2.20	1.76	0.87	1.07
Percent of FAH blue pages matched to receipt	2.97	1.63	0.85	2.12
Percent of FAH expenditures at supercenter	7.87	2.11	1.49	5.45
Percent of FAFH acquisitions obtained for free	1.27	1.36	0.85	0.85
Percent of FAFH acquisitions at top 30 fast food or full service restaurant	1.71	1.41	1.26	0.92
Percent of FAFH with saved receipt	1.72	0.32	1.81	1.61

¹ Average number of school meals is measured per number of children age 5 to 18.
Source: National Household Food Acquisition and Purchase Survey Field Test. Sample includes
completes (respondents to Household Interviews #1 and #3).

MEMORANDUM

TO: FoodAPS TWG Members

FROM: Mathematica Policy Research

SUBJECT: FoodAPS Field Test Findings

DATE: 7/11/2011

This memo summarizes findings from National Household Food Acquisition and Purchase Survey (FoodAPS) Field Test, conducted by Mathematica Policy Research from February through May 2011.

FIELD TEST OBJECTIVES

The primary objectives of the FoodAPS field test were:

1. Assess survey response, including burden and item non-response;
2. Assess data quality, including the consistency of information provided in multiple survey instruments, the completeness of data provided, and the potential to match data to external databases;
3. Assess the impact of alternative incentive levels and survey protocols (single book and multiple book) on participation, burden, and data quality.

This memo addresses each objective. Within the memo, we make frequent reference to findings presented in a set of appendix tables provided as a separate PDF document. Tabulations and were prepared to assess the quality and completeness of data obtained from each survey instrument. All of the appendix tables are weighted by sampling weights and estimated with Sudaan to account for the complex sampling design. The tables show tests of statistical significance for differences between the following subgroups of interest:

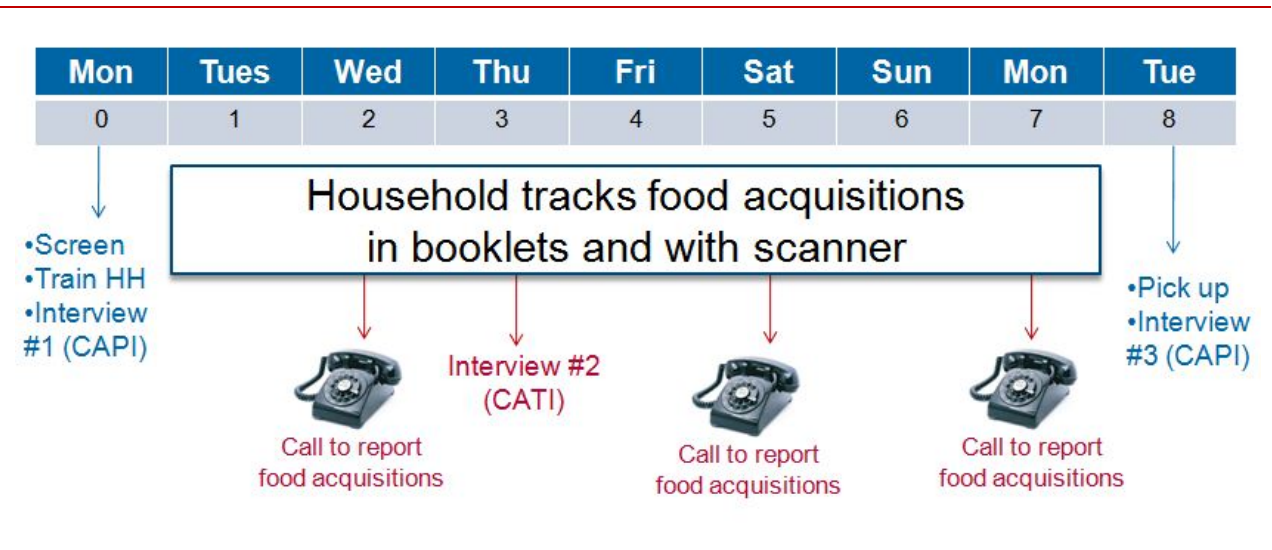
- Single Binder versus Multiple Book survey protocol
- Low versus High incentive level
- SNAP versus Non-SNAP households with income <100% FPL
- SNAP versus Non-SNAP households with income between 100-185%FPL

Appendix tables include asterisks to indicate statistically significant between-group differences, with the asterisks appearing on estimates for the second group listed in each bullet above.

1. SUMMARY OF THE FIELD TEST

The objective of FoodAPS is to obtain a comprehensive picture of household food acquisitions over a 7-day period, including food-at-home (FAH) and food-away-from-home (FAFH) from all sources including purchases and food obtained for free. Addresses were sampled from an address-based sampling frame and a SNAP sampling frame. Field interviewers screened households at sampled addresses to determine eligibility for the study. By design, the 7-day data collection week for the household was intended to begin immediately after screening, when possible. Participating households were in contact with survey staff over a 9-day period with contact at multiple points in the data collection week, as shown in Figure 1.

Figure 1. Data Collection Week for Households



The goal of the field test was to obtain 400 completed interviews with households, with a complete defined by starting and finishing the data collection week with completion of Household Interviews #1 and #3.

The survey collected food acquisition data through multiple instruments and three “food reporting calls.” The instruments are listed in Table 1. Instruments for FAH provided redundant reporting; instruments for FAFH provided some redundancy for respondents (multiple booklet sections and receipt) with the goal of resolving the separate sources of information in a single complete record during the food reporting calls.

As shown in Figure 1, households were asked to complete three household interviews in addition to the initial screener and food reporting. The topics included in each interview are listed in Table 2.

Table 1. Instruments for Reporting Food Acquisitions

Instrument	Method of reporting	Information reported	Data processing
Food-at-home (FAH) Reporting			
Daily List (booklet)	Record in booklet; report by telephone	Place, date, total amount paid	Entered in Food Reporting System (FRS) during telephone interview
Blue page (booklet)	Record in booklet	Place, day, total amount paid, payment methods, use of coupons/loyalty card, “who got the food,” items that could not be scanned	Data entry at Mathematica
Scanner	Scan items	Date/time stamp, barcode for place, barcodes from items or barcode book	Processed to delimit transactions and match UPCs to item descriptions
Receipts	Attach receipt to Blue page	Item descriptions, prices, store savings, coupon savings, weights (if applicable), quantities (if applicable) payment type, total paid	Data entry at Mathematica in system pre-loaded with scanner data
Food-away-from-home (FAFH) Reporting			
Daily List (booklet)	Record in booklet; report by telephone	Place, date, total amount paid	Entered in FRS during telephone interview
Red page (booklet)	Record in booklet; report by telephone	Place, day, total amount paid, payment methods, “who got the food”, items that are not on receipt	Entered in FRS during telephone interview
Receipts	Attach receipt to Red page and report details by telephone	Item descriptions, prices, payment type, total paid	Entered in FRS during telephone interview

Table 2. Household Interviews

Instrument	Expected Burden	Topics
Screener	8 minutes	<ul style="list-style-type: none"> • Confirm address • Identify additional housing units, if applicable • Usual frequency of food shopping • Household size and income • Identification of food shopper and meal planner • For eligible refusals: (a) primary food store; (b) food bank/pantry in last 30 days?; (c) types of stores where household shopped for food in last 30 days; (d) number of household members by age group,
Household Interview #1 (HH1)	19 minutes	<ul style="list-style-type: none"> • Household roster • For each household member: age, race, education, marital status, employment • Nutrition program participation (SNAP, WIC, school meals) • Child care and community meal programs • Usual food shopping location and means of travel
Household Interview #2 (HH2)	25 minutes	<ul style="list-style-type: none"> • Number of jobs and earned income last month • Unearned income by source, last month • Non-food expenditures last month for: housing, utilities, out-of-pocket medical expenses, education, recreation • Vehicle ownership/leasing and expenses • Identification of household assets (no amounts) • Life events: change in household membership or job, of major illness in past 12 months

Table 2. Household Interviews - *continued*

Instrument	Expected Burden	Topics
Household Interview #3 (HH3)	18 minutes	<ul style="list-style-type: none"> • Meals prepared in past 7 days / any guests for meals • General health status of respondent and family • Height and weight of each household member • Nutrition knowledge and attitudes • Special dietary needs • Food security • Previous residence / citizenship

Household Interview #1 (HH1) was to be conducted immediately after screening, however, field interviewers made appointments when respondents were not available immediately or when the interviewer did not have a scanner available for the household to begin the data collection week. Household Interview #2 (HH2) was attempted by telephone in the middle of the data collection week, with attempts continuing until the end of the field period if necessary. Household Interview #3 (HH3) was conducted at the end of the data collection week when field interviewers returned to the household to pick-up survey booklets and scanner, and to distribute the incentive.

Two additional data collections are not shown in Figure 1. First, households were asked to complete a “Meals and Snacks Form” during the data collection week to indicate, for each household member, the meals (breakfast, lunch, dinner) and snacks (morning, afternoon, evening) consumed each day during the week. Second, primary respondents were asked to complete a “Respondent Feedback Form,” self-administered on paper, at the end of the data collection week to report on the ease or difficulty of the data collection and whether they changed behavior because of the survey.

2. SURVEY RESPONSE AND BURDEN

Screening was completed for 1,665 addresses, with 961 households responding to the screener. From the 961 screeners completed, 731 households were determined eligible and 537 agreed to participate. Household Interview #1 was completed with 462 households; 411 households completed the data collection week.

Survey response rates are shown in Table 3 for the overall sample and for subgroups of interest (unweighted). We do not expect differences in the dwelling determination rate (DRR) or screening eligibility determination rate (EDR) for survey groups or incentive levels because addresses were randomly assigned to these groups. The SNAP frame has a higher eligibility rate, as expected.

Screener completion and response rates do not differ significantly by incentive group, suggesting that the incentive has little impact on households’ initial willingness to discuss the survey and determine their eligibility. The SNAP frame had higher screener completion and response rates than the ABS frame.

Response rates for the household interviews do not differ by survey protocol; thus if one protocol was significantly more difficult, the difficulty was not reflected in a significantly greater drop off in participation through the survey week. The higher incentive was associated with higher rates of household interview response, with the difference in response increasing from HH1 to the HH2 and HH3.

Table 3. Weighted Response Rates

Response rate	Overall	Survey Protocol		Incentive Level		Sampling Frame	
		Single Book	Multiple Books	Low	High	SNAP	ABS
Dwelling unit determination rate (DRR)	96.79	96.68	96.90	96.80	96.78	98.16	96.68
Screening eligibility determination rate (EDR) ²	84.93	84.73	85.12	85.15	84.72	89.12	84.57
Screening completion rate (SCR) ¹	70.25	69.11	71.35	67.59	72.73	76.56	69.68*
Screening response rate (SRR = DRR*EDR*SCR)	57.75	56.61	58.85	55.71	59.63	66.97	56.97
Household interview completion rates (CR)							
HH #1	60.98	60.80	61.16	56.02	65.25*	67.02	60.18
HH #2	53.26	53.01	53.51	47.03	58.61*	57.40	52.71
HH #3	55.19	54.62	55.75	49.00	60.50*	59.79	54.57
Household interview response rates (RR = SRR*CR)							
HH #1	35.21	34.42	35.99	31.21	38.91*	44.89	34.29
HH #2	30.76	30.01	31.49	26.20	34.95*	38.44	30.03
HH #3	31.87	30.92	32.81	27.30	36.08*	40.04	31.09

¹ Completed screeners as a percentage of eligible addresses.

² Percentage of screened households eligible for the survey.

Reporting of Food Acquisitions

In addition to the screener and household interviews, respondents were asked to participate in three food reporting telephone calls, track FAH and FAFH acquisitions, and provide scanner data (if applicable). Table A-1 shows the percentages of households responding to each component.

Only half of all households completed at least three food reporting calls, with no statistically significant differences across groups. Ninety percent of households, however, reported

information for all seven days of the data collection week during the food reporting calls. Thus, a significant percentage of households missed a call and recalled data for a longer period during a subsequent call. If households did not complete a call as scheduled, the telephone center attempted to reach them the next day. If the final food reporting call for the week was not complete when the field interviewer picked up materials, the field interviewer initiated the call for the respondent. Sixteen percent of households completed only one call, with most of these households reporting the entire 7-day week during one phone call.

FAH and FAFH acquisitions were reported by 90 and 82 percent of households, respectively, with 71 percent of households providing scanner data (FAH acquisitions may be reported without scanner data if items are written on blue pages). Overall, 55 percent of households reported all components of the data collection: 7 days reported by phone, at least one FAH and FAFH acquisition, and scanner data. This low rate of “complete” reporting partly reflects underreporting during telephone calls. As we discuss later, FAH data was collected using several methods and accounting for all reports indicates that 90 percent of households had at least one FAH acquisition and only 5 households had no food acquisitions.

Burden of Survey Components

Burden estimates are provided in Table A-2. Response burden for the screener was not determined because the screener was administered as a paper instrument. The overall estimate for HH1 was slightly above expectations (24 versus an expected 19 minutes); HH2 was somewhat below expectations (22 versus an expected 25 minutes); and HH3 matched expectations (18 minutes). There were no significant differences by survey type or incentive level. Non-SNAP households had greater burden in responding to HH2 regarding income and expenditures.

Total household burden for food reporting calls was 41 minutes, which is slightly above with our estimate of 13 minutes per call for three calls. The estimate is higher for multiple book households but not statistically significant.

The burden of scanning was measured by the date/time stamps recorded on the scanner at the start and end of each transaction. This estimate overstates actual scanning time if a respondent filled out the booklet as he scanned items; it understates the full burden of FAH reporting if the scanning was done entirely before or after completing the booklet. Thus it is not surprising that the mean time of 20 minutes is below our estimate of 25 minutes for reporting FAH. Multiple booklet households had higher scanning time, possibly because they had to complete forms (Daily List and Blue page) that were located in separate booklets.

Timing of Survey Components

FoodAPS conducted in-person screening. After determination of eligibility, households were asked to complete HH1 and training on the food reporting booklets, which could take up to 1.5 hours. Two-thirds of the households that ultimately completed the survey did not complete HH1 and training immediately after screening (Table A-3). Early in the field period these delays were at the request of the household (with some due to the lack of a scanner); in the last month of the

field period, appointments were also necessary when “screener only” interviewers identified eligible households and referred them to field staff trained for the full survey. Non-SNAP higher income households were more likely to delay HH1 than other groups.

Households were compliant in starting the data collection week on the day of HH1 or the day after (Table A-3). HH2 was completed on the third day of the data collection week, as scheduled, for half of the sample, and completed after the data collection week for only 5 percent of households. HH3 was completed on the day after the data collection week, as scheduled, for 61 percent of households and more than one-week later than scheduled for only 6 percent of households. Compared with the low incentive group, the high incentive group was somewhat less compliant with the study schedule for HH2 and HH3, but differences are not statistically significant.

Item Non-response

Most of the data items from the household interviews and food booklets were tabulated to assess item non-response. Table 4 shows the correspondence between survey instrument and appendix tables. (The consistency and quality of data reported for food acquisitions is discussed in Section 3.)

Item non-response on the household interviews is indicated by a response of “don’t know” or “refused.” Non-response to the Meals and Snacks Form is indicated when no meal or snack is checked off for a person on a given day. Non-response to the Respondent Feedback Form is indicated when no response category is checked for a question.

Overall, item non-response was more likely when the primary respondent was reporting information for other household members than when reporting own information. Item non-response was minimal, except for income measures reported on HH2. Item non-response did not vary significantly by subgroups of interest.

HH1 and HH3 exhibited minimal item non-response. On HH1, age, employment, and citizenship status of household members other than the main respondent were missing for at least one member in 0.7, 4.3, and 3.4 percent of households, respectively (Table A-4). HH3 had item non-response only for the height and weight questions. This missing data resulted in missing weight status (normal, overweight obese) for 2 percent of primary respondents and for at least one household member in 11 percent of households (Table A-14).¹

The paper Meals and Snacks Form was returned by 93.1 percent of households (Table A-16). The form was considered complete if there were no missing days for any household

¹ In an effort to reduce non-response to the question about body weight, we gave respondents the option to enter weight directly into interviewers’ laptops rather than providing the information verbally. Sixteen percent of households used this option to enter weight for one or more household members.

member: 79.7 percent of households returned a complete form and 13.3 percent returned an incomplete form. The survey protocol did not affect the likelihood of returning a complete form (80 percent), but single book households were more likely to return an incomplete form whereas multiple book households were more likely to return no form (10 versus 3 percent).

Among households that returned the Meals and Snacks Form, 5 percent of household members failed to complete the form on one day, and 4.2 percent of household members were missing more than one day of data. The percent of households returning any form was not significantly different by incentive level, but the low incentive group was significantly more likely to have any missing data (13.8 versus 6 percent).

Table 4. Correspondence between Survey Instrument and Appendix Tables

Instrument	Appendix Table
HH1	Table A-4—Characteristics of Households Table A-5—Nutrition Assistance Program Participation Table A-6—Characteristics of Primary Respondents Table A-7—Usual Food Acquisitions
HH2	Table A-8—Non-Wage Sources of Income Table A-9—Non-Wage Income Amounts Table A-10—Employment Data: Persons Age 16 and Over Table A-11—Percent of Households with Any Employed Persons and Missing Earnings Table A-12—Household Income as Percent of the Poverty Level
HH3	Table A-13—Nutrition Knowledge and Attitudes Table A-14—Household Member Health and Weight Status Table A-15—Food Security and Diet Assessment
Meals & Snacks Form	Table A-16—Consumption During the Week As Reported on the Meals and Snacks Form
Respondent Feedback Form	Table A-17—Respondent Feedback

The Respondent Feedback Form was completed on paper and returned to us by 96.8 percent of households. A very small additional item non-response occurred on 3 of the 4 questions, for a total non-response never exceeding 5 percent (Table A-16). There were no differences in response by subgroups of interest.

HH2 Item Non-response and Impact on Household Income

Household income and non-food expenditures from HH2 present the most significant challenge in terms of non-response. The measures of interest are total household income and total monthly non-food expenditures; thus a single missing component of these measures renders the measure of interest missing, or potentially largely understated. Due to time constraints, the non-food expenditure data from HH2 have not been examined.

Sixty percent of households reported at least one source of non-wage income (Table A-8). The most common sources are unemployment compensation, retirement income, disability or welfare payments, and fuel assistance (Table A-8). Of the 26 sources of non-wage income reported by any household, 15 sources exhibit some item non-response on dollar amounts. Overall, 16 percent of households reporting any non-wage income (9.7 percent of all households) had missing data for dollar amounts of non-wage income (not shown in table).²

Seventy-one percent of households reported at least one employed household member (Table A-4), and 43.5 percent of all household members over age 16 were reported to be employed (Table A-10). Survey questions about employment were structured as follows:

- How many jobs (do you/does NAME) work for pay?
- How many hours (do you/does NAME) usually work per week or per month at (your / his / her) (first/second/third) job?
- How often (are you/is [NAME]) paid from (your / his / her) (first/second/third) job?
- What is the amount of pay that (you/NAME) get per check from your (first/second/third) job before taxes and any deductions? IF NEEDED: If you were paid in cash not in a check, that is fine. We are just looking for the amount paid in this item.

Primary respondents were less likely to be employed than spouse/partners (50 versus 57 percent).³ Primary respondents were also less likely to have missing data (13.3 versus 23.5 percent were employed and missing data). Substantially more household members were missing data on amounts of pay than on hours (13.0 versus 5.9 percent for primary respondents; 23.5 versus 3.9 for spouse/partners) (Tables A-10). About one-third of children age 16 and over were employed, but only 8 percent were employed with no missing earnings data.

Overall, 4.3 percent of households did not respond to HH2; 30.4 percent have no employed persons; 38.3 percent have employed persons and no missing earnings data; and 27.1 percent have employed persons and missing earnings data (Table A-11).

² The percentages of households reporting each non-wage source of income in Tables A-8 and the percentages of all households with missing non-wage amounts in Table A-9 are measured among all households, including the 4.3 percent who did not respond to HH2.

³ Seventy-five percent of primary respondents were female (Table A-6).

As a result of missing data for earnings and non-wage income, 35 percent of households are missing total household income as a percentage of the poverty level. However, it is important to assess the accuracy of income reported at screener to determine if the screener properly identified households eligible for the survey and allocated them to the correct strata. Table A-12 provides an indication of the accuracy of income reported on the screener. The top panel shows the mean and median household income as a percentage of the poverty level for: (a) households with no missing data; (b) households with no missing data or missing earnings, with earnings imputed; and (c) all households, with missing earnings and non-wage income imputed. All of these measures indicate that households underreported income on the screener. The bottom panel compares only earnings to the poverty level to assess whether households primarily underreport (or forget to report) non-wage income. These measures provide a closer correspondence of household income with screener category.

3. QUALITY OF FOOD ACQUISITION DATA

The quality of food acquisition data was assessed by examining the following:

1. Consistency of FAH reporting across instruments: Daily List, Blue Page, receipt, scanner data
2. Percent of scanned items matched to item descriptions and prices
3. Completeness of FAFH telephone reporting (examined by comparing food booklets to the telephone data for a sample of households)
4. Item non-response on blue and red pages

One of the main questions for our assessment of data quality is whether or not households report all acquisitions. We will be able to validate SNAP purchases with EBT data, but have not completed that analysis yet. For non-SNAP purchases, the best indicators of whether households report acquisitions may be the degree to which they follow survey protocols (i.e., consistency across instruments) and whether we observe differences in reported acquisitions by survey protocol or incentive level.

Between-group differences discussed below should be viewed with caution because they are not based on multivariate analysis that controls for household characteristics. For example, the Single Book group has a somewhat higher percentage of SNAP households than the Multiple Book group (41 versus 32 percent; Table A-5), and likewise for the low versus high incentive group. In addition, SNAP households are significantly less likely to be single-person households than non-SNAP households (Table A-4), thus influencing results by survey strata.

Consistency of FAH Reporting Across Instruments

The greatest challenge for data processing was the match of FAH acquisitions across the following instruments:

- Daily List – place name and total paid was reported by telephone;
- Blue page – place name, total paid, and other details were recorded on paper;
- Scanner data – electronic file contains a date/time stamp and place category (if respondent scanned the place code);
- Receipt – contains place name, total paid and item details.

Three matching processes were used to match information across these sources:

1. Booklet information was compiled by matching Blue pages with the Daily List
2. Item information was compiled by matching receipts to scanner data during the price entry process
3. A single matched file of FAH places was compiled by matching item information with booklet information

Households reported more acquisitions on Blue pages than on the Daily List (overall, 1,133 versus 1,047). Booklet information was compiled by matching blue pages to the daily list based on household ID, place name, date, and total amount. This match was not exact due to variations in spelling of the place name, and missing total amount on the blue page (respondents were instructed to fill total amount only if they did not save a receipt). Two-thirds of matches were made in the first round of matching by exact match on household ID, date and first three letters of the place name.⁴ After completing multiple rounds of matching, we identified 1314 unique FAH events; 66 percent were reported on both the Daily List and Blue page, 12 percent only on the Daily List, and 20 percent only on a Blue page (Table A-18). Compared with Multiple Book households, Single Book households were more likely to report on both the Daily List and Blue page (71 versus 65 percent of acquisitions) but the difference is not statistically significant. There was no difference in reporting compliance by incentive level.

Households were more likely to scan transactions than save receipts (937 scanned transactions and 763 receipts). The match of receipts with scanner data was a completely manual operation. Scanner files received from the field were processed in batches to: (a) drop items scanned in training; (b) identify separate transactions based on scanned delimiters or date/time stamps; and (c) match barcodes with item descriptions from an external UPC database or the list of barcodes printed in the barcode book. The “cleaned” scanner data was loaded in a web-based entry system, shown in Figure 2.

⁴ We implemented a series of sequential matches wherein each record from the blue page file was joined to each record in the daily list file. Matches were identified if there was a unique match on our match criteria; nonmatches or non-unique matches were passed to the next round of matching.

Figure 2. Interface for the Price Entry System

Screen #1: Home page – Instructions and list of cases

Cases with Scanner data

MPRID	# Trx	# Completed
Complete cases		
10000171	4	4 ✓
10000391	2	2 ✓
10000456	1	1 ✓
10000485	5	5 ✓
10000524	1	1 ✓
10000663	5	5 ✓
10000702	5	5 ✓
10000728	1	1 ✓
10000922	4	4 ✓
10001358	1	1 ✓
10001743	2	2 ✓
10001918	8	8 ✓
10002218	3	3 ✓

National Household Food Acquisition and Purchase Survey EXIT

Select an MPRID from list at left, then follow these steps:

1. Review the list of "shopping trips" that have scanned data in the database
2. Click a place to see the list of items scanned by the household
3. For each trip, determine if you have a receipt and
 - o Enter prices and other information, or
 - o Check "no receipt"

Be sure to click Exit after each session

Screen #2: List of transactions (left) after selecting a case

Acquisitions scanned by MPRID = 10010886

Select a place to see items scanned and enter data from receipts

Trx #	Date	Place	# Items scanned
1	2/13/2011	SHOP RITE	61
2	2/17/2011	SHOP RITE	4
3	2/13/2011	BAGLIANI'S FOOD MARKET	2
4	2/18/2011	BAGLIANI'S FOOD MARKET	2

Return to Case List

National Household Food Acquisition and Purchase Survey EXIT

Select an MPRID from list at left, then follow these steps:

1. Review the list of "shopping trips" that have scanned data in the database
2. Click a place to see the list of items scanned by the household
3. For each trip, determine if you have a receipt and
 - o Enter prices and other information, or
 - o Check "no receipt"

Be sure to click Exit after each session

FoodAPS Price Entry Statistics

Screen #3: Entry screen for item details

Acquisitions scanned by MPRID = 10010886

Select a place to see items scanned and enter data from receipts

Trx #	Date	Place	# Items scanned
1	2/13/2011	SHOP RITE	61
2	2/17/2011	SHOP RITE	4
3	2/13/2011	BAGLIANI'S FOOD MARKET	2
4	2/18/2011	BAGLIANI'S FOOD MARKET	2

Return to Case List

Enter/edit notes for case

NOTES:

MPRID = 10010886, TRX # 2, Place = Not specified

Place name:

EBT/Food Stamp \$:

Match items from receipt to the list below.

1. Section 1 - Enter price and weight/quantity (if applicable) for named items
2. Section 2 - Enter item name, price, weight/quantity for foods/beverages that did not match the UPC database. Items are numbered to correspond to the total number of items scanned by respondent, plus 10 extra rows are provided.

Check **DROP** only if item is not on receipt. Check if no receipt
 Check **NonFood** if named item is not a food or beverage. Check if entry is complete Save

1. Enter price, weight, quantity for scanned items with descriptions

Scanned information				Enter from receipt					
Barcode	Category	Brand	Item description	Quant	Non Food	DROP	Price after savings	Store savings	Coupo saving
076183003749	Bottle Tea	Snapple	Lemon Tea		<input type="checkbox"/>	<input type="checkbox"/>	\$9.99	\$2.80	\$

2. Enter item names and other info. Item #s correspond to number of items scanned, plus 10 extra rows are provided.

Barcode	Brand	Item description	Non Food	Price	Price after savings	Store savings	Coupo saving
Item # 2		GRAPES THOM SDL51PC	<input type="checkbox"/>	\$5.19	\$5.19	\$6.38	\$1.45
Item # 3		NEST PURE WATER 24 PC	<input type="checkbox"/>	\$3.77	\$3.77	\$1.22	\$
Item # 4		TYSN FP BNL	<input type="checkbox"/>	\$8.00	\$8.00	\$8.04	\$
Item # 5			<input type="checkbox"/>				

Using the price entry system, data entry staff could select a case, view the list of scanned transactions identified by date and place type (supermarket, convenience store, liquor store, etc. or not specified)⁵; select a transaction and view the items scanned. Items were listed with an item description, if available, or a blank field for filling the item description. Blank lines were also provided in a separate section for items that may have been missed by the respondent. Information entered for each item, if applicable, included: description, price, store savings, coupon amount, weight, quantity, and indicators that the item is a nonfood or should be dropped because it was not on the receipt.

At the completion of the price entry process, 1,146 unique transactions were identified for the 411 households completing a data collection week: 52 percent of transactions had scanner data and a receipt, 16 percent had only scanner data, and 30 percent had only a receipt (Table A-18). Adherence to survey protocols for scanning and saving receipts did not differ by survey protocol or incentive level. Non-SNAP households with income between 100 and 185% of poverty were more likely to provide both scanner data and receipts, compared with SNAP households and non-SNAP households below poverty (56 percent versus 47 and 49 percent).

The match of booklet information with item information used an algorithm similar to the match of blue pages with the daily list. This match required more sequential rounds of matching because of three complications: the place name was missing on the item file if the respondent scanned items but did not save a receipt; the “price total” from the receipt often did not exactly match the “total paid” because we entered prices only for food items; the SNAP total paid, if available, could be used as an alternative to the “total paid.” This match was manageable primarily because the average number of transactions per household is small and most households have one FAH transaction per date. The number of unique transactions identified in booklets or by item information is 1,616: 55 percent have both booklet and item information, 18 percent have only booklet information, and 27 percent have only item information.

The resulting correspondence of information across sources reflects both non-response to survey protocols and errors in the matching processes. The large percentage of transactions with no booklet information causes us to question the validity of the data: were households playing with the scanner? Did they provide receipts that were not their own or outside the data collection week?⁶

At the same time, collection of redundant information, each with a different level of burden, provides a potential indicator of true acquisition behavior. The end result, however, is that we have a significant amount of missing data when a household reports a FAH acquisition on the

⁵ After receipt information was entered into the system, a transaction was identified by the place name entered from the receipt.

⁶ We will verify the date ranges of receipts within the data collection week prior to the TWG meeting.

Daily List but provides no detail on a Blue page (12 percent of booklet entries); provides scanner data but no receipt with prices (16 percent of transaction with either scanner or receipt), or provides only booklet or item information but not both.

Percent of Scanned Items Matched With Prices and Item Descriptions

Table A-19 describes the scanner and receipt information at the transaction and item level. As noted above, 52 percent of transactions had scanner data and receipt; 30 percent had scanner data only; 18 percent had a receipt only. Table A-19 shows the source of price or expenditure data for transactions. While 70 percent of transactions had a receipt, 65 percent have item-level price data because some receipts did not contain item prices or the item descriptions could not be matched with scanned items (for example, the receipt might say “Grocery” or “Misc item.”). Twelve percent of transactions have a “total paid” amount from the booklets and 22 percent of transactions have no expenditure data. The large percentage of transactions with no “total paid” amount is partly due to the fact that the Blue page instructs respondents to fill “total paid” only if they do not have a receipt, and respondents apparently skipped this item even when they did not attach a receipt or did not know the total if they did not have a receipt.

The bottom portion of Table A-19 presents statistics at the item-level, including all items scanned or obtained from receipts: 50 percent of all observed items were scanned and validated by a receipt; 7 percent were scanned and dropped (a receipt was provided for the transaction but the item was not on the receipt); 14 percent of items were scanned but not validated by a receipt; and 29 percent of items were added to the database from receipts. (The last category overstates the underreporting by scanner because respondents had the option of writing items on the blue page if they could not scan the item and we were unable to supplement the scanner data with the “blue page items” prior to price entry.)

The percent of scanned items matched with the UPC database was lower than we expected: 50 percent of scanned items matched the Gladson database; 10 percent matched the barcode book; 28 percent did not match Gladson but we obtained an item description from the receipt; and 11 percent of scanned items have missing item description. An early review indicated that store brands (private labels) did not match the Gladson database. Gladson confirmed that there is wide variation in the percentage of each retailer’s private label captured in the database, and some retailers do not provide the private label data.

Overall, only 7 percent of items observed in the scanner file or on receipts are missing item descriptions (Table A-19). However, item descriptions entered from receipts are abbreviated, not standardized across retailers, and not linked with information such as package size, manufacturer, and nutrients. In addition, entry of item descriptions from receipts is significantly more costly than obtaining those data from a database.

Completeness of FAFH Telephone Reporting

Respondents were asked to call the survey center three times during the data collection week to report information recorded in their booklets (Daily Lists and red pages). These calls provided an opportunity to reconcile information from the Daily Lists and red pages, and to probe when

information was not complete. Telephone interviewers entered reconciled information into our Food Reporting System (FRS). Thus, unlike FAH data, the FRS data did not include multiple records in need of reconciliation.

Our main concern with FAFH reporting was whether or not the telephone calls captured all of the information recorded in booklets. To address this concern, we selected a random 20 percent of households (from among the 411 completes). Reviewers examined booklets and obtained counts of the number of places on the Daily Lists, the number of red “places” (in some instances respondents used a single red page to record multiple places), and the number of red page receipts. We then compared counts from the booklets with counts from the FRS database.

Table A-20 presents the results of this quality control review of booklets. We found a surprising amount of information in booklets was not reported during telephone calls: 30 percent of households had information on the Daily List that was not reported by phone and 28 percent of households had red page information not reported by phone (Table A-20). This non-reporting by phone does not represent a true loss of data because booklet information may be recovered, however, information not reported by telephone may be less complete.⁷

While some booklets contained more information than was reported by phone, the reverse was also true (and often for the same households): 27 percent of households reported Daily List information by phone that was not in their book, and 22 percent of households reported red places by phone that were not in their book. Overall about 60 percent of households had an exact match of booklet information and phone data and 70 percent of households had phone data that was not missing booklet information.

An exact match of booklets and telephone data was significantly less likely for Single Book households than for Multiple Book households (46 versus 73 percent for Daily Lists; 46 versus 79 percent for red pages). However, this is primarily due to the fact that Single Book households reported more information by phone than they recorded in their book. Thus, much of the information from Single Book households was based on recall.

We examined underreporting of booklet information by a number of household characteristics, such as language or household size, but did not find a significant relationship. Underreporting did increase with the total length of telephone calls, which suggest that respondent fatigue played a role.

⁷ This nonreporting by phone partially explains why the count of FAH acquisitions from Daily Lists was less than the count from blue pages.

Item Non-response on Blue and Red Pages

The data items from blue pages (entered from booklets) and red pages (reported by phone) were tabulated to assess item non-response. These data are presented in Tables A-21 and A-22. Rates of item nonresponse for FAFH are for acquisitions reported by phone and do not account for the underreporting of information in booklets.

Ninety percent of households reported FAH acquisitions, with an average of 3.3 acquisitions per household. Rates of item nonresponse are 4 to 7 percent for information about payment type, use of frequent shopper card, and coupon use. Three percent of blue pages did not have a day of week checked off (this hindered out matching process). Seventeen percent of blue pages did not have a box checked for whether the household scanned all, some, or none of their items. Just over half (52.5 percent) of blue pages had food items written on the page, which were not accounted for in our match of scanned items with receipts.

Only 46 percent of FAH acquisitions were from a supercenter or supermarket (Table A-21). This is an important characteristic of the field test sample, as it is significantly below the national average of 64 percent of SNAP purchases at supercenters and supermarkets.⁸ Nearly all supercenter purchases were from two chains: Pathmark and ShopRite. A large number of FAH acquisitions were from markets or grocery stores that are grouped in the “Other” category in Table A-21.

Eighty-nine percent of households reported FAFH acquisitions, with an average of 8.3 acquisitions per household. Item nonresponse was under 2 percent for payment type and purchase amount (Table A-22); there was no missing data for date, and almost no missing data for place name. Less than half of acquisitions were associated with a saved receipt, but this should be examined conditional on whether the acquisition was free and/or obtained from a source that typically does not provide receipts (school or workplace). For 7 percent of acquisitions, respondents did not report whether they purchased food for non-household members; however, these purchases were rare, when reported.

Nearly half of FAFH acquisitions were obtained for free (mostly school meals). Less than 30 percent were obtained from a “top 30” fast food or “top 30” full service restaurant. (Our Food Reporting System was preloaded with menu items from these 60 establishments to standardize food item data upon entry.) The precise location of FAFH establishments was obtained for all acquisitions that were not school meals, workplace meals, or meals in a private home.⁹ After the completion of the field period, the establishment locations were used together with household

⁸ We did not tabulate the distribution of FAH purchase amounts by store type due to missing data on purchase amount. The PSUs were chosen on the basis of this characteristic so that we could test data collection procedures in a diverse retail setting.

⁹ When a precise location was not obtained, the town name was recorded.

addresses to measure driving distances via a Google map API. The distribution of driving distances is shown in Table A-22.¹⁰

4. WEEKLY FOOD ACQUISITIONS AND EXPENDITURES

Table A-23 provides a summary of food acquisitions and expenditures. As noted above, 90 and 89 percent of households reported FAH and FAFH acquisitions, respectively. The 10 percent with no FAH and FAFH acquisitions were mostly mutually exclusive with only five households reporting no acquisitions during the data collection week. Average weekly spending per household was \$38.30 for FAFH and \$93.20 for FAH. The only statistically significant between-group difference is in the percentage of households with no reported FAH acquisition, which is significantly greater for households in the low incentive group. However, among households reporting acquisitions, there were no significant between-group differences in number of acquisitions or dollar amounts. Therefore, we cannot conclude that survey protocol or incentive level significantly affected reporting.

Household acquisitions and expenditures, normalized for household size (per adult male equivalent or AME), are shown in the middle section of Table A-23. The only statistically significant differences in the normalized data are for income groups, with non-SNAP, higher income households having more food acquisitions and expenditures than SNAP and lower-income non-SNAP households.

Total weekly expenditures on FAH and FAFH, per AME, is \$76.30 for higher income households, \$49.60 for SNAP households, and \$43.90 for non-SNAP, lower income households. For comparison, the 2008 Consumer Expenditure Survey (CEX) reported average annual food expenditures of \$6,443 per household (= \$124/week), and average household size of 2.5 persons, which yields an estimate of \$50 per person per week without accounting for male equivalence.¹¹ Thus, it appears that the FoodAPS field test captured more food spending per household, on average, than the CEX.

5. SUMMARY

The FoodAPS field test revealed several challenges for a 7-day collection of food acquisition data. The greatest challenge may not be an underreporting of acquisitions, as feared, but our ability to manage the several redundant sources of information so as to assemble a

¹⁰ Precise locations were obtained during the telephone interviews via an automated Google search for place name within the vicinity of a household's address. We experienced problems with this feature of our Food Reporting System during the first weeks of the field period. Similar distances to FAH locations could not be obtained due to an error in the CAPI programming for the questions about usual "food stores" on HH1.

¹¹ The Consumer Price Index has risen just over 2 percent from the 2008 annual average to March 2011.

complete record for each household. Retooling our processing of FAH data may yield significantly more consistent data across reporting instruments. Adding a formal review of booklets will eliminate underestimation of FAFH.

An important question for the field test was “Would respondents be willing to participate?” We will complete our analysis of nonresponse bias in time for presentation at the TWG meeting.

Other questions were “Could respondents follow the survey protocols?” and “Would participation in the survey change household acquisition behavior?” Table A-17 presents data collected from the self-administered respondent feedback form at the end of the data collection week. Seventy percent of respondents reported that it was easy or very easy to keep track of foods during the data collection week, with no statistically significant differences between groups. Two-thirds of respondents reported that it was easy or very easy to get other household members to participate, with a significant difference between Single Book and Multiple Book households (72 versus 60 percent). Finally, nearly 80 percent of households reported that they did not change the way they got food because they were participating in the survey, however, those in the higher incentive group were less likely to report no change (75 versus 85 percent).