Updated Study Plan

Healthy Incentives Pilot Evaluation

Task Order # AG-3198-D-10-0044

June 27, 2011

Prepared for:
Kelly Kinnison
U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

Prepared by
Abt Associates Inc.

Table of Contents

List of Exhibits

List of Abbreviations

AAPOR	American Association for Public Opinion Research
ACS	Affiliated Computer Services
BAFO	Best and Final Offer
CAI	Computer Assisted Interview
CATI	Computer Assisted Telephone Interview
CBO	Community Based Organization
DTA	Massachusetts Department of Transitional Assistance
EBT	Electronic Benefit Transfer
FNS	Food and Nutrition Service
HIP	Healthy Incentive Pilot
HoH	Head of Household
IAPD	Implementation Advanced Planning Document
IECR	Integrated Electronic Cash Register
MDD	Minimum Detectable Difference
MOU	Memorandum of Understanding
MTFV	Modified Targeted Fruits and Vegetables
NCI	National Cancer Institute
NHANES	National Health and Nutrition Examination Survey
POS	Point of Sale
RFP	Request for Proposals
SNAP	Supplemental Nutrition Assistance Program
SSN	Social Security Number
STARS	Store Tracking and Redemption Subsystem
TFV	Targeted Fruits and Vegetables
TPP	Third Party Processor
TWG	Technical Working Group
USDA	U.S. Department of Agriculture
USDHHS	U.S. Department of Health and Human Services

1. Evaluation Overview

Reducing the prevalence of obesity is a key national health objective. In response to the increase in obesity and chronic disease in the United States, the nation's scientific authorities and policy leaders have emphasized the goal of increasing fruit and vegetable intake (USDHHS, 2010; USDHHS and USDA, 2010). Most U.S. adults fail to meet the fruit and vegetable intake goals of the *Dietary Guidelines for Americans*. Fruit and vegetable intake shortfalls are comparatively large for low-income Americans and participants in the Supplemental Nutrition Assistance Program (SNAP, formerly known as the Food Stamp Program) (Cole & Fox, 2008).

To address this problem, the Food, Conservation, and Energy Act of 2008, also known as the 2008 Farm Bill, authorized funds for pilot projects to evaluate health and nutrition promotion in SNAP to determine if incentives provided to SNAP recipients at the point of sale increase the consumption of fruits, vegetables, or other healthful foods. On the basis of this legislative authority, USDA designed the Healthy Incentive Pilot (HIP).

Under HIP, SNAP participants will be offered an incentive of 30 cents for every dollar of expenditures on targeted fruits and vegetables (TFVs). For every SNAP dollar spent on TFVs, the household will receive an additional 30 cents on its SNAP Electronic Benefit Transfer (EBT) card. The incentive payment may then be spent on any SNAP-eligible foods and beverages. The incentive payment is capped at $60 per household per month, which is a level sufficiently high that not many households are expected to be capped (see Appendix A). TFVs are defined as fresh, canned, frozen, and dried fruits and vegetables without added sugars, fats, oils, or salt. Fruit juices and white potatoes are excluded, but yams and sweet potatoes are included (see Section 7.1.2).

An evaluation of HIP's impact and implementation is being conducted in Hampden County, Massachusetts. Hampden County was selected by FNS through a competitive application process. In this process, State SNAP agencies proposed local sites that met criteria established by FNS. The applications were judged based on the proposed pilot design, staffing, management, and budget. The proposal from the Massachusetts Department of Transitional Assistance (DTA) was selected. Although studying HIP in a single site severely limits the ability to generalize the findings to the national context, the compelling offsetting advantage is that this approach permits the evaluation to have a strong random assignment research design.[1]

The evaluation team includes several partners:

- Abt Associates Inc., the primary evaluation contractor, is involved in evaluation design, all stages of data collection, and analysis.

[1] Due to resource limitations, it is impossible for the HIP evaluation to have both random assignment and national representativeness (for example, by piloting the healthy incentive in a representative national sample of sites). Similarly, due to resource limitations, participating households may earn incentives for 12 months. It is not possible to know if results in the HIP evaluation will differ from what one would observe if participants were assured that the policy change was permanent.

- Westat, Inc., is leading the survey instrument design and the collection of participant survey data.
- The Center for Weight and Health, University of California at Berkeley is providing nutrition expertise.
- Maximus is contributing to instrument design and data collection from stakeholders.

In addition to the evaluation team, a Technical Working Group (TWG) was convened to offer external expertise. The five external TWG members are Tom Baranowski, Children's Nutrition Research Center, Baylor College of Medicine; Simone French, University of Minnesota; Joel Gittelsohn, Bloomberg School of Public Health, Johns Hopkins University; David Just, Cornell University; and Diane Whitmore Schanzenbach, Northwestern University. The TWG's three Federal government experts are: Margaret Andrews, USDA Economic Research Service; Sue Krebs-Smith, National Cancer Institute; and Alanna Moshfegh, USDA Agricultural Research Service. The first meeting of the TWG took place in September, 2010.

This Updated Study Plan is based on the evaluation team's Best and Final Offer (BAFO) proposal, with revisions to reflect discussions at the orientation meeting, meetings with the Massachusetts and Hampden County staff, the first meeting of the Technical Working Group (TWG), and extensive interaction with the Food and Nutrition Service (FNS) through May, 2011. The remainder of Section 1 describes the evaluation objectives and presents an overview of the evaluation design and data collection activities. Section 2 describes the process for randomly assigning SNAP households to the HIP demonstration. Section 3 discusses sampling procedures for the evaluation. In Section 4, we discuss the development and pretesting of survey instruments and Section 5 provides the details of all data collection activities. Section 6 describes the data files that we will create from data collected. Section 7 describes the analyses we will conduct to answer the evaluation objectives and Section 8 describes reporting activities. A schedule of tasks and deliverables is provided in Appendix B. This Updated Study Plan is designed to be useful as a reference throughout the execution of this project.

1.1 Objectives

The five evaluation objectives are:

- Objective 1: Assess the causal impact of HIP on fruit and vegetable consumption by SNAP participants, and on other key measures of dietary intake.
- Objective 2: Identify and assess factors that influence how HIP impacts participants.
- Objective 3: Describe the processes involved in implementing and operating HIP.
- Objective 4: Assess the impact on the HIP grantee (the State SNAP agency), the local SNAP agency, and their team of partners (including retailers, EBT processors, and community organizations).
- Objective 5: Quantify, to the extent possible, the Federal, State, and local administrative and benefit costs of the pilot.

Exhibit 1.1 lists the five objectives, the corresponding data sources, and the reports that will address each objective. The data sources in the second column are introduced briefly in Section 1.2 and then described in detail in Sections 3 through 6 of this document. The reports in the third column are described in Section 8 of this document.

Exhibit 1.1: Objectives, Data Sources, and Reports

Objectives	Data Sources	Reports
1. Assess the causal impact of HIP on fruit and vegetable consumption by SNAP participants, and on other key measures of dietary intake.	Round 1 (baseline) and Round 2 of participant survey, Round 1 focus groups, SNAP caseload data, EBT transactions data	Interim
	Round 1, Round 2, and Round 3 of participant survey, focus groups, SNAP caseload data, EBT transactions data	Final
2. Identify and assess factors that influence how HIP impacts participants.	Round 1 (baseline) and Round 2 of participant survey, Round 1 focus groups, SNAP caseload data, EBT transactions data	Interim
	Round 1, Round 2, and Round 3 of participant survey, focus groups, SNAP caseload data, EBT transactions data	Final
3. Describe the processes involved in implementing and operating HIP.	Retailer survey, on-site observations, key informant interviews	Implementation
		Final
4. Assess the impact on the State SNAP agency, the local SNAP agency, and partners (including experiences and satisfaction of retailers, EBT processors, and community organizations).	Retailer survey and on-site observations, key informant interviews, SNAP caseload data, EBT transactions data	Interim
		Final
5. Quantify, to the extent possible, the Federal, State, and local administrative and benefit costs of the pilot (including cost to implement HIP nationwide).	Retailer survey and on-site observations, key informant interviews, SNAP caseload data, EBT transactions data	Final

1.2 Overview of Design, Data Sources, and Study Plan

SNAP households in Hampden County will be randomly selected to participate in HIP. A random-assignment research design offers the highest level of internal validity, because it provides reassurance that HIP participation is the cause of observed differences in the follow-up period between the HIP and non-HIP groups. Within the HIP and non-HIP groups, participants will be randomly chosen for three rounds of participant surveys. The random assignment of HIP status is discussed in detail in Section 2, and the random sampling for the participant survey is discussed in Section 3.

The main impact measure for the study is the difference in fruit and vegetable intake for HIP and non-HIP participants. The main data source for the impact estimate is the participant survey. The three rounds of the survey are: Round 1 (baseline), Round 2 (after 3 months of participation) and Round 3 (after 11 months of participation). The HIP/non-HIP difference in fruit and vegetable intake will be measured using a 24-hour food recall instrument at Rounds 2 and 3.[2] Combining data from Rounds 2 and 3, we expect to be able to detect an effect smaller than a quarter cup of fruits and vegetables, the design specification set by FNS (see Section 3.1 for details).

In addition to the participant survey, several other data sources will provide valuable information about HIP participants, other stakeholders, the implementation process, and the costs of HIP. These data sources are listed in Exhibit 1.2, along with summaries of the research samples for which each data source will be collected. Further details on sampling and data collection appear in the report sections indicated in the final column of the exhibit. Thus, for each data source in turn, Section 3 of this document describes sampling, Section 4 describes survey instruments and interview protocols, and Section 5 describes data collection. Section 6 describes how all the data will be processed and compiled into analysis files. Section 7 explains the analysis plan for addressing each objective, and Section 8 describes how results will be reported.

[2] As explained in Section 7.1, the main confirmatory outcome will be the HIP/non-HIP difference in modified targeted fruits and vegetables (MTFVs), where the modification to the definition of the food category allows intake to be measured with a 24-hour recall instrument.

Exhibit 1.2: Data Sources

Data Sources	Sample	Document Sections *
1. Participant survey: • Round 1 (baseline) • Round 2 (after 3 months of participation) • Round 3 (after 11 months of participation)	Respondents will be randomly sampled from a sampling frame of individuals in SNAP participant households in the HIP and non-HIP groups. If multiple adults from one household are sampled, one adult will be randomly chosen to represent the household. The initial sample size in Round 1 (2,535 participants) will provide at least 750 HIP and 750 non-HIP completed interviews in Rounds 2 and 3.	3.1, 4.1, 5.1
2. Participant focus groups: • Round 2 (3 groups) • Round 3 (3 groups)	For each of six focus groups, a convenience sample of about 10 HIP participants will be recruited using SNAP administrative records.	3.2, 4.2, 5.2
3. Retailer survey and on-site observations	From SNAP authorized retailers, a random sample will be selected to complete written surveys at Round 1 and Round 3. In each round, the sample will have 60 retail stores that participate in HIP and 15 that do not participate. On-site visits will be made to 10 retailers in each round.	3.3, 4.3, 5.3
4. Key informant interviews	There will be interviews with at least 4 State SNAP staff, 3 local SNAP managers, 18 other local SNAP staff or community organizations, and 5 EBT staff at the State Agency, the EBT vendor, or third party processors.	3.4, 4.4, 5.4
5. SNAP/EBT data	SNAP administrative records and EBT transactions data will be acquired for the full HIP group (7,500 households) and non-HIP group (approximately 45,500 households).	3.5, 5.5

* Section 3: sampling; Section 4: instruments; Section 5: data collection

.

2. Random Assignment to HIP and non-HIP Status

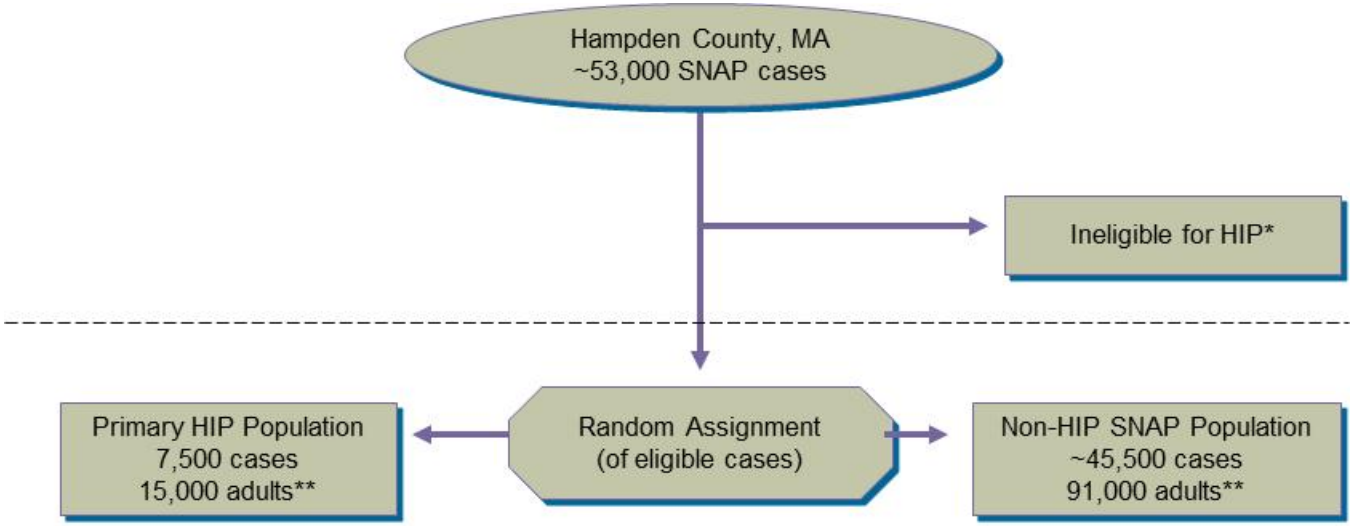
The random assignment of SNAP participant households to HIP and non-HIP status is central to the evaluation design and HIP operations. Using administrative files of SNAP participants provided by DTA, we will randomly select 7,500 SNAP households to participate in HIP. DTA will then notify these households, provide specially marked EBT card sleeves to identify them as HIP participants, and provide training on HIP procedures. The balance of SNAP households in Hampden County (about 45,500) will not receive HIP. As shown in Exhibit 2.1, random assignment creates two groups—one of adults in HIP-designated households and another of adults in non-HIP households—that are balanced with respect to key participant characteristics.

2.1 Eligibility

As shown in Exhibit 2.1, implementing our design requires that we divide Hampden County's SNAP population into three groups. First, we need to exclude households that are ineligible for HIP; we expect very few households to be in this group (see definition of ineligible households below). Then, we will divide the households that are eligible for HIP randomly into a HIP (treatment) and non-HIP (control) group.

The evaluation attempts to estimate the impact of HIP on adult food intake. Households without an adult are therefore not eligible for HIP. Furthermore, only households that do their own shopping are eligible for the evaluation. SNAP participants who sign over their benefits to a residential or treatment facility are not eligible for the study. On the other hand, homeless participants who retain the use of their own benefits are eligible for the study.

Exhibit 2.1: Random Assignment



*Child only cases; households that sign over benefits to treatment facility
**Based on average household size of 2 in DTA case file data.

2.2 Household and Individual Classifications

DTA will generate and send the Hampden County SNAP case extract file of HIP-eligible households to Abt in July 2011. All households in that file will be considered to be in the HIP universe for the duration of the pilot. Each household will receive a HIP flag identifying it as one of the following four groups:

1) HIP non-survey (group H)
2) HIP survey (group I)
3) Non-HIP survey (group J)
4) Non-HIP non-survey (group K)

Following random assignment at the household level, these flags will be applied to all members of the original households in the HIP universe.

As household composition and program participation change over time, the following rules determined by DTA will have implications both for HIP operations and for survey sampling (see Section 3.1). The SNAP case is tied to the Head of Household (HoH); therefore the HIP flag and the HIP incentives will also be tied to the HoH. If the original HoH leaves the SNAP household, by DTA rule that SNAP case closes. Other household members may form a new case, but that new case will not get the HIP flag and thus will not earn HIP incentives even if their prior flag were H or I. Similarly, if a member of a HIP household other than the original HoH leaves the household, that person will not be given a HIP flag and will not be eligible to earn HIP incentives; but the household with the original HoH will retain the HIP flag and HIP incentives if the HIP flag were H or I.

The SNAP case may also close without any changes in household composition. Regardless of how a SNAP case closes, if the SNAP case reopens with the original HoH, the household will once again receive the HIP flag and resume earning HIP incentives if the HIP flag were H or I.

2.3 Random Assignment Schedule and Procedure

2.3.1 Testing the Random Assignment Algorithm

We expect that the DTA will provide a test extract of the SNAP case file for Hampden County by June 1, 2011. We will then verify the file layout and fields, and construct and test the random assignment algorithm.

To test the random assignment algorithm, we will proceed as follows. We will procure the DTA case file for 2 months prior to scheduled randomization. Using this file, we will randomly assign cases to experimental or control status. We will verify that the random assignment algorithm is working properly by comparing observable characteristics of the experimental and control groups. Characteristics to be compared are household size, race/ethnicity, age, receipt of other cash assistance, gender, benefit as a percentage of the maximum, and employment status. We will conduct both characteristic-by-characteristic tests of equality and a joint test (i.e., an "F-test"). Proper randomization implies the strong null hypothesis of no differences between the HIP and non-HIP groups that cannot be explained by chance. We will repeat this process multiple times (tentatively 100

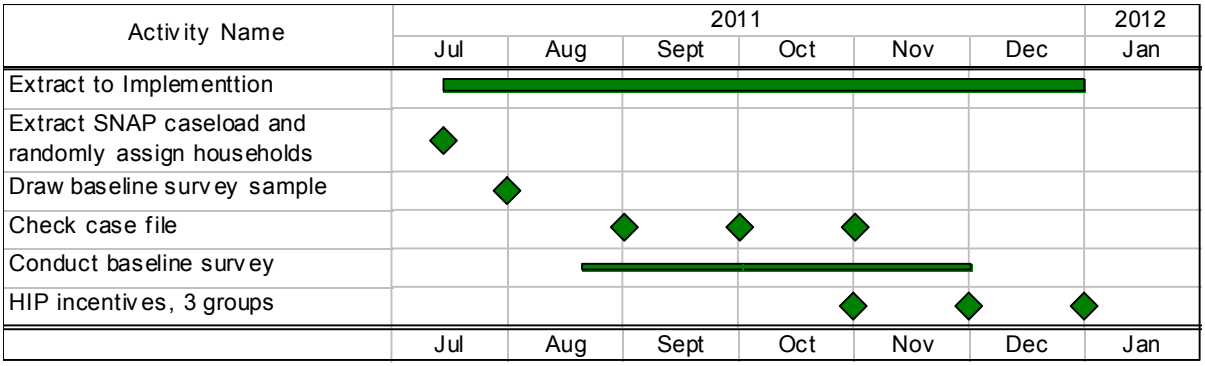
times) during testing to verify that it works well. We will report the results of testing to FNS by July 26, 2011.

2.3.2 Implementation of Random Assignment

Once FNS has approved the random assignment procedure and we have received a new case file suitable for main sample random assignment, we will execute the procedure and notify DTA of the households selected to be in HIP. This information also will be used for sampling for the different data collection efforts (see Section 3). As a result of the random assignment, the two groups will have similar characteristics except for random variation. To improve the precision of the estimates, we will use a blocked random assignment design, i.e., we will create "blocks" with the same characteristics and then do random assignment within the blocks.[3] The result is exact treatment/control balance on the blocked characteristics. Specifically, the blocking characteristics we are considering will be geography (exact geography still in discussion with DTA), household size (1, 2, 3, 4+ household members), race/ethnicity, presence of elderly and children in the household, and length of stay on SNAP. We chose these characteristics because they are available in the administrative data files and are likely to be correlated with food intake behaviors. We will make final decisions about appropriate cells after preliminary data analysis of the test file.

We will draw the sample one time, in July 2011. Exhibit 2.2 shows the 5½ -month process from extraction of the sample to completion of HIP implementation, when all cases assigned to receive HIP are able to begin earning incentives on purchases of target fruits and vegetables. The 7,500 households sampled to receive HIP benefits will begin receiving the incentive in three rounds—at the beginning of November, December, and January. DTA will recruit and train households to participate in the pilot on this same schedule. Similarly, survey cases (both HIP and non-HIP) will be extracted and sampled in August 2011, and the baseline (Round 1) survey will be fielded for 3 months from September through November (see Section 3.1 for details on survey sampling).

Exhibit 2.2: 1-Month Sample Draw



Drawing the sample one time has several advantages over a multi-month approach. First, with a one-time sample draw we send only one file of randomly selected HIP households back to the State. Thus, this approach requires fewer file extracts and transfers, and it gives the State more control over the

[3] The BAFO proposal discussed sorting cases by a list of characteristics and then randomly sampling households. The procedure for blocking, described in this paragraph, is a refinement that achieves the same purpose.

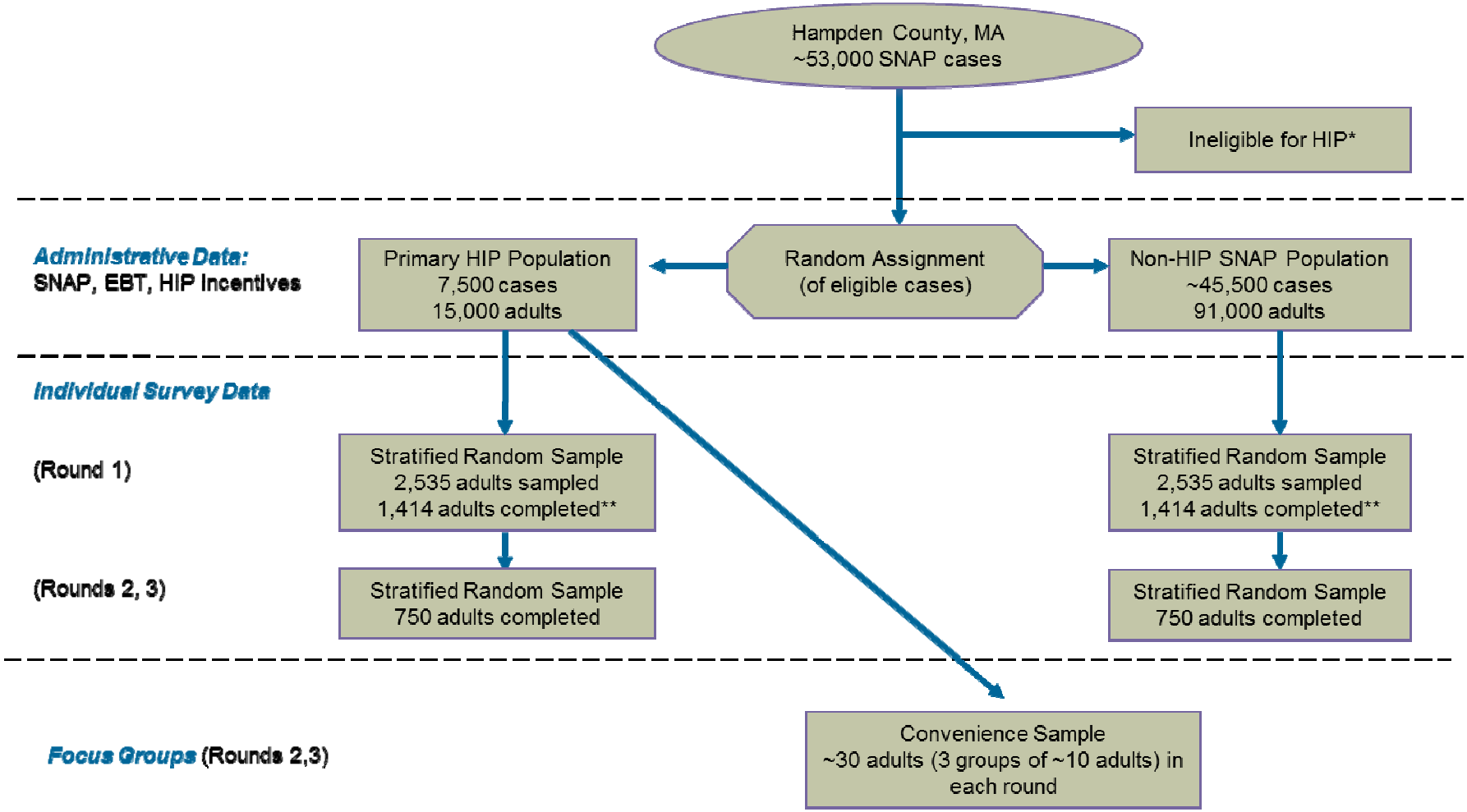
timing and process of HIP implementation. Second, the one-time draw also gives us more flexibility in fielding the baseline survey to ensure the highest possible response rates. And, third, with fewer file transfers, there is less chance of error or delay in the pilot or the evaluation.

3. Sampling Procedures

This section discusses sampling procedures for the different data sources to be used in the evaluation. The largest subsection addresses the participant survey, including sampling procedures for all three rounds, determination of eligibility for the survey, expected response rates, and power calculations (Section 3.1). The subsections that follow describe sampling for the participant focus groups (Section 3.2), retailer survey and on-site observations (Section 3.3), key informant interviews (Section 3.4), and SNAP/EBT administrative data (Section 3.5).

Exhibit 3.1 presents an overview of the participant sampling process and shows how three of these data sources are related to the randomly assigned HIP and non-HIP participant groups described in Section 2: (a) the SNAP/EBT administrative data will be available for the full HIP and non-HIP SNAP populations, (b) the participant survey will have a sample size sufficient to generate 750 completed interviews from the HIP group and 750 completed interviews from the non-HIP group in Rounds 2 and 3, and (c) focus groups will provide greater narrative detail about the experiences of a smaller number of HIP participants.

Exhibit 3.1: Random Assignment, Random Sampling, and Participant Data Sources



*Child only cases; households that sign over benefits to treatment facility
**See Exhibit 3.2 for details about attrition rates and response rates.

3.1 Participant Survey

To estimate the impact of the HIP incentive, we will survey random samples of SNAP HIP participants and SNAP non-HIP participants. Specifically, using a computer-assisted interview (CAI), we will survey at three points in time: baseline prior to HIP implementation/participation (Round 1), 3 months into HIP implementation/participation (Round 2), and 11 months into HIP implementation/participation (Round 3).

3.1.1 Sample Sizes

This subsection summarizes the expected sample sizes for the three survey rounds. These sample sizes were chosen so that the estimated number of completed surveys would be 750 in the HIP sample and 750 in the non-HIP sample at Rounds 2 and 3. Having equal sample sizes maximizes the statistical power of HIP/non-HIP comparisons. The Minimum Detectable Difference (MDD) for this sample size is discussed below in Section 3.1.6.

Exhibit 3.2 shows the number of individuals who will be sampled at each round, in order to achieve the desired completed surveys at Rounds 2 and 3. The number of completed surveys depends on:

- The number of households initially sampled in each round
- The assumed exit rate percentage from SNAP during the survey round
- The assumed response rate percentage among those who were sampled
- The assumed exit rate percentage between one survey round and the next

The exit rate assumptions are explained later in this subsection, and the response rate assumptions are explained in Section 3.1.5.

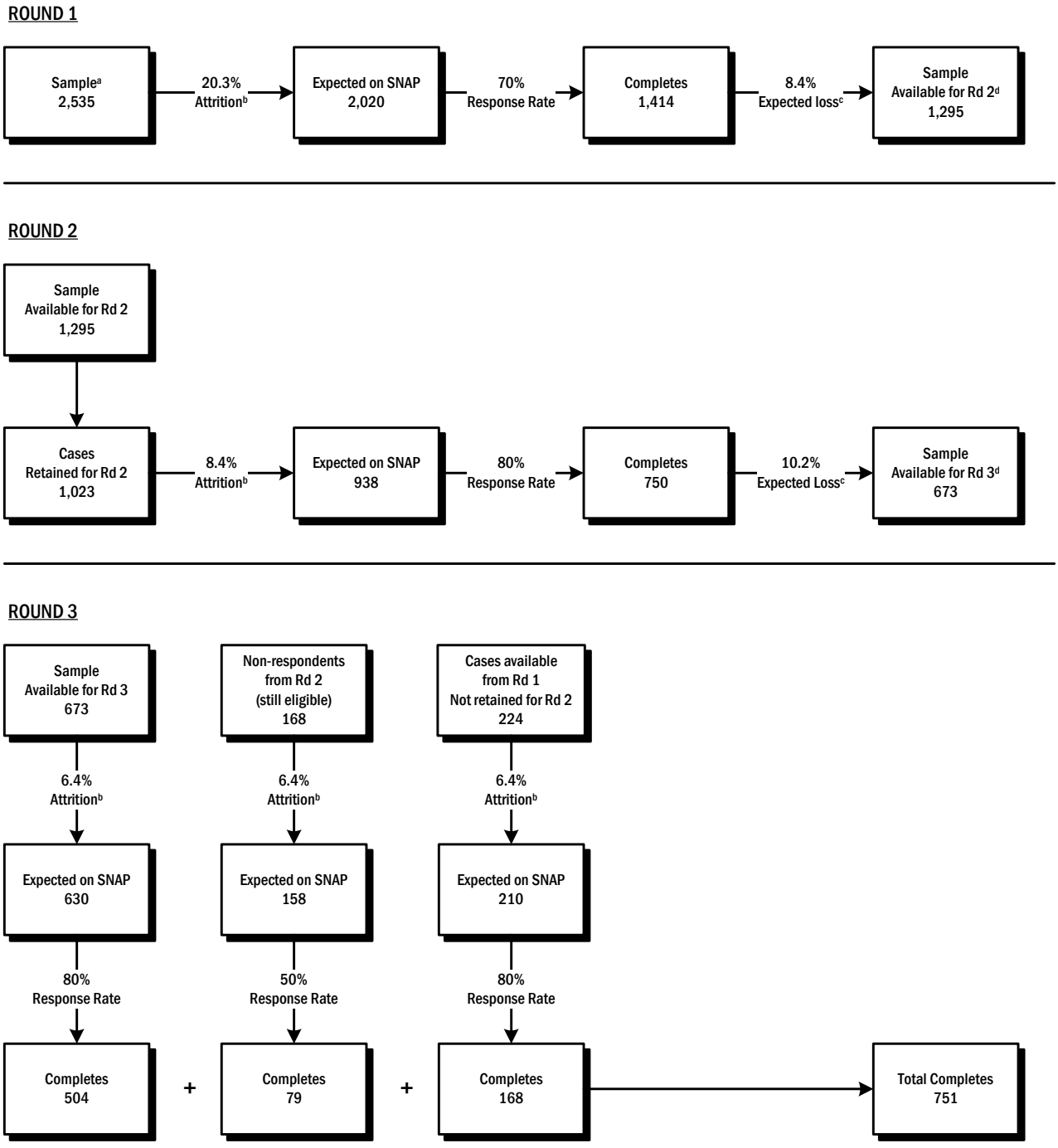
Based on these assumptions, Exhibit 3.2 shows that 2,535 participants in each group will be initially sampled for Round 1. After attrition and non-response, 1,295 in each group are estimated to be available for sampling in Round 2. Of these, 1,023 in each group will be sampled for Round 2 (enough to generate 750 completed surveys). The Round 3 sample is slightly more complicated. As Exhibit 3.2 shows, the initial Round 3 sample will come partly from Round 2 respondents and partly from Round 1 respondents who did not complete Round 2 surveys. Together, these respondents suffice to provide 750 HIP and 750 non-HIP interviews in Round 3. We tolerated this additional complexity in the Round 3 sample design, in order to avoid unnecessary costs through a larger sample size at Round 1.

We decided to interview sample cases only if they are on SNAP at the time of the interview. This decision was based on two factors. First, we judged that any differential exit from SNAP between HIP and non-HIP cases is likely to be small (confirmed by our Technical Working Group). Second, the primary impact of SNAP will be through a price effect (the rebate implicitly lowers the price of fruits and vegetables) and an income effect (any rebate earned, even in the absence of a behavioral response). Neither the price effect, nor the income effect, will be operative for those off SNAP. It is possible that short-term participation in SNAP will have effects on fruit and vegetable intake even after individuals leave SNAP, but any such impacts seem likely to be quite small.

The expected attrition rates are based on published national SNAP exit rates (Cody et al., 2007). Massachusetts statewide exit rates provided by the Massachusetts DTA are broadly similar to the national rates.[4] Our analysis of exit rates in Hampden County did not lead us to change the desired sample sizes. According to analysis done by DTA, exit rates in Hampden County are less than national SNAP exit rates and also less than Massachusetts statewide exit rates. This suggests that the initial sample sizes included in the BAFO proposal will allow us to achieve the desired number of completed interviews in Rounds 2 and 3.

[4] For example, after 12 months the cumulative exit rate for Massachusetts is 38.5 percent compared to 40.0 percent nationally. In general, the cumulative exit rates for Massachusetts are somewhat lower than the corresponding national rates up to month 13, and then begin to increase at a somewhat higher rate than the national rates after month 14.

Exhibit 3.2: Expected sample sizes of each survey group (HIP and non-HIP)



[a] For Round 1, the number of participants in sample is the number of persons to be sampled from SNAP administrative files. By the time data collection for Round 1 starts (approximately 1 month later), an estimated 5.5 percent of the original sample will no longer be in SNAP. Any sampled participants who are determined to be no longer in SNAP will not be fielded in Round 1.

[b] SNAP exit rates derived from hazard rates in Table II.14 of Cody et al., (2007).

[c] At the beginning of the data collection round, we assume that everyone is in SNAP. The exit rates in this column reflect the expected cumulative losses by the end of the data collection period.

[d] Losses determined by matching survey respondents against SNAP administrative files just prior to fielding next round.

3.1.2 Sample Frame and Stratification for Round 1 of the Participant Survey

The samples for the survey (to be equally distributed between the HIP and non-HIP groups) will be selected from SNAP case record files for Hampden County. Only currently eligible cases at the time the Round 1 sample is selected will be included in the sampling frame. The names, addresses, and telephone numbers of all eligible participants will be required for survey administration. In addition to contact information, demographic data, household-level characteristics, and SNAP certification status (regular issuance and recertification dates) may be used for sample stratification purposes (as discussed below).[5]

The general approach for identifying participants for the participant survey will be to select stratified samples from the two sampling frames (SNAP HIP and SNAP non-HIP) created by random assignment. The purpose of stratification under the proposed design is to ensure that the two samples are balanced with respect to demographic and other characteristics, yielding more precise impact estimates. We plan to stratify or sort samples by the same case-level and person-level characteristics used during the random assignment of cases to HIP.[6] Since the number of cells obtained by a complete cross-classification of all of the specified characteristics will be large relative to the sample size, many cells will result in very few or no sampled cases. Thus, we plan to define explicit strata based on a subset of the characteristics (e.g., geography, household size, and race/ethnicity) to ensure reasonable stratum sample sizes. Within the primary strata, cases will be sorted by the remaining characteristics. The sorting induces an implicit stratification of the sample when random systematic sampling is used to select cases.

Because we are sampling individuals instead of households, depending on the individual SNAP participants listed in the frame, we could occasionally sample more than one person in the same household. However, the likelihood of this occurring is minimal because of the way we intend to prepare the sampling frame and select the sample. First, we will sort individuals in the sampling frame by household so those in the same household will be listed together. Then we will systematically select persons in the frame (for example every 5th or 10th person depending on the size of the list) which will generally prevent multiple individuals in the same household from being chosen. If, despite this systematic approach, we select more than one individual in a household, we will randomly pick one adult household member to represent the household.

3.1.3 Sampling for Rounds 2 and 3 of the Participant Survey

In subsequent rounds, eligibility for the participant survey corresponds to the rules for continued eligibility in the pilot itself (Section 2.1 describes the eligibility rules for the pilot). In addition to SNAP case closures, changes in household composition will also affect the analysis sample. Section 2.2 described how a HIP flag will identify households as one of four groups:

---

[5] Participants with missing values of the stratification variables will not be excluded from the sampling frame. Instead, if a stratification variable has a large number of missing observations, then "response missing" will be included as an additional category.

[6] Random assignment stratification variables will include some or all: geography, household size, benefit as a percentage of the maximum benefit, and race/ethnicity, presence of elderly and children in the household, and length of stay on SNAP.

- HIP non-survey (group H)
- HIP survey (group I)
- Non-HIP survey (group J)
- Non-HIP non-survey (group K)

This HIP flag is attached to a household as long as its head is the original head of household (from the time of the baseline draw). When the household composition changes, sampled adults (and other household members) may move from one of the four HIP groups to another HIP group or become part of a household without a HIP flag.

Most adults sampled in Round 1 will **remain in a household of the same type** at the time of Rounds 2 and 3. This situation happens if:

- A HIP-sampled adult (from group I) remains in the same household
- A HIP-sampled adult (from group I) moves to a different group I household
- A non-HIP sampled adult (from group J) remains in the same household, or
- A non-HIP sampled adult (from group J) moves to a different group J household

A smaller number of adults sampled in Round 1 will **move to a household of a different type** at the time of Rounds 2 and 3. This situation happens if:

- A HIP-sampled adult (from group I) moves to a household in group H, J, K, or not in the HIP universe, or
- A non-HIP sampled adult (from group J) moves to a household in group H, I, K, or not in the HIP universe

In light of this, we have determined the need to refine the definition of the analysis sample. Sampled adults from Round 1 will remain eligible for Rounds 2 and 3 if they are receiving SNAP benefits at the time of the survey or received benefits in the month prior to the survey and remain in a household of the same type. In the less common situation where sampled adults from Round 1 move to a household of a different type at the time of Rounds 2 or 3, these sampled adults will be dropped from the sample and the analysis in the later rounds.[7]

Since some large retailers will implement EBT changes in areas outside Hampden County (and outside Massachusetts), it will be possible for some households to receive the incentive even if they no longer reside in the county. Thus, households that leave Hampden County but remain on SNAP in Massachusetts will continue to be surveyed. However, households leaving the State will not be surveyed as we will not have any information about where they are living and whether they are still receiving SNAP.

[7] We considered but rejected the alternative of dropping all sampled adults who were not in the original household, which would have reduced the sample yet further.

Estimating usual intake

Dietary intake estimates based on a single day of recall data do not accurately represent long-term average intake for that individual, also known as "usual intake," as there exists substantial within-person variation in consumption patterns from day to day. The mean of multiple days of intake for an individual is a better measure of usual intake than a single day; however, it is often not practical to collect more than one day of intake on the entire sample, without either dramatically increasing the cost of the study and/or reducing the sample size and power to detect differences in intake. Therefore, researchers have developed statistical methods for estimating usual intake of foods and nutrients for samples in which only a subset of respondents report a second day of recall data. Based on guidance provided by recent research, we will collect a second day of intake data for a 10 percent subsample, or 75 individuals in the treatment group and 75 in the control group.[8] This sample size will be sufficient to estimate usual intake for the treatment and control groups overall. However, a larger sample size would be required to estimate usual intake for any smaller subgroups or to estimate the usual intake of certain food groups (such as dark green vegetables) that are episodically consumed. The 10 percent sample who complete a secondary 24-hour will be selected at random, so their ratio of between-person to within-person variation may be used in the estimation of usual intake.

3.1.4 Determining Survey Eligibility

Because we only want to survey individuals who are eligible for the analysis sample (as defined above) at the time of the interview, we will use case file extracts to screen SNAP participants who were randomly assigned to the HIP and non-HIP groups. These data will be obtained starting in July 2011 (for random assignment) and thereafter on a monthly basis.[9] Prior to each month of the survey, we will verify that each sampled participant (a) is a member of the same type of household as when sampled (HIP or non-HIP), and (b) received benefits in the month represented by the case file extract. We will need to obtain case file extract data in time to process the data and update the sample file prior to the first of the month. Depending on the timing of when case file extracts become available, we may request that extract files are produced once all regular monthly SNAP benefits have been issued (i.e., around the 16th of the month).

3.1.5 Expected Response Rates

This subsection explains the expected response rates in Exhibit 3.2. Weighting to account for non-response is discussed in Section 3.1.6.

We have assumed response rates of 70 percent for any Round 1 interview, 80 percent for Round 2 and 75 percent on average (ranging between 50 and 80 percent depending on disposition from previous

---

[8] The 2002 New Zealand National Children's Nutrition Survey successfully used a 10 percent replicate subsample to adjust for usual intake (Ministry of Health, 2003). According to current Institute of Medicine (IOM) dietary assessment guidance, the *number* of replicate observations is more important in estimating usual intake than the *proportion* of replicate observations relative to the full sample. Nusser et al. (1996) recommend that replicate data be collected on not fewer than about 50 or 60 subjects. IOM guidance notes that replicate subsamples consisting of fewer than 70 to 80 individuals have been successfully used in the past to obtain usual intake estimates (IOM, 2000).

[9] Monthly case extract data are also needed for analysis of EBT transaction data, which will be collected for every month of the demonstration, as discussed in Section 4.6.

rounds) for Round 3. Under these assumptions, a subsample of 1,023 selected from all who completed Round 1 interviews must be retained for the first intake interview in order to obtain the 750 required at Round 2 in each group. After sorting the full sample of completes by the blocking variables, this subsample will be selected systematically from the sorted list using a random start.

All of the completed cases at Round 2 will then be re-fielded in Round 3. To compensate for attrition and non-response losses among these cases, an additional 168 cases that were selected for Round 2 but did not complete the interview, and 224 cases that completed the Round 1 interview but were not fielded in Round 2, will be selected for Round 3. Thus, an expected 504 of the 750 Round 3 interviews per group will also have a completed Round 2 interview. The expected sample sizes are based on assumptions about attrition and non-response rates shown in Exhibit 3.2. Finally, we note that the design summarized in Exhibit 3.2 indicates that a portion of the sample in Round 3 will come from still-eligible participants who did not complete the Round 2 interview (but did complete Round 1).[10] The absence of these individuals from the Round 2 data set will not bias estimates based on those data, since the omitted cases will be chosen using a random process.

Using a combination of telephone interviewing and in-person tracking (as much as 50 percent of the sample in Round 1), we estimate that the Round 1 interview will achieve a response rate of 70 percent. This projection is based on the NCI Food Attitudes and Behavior Survey (FAB) (Li et al., 2007), where more than 25 percent of the sample was African-American and 55 percent had income below $50,000 and more than 13 percent below $17,500. The survey included three rounds of telephone data collection to conduct a 24-hour dietary recall; participants were paid $5 at each round for their time. Twenty-five percent of the sample had bad phone numbers; of the remaining, 64 percent were recruited into the study. Given that we plan to use in-person tracking to follow-up bad phone numbers and non-responding households, and will reimburse participants $20 for their time, we believe 70 percent is a reasonable response rate for Round 1. A lower response rate for the comparison group could be an issue if comparison group members are aware of being excluded from a desirable benefit. That should not be an issue for this pilot where fewer than 25 percent of SNAP households are being included in the pilot project to receive the HIP incentive in Hampden County and where the control group will continue to receive their normal SNAP benefits. At the time of the Round 1 interview neither group will be aware of their assignment so response propensity should be the same for the two groups.

Once respondents have participated in the Round 1 interview, it will be much easier to get them to commit to the follow-up. Reasons for non-response include being unable to locate the respondent and respondent refusals due to perceived burden of the instrument. Using procedures to minimize non-response, we expect to interview at least 80 percent of the sample at Round 2 and 75 percent at Round 3. As with Round 1, these estimates are based on the FAB data where 92 percent of the sample completed the second interview and about 82 percent completed the third interview. We believe that we will obtain a similar response rate of 75 percent in Round 3 with both those who completed the Round 2 interview and those who were not selected for the Round 2 interview. Respondents who completed the Round 2 interview will have demonstrated their willingness to be interviewed multiple

[10] This design for the Round 3 sample is unchanged from the BAFO proposal. Larger samples for Round 1 and Round 2 surveys would have been needed if we had sought to have 750 Round 3 respondents with completed Round 1 and Round 2 interviews.

times. However, those who were skipped for Round 2 will be less likely to experience study fatigue. The larger incentive ($40) will provide a similar motivator for both groups. We assume that we will achieve about the same response rates for the evaluation experimental and control groups at Rounds 2 and 3.

We will release the sample for the two groups on a rolling basis to ensure that we have at least 750 completed evaluation experimental and control group interviews at each of Rounds 2 and 3. During Round 3 interviewing, preference will be given to those respondents who participated in Round 2. Due to sample attrition from SNAP ineligibility, we anticipate that there will be an insufficient number of Round 2 households who can be interviewed at Round 3, thereby requiring the inclusion of some members of the Round 1 sample who were not interviewed at Round 2.

We will use American Association for Public Opinion Research (AAPOR) Response Rate 4 to calculate the response rates for each round. The AAPOR Response Rate 4 includes both full and partial completes in the numerator (partial completes are instruments where critical items are completed, though the respondent may have broken off before the very end). The AAPOR Response Rate 4 denominator includes full and partial completes, refusals, non-contacts, and an estimate of the proportion of cases of unknown eligibility that are eligible. We anticipate that the number of cases of unknown eligibility will be fairly small, due to the use of monthly case extracts to verify both SNAP receipt and household status. Thus, we will assume conservatively that all unknown eligibility cases are eligible. The impact on response rates will be small given the small number of cases involved. We will exclude ineligible households from the calculation of response rates.

3.1.6 Expected Levels of Precision

The evaluation was designed to satisfy two quantitative targets. The evaluation should be able to detect a HIP/non-HIP difference in Modified Targeted Fruits and Vegetables (MTFV)[11] intake that is (a) less than or equal to the target difference in the FNS Statement of Work (0.250 cups of fruit and vegetables per day), and (b) less than or equal to the target difference we estimate would be most probable based on economic theory.

Our theoretical estimate of the difference that is most probable is necessarily approximate. The premise of the HIP evaluation is that a more authoritative estimate of the HIP/non-HIP difference requires empirical research with a strong random-assignment research design. Our advance estimate is based on the size of the incentive and the relevant economic theory. The economic theory is summarized briefly in Appendix C. Estimates from the 1999–2002 National Health and Nutrition Examination Survey (NHANES) suggest that among near poor households (less than 130 percent of poverty line), average fruit and vegetable purchase is 2.0 cups.[12] Mean estimates of the elasticity of fruit and vegetable consumption with respect to price are 0.70 and 0.58, respectively (Andreyeva, Long, & Brownell, 2010). Together, these figures suggest an impact of HIP on MTFV of about 0.384

[11] MTFV is identical to TFV except that it does not incorporate the restriction against added sugars, fats, oils, and salt. We make this modification because the 24-hour recall instrument cannot always identify whether such ingredients were included in a purchased product or added later as part of a recipe.

[12] Computed from Dong and Lin (2009), Table 1. 2.0 cups = 1.43 Total Vegetables + 0.96 Total Fruits – 0.39 White Potatoes. This estimate is close, but not ideal. MTFV also excludes fruit juices, consumption away from home, and some fruits and vegetables consumed in combination with other foods.

cups.[13] This impact is larger than the FNS target of 0.250 cups. Hence, a design whose minimum detectable difference (MDD) suffices to achieve the target in the FNS Statement of Work also suffices to detect the differences we considered most probable.

The estimated MDD for our design is stated in terms of MTFVs, which is an empirically measurable food grouping defined formally in Section 7.1.2. First, we estimate the MDD using a single round of the survey (Round 2 or Round 3). Second, we estimate the MDD using Round 2 and Round 3 combined. Third, we estimate the MDD for subgroup analysis.

First, we estimate that our design has an MDD of 0.164 cups of MTFVs at either follow-up interview (i.e., Round 2 or Round 3). This power calculation assumes, in a one-sided test:

- 1,500 completed interviews (750 in the HIP Group and 750 in the non-HIP Group) at Round 2 and Round 3
- An R-squared of 12 percent in a regression equation using data from the Fruit and Vegetable Screener to predict MTFV intake (a conservative value based on Li et al., 2007)
- A design effect (deff) of 1.05 (for non-response)
- The variance parameter for TFV from Mabli, Gleason, and Hall (2009; FNS Statement of Work, Appendix B)
- Conventional power parameters $1-\alpha$=80 percent; β=5 percent

A one-sided test is appropriate, because we would treat a negative impact as equivalent to no impact.

Second, using average estimates for Round 2 and Round 3 combined, the sample size for each estimate is twice as large. As explained in Section 7.1.5, our analysis strategy treats the average impact across Rounds 2 and 3 as the main confirmatory outcome. With the larger sample size, the MDD is 0.116 cups (computed from 0.164 cups/the square root of 2).

Third, pooling the data across the two waves and using the 0.164 cups estimate for the MDD for that analysis, we can estimate MDDs for subgroups. MDDs in a subgroup of half the population will be approximately 0.164 cups (0.116 x the square root of 2). For a subgroup of a quarter of the population, the MDD would be about 0.232 cups (0.116 x the square root of 4). MDDs to test the difference in impact across two subgroups, each with half the sample, will be about 0.232 cups (0.116 x 2).

---

[13] An elasticity gives the percentage change in purchases with a percentage change in price. This estimate is computed given the elasticity and the estimate of baseline fruit and vegetable consumption; i.e., 0.64 elasticity x 2.0 cups x 30% price cut ~ 0.384 (where 0.64 is the simple average of the mean elasticity estimates for fruits and vegetables). The actual impact of HIP may be higher or lower than suggested by this forecast based on existing elasticity estimates. The impact of HIP could be lower than this estimate if participants are less responsive to price signals, or if they do not fully understand the incentive program, or if the incentive cap is binding for more participants than expected. The impact of HIP could be higher than this estimate if the fruit and vegetable discount shifts participants' tastes and preferences toward fruits and vegetables in addition to changing the budget constraint, just as the past SNAP research has found that program benefits have a larger effect on food spending than one would expect from the change to the budget constraint alone (Fraker, 1990). The uncertainty of forecasts based on historical estimated elasticities provides a strong motivation for this new study, which uses a strong random-assignment research design to evaluate HIP.

3.2 Focus Groups

Participant focus groups will provide additional qualitative information and narrative detail, for use in interpreting the participant impacts and assessing the experiences and satisfaction of SNAP recipients as HIP stakeholders. Six focus groups will take place at about the same time as Round 2 and 3 of the participant survey data collection (three focus groups in each round), using a stratified convenience sample of HIP participants.[14] All 7,500 HIP participants will be classified by demographic strata of race and age so that the diversity of HIP participants will be included in the focus groups. In each stratum the sampling frame will be randomly sorted into a list by a statistician. We will recruit from the list for each stratum by sequentially contacting each person on the list and inviting them to participate until we recruit 14 people. HIP participants who are in the survey sample will be excluded from the focus groups to minimize respondent burden and to avoid biasing responses on the Round 2 and Round 3 surveys due to the focus group discussions.

Focus groups will include a minimum of 6 individuals and a maximum of 10 or 11. We will recruit 14 individuals for each of the six focus group sessions to account for no-shows. If more than 11 participants come to a focus group session, the first 11 to arrive will be asked to participate and the others will receive their incentive and be released. Focus groups with more than 11 participants do not allow enough time for everyone to participate fully and may leave participants feeling frustrated that their views were not heard.

Because the focus groups use a convenience sample, they will not be statistically representative of the entire HIP population. The offsetting advantage is that focus groups provide greater narrative detail, illustrating viewpoints and suggesting hypotheses about the HIP population (see Section 7.2 for a discussion of the analytic uses of the focus group data).

3.3 Retailer Survey and On-Site Observations

Surveys and on-site observations of food retailers offer insight into the experiences and satisfaction of an important HIP stakeholder group, and they provide useful information about the pilot's implementation process and costs. We will conduct a surveys of SNAP retailers participating in HIP in Round 1 (pre-implementation) and in Round 3 (late implementation). Retailers will be given two opportunities to join HIP. If they choose not to participate initially, they will be able to participate at a later date. The sample of participating retailers will be selected from both the first and second "go-live" dates. A sample of retailers that declined to participate in HIP at the first opportunity will be surveyed in Round 1, and a sample of retailers that dropped out after participating will be surveyed in Round 3.

As discussed in Section 4.3, all surveys will be conducted by mail with telephone follow-up.

[14] The focus groups select HIP participants, rather than HIP and non-HIP participants. With a small and fixed number of focus groups, the high value of qualitative information about the new incentive program overrode the offsetting value of having information from non-HIP participants.

The samples for the SNAP retailer survey will be drawn from two discrete populations of retailers: those that choose to participate in HIP (participating retailers) and those that are eligible to participate but do not (non-participating retailers). Hampden County has approximately 455 SNAP retailers eligible to participate in HIP. The recruiting process is in progress, so the final number of participating retailers is unknown.

For both participating and non-participating retailers, we will stratify stores participating in HIP by store type. FNS has official store types, which we will combine into superstores, supermarkets, other (small/medium) grocery stores, other stores (including convenience stores), and farmers markets. Exhibit 3.3 provides the distribution of all SNAP retailers in Hampden County and the expected sample size by store type for participating and non-participating retailers. DTA currently has commitments to participate in the pilot from 74 retail locations, including major national chains, local chains, independent grocers, convenience stores, and farmers markets.

For the purposes of this sampling plan, we project that 150 retailers will participate. As a survey of all participating retailers would be excessively burdensome, we will select a sample of 75 participating retailers (one-half of all participating retailers). This sample will be proportionately allocated across strata, and retailers will be randomly selected within strata.

We expect that much of the variance in responses from stores with the same owner (superstores, supermarkets, and convenience stores) will be at the chain level (though there will be variance within chains on some outcomes that are related to differences among stores within the chain, such as location). For the strata that include chain stores, each chain will constitute a stratum. Assuming that the number of stores to be sampled exceeds the number of chains, we will select at least one store from each chain. If the desired sample is large enough to require more than one store per chain, we will allocate the sample proportionately across chains within their strata. For example, if there are three supermarket chains and we need to select 10 supermarkets, we will select between one and seven stores per chain, based on the number of supermarkets owned by each chain. This strategy will yield approximately equal weights for all stores, while assuring that all chains are represented.[15]

With an expected response rate of 80 percent, this sample will yield 60 completed surveys, with at least 10 per stratum (except for farmers markets, where we will survey all participating markets). Assuming the population and sample sizes in Exhibit 3.3, we estimate that, for a sample proportion of 50 percent, the standard error will be approximately 5 percent and the half-width of the 90 percent confidence interval will be 8.4 percent (i.e., from 41.6 percent to 58.4 percent).[16] While the sample sizes within strata will probably be too small to produce usable stratum-level estimates or detect differences between strata, they will be sufficient for qualitative analysis.

---

[15] We note that the original sampling plan called for sampling up to 15 chains and one store per chain, as specified in the RFP. This strategy was intended to reduce burden on corporate officials at chains; it also might be more efficient if there is no variation among stores within chains. However, the original strategy would result in larger weights for chain stores and less precise estimates if there is variation among stores within chains. Therefore, we have modified our approach as described above. Once we have more definitive data on the population of likely participating stores, we will reassess these plans.

[16] Details of calculations available if desired by FNS.

Data to construct the sampling frame will come from two sources: FNS will provide Abt the Main Retailer Files, and DTA (in conjunction with ACS) will provide Abt with Supplemental Retailer Files that will have HIP participation information and local contact information. FNS and DTA will provide the retailer files monthly; files from August 2011 and August 2012 will be used to draw the retailer samples. At each round, we will randomly draw a representative sample of 75 retailers participating in HIP. Given the size of the samples relative to the universe, there will be some overlap between the samples at the two rounds. However, we will not attempt a panel design, because we expect that there will be sufficient turnover among smaller stores, and we would therefore need a much larger baseline sample to assure the targeted sample at follow-up. Instead, we have designed the sample to be representative of the participating retailer population at each wave. In addition, this approach allows us to include retailers that do not participate at the start of the demonstration but later choose to participate ("drop-ins").

In addition to the sample of participating retailers, we will select three other samples: 19 retailers who choose not to participate in HIP, 19 retailers who initially choose to participate but drop out prior to the end of HIP, and 19 retailers who initially choose not to participate but later enter the HIP demonstration (assuming there are 19 retailers in each of the three groups). As in the participating retailer survey, the unit of sampling will be the firm, so we will select one store to represent each chain. Due to the small number in each sample, approximately equal numbers will be selected from each stratum. The goal is to obtain 15 completed surveys of non-participating retailers at Round 1, 15 completed surveys of dropouts at Round 3, and 15 surveys of the late-entering retailers ("drop-ins") at Round 3. Additional retailers will be selected from each population as replacements (where possible) in case the response rate is lower than expected. If one of the survey populations is smaller than 19 retailers, we will attempt to survey all of the retailers in that population.

Exhibit 3.3: Hampden County SNAP Retailers: Population and Expected Sample Sizes

Store type	Population	Pct. Of population	Expected participants	Pct. of partic. retailers	Survey partic. sample*	Non-part. total	Pct. of non-partic. retailers	Survey non-partic. sample**
Superstore	35	8%	22	15%	11	13	4%	5
Supermarket	19	4%	19	13%	10	0	-	0
Other grocery	88	19%	44	29%	21	44	14%	5
Convenience/ other store	306	67%	61	41%	29	245	80%	6
Farmer's market	7	2%	4	3%	4	3	1%	3
TOTAL	455	100%	150	100%	75	305	100%	19

* Participating retailer sample allocated proportionately by strata, except sample will include 4 farmer's markets.

** Non-participating retailer sample allocated approximately equally by strata.

In addition to the retailer survey, Abt staff will conduct in-person, on-site observations in a purposively selected subsample of 10 participating stores that are also selected for the survey. We will conduct observations at three points in time, corresponding to the rounds of key informant interviews (discussed below in Section 3.4). In selecting the Round 1 observation subsample from the Round 1 survey sample, the selection criteria will include store type, payment system, and location. Stores in the Round 1 observation sample will be selected for Round 2 observations if they still

participating in HIP. For Round 3 observations, the observations will again be conducted in a subsample of the survey sample. The stores from the Round 2 observation sample that are in the Round 3 survey sample will be selected again for observations; otherwise additional stores from the same strata of the survey sample will be selected as replacements for the observation sample. Back-up stores will be selected within each store type in case a store in the initial sample does not cooperate. Abt staff will observe retailer checkout procedures and collect price data for a market basket of commonly purchased TFVs, including apples, bananas, carrots, and tomatoes. As discussed in Section 7, the data collected will address research questions on implementation processes and impacts of HIP on retailers.

3.4 Key Informant Interviews

Insight from key informant interviews will be used in several analyses: project implementation, the cost analysis, and the feasibility of expansion. The interviews will be conducted in person to gain insight into the implementation of HIP, impacts on DTA and partners, and administrative and benefit costs of the pilot. The final list of informants is still under development, and will be based on consultation with DTA and FNS.

There will be three rounds of in-person interviews with these stakeholders.

- Round 1: approximately 2 months before the implementation of HIP. These interviews will focus on the design and development of the EBT system modifications for HIP, retailer recruitment, and preparations for notifying and training participants. These interviews will coincide with the Round 1 participant interviews.
- Round 2: approximately 4–5 months after HIP implementation, focusing on early operations and coinciding with Round 2 participant interviews.
- Round 3: approximately 11–15 months after HIP implementation, focusing on mature operations and the feasibility of expanding HIP, and coinciding with Round 3 participant interviews.

Based on the FNS Statement of Work, the original plan presented in our BAFO provided for each round to include 4 interviews with State SNAP staff, 3 interviews with local SNAP managers, 18 interviews with other local SNAP staff or community-based organizations (CBOs), and 5 interviews with EBT staff at the State Agency, the EBT vendor, or third party processors. After attending the Orientation Meeting with FNS and DTA, the Abt team expanded the group of stakeholders from the original plan to include additional groups: retailers that are undertaking system development to implement an integrated HIP solution in-lane (referred to as integrated retailers), and third party processors (TPPs) that serve as transaction processors and/or acquirers for retailers that process transactions through the credit/debit network.

Exhibit 3.4 provides a list of interview groups and subgroups and the expected method of conducting the interview. This list reflects our current understanding of the staff at DTA and partner organizations who will be involved in HIP. Some uncertainty remains—for example, what CBOs will be involved in HIP and what their roles will be (retailer outreach, participant outreach and training, etc.). Thus, we expect this list to change prior to Round 1 of the interviews.

Exhibit 3.4: Groups and Subgroups for Key Informant Interviews

	Interview Group	Interview Subgroup	Location of Interview
1	DTA Executives and other Massachusetts Executives Involved in HIP	DTA Executives and other Massachusetts Executives Involved in HIP	On-site in Boston
2	DTA Staff Involved in HIP	DTA Management	On-site in Boston
3		HIP Senior Management	On-site in Boston
4		HIP Management Team	On-site in Boston
5		HIP Financial Management	On-site in Boston
6		HIP Management Team	On-site in Hampden County
7	Executive Office of Health and Human Services (EOHHS) Technical Staff	EOHHS Technical Staff	On-site in Boston
8		Novo Dia Group (NDG; consultant to DTA)	On-site in Austin, TX or via phone interview
9	Local DTA Staff	Holyoke DTA Office	On-site at Holyoke Office
10			On-site at Holyoke Office
11			On-site at Holyoke Office
12		Springfield State DTA Office	On-site at Springfield State Office
13			On-site at Springfield State Office
14			On-site at Springfield State Office
15		Springfield Liberty DTA Office	On-site at Springfield Liberty Office
16			On-site at Springfield Liberty Office
17			On-site at Springfield Liberty Office
18	EBT Vendor (ACS)	Management, Systems, Administrative Staff	On-site at ACS location in Austin TX, additional interviews by phone
19		Management and Retailer Staff	Hampden County, location TBD
20	Retailers With Systems Modified to Accept HIP	National Chain	Phone interview
21		Regional Chain	Phone interview
22		Additional retailers, as applicable	Phone interview
24	Third Party Processors	First Data	Phone interview
25		Fifth Third	Phone interview
26		FIS (EBT Gateway)	Phone interview
27		RBS (if applicable)	Phone interview
28	Community-Based Organizations (CBOs)	HIP Steering Committee	Hampden County, location TBD
29		SNAP-Ed implementing agency	Hampden County, location TBD
30		Western Massachusetts Food Bank	Hampden County, location TBD
31		Health Center TBD	Hampden County, location TBD
32		Senior Center TBD	Hampden County, location TBD
33		Legal Services Advisor TBD	Hampden County, location TBD
34		Farmers' Market Sponsors TBD	Hampden County, location TBD

Interviews will be arranged through the DTA HIP Project Manager or Project Team. Due to the number of subgroups and participating executives and staff, we will confer with the DTA Management Team to ensure we are only conducting interviews with stakeholder participants with specific involvement in HIP and that we identify the best respondent for each topic so as to avoid redundancy and minimize respondent burden. In most groups, there will be a single person who is the best respondent for each topic, based on assignments of responsibility. Where there are multiple staff who would be equally good respondents (e.g., line staff and their supervisors in local DTA offices), we will interview the same respondents in each round (to the extent possible), in order to assure continuity and capture changes over time.

3.5 Massachusetts SNAP Data and EBT Transactions Data

The acquisition of SNAP and EBT administrative data is described in this subsection.

3.5.1 SNAP Data

DTA maintains information on SNAP households and participants (including demographics, income, contact information, and benefits) in its BEACON eligibility system. This system will be used for the initial random assignment of eligible households to HIP and for several additional evaluation purposes:

- Survey sampling frame, stratifiers, sorting variables, and demographic analysis variables for the participant survey
- Updates to contact information and demographic data for participants sampled for Rounds 2 and 3 of the participant survey
- Characteristics of HIP and non-HIP households to be matched with monthly EBT transaction data for analysis

Important demographic variables to be obtained from the SNAP eligibility system include age of the head of household, size of household, relationships of SNAP household members, number of adults in the household, number and ages of children in the household, presence of an elderly member, employment status and earnings, and presence of cash income from other sources.

The SNAP data will be obtained for all HIP-eligible households (approximately 53,000) in Hampden County receiving SNAP benefits in July 2011. As noted in Section 2, the number of participants in the SNAP data is much greater than in the number of survey respondents. The SNAP data will be obtained in April 2011 (test extract) and monthly from July 2011 through February 2013.

3.5.2 EBT Transactions Data

The EBT vendor for Massachusetts collects and maintains data pertaining to the SNAP EBT transactions. This data will be used for the following analyses:

- Take-up rate of HIP incentive by HIP participants
- Difference in redemption patterns between HIP and non-HIP households

- Impact of HIP on retailers' SNAP redemptions

We will obtain daily EBT transaction data for the full pool of HIP (7,500 households) and non-HIP participants (about 45,500 households). Similar to the SNAP data, the number of participants in the EBT administrative data is much larger than in the participant survey. These data will cover the period beginning 2 full months prior to the HIP implementation (September 2011) and ending 2 months after the Round 3 survey (February 2013). The EBT vendor will provide the EBT transaction data to FNS daily and FNS will transfer the data without SSNs to Abt on a daily basis using the FNS MOVEIT transfer facility. Daily transmission of data will allow any issues in the EBT transaction data to be identified in time for the EBT vendor to address them.

4. Develop and Pretest Instruments

This section discusses the development of instruments. The discussion is organized by data source: the participant survey (Section 4.1), participant focus groups (Section 4.2), the retailer survey and on-site observations (Section 4.3), the key informant interviews (Section 4.4), and cost data reports and forms (Section 4.5).

4.1 Participant Survey

The participant survey is designed to capture relevant and high quality data, using existing items to the extent possible, while minimizing respondent burden. The random assignment of SNAP participants to HIP and non-HIP groups, and the random sampling design for the participant survey sample from each group, are described in Section 3.1. The participant survey includes a questionnaire section for the sampled participant and a section for the primary shopper. In most cases these will be the same person. The instrument will be administered in English or Spanish. The final participant survey reflects input from the TWG, FNS, and cognitive interviews conducted with SNAP participants.

Exhibit 4.1 presents the focal domains in each round of data collection for both the sampled participant and the primary shopper sections of the survey.[17] As shown, the sampled participant portion of the Round 1 instrument includes a domain on respondent characteristics; questions on attitudes, perceptions, and barriers to consuming fruits and vegetables; and a fruit and vegetable screener. The primary shopper portion of the Round 1 instrument includes domains on household characteristics, nutrition assistance programs, family food environment, general shopping patterns, and food expenditure. Rounds 2 and 3 will collect data on all the same domains as Round 1, except for respondent characteristics and participation in nutrition assistance programs. Of note, Rounds 2 and 3 will collect information on household characteristics to identify any changes in household composition and employment status. Rounds 2 and 3 also include domains on exposure to nutrition education and participation in HIP. Finally, Rounds 2 and 3 include a 24-hour dietary recall (24HR) on the sampled participant using USDA's Automated Multiple Pass Method (AMPM).

[17] The FNS Statement of Work called for a measure of household food security. After subsequent discussions with FNS and the TWG, the decision was made to omit the food security module, as this topic is not directly related to the study objectives and would contribute minimally to addressing the research questions.

Exhibit 4.1: Domains in Each Round of the Participant Survey's Sampled Participant and Shopper Sections

Domains	Round 1	Round 2	Round 3
Sampled Participant Section			
Respondent Characteristics	√	-	-
Exposure to Nutrition Education	√	√	√
Attitudes, Perceptions, and Barriers to consuming Fruits and Vegetables	√	√	√
Barriers to consuming fruits and vegetables	√	√	√
Fruit, Vegetable Screener	√	√	√
24-hour dietary recall		√	√
Primary Shopper Section			
Household Characteristics	√	√	√
Participation in Nutrition Assistance Programs	√	-	-
Family Food Environment	√	√	√
Participation in HIP	-	√	√
General Shopping Patterns	√	√	√
Food Expenditure	√	√	√

4.1.1 Sampled Participant Section

Respondent Characteristics

These data will be collected in Round 1, and include age, gender, marital status, language spoken at home, race/ethnicity, education and employment status of the respondent. As noted below, in Section 4.1.2, characteristics of the respondent's household will be collected in all three rounds. We reviewed the SNAP case data file in Massachusetts to identify respondent variables that may be available and complete from this source. We concluded it would be best to collect this information directly from respondents. (We will review administrative data when they are available and assess whether to incorporate this information.)

Exposure to Nutrition Education

The instrument asks sampled participants two questions to assess their exposure to nutrition education programs and nutrition messages promoting fruit and vegetable intake. These data will be collected in all three rounds.

Attitudes, Perceptions, and Barriers to Consuming Fruits and Vegetables

The instrument asks sampled participants a series of questions to assess their preferences and beliefs as well as barriers to consuming fruits and vegetables. These data will be used to assess if food preferences and beliefs differ between the HIP and non-HIP samples in Rounds 2 and 3, due to participation in HIP. The TWG members stressed the importance of capturing data on these domains separately for fruits and vegetables, and where appropriate, items have been split to capture preferences, beliefs, and barriers separately for fruits and vegetables.

Fruit and Vegetable Screener (FVS)
The FVS provides data on intake of fruits and vegetables over a longer timeframe (i.e., one month) than the 24-hour recall. This information will be collected at all three rounds. For analysis of the impact of HIP on MTFV intake, information from the Round 1 FVS will be included as regressors to increase the precision of outcome estimates in multivariate analyses comparing the HIP and non-HIP samples. The original 16-item FVS has been modified, with input from Dr. Fran Thompson of the National Cancer Institute (NCI), who designed the original instrument. The NCI will provide a scoring algorithm for the revised instrument.

Twenty-Four Hour Dietary Recall (24HR)
As planned in the BAFO proposal, these data will be collected using the AMPM from sampled participants in Rounds 2 and 3. In addition, a second dietary recall will be collected in Rounds 2 and 3, from a 10 percent subsample, to obtain estimates of usual dietary intake. The TWG recommended that these data also be collected at Round 1, but this was not possible due to budget and respondent burden constraints.

4.1.2. Primary Shopper Section

Household Characteristics
These data will be collected in all three rounds. The questions in this domain include household size and composition, employment status of the shopper, and number of adults as well as youth who are unemployed, employed full-time and part-time.

Participation in Nutrition Assistance Programs
These data will be collected only in Round 1, per input from the TWG.

Family Food Environment
The instrument asks the primary shopper about availability at home of fruits, vegetables, salty snacks, soda and low-fat milk. These data will be collected in all three rounds.

Participation in HIP
This module will only be administered to the HIP sample. These data will be collected in Rounds 2 and 3. These items will provide data on primary shopper understanding of the HIP program, tracking HIP receipts and incentives, and barriers or facilitators to earning the HIP incentive. Additionally, the primary shopper will be asked if they have changed their shopping behavior and increased their purchase of fruits and vegetables due to HIP.

General Shopping Patterns
These data will be collected in all three rounds. The instrument asks the primary shopper to report the frequency of shopping, location of usual shopping, and if they make special trips to purchase fruits and vegetables. Primary shoppers will be asked if limited transportation or distance to grocery store was a barrier to grocery shopping.

Food Expenditure

The instrument asks primary shoppers to report their total food and non-food expenditure at stores, eating out, and fruit and vegetable expenses in the past week. The expenditure questions were adapted from the first three items on the Bureau of Labor Statistics Consumer Expenditure Survey (CES). These items pertain to monthly expenses for food items; non-food items; and meals or snacks from restaurants, fast food places, cafeterias, and carryout.

The evaluation team and FNS considered, but rejected, several alternative approaches to measuring food expenditure:

- An option of collecting a complete expenditure diary from sampled participants, in addition to the 24-hour food recall, was rejected because it would pose an excessive burden on respondents.
- A complete expenditure diary from a different sample of HIP and non-HIP SNAP participants was included as an option in the evaluation team's BAFO proposal, but it entailed significant expense and the option was not exercised.
- An option of using an expenditure diary in place of the 24-hour food recall instrument would have been feasible and affordable, but it would have left unanswered the central question of how HIP affected actual food intake.

These alternatives for measuring food spending were discussed in the first TWG meeting, but the discussion indicated no way of avoiding the challenging tradeoffs. In conclusion, the evaluation team and FNS decided that it was essential to have some food spending questions, adapted from an existing survey source, but we recognize the limitations of the chosen short set of broad spending questions.

4.1.3. Pretesting the Participant Instrument

We conducted two rounds of cognitive testing on the participant survey questions. Cognitive testing focused most heavily on survey questions that were new or revised in this participant survey.

We translated the selected questions into Spanish and prepared an interview guide for each questionnaire. The English and Spanish versions of the respondent and shopper sections went through two rounds of cognitive testing. For Round 1, the respondent and shopper sections were combined into one interview guide, and a Spanish version was created. For Round 2, the respondent and shopper sections were separated and tested in both English and Spanish. The first round of testing was conducted in late January 2011, and the results were compiled and reviewed. Two major issues were identified—respondents had different concepts of "household" and they did not understand the food expenditure questions. These issues were discussed among the study team and the questions were revised based on the group recommendations. The second round of testing was conducted in early February 2011. The objectives of the second round were to re-test the revised questions and to assess the administration time for the respondent and the shopper sections.

To recruit participants for cognitive testing, we identified five organizations in the greater Washington area that serve individuals who receive Supplemental Nutrition Assistance Program (SNAP) benefits. These organizations were: Manna Food Center; Montgomery County Health and

Human Services in Maryland; Casa De Maryland at Shady Grove and Wheaton in Maryland; Neighborhood Consejo in Washington, DC; and Saint Martin's Church in Gaithersburg, MD. These organizations distributed flyers in Spanish and English to individuals who visited their office or food pantry locations. All participants were recruited from either the Manna Food Center or Montgomery County Health and Human Services. These individuals called the Westat phone number listed on the flyer to schedule an interview. To supplement this recruitment, we used an in-house list of 80 Spanish-speaking individuals who had participated in previous Westat studies. For Round 1, a total of 9 English-speaking and 14 Spanish-speaking respondents were recruited for the first round of cognitive testing.

Data were collected using a semi-structured interview protocol to gather insights into the cognitive sources of potential misinterpretation (Willis, 2005; Groves et al., 2004; Sirkin et al., 1999). This methodology focuses on examining the question-answer process using a cognitive psychological perspective. According to Tourangeau et al. (2000), survey response unfolds through four stages: comprehension, retrieval, judgment, and responding. Using this information about respondents' thought processes, survey researchers can identify question wordings that are either misunderstood or understood differently by different respondents; recognize instructions that are insufficient, overlooked, misinterpreted, or difficult to understand; ascertain vague definitions or ambiguous instructions that may be interpreted differently; determine items that ask for information that the respondent does not have access to; and distinguish confusing response options or response formats. The cognitive interview protocol consisted of a series of predetermined general probes, such as "What, in your own words, is this question asking?" and "What, if anything, did you find confusing or unclear about this question?" as well as more tailored probes that focused on assessing comprehension of specific words, phrases, or concepts of concern.

There are multiple approaches to cognitive interviewing, including "think aloud," concurrent, and retrospective probing. Multiple methods are often applied within the same interview (Beatty and Willis, 2007). We variously applied these cognitive testing methods to the respondent and shopper sections to determine how well potential participants understood certain words, phrases, and concepts used in the survey items, as well as to assess any difficulties they had answering the questions. We also tested the Spanish version of the survey for clarity of translation. This allowed an assessment of the thought processes respondents use to understand the survey instructions, interpret questionnaire items, and select responses. For the English and Spanish survey testing, we used concurrent probing, whereby the interviewer asked the cognitive probes immediately after each item or set of items that had been selected for testing.

The cognitive interviews were scheduled for an hour. The subset of questions was varied across all respondents to ensure that all of the selected survey questions were adequately tested In the first round we tested more than 50 questions and responses; in the second round we re-tested the questions that were revised and monitored the timing of the interviews at each HIP study round (Rounds 1, 2, and 3). In Round 1, 10 cognitive tests were conducted in person (5 English and 5 Spanish) and 8 were conducted by phone (4 English and 4 Spanish). In Round 2, 4 tests were in person and 5 were by phone. Respondents who completed the testing in-person received $50, and respondents who completed the testing by phone received $35.

4.2 Participant Focus Group Protocol

Two rounds of focus groups will be used to gather information from HIP participants regarding their experiences with the pilot. The same guide is being used for both rounds of focus groups. The first round of focus groups will collect information on participants' early impressions of HIP after 3-4 months of HIP participation. The second round will gather information on participants' perceptions of HIP once they have 9-11 months of experience with HIP. The guide contains the instructions and questions for the moderator to use during the focus groups along with specific probes to elaborate upon participants' statements. The guide includes the following topics:

- How and when participants learned about HIP
- How much and what type of information/training participants received regarding HIP
- Participants' levels of understanding about how the incentive can be used
- Changes in food spending habits and changes in food eating habits
- Level of ease associated with use of HIP
- Level of understanding regarding what foods in what forms earn the incentive, how much of the incentive money participants have earned and spent
- How HIP did not meet, met, or exceeded the participants' expectations
- Which stores' participants use their HIP EBT card and the level of ease associated with HIP participation at these stores. This will include the attitudes of the stores' staff and availability of HIP-eligible products.
- Overall thoughts about HIP

4.3 Retailer Survey and On-Site Observations

We will conduct a mail survey of retailers to solicit information from different retailer types (e.g., chain retailers and independent retailers) and different employees within a retailer (store owner/managers and clerks/clerk managers). In addition to the survey, we will develop a form for on-site observations in a subsample of participating HIP retailers. The survey will be conducted in Rounds 1 and 3, while the observations will be done in all three rounds of the evaluation.

4.3.1 Retailer Survey and Observation Instruments

In the survey of retailers participating in HIP, the first section of the questionnaire will address store characteristics, recruitment to HIP, implementation and operation of HIP, and fruit and vegetable inventory. The second section of the survey will address check-out procedures. The third section of the survey will address training for HIP. For large stores, managers will be asked to complete the first section, and to select the employee most knowledgeable about check-out procedures and training for the second and third sections. For smaller stores, all sections can be completed by the manager.

We will use a different survey instrument for stores that do not participate in HIP. (See Section 3.3 for information on sample design.) For those stores, only the store manager or owner will need to complete the survey.

On-site observations will give us first-hand information about the retailers and validate the survey responses. The on-site observation form will include a protocol for collecting data in three domains: transactions, inventory observations, and other environmental factors. We will observe simulated HIP transactions for a specified bundle of goods. We will use a different module for stores with integrated electronic cash registers (IECRs) which automatically identify HIP-eligible items and for stores which do not have IECR and where items must be separated manually (non-IECR). Observers will also collect retailer feedback about or issues with HIP. The transaction observations will inform the description of the transaction process and may give rise to questions not anticipated in developing the survey.

Independently of store personnel, evaluation staff will also complete a module that will provide information on fruit and vegetable inventory and other environmental factors. The evaluation staff member will observe the location of fruits and vegetables (fresh, canned, frozen, dried), shelf labeling, other signage, quality of fruits and vegetables, informational handouts, and other factors that may help or hinder participants' use of the HIP incentive. Information on available fruits and vegetables will be collected both in the observation and the survey, as a check on the accuracy of the responses in the retailer survey. Data collected in the first round of site visits will help refine the retailer follow-up survey.

A list of data elements for the survey and observation form is provided in Exhibit 4.2. The list indicates the respondent and the items to be covered in each round of data collection.

The instruments submitted for OMB approval reflect feedback from FNS, DTA and the pretest of the instruments. (The pretest is described in the next section.) The retailer survey questionnaire and advance letter will be translated into Spanish, using the Spanish dialect that is predominant in Hampden County. The preferred language for stores where observations will be held will be determined at the time when the visit is arranged. Bilingual store observers will be sent to stores where Spanish is the preferred language.

The retailer survey is used in several analyses: a description of retailer impacts and satisfaction (Section 7.2.3), as part of the implementation analysis (Section 7.3), and as one source of implementation and operation costs in the cost analysis (Section 7.4).

Exhibit 4.2: HIP Retailer Instrument Topics

Domain	Respondent[a]	Participating Retailer Survey		Store Observations	
		Rd 1	Rd 3	Rd 1	Rd 2&3
Training screener	M	√	-		
Store characteristics[b,c]	M	√	√		
Recruitment[b,c,d]	M	√	-		
Preparing for HIP					
Problems[c]	M	√	-		
Costs to update checkout lanes	M	√	-		
Training	C	√	√		
How HIP has affected your store					
Problems[c]	M	-	√		√
Stocking of fruits and vegetables	M	-	√		
Costs, sales and profits[c]	M	-	√		
Observe simulated HIP transaction	C				√
Questions from CUSTOMERS	C	-	√		
Fruit and vegetable inventory	M/O	√	√	√	√
Store environment	O			√	√

[a] M= store manager/owner; C= checkout supervisor; O=observer (from evaluation team).
[b] Items to be included in the Round 1 survey of non-participating retailers.
[c] Items to be included in the Round 3 survey of retailers who have terminated HIP participation ("drop-outs").
[d] Items to be included in the Round 3 survey of retailers who began participation after the start of HIP operations ("drop-ins").

4.3.2 Pretesting of Retailer Survey and Store Observation Form

With the assistance of DTA and FNS, Abt recruited nine stores to pretest items from the retailer survey that do not refer specifically to HIP. These stores included three supermarket chains, two convenience store chains, and four small/medium grocery stores. Abt sent the pretest version of the retailer survey to the nine stores, with a cover letter explaining the pretest, letters from DTA and FNS explaining HIP and the importance of the survey, a form for comments on the survey, and a prepaid FedEx form and envelope to return the survey. Store managers/owners provided the start and end times for each section, circled confusing language or formatting, and provided feedback on the comments form after completion of the survey. Telephone debriefings were conducted with the retailers to determine any unclear or difficult questions, missing questions, or other recommended changes.

The observation form was pretested on site in three stores—one supermarket, one chain and one independent store. For survey and store observation questions referring to HIP, and thus not included in the pretest, we reviewed the content and wording with DTA, retailer corporate contacts, and retailer association representatives. Instruments were revised to reflect the comments from the pretest and other reviews.

4.4 Key Informant Interviews

The Abt team prepared the interview instruments for the specified stakeholder groups, incorporating comments from FNS. Abt will prepare for the interviews by reviewing applicable documents, such as the Massachusetts DTA Grant Application, project status and progress reports, and internal communications.

In addition, members of the Abt evaluation team have attended two key technical meetings in preparation for HIP implementation: the orientation meeting for DTA and ACS (conducted in November 2010 at FNS) and the application design sessions (conducted by DTA and ACS in December 2010 at DTA offices). These meetings provided opportunities to gather information about implementation plans and processes, and also to discuss evaluation requirements. Finally, evaluation team members will participate in regular status calls among FNS, DTA, and ACS to learn about progress and issues in the planning and implementation of HIP, and to coordinate evaluation plans with implementation plans.

The detailed interview guides provide questions that lead informants through the various aspects of their involvement with HIP. These questions are organized by general topics. Exhibit 4.3 provides the list of topics. Each " √" indicates that a topic will be covered by at least one interview in the group of key informants identified by the column heading. The draft protocols for the key informant interviews are modular, so not every topic will be covered in every interview. In selecting topics for inclusion, the objective is to acquire complete information for the analyses of project implementation and costs, without excessive duplication and unnecessary respondent burden.

We will be specifically interested in learning from DTA if implementation is being rolled out successfully and with minimal disruptions. We will pay close attention to ACS operations and what happens at the retailer point of sale. Integrated retailers and TPPs will be asked about the complexity and risk of making changes to IECR/POS systems, acquiring and processing applications, and the constraints of time and resources for making system changes. Finally, we will talk with the HIP Steering Committee and community-based organizations (CBOs) about being included in planning and design of the pilot, whether HIP functions are user-friendly (e.g., training, access to balance information), if participants have adequate access to retailers participating in HIP, if HIP participants are being singled out by having to identify themselves or separating their TFV, and if there are any other issues or positive results seen with HIP operations.

As discussed in Section 4.5, key informant interviews will be used to collect data on costs that are not identified by reporting procedures for the demonstration. Cost questions in key informant interviews are designed to supplement these documentary information sources in an efficient manner, filling in gaps and providing further detail about project costs that are not reimbursed by FNS.

Exhibit 4.3: Topics for Key Informant Interviews

Topics	DTA & MA Executives	DTA HIP Staff	DTA Technical Team	Local Office Staff	EBT Contractor (ACS)	Integrated Retailers	Third Party Processors	Community-Based Organizations
General Information (Name, Role, etc.), Including General HIP Information	√	√	√	√	√	√	√	√
HIP Evaluation	√	√	√					
HIP Grant, Including Accounting and Reporting	√	√						
HIP Implementation	√	√	√		√			√
HIP Management and Coordination		√	√	√	√			√
HIP Participant Feedback				√		√		√
Development of DTA's HIP Operational Design		√	√					
HIP Retailer Recruitment, Participation and Support	√	√	√	√	√		√	
HIP System Design and Development and Technical Support		√	√		√			
HIP Staff and Participant Training		√		√	√			
Third Party Processor (TPP) Support			√		√		√	
Card Issuance, Cardholder Customer Service, Cardholder Training				√	√			
Retailer Operations, System Design, and System Operations						√		
Third Party Processor Implementation, Operations, and Support of Retailers							√	
Community-Based Organization Coordination, Operations, and Readiness	√	√	√					√
Issues/Lessons Learned	√	√	√	√	√	√	√	√
Non-Reimbursed Expenses (Non-Grant)	√	√	√	√	√	√	√	√

4.5 Cost Data Forms and Protocols

The evaluation team will identify and quantify start-up and operating costs of participating in HIP from major stakeholders in HIP including: the State's SNAP program (State and local levels), community-based organizations (CBOs), EBT processors, third-party processors (TPPs), and retailers. We will report on actual pilot costs and estimate costs of a nationwide rollout of Healthy Incentives as part of SNAP. The cost analysis will consider (a) administrative costs reimbursed by the HIP grant from FNS, (b) HIP incentives received by participants, and (c) uncompensated costs incurred by stakeholders. Costs will be organized three ways: by organization, function (defined below), and object (personnel, travel, supplies, postage, etc.). This structure will facilitate summarizing costs as desired by FNS.

As part of the evaluation reporting requirements, the Abt team is working with stakeholders to establish or enhance procedures to capture actual costs of HIP implementation and operations as they are incurred. As part of this process, we have provided specifications and formats for the cost data that are needed for the evaluation. The status of these discussions as of the date of this document is specified below. Where possible, we are communicating directly with the stakeholders; the process will be different for TPPs and retailers, as described below. By notifying stakeholders in advance that we will be requesting and gathering start-up and operating costs, we will assure that stakeholders will provide more accurate and complete data on their actual costs throughout the project.

Key informant interviews will be used to collect data on costs not documented by the cost reporting procedures for the evaluation. Cost questions in key informant interview guides are designed to fill gaps in the documented costs in an efficient manner. In the final interviews, we will obtain data on possible impacts on costs during Statewide or national rollouts. The cost data collection is described below in Section 5.5, and the analysis of cost data is discussed in Sections 7.4 and 7.5.

State SNAP Cost Reports

The data elements we will capture will be modeled on those required for submittal of an EBT Implementation Advanced Planning Document (IAPD), including staff time, pay rates, benefits, administrative overhead, materials, travel, and outside support (such as a quality assurance contractor). The estimated costs for the State's SNAP program were provided in the grant application, including costs at the State and local level. DTA headquarters staff who have significant roles in HIP record demonstration time on a weekly basis as part of their routine time reporting. For evaluation purposes, the time sheets include a breakdown of time by the major functions for HIP implementation and operations:

- Design, develop, test, and operate payment processes
- Household recruiting and customer service
- Retailer recruiting and relations
- Training and related materials
- Community relations
- General and administrative
- Evaluation support

On a monthly basis, DTA will compile the timesheet data and associated salary data into a spreadsheet, and submit the detailed weekly data (at the person level) to the evaluation team. Non-personnel expenditures will be reported on a quarterly basis, when DTA submits its expenditures to FNS for reimbursement. DTA will provide monthly reports of total HIP incentives earned from ACS settlement data. DTA interview guides will include items to confirm the reported expenditures and probe for any significant gaps. DTA has begun using the timesheets.

Community-Based Organization Cost Reports

For CBOs that have a major role in HIP as part of the Steering Committee, we have developed a worksheet for tracking project-related expenditures. These organizations do not receive reimbursement from the FNS grant, but they are active in recruiting retailers and are expected to play

roles in participant training. The worksheet provides a template for tracking time spent on the functions listed above. Individual agencies will be asked to record staff time and estimated costs on a monthly basis, with data to be collected in each of the three rounds of stakeholder data collection. Participation in this cost reporting will be voluntary but will be encouraged by DTA as a means to document the full cost of HIP. To fill gaps in these data, community partner interview guides will include items to gather data on any HIP-related costs incurred by CBOs that do not complete the cost worksheets. The worksheet and instructions have been provided to DTA. We plan to discuss the worksheet and instructions with members of the HIP Steering Committee in June 2011.

EBT Processor Costs

For DTA's EBT processor (ACS), implementation costs will include system design and development, and working with the retailers and TPPs to coordinate development and testing of system upgrades. Increases to operating costs will be due to increased data processing and storage, reporting responsibilities, and helpdesk support. New costs are normally passed to the State in one of three ways: (1) an increase in the cost per case month (CPCM) fee paid by the State (covering implementation and operations); (2) a fixed price for the implementation and an increase in the CPCM for operations; or (3) hourly fees for implementation work (usually with an estimate of hours) and an increase in the CPCM for operations. However, the EBT processor may absorb some of the one-time costs. DTA has also retained a project management contractor (Novo Dia Group) to work with retailers and TPPs on new agreements, and to coordinate design and testing among all of the organizations involved in system development (including DTA, ACS, retailers, and TPPs).

Data on EBT processor costs will be collected from three sources. First, original EBT processor pricing will be obtained from the grant budget. Second, the EBT processor and EBT project management contractor will report staff hours and approximate labor costs (with loadings for overhead, based on standard billing rates) and travel costs on a monthly basis, broken down by the same functions as the DTA and CBO expenditures. Third, the interview guides for these firms will include items to probe for whether the actual costs were more or less than the billed costs, and whether other costs (or cost reductions) would be expected for future rollouts. The first report is expected by May 31, 2011.

TPP and Retailer Costs

TPP and retailer implementation costs are those costs associated with upgrading TPP software to accept the differentiation between HIP and non-HIP purchases, and the cost to develop and implement the software changes to IECR systems so that the differentiation of HIP and non-HIP items is automated. DTA's payments to TPPs and retailers for these costs will be obtained from DTA's expenditure reports to FNS. Additional detail to break these costs down by object or function will be obtained, to the extent available, from the budgets negotiated between DTA and the TPPs/retailers. The extent to which retailers will agree to release these details to the evaluation is under discussion.

Costs for retailers without IECRs—meaning those retailers that have to separate HIP-eligible items from non-HIP items—will not include development costs but will include training and possibly other staff time impacts. These changes are small and are not expected to materially affect staffing levels, which would be the only true cost impact to HIP.

The retailer survey will gather data qualitative data on retailers' unreimbursed costs for HIP implementation (for independent retailers only, in Round 1), and on the impacts of HIP on time for operations, including checkout, settlement, reconciliation, and returns (for all retailers, in Round 3). The major retailer interviews will be designed to gather similar data.

National implementation of HIP will require upgrades to all TPP systems as well as to all proprietary and universal ECR software. Much has been done in estimating the impacts of WIC EBT on retailers and how ECR software providers can push an enhancement to all retailers using its system. However, these impact studies are on a State-by-State basis, and some IECR firms are regional. Abt will need to identify and estimate the number of IECR systems that will require upgrades on a national basis. In addition, nationwide cost estimation will require estimation of the number of proprietary systems that are available. To generate these estimates, Abt will consider the number of shared and proprietary systems identified in the WIC EBT implementation and planning activities in various States such as Wyoming, Texas, Kentucky, Michigan, Wisconsin and Florida.

Development and operating costs for which data will be collected are listed in Exhibit 4.4.

Exhibit 4.4: Start-Up and Operating Costs to be Collected

Stakeholder	Start-up Costs	Operating Costs
State SNAP program (State and local agencies)	• Pilot project planning • Negotiation with EBT processor • Participation in system design, testing • QA Contractor • Retailer outreach and recruiting • Training for local staff • Implementing State and local procedures • Participant outreach and recruiting • Funding for CBOs (if applicable)	• CPCM increases (see below) • Project management • Data systems operations • Providing support to participants • Funding for CBOs (if applicable)
CBOs	• Retailer outreach and recruiting • Training staff • Implementing procedures • Participant outreach, recruiting, and training	• Providing support to participants
EBT Processor	• Passed to State: • System design, development, testing • New TPP, retailer agreements	• Some or all of the following included in the CPCM: • Increased stored data • Increased customer service • Increased reporting
TPPs	• System design, development, testing (may be passed to retailers in the form of fees)	• Passed to retailers in the form of fees
Retailers	• IECR system design, development, testing • Training	• Manual retailers—time in lane • Possible increased transaction fees (paid to TPPs)

Definitions: CBO=Community-based organization; may include other non-profit entities. CPCM=cost per case month fee paid by DTA to EBT processor. IECR=integrated electronic cash register (i.e., integrated with payment system)

.

5. Collect Data

This section describes the procedures for data collection in several steps: recruitment, training of data collectors, and field operations for the participant survey (Section 5.1), conducting the focus groups (Section 5.2), conducting the retailer survey and on-site visits (Section 5.3), and conducting the key informant interviews, including collecting information about costs (Section 5.4). The instruments that will be used in data collection were described in Section 4.

5.1 Participant Survey

5.1.1 Recruit and Train Data Collectors

All of the interviews will be conducted over the telephone. When we have difficulty reaching respondents because they have no phone or published number, we will send field recruiters to respondents' homes. The field recruiters will explain the study to the participants and give respondents a cell phone to use to call the Westat Telephone Research Center (TRC). We anticipate we will need to hire and train approximately 24 telephone interviewers for Round 1, 15 for Round 2, and 15 for Round 3. We anticipate retaining the Round 1 interviewers for Round 2 and Round 3 data collection. Because we expect that nearly half of the Round 1 interviewees may have missing or bad phone numbers and will require in-person contact, we will use approximately 32 field recruiters at Round 1 to go to homes and provide a cell phone for the interview. We anticipate having better phone numbers after Round 1 so will need fewer field staff for Rounds 2 and 3 and plan to hire 8 field recruiters for those rounds. We will hire both English-speaking and English/Spanish bilingual interviewers and recruiters. If the pilot site requires languages other than English and Spanish, we will hire interpreters.

Recruiting Telephone Interviewers

The AMPM is a complex instrument to administer. An important consideration for telephone interviewers will be their ability to conduct the AMPM interview. Westat maintains a pool of interviewers who have experience conducting 24-hour dietary recall using the AMPM who formerly worked on NHANES. Given that the AMPM is a very systematic and labor-intensive instrument to learn to administer, we will make every effort to use the skilled NHANES interviewers who are already trained.

To meet the needs of this study, telephone interviewers must be able to:

- Represent the study to respondents knowledgeably and professionally
- Effectively encourage sampled respondents to participate in the interview
- Administer the survey skillfully
- Record respondent data accurately

Training Telephone Interviewers

To train interviewers to master each of these tasks, we will use a rigorous training protocol that emphasizes all of these skills. Because the Round 1 interview will not require administration of the specialized AMPM instrument, training will consist of a relatively standard training protocol that will

include about 8 hours of interviewer training. This 8-hour training will be specifically dedicated to the details of the HIP evaluation and will be in addition to the 8-hour initial general interviewer training session that teaches all new interviewers the basic skills of telephone interviewing and how to use the Computer Assisted Telephone Interview (CATI) platforms that will be used for the interviews. The Round 1 interview will be respondents' first interaction with the interviewers who will represent this study. To make sure we create a good initial impression and make the interview a positive experience to encourage participation in future interviews, we will select interviewers who have shown a strong ability to gain respondents' trust and cooperation on previous Westat studies.

Westat's computing and telephony capabilities enable a networked combination of geographically diverse data collector locations to operate as a single and secure "virtual" TRC managed from the home office in Rockville. Thus many interviewers work from their homes and are monitored via the telephone by TRC supervisors. While the training will originate from the Westat home office, all interviewers will be trained remotely through web-based training and WebEx conference calls. The interviewers will receive the training at either a Westat TRC computer or their home computer, depending on their designated workplace.

The Round 1 training will begin with 3–4 hours of training with self-paced materials presented via web that give interviewers the chance to familiarize themselves with the study background and questionnaires. It will continue with a 4–5 hour trainer-led WebEx session that focuses on gaining cooperation, questionnaire delivery, accurate coding, effective neutral probing, and appropriate contact procedures. Each element of the trainer-led session will require trainees to read portions of the survey and to respond to questions and concerns just as they would in live interviewing. The Round 1 training will conclude with monitored practice interviews that will serve as a complete "dress rehearsal" using the same program and environment as the actual live data collection. The interviewers will read the questions from their CATI screens at their work stations (whether from a physical Westat TRC or their homes) and record the answers of another interviewer acting as the respondent. Experienced supervisors will monitor the practice interviews and assess each interviewer's readiness to conduct live interviews. Trainees who are not fully comfortable with the instruments and procedures will receive further coaching and evaluation or will be replaced.

For the experienced NHANES telephone interviewers who will be administering the AMPM in Rounds 2 and 3, we will provide a refresher training that describes the study and informs the interviewers of any new aspects of the data collection. This training will be delivered remotely and interviewers can complete their training sessions by logging on to our Learning Management System through the Internet. If we need to train telephone interviewers who are not experienced in administering the AMPM instrument, we will provide a standardized remote training. Westat has recently restructured the 40-hour NHANES AMPM in-person training for off-site (i.e., virtual) telephone interviewers. The revised training structure can be completed in 30 hours. The first component of this training is a series of self-paced tutorials that are placed online in our Learning Management System. These tutorials provide interviewers with background information about the study and familiarize them with the instrument and interviewing protocols through a series of interactive presentations, including several that familiarize them with the actual CATI program. After successfully completing the self-paced training, interviewers are invited to the trainer-led session. This 8-hour session allows the trainer to evaluate the trainees and to focus on the more complex aspects of the interviewing process such as probing and portion size estimation with additional

explanation and examples. This trainer-led session covers contact procedures to teach interviewers how to accurately assign result codes that will help managers track survey progress and ensure that future call attempts are made at the right time. This session also includes training and exercises that teach interviewers how to knowledgeably answer questions about the study and establish rapport that encourages cooperation and survey participation. Training concludes with 7 to 8 hours of practice interviewing. This portion of training begins with interviewers conducting interviews with each other using scripted role-play scenarios to make sure everyone gets to experience some standard survey paths and issues. After five role-play sessions, interviewers conduct two practice interviews with volunteer respondents to simulate as closely as possible the actual interviewing experience. These volunteers will be called at home by the interviewers. Supervisors and project managers monitor the role-play exercises and practice interviews to provide coaching to trainees, and evaluate and confirm their readiness for live data collection.

Recruiting Field Recruiters for In-Person Contacts

Given the need to obtain approximately 2,800 completed interviews at Round 1 and our projection that approximately half those interviews will require home visits, we expect to field 32 in-person field recruiters for the Round 1 interview.

Westat employs well-established methods for hiring in-person data collection staff. To assess the extent to which potential personnel hired meet study needs, we rely on personal interviews, contact with professional references, assessment of previous experience, and observations during training sessions. Westat maintains a database of over 5,000 field workers. We will first access Westat's Field Files database to determine the number of qualified field data collectors living within or within 50 miles of Hampden County. Following is a preliminary outline of basic requirements:

- Experience working with low-income urban and rural residents
- Experience tracking hard-to-reach populations
- Favorable or highly favorable reviews after finishing their last project at Westat or at their most recent employer
- At least two excellent references (for new hires)

Because of our intensive recruiting and screening process, we will hire staff members who meet the skill level required for this project. In addition, there will be an ongoing screening at the recruiter training sessions, and we will carefully observe all recruiters. If the training staff is uncomfortable with a particular recruiter, we will take the necessary actions to replace that recruiter with someone more suited for the work.

Given the population size of the grantee area, we expect that we will be able to hire 22 local recruiters. The remaining 10 recruiters will be travelers; that is, we will need to hire recruiters living beyond 50 miles of Hampden County. Westat has experience in hiring experienced field data collectors who are willing to travel to conduct data collection.

Training Field Recruiters

The 32 recruiters will be trained onsite in Hampden County for the Round 1 interview. The 2-day training will cover contact procedures, study procedures, and establishing rapport with respondents. We believe that the quality of the data is directly related to the quality of the training of our field staff.

Recruiters will be thoroughly trained on all aspects of data collection, from initial contact procedures, to refusal avoidance and conversion.

The training materials that will be developed for the study's recruiters will include, but will not be limited to, the following:

- ***Field Recruiters' Manual.*** This manual documents all study procedures for the recruiter. The manual serves as the reference document and training device that provides an overview of the study, detailed descriptions of study procedures, handling of respondent questions and refusals, and use of the study cell phone. The manual also provides space for memos and procedural updates to be filed.
- ***Training Agenda.*** This agenda will be a detailed document that divides the training into timed sessions on specific topics. It lists both the trainer and trainee materials that are required for each, and indicates the overhead transparencies or PowerPoint presentations that will be used in each presentation.
- ***Introduction to the Study.*** Generally (but not always) our clients provide this introduction and welcome. Clients are always welcome at training and their presence and enthusiasm for the study has a positive effect on the field staff.
- ***Demonstration of the Contact Procedures.*** The trainers will complete a contact scenario; one trainer will act as a respondent and the other as a field recruiter. This is done to give the field recruiters a feel for the types of questions being asked about the study and the interview. Interviewers will also be trained on the best methods on approaching respondents to ask them to use a Westat cell phone to call into the TRC and complete the interview.
- ***Commonly Asked Questions.*** This training session will concentrate on refusal avoidance. We will give the recruiters responses to questions and objections respondents usually have to participating in a survey.
- ***Dyad Role Plays.*** In this part of the project-specific field recruiter training, each recruiter trainee acts as a respondent, then a recruiter, with a partner to practice contacting respondents and refusal conversion. Each dyad, or set of partners, works through six to eight role plays. Each partner must act as the respondent and the field recruiter for each role play. Project staff and field supervisors listen to the field recruiters and evaluate their performance during the role plays.
- ***Administrative Matters and Case Updates.*** Trainees will also learn how to assign case status outcomes to study participants using texting on cell phones to the TRC, and how to document case status on hard-copy study materials.

Training for the Round 1 interview will occur August 22–23, 2011 and recruiters will begin work immediately following training. Interviewers retain training knowledge best when it is put into practice most quickly. For the Round 2 and Round 3 field efforts, only eight interviewers will be needed to do the in-person tracking of respondents, so we will employ a subsample of the best local interviewers from Round 1.

5.1.2 Interview Schedule and Procedures

The Round 1 survey will be conducted between August 22 and November 30, 2011. HIP implementation will begin November 1, 2011, with a staggered rollout lasting for 3 months. DTA will send a notification letter to HIP participants approximately 3 weeks before their HIP participation begins. All baseline interviews with treatment participants need to be completed before they receive the notification letter.

Contact Information

The SNAP case file will provide contact information for each sampled survey participant. We will review the file to identify those households that have no listed telephone number. These cases will be submitted to Lexis-Nexis, a vendor specializing in identifying telephone numbers associated with addresses and/or individual names provided. We may also use directory assistance and other Internet tracking methods to identify as many telephone numbers as possible. Tracking requires the collection and use of respondents' personal identifiers. We will carefully protect such data, both within the Abt team and when submitting information to outside vendors for tracking information. Despite our best attempts to obtain working phone numbers where they are missing from the SNAP case file, we know that some cases will not have phones and will require in-person contact.

We expect the percentage of households who cannot be reached by telephone will decrease in Rounds 2 and 3. This is because we will collect contact information including cell phone numbers for the respondent and primary shopper, as well as for alternate contacts, at the end of the Round 1 and Round 2 interviews. We will send a mailout requesting respondent's current contact information between the Rounds 2 and 3 interviews (described in more detail below).

Advance Letter

The advance letter, mailed approximately 5 days before the start of interviewing, will explain the purpose of the study and mention that a telephone interviewer will be in contact shortly. We will print the letter on the sponsoring agency's letterhead or on a letterhead especially designed for the study. We will include an endorsement letter from the agency Project Officer, as well as our State contact. Important features of the advance letter are described below.

- ***Tailored content.*** Given the strong correlation between topic salience and respondent cooperation (e.g., Groves, Presser, & Dipko, 2004), we will tailor the content of the letters to be appropriate to SNAP participants living in Hampden County. We will develop the text of the letter so that the targeted sample is not aware that the study is capturing information about the quantity of fruits and vegetables they consume.
- ***Assurances of confidentiality/other content.*** The letter will include assurances that the survey information will be kept confidential and that participation is voluntary. A toll-free number will also be provided that respondents can call with questions about the survey. The letter will highlight the expected time it will take to complete the interview and mention the survey incentive.
- ***Type/Design of notification.*** Although survey methodologists tend to agree about the value of advance letters in increasing response rates, they also acknowledge that many recipients do not read them. Discussions and experiments about how to improve readership of cover letters

and advance letters focus not only on content (keeping them brief yet compelling), but also on graphic presentation, such as using more eye-appealing fonts, headings, and graphic images, and whether to use a letter or a postcard, or a letter accompanied perhaps by an eye-catching insert. We will work with FNS to identify an appropriate format to ensure that the letters are opened and not tossed out.

Case Flow

Participants will be enrolled in HIP in three waves, with Wave 1 starting November 1, Wave 2 December 1, and Wave 3 January 1. The 5,070 respondents sampled for the household survey will be evenly distributed among the three waves with 845 treatment and 845 control cases fielded in each wave.[18] Field protocols for treatment and control cases will be as similar as possible.

Consent Procedures

Verbal informed consent will be obtained from all participants at each round. Interviewers will explain the purpose of the study, study components, and approximate length of the interview. They will also inform the participant that participation is completely voluntary, and review our procedures for ensuring confidentiality, emphasizing that the participant's SNAP benefits will not be affected if they decline participation. If the primary food shopper is someone other than the sampled participant, the interviewer will explain the study consent and confidentiality procedures to the primary shopper when that individual comes on the telephone.

Consent for Sampled Youth Ages 16–17

We will obtain parental consent for respondents under age 18, except for emancipated minors who have been granted the status of adulthood by a court order or other formal arrangement.

Should minors be sampled, our approach to recruitment includes first identifying participants in this age group who are not heads of household through information provided on the SNAP case file. We expect this segment to be small (~150 for Round 1, ~80 for Round 2, and ~40 for Round 3). In Hampden County, SNAP participants aged 16 and 17 years old are only 5.6 percent of all participants aged 16 and older. The parents or guardians of these youth will be sent a customized advance letter detailing our plans to interview the youth and then interview the parent/guardian as the primary shopper. We will first ask to speak to the youth's parent/guardian to obtain consent for the youth's interview. We will then obtain the youth's assent and interview the youth. If the parent is the primary shopper, we will then complete the primary shopper portion of the interview with him/her. If the parent is not the primary shopper, we will interview the individual who is. If the parent/guardian is not available, the interviewer will ask for contact information and an optimal time to contact them.

Scheduling the Interview

Since we estimate that the Round 1 interview will take approximately 30 minutes to complete, we will attempt to interview the respondent if he or she is available when the telephone interviewer calls or the field recruiter visits the respondent's home. We will schedule the interview for another time, if the respondent is not available or the time is not convenient.

---

[18] Some respondents will have exited SNAP between the time the sample is drawn and the date of the interview. We will check SNAP enrollment monthly using the case files. Those respondents who are no longer receiving SNAP will be excluded from the baseline survey.

Completing the Primary Shopper Portion of the Interview

At the end of the individual portion of the interview, the interviewer will ask the sampled respondent whether he or she considers himself or herself one of the primary shoppers for the household. If the sampled respondent is a primary shopper, the interview will continue with the shopping behavior questions. Because shopping roles may change within a household, we will identify the primary shopper at each round. If the sampled respondent does not self-identify as a primary shopper of the household, the interviewer will ask to speak to the primary shopper. We estimate that 72 percent of the selected respondents will also be the primary shopper in the sampled households.[19] The remaining 28 percent will have a second person who will be identified as the primary shopper. As noted above, the interviewer will explain the purpose of the interview, review study confidentiality procedures, and continue the interview with the primary shopper if that individual agrees to the interview. If the primary shopper is not available, the interviewer will verify the primary shopper's name and contact information and try to set an appointment.

Collecting Contact Information

We will collect contact information at the end of the Round 1 and Round 2 interviews from the sampled participant, for help in locating the participant for Rounds 2 and 3. This will include collecting contact information for up to three alternate contacts who should always know how to reach the participant. If the sampled participant is the primary shopper, this information will be collected at the end of the survey. If the sampled participant is not the primary shopper, this information will be collected at the end of the sampled participant section of the interview, before the primary shopper is interviewed.

Payment of Research Incentives

Incentive checks with an accompanying thank-you letter will be mailed to respondents after completion of the respondent and primary shopper interviews. We will add the contact information received from the interviews for respondents into our Study Management System (SMS). The mailing address information will be provided by the sampled respondents at the end of their interview so that they can receive the research incentive check.

Tracking between Rounds

A standard practice for surveys with multiple interviews over time is to communicate with study respondents between interviews, in order to maintain up-to-date contact information as well as interest in and cooperation with the study. There is relatively little time (3 months) between Round 1 and Round 2, so we are more concerned about losing track of respondents between Rounds 2 and 3 as they move or change their telephone number. We will send out a mailing between the two follow-up interviews, approximately 6 months after the Round 1 interview is completed. The mailing will include a card that respondents will be asked to use to provide current contact information (and an alternate contact) and mail in to Westat. The mailing will also contain information about the Round 3 interview and the approximate date of the interview, as well as $5 to thank respondents for their time.

[19] This estimate is based on data received from DTA on the number of SNAP households in Hampden County with more than one adult where the probability that the sampled respondent is the primary shopper decreases with larger household size.

We will use USPS address return service to determine if a respondent has moved to a new address, in lieu of the letter being forwarded.

Telephone Interviewer Supervision and Quality Control

With careful selection and an extensive and tested training protocol, the HIP evaluation will have a team of telephone interviewers who can represent this study well and collect accurate data. However, our efforts to ensure the highest quality of data continue after training. From any location, supervisors can monitor any interviewer with full and undetectable audio and visual monitoring of the interviewer's screen and both sides of the telephone call. Supervisors will monitor about 10 percent of the interviews. Supervisors and interviewers communicate via instant messaging when calls are in progress and are debriefed by telephone after monitoring sessions. Supervisors offer constructive feedback and coaching and document all monitoring sessions in our monitoring database. At a broader level, shift coordinators are responsible for allocating supervisor hours and assigning supervisor tasks to make sure all interviewers and all projects receive appropriate levels of monitoring. In addition to supervisors, newer interviewers have mentors who provide a more consistent and personal point of contact and provide interviewers with additional support and coaching. Project coordinators and/or operations managers review a variety of reports and review data to identify interviewers who have low cooperation, low productivity, and/or questionable data in their completed surveys.

Field Recruiter Supervision and Quality Control

For Round 1, Westat will hire two field supervisors to oversee the work of the 32 field recruiters. The field supervisors will be responsible for assisting with in-person training of recruiters. The supervisors will speak with recruiters daily at the beginning of data collection and weekly (or more frequently, if needed) thereafter during the data collection period to collect information on response rates and provide guidance on field issues.

Operational processes will be monitored, controlled, and adjusted based on various quantitative reports that will be regularly reviewed and analyzed by the field manager, the field director, and other project staff. These will cover such areas as the following:

- Interviewing production and survey progress
- Sample yield performance overall and by jurisdiction/travel day stratum
- Response rates
- Recruiter performance
- Tracing workload

5.2 Participant Focus Groups

Focus groups during Round 2 and Round 3 will be used to listen to and gather information from the participants regarding their perceptions of HIP. The project team will conduct three in-person focus group sessions in Hampden County at two points in time during the HIP evaluation data collection. Each session will run about 1.5 hours. The first three focus group sessions will be conducted between 1 and 2 months after the Round 1 interview, and the second group of three sessions will be conducted between 9 and 10 months after the Round 1 interview. During each round, two focus groups will be conducted in English and one in Spanish.

5.2.1 Location of the Focus Groups

We will secure locations in Hampden County that are centrally located and comfortable for participants. In addition, we will choose a physical environment that will enable effective group moderation and participation in discussions, and allow for effective, and minimally distracting, data collection. Participants will be paid $75 for their time and transportation costs. As food can positively influence focus group discussions because eating together tends to promote conversation, we will provide a light meal such as pizza or sandwiches during the focus group session. To ensure clear audiotapes of the sessions, we avoid serving "noisy" foods.

5.2.2 Focus Group Field Methods

Four study team members will travel to the site to conduct three focus groups over a 2–3 day period. The team will be made up of an English-speaking moderator, a Spanish/English bilingual moderator, a bilingual notetaker, and a staff person assigned to greet participants, ensure participants get to the correct room, obtain informed consent, and distribute incentive payments. A moderator will facilitate each session and a notetaker will record major themes, relevant gestures, etc., from the session. All sessions will be audiotaped. Moderators fluent in Spanish will conduct focus group sessions with Spanish-speaking participants.

Identifying transitions between speakers and whether an idea is being generated multiple times by one speaker or one time by multiple speakers is important during the analysis stage. Therefore, each participant will be identified by first name, pseudo-name, or initials. The note taker will track the individuals as they speak. In addition, the moderator will either call on people by name or thank them for their contribution.

Audiotapes will be transcribed verbatim, and the moderator will review for gaps.

5.3 Retailer Survey

5.3.1 Field Methods

The retailer survey will be conducted by mail, with telephone follow-up. Below, we describe the specific data collection and processing procedures that we will use to collect survey data from store managers and clerks/clerk managers.

Notification of Survey Respondents

The data quality assurance process begins with providing clear instructions to respondents. We will notify the selected retailers about the study via a telephone call. That call will describe the study, solicit participation on behalf of FNS and DTA, and confirm the store manager's name and mailing address. That call will also identify retailers requiring the Spanish version of the survey. The data collection staff conducting calls to retailers and answering questions from respondents will include a member who is fluent in Spanish. Surveys will be mailed to store managers/owners. All sampled retailers will be assigned an ID code. Each survey form will have a specific pre-printed respondent ID code that indicates the store type and number.

The survey mailing from the evaluation team will contain the following materials:

- An introductory letter signed by an FNS official that explains: 1) the purpose and the importance of the study; 2) the expected use of study findings at the Federal, State, and local levels; and 3) that all information collected from participants will remain confidential.
- A letter from DTA, with support of the retailer association, requesting retailer participation and explaining participation requirements.
- A fact sheet introducing the HIP evaluation and describing the objectives and importance of the study, data collection activities and schedule, and the evaluation team's contact information.
- The survey forms, divided by questions about the financial effects of HIP participation and questions about effects of HIP on checkout processes.
- A prepaid FedEx envelope addressed to Abt, for retailers to return the participation agreement and survey(s).

Technical Assistance for Survey Respondents

To ensure that all survey questions are clearly and uniformly understood by respondents, the introductory letters will include an 800 technical assistance number that respondents can call to request clarification on any of the survey questions or procedures.

Monitoring Survey Progress—Data Receipt Database

Survey packets will be sent by overnight delivery. For retailers participating in HIP, the Round 1 surveys will be sent prior to the start of HIP. For those who do not participate, the surveys will be sent in October 2011. For those retailers who initially choose to participate but drop out, the Round 3 surveys will be sent approximately 11 months into HIP, in November/December 2012. Responses will be due to Abt Associates within 2 weeks of receipt for all rounds of data collection. The managers and clerks, where applicable, will seal their completed survey forms individually in envelopes provided, and return the envelopes by overnight mail using a pre-paid package label.

An Access-based relational data receipt database will be used to store and manage evaluation-related information on all retailer survey participants by their ID codes. Each respondent record in the database will be identified by store type and ID.

The data receipt database will also contain all data collection activity information for each sampled retailer, such as the dates of initial survey mailing, any follow-up activities conducted, and receipt of completed survey; conditions and final dispositions of the returned survey form; and status in survey data processing and research incentive payment. The data receipt database will be used to generate regular reports on data collection activities such as survey distribution, receipt, and data processing status as well as to conduct real-time assessments of survey progress at any time, as needed.

Survey Follow-Up Procedures

During the data collection period, we will monitor the return rates of surveys, and conduct follow-up activities to facilitate timely completion of the surveys. Two weeks after the initial mailing of the survey packages, we will make reminder calls to all non-responders, asking them to mail in their

survey or to complete the survey by telephone. Staff assigned to follow-up will be trained to administer the entire survey using the paper questionnaire, or to obtain responses to incomplete surveys submitted by retailers. Training will include background, confidentiality, contact protocols, and responses to frequently asked questions. Additional telephone calls to remaining non-responders will continue every week starting with the third week of the survey period.

Progress in survey return rates, any unexpected data collection problems, and how those problems were resolved will be reported to FNS.

Data Receipt and Survey Checking

Completed survey forms will be returned to Abt Associates on a rolling basis. On receipt, we will make an initial pre-processing check to ensure that all documents and forms have been returned, or that an explanation is available for incomplete submission. If the explanation for an incomplete submission is missing or inadequate, the data receipt staff will contact the respondent to request the missing forms. The outcome of this follow-up contact will be entered in the data receipt database.

Data coding and data entry will proceed as data are received. We will use double entry. Data will be edited for valid values and internal consistency, and response rates will be tracked on an ongoing basis. Inconsistent and/or incomplete data will be resolved by callbacks to respondents.

5.3.2 Maximizing the Response Rates

By carefully and convincingly explaining the importance and potential usefulness of the study findings in the introductory letters from FNS and DTA, and by implementing a series of follow-up reminders and offers to complete the survey by telephone, we expect to achieve an overall survey response rate of 80 percent for participating retailers.

Securing the cooperation of the non-participating retailers and those who dropped out of the pilot may be more challenging than for participating retailers. Prior to participating in the HIP, retailers will be required to sign a letter or memorandum of understanding (MOU) with the DTA, agreeing to comply with the terms of the MOU and adhere to the procedures specified by FNS. The DTA is developing a three-party memorandum of understanding among DTA, Affiliated Consumer Services (ACS, the EBT vendor), and retailers. The MOU will include language about evaluation participation requirements.

Lacking this arrangement with non-participating retailers, we will encourage non-participating retailers or those who dropped out of the pilot to cooperate by appealing to their interest in improving the health and nutrition of Americans participating in SNAP. We will also offer a modest monetary incentive of $40 to the non-participating retailers for completing a survey.

Procedures for On-Site Visits

We propose to visit approximately 10 purposively selected retailers during each of the three rounds of site visits for key informant interviews (described in Section 5.4). The store visits will be scheduled in advance with owners or managers, and will use the observation form described in Section 4.3. The preferred language of a selected store's owner or manager will be determined at the time the appointment is made. Store visits will be conducted by teams of two staff, to reduce the time in store

and for safety reasons. If the preferred language for a store owner or manager is Spanish, a member of the team will be bilingual. In the store, the data collectors will first introduce themselves to the store manager. One of the data collectors will observe while a checkout clerk or supervisor conducts a simulated HIP transaction using pre-specified items (including target fruits and vegetables, other foods, and non-food items). The team will collect data on the offerings and prices of selected target fruits and vegetables, and on signage related to HIP and other aspects of the store environment.

The store observation teams will be led by one of the two senior staff who developed the instruments and conducted the pretest. Before each round of observations, they will conduct a 2-hour training for the other data collectors. The training will include: study background, objectives and procedures for store observations, projecting confidentiality, and responses to potential problems.

5.4 Key Informant Interviews

Key informant interviews provide essential information for the implementation analysis (see Section 7.3), the cost analysis (see Section 7.4), and the feasibility of extending HIP nationwide (see Section 7.5). Interview topics were described earlier in Section 4.4. Informants will represent all stakeholders: DTA (headquarters and local), CBOs, EBT processor, EBT project management contractor, major retailers, and TPPs (see Exhibit 3.4 for complete list.).

5.4.1 Interviewer Training

Experienced interviewers, familiar with monitoring EBT implementation and SNAP operations, will conduct the in-person interviews with stakeholders. To prepare them for this task, interviewers will meet to review the topic guides and the study context prior to each round of interviews. So that data collection is consistent and coordinated among the interviewers, evaluation staff will conduct a single half-day training session for all interviewers. In the training, we will explain the purpose and context, the interview objectives, the data collection process, the interview methods, respondent organizations, and current status of the demonstration. We will develop background materials specifically for the stakeholder interviewers that include:

- Overview of the project
- An explanation of why the study is being conducted and what it is designed to accomplish
- The research questions the study addresses
- A description of the main goals in terms of interview targets and expected time frame for data collection
- A discussion of stakeholder recruitment procedures
- A discussion of interview questions

Interviewers will be trained to encourage stakeholder candidness while assuring confidentiality. In support of this goal, we will include scripts that explain that no names or identifying information will be used in summary reports or in a public database.

5.4.2 Data Collection

It is the intent of the Abt evaluation team to maintain a focus on applicable topics for each informant group. Prior to conducting the interviews, we will review the interview guides with the DTA HIP management team to confirm which subject areas remain applicable to informant groups and to determine if additional questions or probes are required. We will send materials to each interviewee which will include an introductory letter requesting consent and providing a brief overview of the key questions and topics to be discussed in the interview. This material will help informants identify in advance any extant information that could be useful in answering the interview questions. After sending the introductory information, interviewers will call or email each respondent to schedule an interview, or, if applicable, will request DTA HIP project management support in setting up interview schedules and locations.

During each round, teams of two researchers will conduct in-person interviews with DTA State and local staff. Each team will consist of a senior researcher and an analyst, who will record each interview. This configuration will allow two-person interviews with State staff and local managers, and one-person interviews with local line staff and community organizations, balancing the need for senior interviewers to connect multiple perspectives with the efficient use of senior researchers. State staff interviews will be approximately 2 hours in length, depending on the respondent, the round, and the documents and other information received prior to the interview. We expect that the team will be on site in Boston for one or two days for each round. Local interviews with managers and community organizations will be 1–2 hours in length; line worker interviews will be ½ to 1 hour in length. The schedule allows for 3–4 days of interviews in Hampden County in each round, taking into consideration travel time between interview sites.

A team of two EBT experts will conduct three rounds of EBT staff interviews. In coordination with State-level DTA interviews, an EBT expert will interview DTA EBT staff and Executive Office of Health and Human Services (EOHHS) technical staff; this expert also observe the interview with the DTA HIP Project Manager. An EBT expert will interview the EBT vendor (ACS) staff on-site at their headquarters in Austin, Texas. While in Austin for two days, the EBT expert will also interview the Nova Dia Group, which is serving as technical support for DTA. If additional ACS HIP staff are located outside of Austin, they will be interviewed by phone. Integrated retailers and TPPs are likely to be numerous and widely dispersed. Therefore, the EBT experts will conduct telephone interviews (approximately 1–2 hours in length) with participating integrated retailers and TPPs.

After each round of interviews, teams will conduct telephone follow-up, as necessary, to clarify responses and speak to additional contacts identified during in-person interviews. Interview data will be compiled in Word documents structured according to the interview guides. Database software will be used to integrate and summarize the interview data.

5.5 Cost Data Collection

The research strategy will follow the "resource inventory" method that the Abt evaluation team has used in numerous EBT studies to measure or estimate costs of all pilot activities, regardless of who pays for the costs. The pertinent activities for HIP include all steps necessary to recruit and train retailers and participants, establish payment procedures, and implement and operate the pilot. From

DTA's monthly reporting of HIP grant expenditures, we will extract the cost of DTA and technical staff supported by grant funds and the funds used to support the ACS HIP implementation, the services of Nova Dia, DTA's technical consultant, and the funds paid to TPPs and SNAP retailers to prepare their systems for HIP transactions. Funds paid to ACS, Nova Dia, TPPs and retailers should also be supported by detailed budgets provided to DTA. The Abt evaluation team expects this information on grant expenditures to be available by organization and object, as discussed in Section 4.5.

The team will gather additional HIP costs through various avenues and will use these costs to develop an overall cost for the pilot as well as the estimated costs for a national rollout of Healthy Incentives to all States' SNAP programs. The approach for gathering costs will depend on the stakeholder group.

- ***DTA State and Local Staff.*** DTA will track time spent on HIP by all staff with direct and substantial involvement in the project. This time will be tracked via weekly time sheets indicating total HIP time, total non-HIP time, and a percentage breakdown of HIP time by function (using functions identified in Section 4.5.6). DTA will provide a quarterly report showing the actual hours reported on HIP and the hours charged to HIP. Upon completing each round of interviews, Abt will compare DTA HIP resources identified in the interviews against grant reports and time sheets to ensure that the time and cost for each resource is captured. If a HIP resource is identified during the interview whose time/cost has not been captured, Abt will follow up with DTA to ensure appropriate costs are captured. During the final interview, the Abt team will ask questions concerning anticipated costs for implementing the system statewide.
- ***ACS (the EBT Processor) and Nova Dia (the EBT Consultant).*** Each of these entities is being paid for their services through grant funds. The cost analysis will use the budgets submitted by these firms and the quarterly grant reports to ascertain the grant funds spent on services. We will inquire whether these firms are tracking actual time spent and are willing to share this information; if so, we will obtain the data and compare to the level of effort indicated in the firms' budgets. During interviews, the Abt team will inquire as to whether the HIP project costs were greater than anticipated and if so, what additional costs were incurred by the firm. During the final interview, the Abt team will ask questions concerning whether the cost per case month (CPCM) would fall with an increase in the caseload during a statewide rollout, additional costs for implementing statewide, and potential costs, or cost savings, if implementing in other States. We will also ask questions of ACS and Nova Dia concerning the estimated number of IECR systems (off-the-shelf and proprietary) that would require modifications if HIP were deployed nationally.
- ***Third Party Processors.*** Participating TPP system modifications are being supported through grant funds. The cost analysis will use the budgets submitted by these firms and the monthly grant reports to ascertain the costs of TPP system modifications. During interviews, the Abt team will inquire as to whether the HIP project costs were greater than anticipated and if so, what additional costs were incurred by the firms. During the final interview, the Abt team will ask questions concerning whether HIP transactions will impact the per transaction fees currently being assessed to retailers for SNAP EBT transactions. The team will also inquire as to whether additional effort would be required to expand HIP statewide or to capture HIP transactions if HIP were implemented in other States. We will also ask questions about

whether other TPPs known to the respondents would need to modify systems to accept TPP transactions if HIP were deployed nationally.

- ***SNAP Retailers with Integrated Systems.*** Participating SNAP retailers with IECR/point-of-sale (POS) systems are undergoing system modifications to accept HIP; these activities are being supported through grant funds. The cost analysis will use the budgets submitted by these firms and the monthly grant reports to ascertain the cost of IECR/POS system modifications. During interviews, the team will inquire as to whether the HIP project costs were greater than anticipated and if so, what additional costs were incurred by the firms. During the final interview, the team will ask questions concerning whether HIP transactions impact retailer costs of operations and whether additional efforts would be required to expand HIP statewide or to capture HIP transactions if HIP were implemented in other States where the retailer has stores.
- ***Community-Based Organizations.*** Prior to implementation, CBOs will be provided with a template to capture resources, time and costs related to supporting HIP in Hampden County. As CBOs have not been supported by grant funding, participation in data collection will be voluntary. Time sheets will be gathered during on-site interviews and reviewed with the primary contact person at each CBO. It is understood that these costs are ancillary to the cost of HIP as the support of CBOs may or may not be used by States and local areas if HIP were deployed statewide or on a national basis.

5.6 EBT Transaction Data

Transaction data will be used to examine take-up of HIP incentives by participants and how SNAP redemptions will differ at participating and non-participating retailers. These data will include benefit issuance, purchase transactions conducted at the POS, and HIP incentive accrual resulting from eligible purchases.

In Massachusetts, SNAP issuances are distributed to SNAP cases during the first 13 days of the month. The State updates the EBT system with SNAP issuance data on a daily basis, which in turns activates the monthly SNAP allotment on the EBT cards of SNAP households. The EBT system will capture all transactions of SNAP households carried out using their EBT cards, such as purchases and returns. After each HIP-eligible transaction, the EBT card of a HIP participant will be awarded the appropriate HIP incentive accrual up to the monthly maximum incentive allowed. All transaction information will be recorded in the EBT transaction system.

The daily EBT transaction data from the EBT processor will include:

- Unique case number (encrypted)
- HIP participation flag (H, I, J or K)
- EBT card number
- Retailer ID (where applicable)
- Transaction Type (Issuance, Purchase, etc.)
- Transaction Date
- Transaction Time

- Transaction Amount
- Account Balance
- Transaction HIP-eligible Amount
- Transaction HIP Incentive Amount
- HIP Incentive Earned Month to Date

From this data, we will be able to construct the following household-level data:

- Issuance Data for each issuance:
  - Amount of issuance
  - Date and time of issuance
  - Type of issuance (monthly/non-recurring)
- Transaction Data for each purchase transaction:
  - Type of transaction (HIP-eligible or not)
  - Date and time of transaction
  - Retailer ID associated with the transaction
- Incentive Data for each incentive credit:
  - Amount of incentive credited to the case
  - Date and time of incentive credit
  - Retailer ID associated with the purchase for which credit is awarded

We will be able to manipulate the EBT transaction data to view transactions at the retailer level, using data elements from the retailer data files such as retailer type, how the retailer processes SNAP transactions, and whether the retailer is participating in HIP. In addition to the EBT transaction data, we will obtain the retailer data files from FNS and DTA to identify the universe of SNAP retailers in the project site. FNS data will provide the REDE file of authorized retailers and corporate data from STARS. These data will be the primary source of contact information, and will also identify FNS store type (also used for stratifying the retailer survey sample). The information from the FNS retailer database will be supplemented by the retailer files from DTA, which will include supplementary contact information, information on which retailers are participating in HIP and whether participating HIP retailers are manual (separating items in-lane) or whether their systems have been enhanced to automatically identify HIP items. Use of these retailer files will facilitate the survey and EBT analysis, and reduce survey burden.

6. Create Data Files

This section provides our data management plan in six sections: data overview (Section 6.1), data flow and creating the analysis database (Section 6.2), participant data cleaning and editing (Section 6.3), weighting procedures (Section 6.4), survey data entry, cleaning, and editing (Section 6.5), key informant and cost data process (Section 6.6), and creating data files for FNS and public use (Section 6.7).

6.1 Data Overview

We will obtain administrative data from three sources: 1) DTA, Massachusetts's SNAP agency; 2) ACS, DTA's EBT vendor; and 3) FNS. However, data originating from ACS will be transferred to Abt through DTA or FNS. In addition, the Abt team will also collect participant survey and retailer survey data. A brief description of the seven data files to be used in the project is presented below and an overview of data collection is presented in Exhibit 6.1. (Note: this discussion excludes qualitative data, such as key informant interview notes and summaries, and demonstration documents.)

1. ***SNAP Case Files:*** The SNAP Case Files obtained from DTA will contain demographic, household, eligibility, recertification and issuance information for persons in households receiving SNAP benefits. Households without adults and households that do not purchase food for themselves will be excluded from the Case Files. The first SNAP Case File obtained in August 2011 will be used in conjunction with the subsequent daily EBT Transaction Files to identify the sampling frame and conduct random assignment prior to the Round 1 survey as well as a source of characteristics for household analysis.
2. ***Admin Costs Files:*** These files will be obtained from DTA and will contain data on the HIP-related implementation and operating costs for the State and local SNAP agencies, and costs for the EBT vendor and retailers reimbursed by the State.
3. ***Daily EBT Transaction Files:*** On a daily basis, ACS will provide the EBT Transaction Files to FNS, which in turn will provide the same data to Abt without the SSNs. These files will contain data on SNAP issuance, HIP and non-HIP purchase transactions, HIP incentive credits, and other transactions for each EBT card issued to a SNAP household in the project site. These files will be used for the transaction analysis as well as to confirm SNAP participation prior to the participant survey.
4. ***Retailer Files:*** The Retailer Files will contain retailer characteristics. These files will be used for the analysis of the retailer survey and the transaction data. FNS will provide Abt the Main Retailer Files and DTA will provide Abt with Supplemental Retailer Files in conjunction with ACS that will have additional HIP information and local contact information.
5. ***Household Survey Data:*** A sample of SNAP HIP participants and SNAP non-HIP participants will be interviewed three times. The survey will include household and respondent characteristics, a fruit and vegetable screener, and questions on food knowledge and attitudes and home food environment.
6. ***Intake Data:*** The 24-hour dietary recall data in Rounds 2 and 3 of the participant survey will be combined with the food and nutrient codes to create the Intake Data.

7. ***Retailer Survey Data:*** The retailer survey will identify retailers' perspectives and the opportunities and challenges they encounter as a result of the pilot. It will also include data on costs of participating in HIP and on fruit and vegetable stocks and prices. The retailer survey will be conducted on a representative sample of store types and multiple employee perspectives will be used to address all required questions.

In Exhibit 6.1 below, we present an overview of the data files. We have outlined the timing of data collection for each data file, the primary owner of the data after data collection, and identifier variables in each data file in order to link files with each other. The primary owner of the data is the entity responsible for acquiring and processing the data. In Section 6.2, we will discuss the linkages and elaborate on identifiers where necessary.

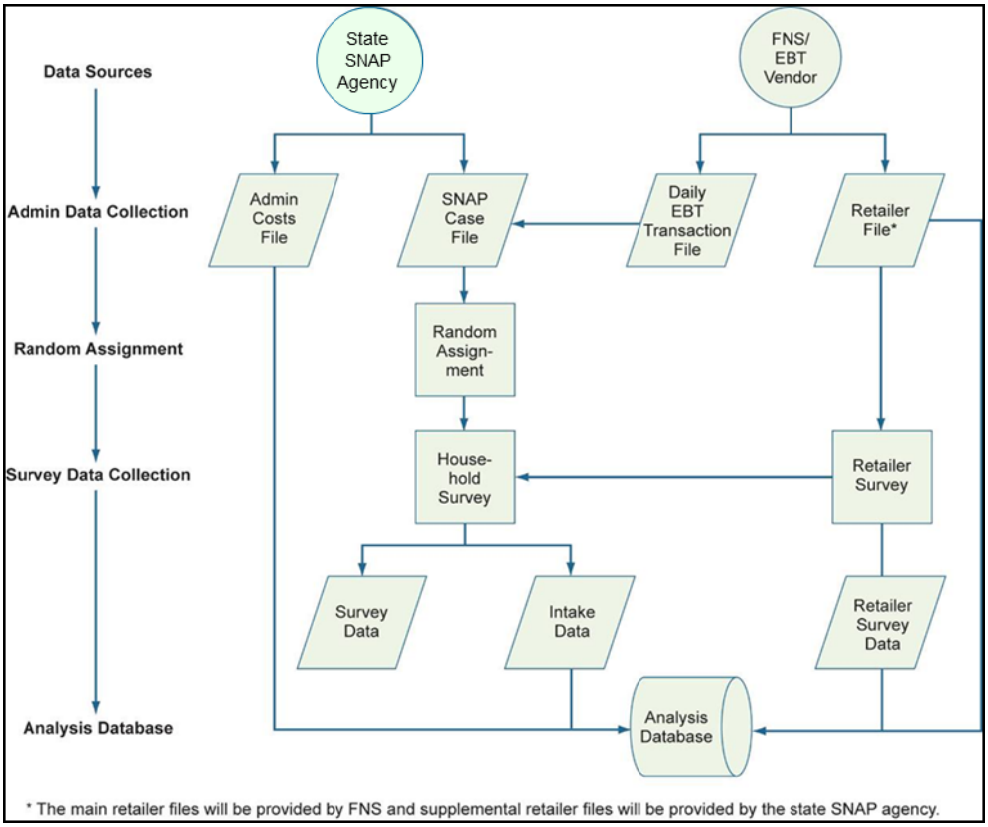
Exhibit 6.1: Data Files Overview

Data	Source	Timing of Collection	Primary Responsibility	Identifiers
SNAP Case Files	DTA	May 2011 (test extract) July 2011 through January 2013 monthly extracts)	Abt	Case and person
Admin Costs Files	DTA	Quarterly until close-out of the HIP grant	Abt	State
Daily EBT Transaction Files	FNS	Daily from September 2011 through February 2013	Abt	EBT card, case and transaction
Retailer Files	FNS and DTA	Monthly from August 2011 through February 2013	Abt	Retailer
Household Survey Data	Household Survey	Prior to implementation, 3 months into operation, and 11 months into operation	Westat	Case and person
Intake Data	Participant Survey and AMPM	3 months into operation, and 11 months into operation	Westat	Case and person
Retailer Data	Retailer Survey	2 months prior to implementation and 11 months into operation	Abt	Retailer

6.2 Data Flow and Creating the Analysis Database

Exhibit 6.2 outlines the flow of data from the primary sources to the creation of an analysis database. However, the discussion excludes qualitative data, such as key informant interview notes and summaries, and demonstration documents.

Exhibit 6.2: Data Flow From Primary Sources to Analysis File



The complete analysis database will be a relational database with multiple SAS data files, each with records at varying levels. While the data from the SNAP Case File will be at the case and person level, the data from the EBT Transaction File will be at the transaction level. The SNAP Case Files will be collected on a monthly basis, and daily EBT Transaction Files with household identifiers and card numbers will be collected.[20] The Household Survey Data will have up to three time points per respondent, and the Retailer Survey Data will have up to two time points per retailer. Thus the analysis database will be longitudinal in nature capturing multiple entries over time in different data files.

As shown in the flowchart, the person-level data from the first SNAP Case File, in conjunction with the Daily EBT Transaction Files, will be used to create a sampling frame and randomly assign persons in SNAP households to HIP. Abt will send the SNAP Case File and random assignment results to Westat so that they can draw the sample of persons in households that will be interviewed at Round 1. These files will be linked by the person identifier. Westat will assign a respondent ID number for survey use. The full SNAP Case Files will be linked with the Daily EBT Transaction Files for analysis using an encrypted identifier that identifies the initial SNAP case household from the sample frame.

[20] The household ID and card number do not have a persistent 1:1 relationship. A household occasionally has more than one active EBT card. When a card is replaced, the new card has a new number.

The SNAP Case File for the most recent month will be used to verify that the survey sample is still eligible for SNAP prior to each onth of the Round 1 survey. The Round 1 survey will be conducted by Westat using CATI or CAPI (for in-person interviews). CATI and CAPI data will be combined to create a standard Round 1 survey data file. Similar CATI files will be created for Rounds 2 and 3. The use of the SNAP Case Files to verify SNAP eligibility of the sample prior to interview will be repeated for Round 2 and Round 3 household surveys. The 24-hour dietary recall data will be collected using USDA's AMPM at Round 2 and Round 3. Nutrient codes will be attached to the AMPM data in SurveyNet to construct the Intake data. The survey data sets will be linked with the SNAP Case Files using the SNAP person identifier. These data sets will be used for the participant descriptive and impact analysis.

The Retailer File with records for retailers participating in the HIP program will be used in conjunction with the EBT Transaction File for the transaction analysis. The Transaction File will identify each EBT card used by the SNAP household. These transactions will be linked to SNAP cases using the case identifier and to the retailer data using the retailer identifier. The Retailer File will be used to sample retailers for the retailer survey. This survey will be conducted 2 months prior to the start of the pilot and 1 year later (after 9 months of operations). This data set will be used to conduct the retailer analysis.

All analysis files created will have flags indicating rounds of data collection and disposition status when necessary. These files will undergo a systematic review process that identifies outliers and tests for internal consistency.

6.3 Participant Data Cleaning and Editing

Abt and Westat will apply rigorous data cleaning methods to the SNAP participant data. First, we will scrutinize the case files provided by the State SNAP client eligibility system offices to ensure they contain the expected variables in a readable format and will compare the contents of multiple SNAP sample frames in the event that we receive more than one. Second, we will implement standard procedures to edit and clean the response data describing the non-food variables; and third, we will administer the USDA Survey Net coding process for dietary recall data.

6.3.1 Sample File Verification

When we receive the sample frame files, we will initiate a series of checks to verify the data. These include the following:

- Evaluate documentation provided by DTA to see if there is a suitable file description, case counts, record layout, and a codebook showing value ranges and missing values.
- Verify the record layout by printing a sample of cases from different parts of the file in SAS.
- Produce frequencies and means to confirm categories and value ranges, and check for missing value indicators.
- Compare expected case counts to the SAS log.

If any of these tasks expose unexpected results, project management will submit a request documenting the problem along with a PDF file of the SAS output. Once the file and record layout have been confirmed, we will proceed with the sampling process. At the completion of sampling, Westat will examine the contact information and the address location for the sampled cases to:

- Identify all cases with missing addresses or elements in the address
- Identify all cases with missing or incomplete telephone numbers

For those cases with missing telephone and address information, Westat will use a variety of tracing mechanisms to locate respondents (these are described in detail earlier). Using the data from these additional sources, Westat will update the sample files.

6.3.2 Edit and Clean Non-AMPM Response Data

All new CATI and CAPI data collection instruments will be programmed in Blaise and will take advantage of a full range of data editing capabilities during the interview, which will help to ensure data integrity. These edits will be specified as part of the development process. This approach has been successfully applied to the AMPM dietary recall instrument where Blaise checks all appropriate edits when an interviewer updates the response to a question, not just those that pertain to the current question. These features result in high-quality data since edit rules are enforced and corrections made as the data are captured. Edits that are typically used in an interview include:

- "Hard" range checks—indicate entries outside of the possible range for the items and normally require the interviewer make a correction before continuing.
- "Soft" range checks—indicate unusual or unlikely situations so that the interviewer may verify them.
- Cross variable checking—triggers hard and soft consistency checks between multiple variables.

Range and consistency edits to be included in the interview will be incorporated into the instrument specifications; each check will be given a reference number that will aid in problem resolution. Exhibit 6.3 presents an example of range and consistency check specifications.

Exhibit 6.3: Example of Edit Specifications

HIP1035	
Item Type:	Check
Item Class:	Soft [Number of years lived in the United States (HIP1030) cannot exceed age calculated from date of birth (HIP1032)]
Item Text:	If HIP1030 = 5 and age based on HIP1032 is less than 21 or
	If HIP1030 = 6 and age based on HIP1032 is less than 31 or
	If HIP1030 = 7 and age based on HIP1032 is less than 41 or
	If HIP1030 = 8 and age based on HIP1032 is less than 51, then display the following edit:
	"I need to verify what I entered. I recorded that you have lived in the United States for {HIP1030 RESPONSE} years, but you are {AGE FROM HIP1032} years old. Which is correct?
HIP1522	
Item Type:	Check
Item Class:	Hard
Item Text:	Trigger hard edit if HIP1515 is greater than 7.
	CANNOT WORK MORE THAN 7 DAYS PER WEEK.

To handle unforeseen situations where edits may not completely cover the situation or extreme out-of-range values are encountered, the interviewer may also enter a comment or remark at any item. These are typically used by the interviewer to handle situations where the response(s) are out of the established hard range or to indicate problems with an item as they are encountered.

Post Data Collection Editing and Cleaning

This cleaning and editing will include:

- List of edits—edits to be performed on each data item
- Specifications—edit specifications for each data item
- Edit checks—response that will follow failed edit checks
- Open-ended items—coding approach including coding specifications for open-ended items
- Reliability—procedures for checking and calculating reliability during data coding

Since range checks, skip edits, and logic edits are built into Blaise instruments, we anticipate the quality of the data will be high. Blaise enforces standard skip patterns and edits, such that missing data are normally the result of "don't know" or "refusal" responses in which the respondent is unable or unwilling to provide a response. However, to verify the level of quality and to detect situations that may not be covered by the edits incorporated into the instrumentation, we will perform automated, post-collection editing using features internal to Blaise in conjunction with Westat's automated data editing systems and procedures. Using this approach preserves the variable ranges, skip patterns, and edits inherent to the interview while also allowing for the addition of edits and re-coding of open-ended responses within the editing process. The editing system also maintains a keystroke audit trail for each case.

Westat's editing system is composed of two levels: a management component and an editing component. The editing supervisor will batch completed interviews to simplify the editing assignment

process and use utilities available to produce reports on editing assignments, editing progress, and batch contents. As cases are completed in the field, they will be added to editing batches and assigned one of the following status codes:

- Clean—indicating no rule violations exist and all questions on the path have responses.
- Suspect—indicating only errors associated with soft edit checks.
- Dirty—indicating that either a hard edit has been violated and not corrected, or that not all of the questions have responses.

This process allows for multiple editing phases to refine the editing process. The first phase is designed to create manageable editing tasks, review data within the construct of the CATI or CAPI interview, and resolve outstanding data issues found in the field during the actual interview. It allows the editor to review the interviewer's comments, to review all suppressed edits, to check for hard edit violations, to review all questions with a response of "don't know" or "refused," to review all open-ended responses, to determine possible errors in response entry, and to identify possible responses that could be coded. The purpose of the second phase is to handle other-specify coding, coding of verbatim text, expanded ranges for unexpected responses, additional consistency edits that may have been too complex to handle during administration of the interview, and resolution of interviewer comments. In phase three, SAS datasets are created from the edited files and installed on the SAS*IntrNet server, and multiple case edits are performed using an automated edit cleaning utility. This web-based product enables editors to select any subset of cases and run cross-tabulations on one, two, or three variables in an interactive mode. The tool enables the editors to search for systematic inconsistencies, outliers, and routing issues without having to make a request to the programming staff. Data anomalies will be recorded and reported to data and project management staff on a weekly basis. Managers will review these reports in a weekly meeting before the application of any systematic changes. Edit problems will be addressed using the Blaise editing features or in SAS. In either event, any updates to the data will be recorded and reported to data management showing edits that have been applied, edits due to be applied and edits awaiting review.

6.3.3 Edit and Clean AMPM Data

In conjunction with the USDA Agricultural Research Service (ARS), Westat has well-established procedures for reviewing, consolidating, editing, and coding AMPM dietary recall responses. The dietary data processing tasks will take place outside of the Blaise editing process described above. This method has been applied to almost 10 years of NHANES collections, as well as other studies using the AMPM. During the data collection field period, Westat will receive Blaise data from the field on a daily basis; we will prepare and send cases in batches to USDA/ARS for processing. The USDA Post-Interview Processing System (PIPS) will be used to convert the Blaise files to MS Access and apply food codes to about 60 percent of the foods. Once the file has been converted, USDA/ARS will return the file to Westat's dietary coding shop, where the contents will be confirmed and tracked. Dietary coders will assign codes to foods that were not automatically coded during the PIPS process. Trained dietary coders will use standard procedures and established guidelines to code the intake data in the USDA SurveyNet coding system through a multiuser tracking system that assigns cases and enforces workflow rules to manage the internal editing and nutritional coding and third-party validation.

In our experience with the AMPM and SurveyNet, there are very few foods that cannot be coded using the USDA Food and Nutrient Database for Dietary studies (FNDDS). This is because most new foods on the market are derivations of existing foods (e.g., new cereals, juice drinks, power bars, etc.). When a new food name cannot be found in the FNDDS, a similar food with the same nutrient profile can typically be matched to the new food. Therefore, these foods can be coded using existing food codes. This process involves researching the nutrient content of the new food (by searching the USDA Survey Database, manufacturer's websites, websites of other databases) and assigning an acceptable food code on the basis of macronutrients (kcal, protein, carbohydrate, fat) and selected micronutrients (e.g., carotenes, vitamin C). The match is considered acceptable if the nutrients are within 10 percent. We expect that acceptable matches can be determined for almost all new foods. In instances when an acceptable match cannot be determined, we will assign the closest available food code and flag it as needing modification. In the nutrient output file, we will replace the nutrients for the flagged food item with the nutrients from USDA's database or a comparable nutrient database.[21] For these modified food codes, we will also review the My Pyramid values assigned to the original food code and determine if any servings need to be changed.

The coding shop will hold regular coding meetings to answer questions, resolve problems, and disseminate information to the coding team. Quality control procedures require that all coders maintain 98 percent accuracy based on independent coding of 10 percent of their work. The coding supervisor will review, document, and adjudicate discrepancies.

Additional steps to ensure the quality of the dietary data include significant review and editing efforts by senior coding and/or supervisory staff. Senior staff will conduct a 100 percent review of coded data, examining the data line-by-line to verify the accuracy of the work. Supervisory staff are responsible for the final quality control edits and reconciling intake data that are outside normal limits. Reports used to identify outlier values include the portion outlier report, which highlights unusually large or small portion sizes. This step is of key importance to HIP pilot study. The process also generates SAS frequencies for selected nutrients and can be limited to MTFV.

6.3.4 Focus Group Data Cleaning

The focus groups will be transcribed as near verbatim as possible with the intended effect of allowing others to "hear" exactly what transpired. Westat works with several small businesses that quickly and accurately transcribe English and Spanish-speaking focus group audiotapes. As soon as a focus group session is transcribed, the moderator will read through the transcript for that session to fill in any gaps. Once the transcription is final, the data will be analyzed as described in Section 7.2.2.

6.4 Participant Survey Weighting Procedures

To prepare for analysis of participant survey data, the Abt team will develop sampling weights. The required weights will be computed at the end of each data collection round. Sampling weights will be

[21] Foods for which an acceptable match cannot be found could be sent to USDA Food Survey Research Group (FSRG) for review before final processing of the data. We have not, however, budgeted any costs for this service.

computed for the completed cases in the sample. In general, weights are needed to compensate for differential probabilities of selection and non-response. Based on the required analyses, we will develop two types of weights for each of the three rounds of interviews: person-level weights and household-level weights. In addition, a set of longitudinal weights will be developed to analyze change between rounds for those respondents who completed intake interviews at both Rounds 2 and 3.

Person Weights

The base weight for a sampled person equals the reciprocal of the probability of selecting that person from sampling frames derived from SNAP administrative files (see Section 2.2). Up to non-response, the sample will be self-weighting for analyses of individuals (e.g., the impact on dietary intake), and we will construct appropriate weights for case-level outcomes.

Next, the base weights will be adjusted for non-response within adjustment cells that are expected to be homogeneous with respect to response propensity. We will conduct a non-response bias analysis to determine characteristics that are correlated with non-response (defined below). For the Round 1 survey, the non-response adjustment cells will be defined using both household-level and person-level characteristics that are available from SNAP administrative files and use those characteristics to form adjustment cells. At the case level, candidate variables for defining cells include: case size, race/ethnicity of head, gender of head, age of head, concurrent receipt of TANF, employment status. At the individual level, candidate variables for defining cells include: gender, age, race/ethnicity. Within these cells, a weighted response rate will be computed and applied to the person base weights (i.e., we will multiple by the inverse of the probability of response) to obtain the corresponding baseline non-response adjusted weights. These weights will then be adjusted to account for subsampling prior to fielding in Round 2 (see the discussion in Section 3.1.3), and carried over as the "base weights" for Round 2 since only interviews with Round 1 respondents will be attempted in Round 2. To compensate for Round 2 non-response, data from the Round 1 survey, in addition to SNAP administrative data, will be used to construct appropriate non-response adjustment cells. Finally, the non-response adjusted weights for Round 2 (which correct for subsampling at Round 2 and non-response at Round 1 and Round 2) will be carried over as the base weights for Round 2 respondents as they enter Round 3, and adjusted for subsampling prior to fielding of Round 3 (again see the discussion in Section 3.1.3) non-response in Round 3 using SNAP data collected in Round 2 to form the required adjustment cells.

Household Weights

At each of the three rounds of data collection, household-level data will be collected from a household member knowledgeable about food purchases and other household characteristics. The sample arising from the proposed sample design is not self-weighting *for household level measures* because large households (defined for sampling purposes by the number of persons 16 years of age or older in the household) will have higher probabilities of selection than small households. As described above for the person-level weights, household weights will be produced in a series of steps. The first step will be to create a base weight that reflects the probability of selecting the sampled household. For example, assuming that all household members are sampled at the same rate, the corresponding household weight would be inversely proportional to the number of persons 16 years or older in the household. Next, the household base weights will be adjusted for non-response within

adjustment cells that are expected to be homogeneous with respect to response propensity. For the Round 1 survey, the non-response adjustment cells will be defined using household-level characteristics that are available from SNAP administrative files. For Round 2 non-response, SNAP administrative data and Round 1 survey data will be used to construct the non-response adjustment cells. Finally, data from the SNAP administrative data, Round 1, and Round 2 will be used to form adjustment cells for Round 3 non-response.

Longitudinal Person Weights

To facilitate estimation of change in outcomes among those persons who responded in both Rounds 2 and 3 (see Exhibit 3.2), a separate set of (person-level) "longitudinal" weights will be provided. These weights will include an adjustment to compensate for the loss of persons who responded in Round 2, and were still eligible (in SNAP), but did not respond in Round 3; and vice versa.

Replicate Weights for Variance Estimation

In addition to the full sample weights described above, a series of jackknife replicate weights will be created and attached to each data record for variance estimate purposes. Replication methods provide a relatively simple and robust approach to estimating sampling variances for complex survey data (Rust & Rao, 1996). Under the proposed replication approach, 100 jackknife replicates will be formed by deleting selected cases from the full sample and adjusting the weights of the retained cases accordingly. The entire weighting process developed for the full sample will then be applied separately to each jackknife replicate, resulting in a series of replicate weights. The replicate weights can be imported into variance estimation software (e.g., SAS, SUDDAAN, WESVAR) to calculate standard errors of the survey-based estimates.

In addition to the replicate weights, stratum and unit codes will also be provided in the data files to permit calculation of standard errors using Taylor series approximations if desired. Note that while replication and Taylor series methods often produce similar results, jackknife replication has some advantages in reflecting statistical adjustments used in weighting such as non-response.

6.5 Retailer Survey Data Entry, Cleaning, and Editing

Data Coder Training

Data coders will be trained to process the HIP retailer survey data after an adequate number of survey forms have been received. Prior to training, a coding supervisor will work with the evaluation team to develop all coding rules. The coding supervisor will then code a test batch of survey forms to become thoroughly familiar with the survey instrument and to ensure that the coding instructions are clear and complete.

At the training, coders will be briefed on the purposes of the survey and will be provided with a "coder version" of the survey form as well as a detailed list of codes and code definitions. The coder version of the survey form details the instructions for the allowable values, field formats, and any special instructions on a question-by-question basis. Each of the surveys will include some open-ended response questions, and may include questions that use an "other, please specify" response. Coders will be briefed on the intent of each such question and trained on how to code these responses.

Data Entry Quality Control

Once paper instruments are edited and coded and the required level of quality control achieved, the surveys data will be double-entered in an Access database and 100 percent verified. All discrepancies identified in this process will be examined and reconciled.

Upon completion of data entry, we will run Access reports to check for logical inconsistencies and errors. As the final step of survey data processing and quality control, the data entry summary report will be compared to the data receipt database to ensure that data for all instruments that were received have been entered and that no duplicate records were created.

Weighting

In order to provide unbiased estimates for the population of participating retailers, tabulations of the retailer survey will use weights based on the sampling design. The retailer weights will take into account each retailer's probability of selection for the survey and the rate of non-response for each stratum. For stores selected within a chain, the weight will take into account the sampling probability of the chain and the store's sampling probability within the chain. This weighting approach will eliminate the bias that might otherwise be introduced by the sampling design and any differential rates of non-response across strata. Calculations of sampling variances will also take into account the sampling design as well as the proportion of the population sampled in each stratum (i.e., a finite population correction will be applied). Our preliminary calculations suggest that, if the population of participating retailers is as assumed in Exhibit 3.3, an estimate of 50 percent for all retailers will have a 90 percent confidence interval of plus or minus 8.4 percentage points. While we originally considered the idea of taking redemptions into account in weighting (i.e., giving larger retailers more weight), this strategy is not appropriate because our objective is to present results for the average retailer. Due to the small size of the non-participating retailer sample, we do not plan to use weights in the analysis, and the tabulations will be limited to frequencies of responses among the respondents.

6.6 Key Informant and Cost Data

All of the stakeholder interview data will be stored in an Access database. The team members will enter data from their interviews directly into electronic forms and will audio-tape interviews (with the interviewee's consent) as back-up to the interview notes. Interview notes will be reviewed by both members of the data collection team for accuracy. If the two staff differ in their understanding of the respondent's answers, the senior data collector will re-contact the respondent and obtain clarification. Senior staff members of the interview teams will prepare summaries of the highlights and key themes of each interview.

Cost data from DTA reports and time sheets, contractor data, the retailer survey, and key informant interviews will be entered into a master cost database. The cost database will include the following data elements:

1. Unique cost identifier
2. Amount of the cost in dollars

3. Date cost was entered into database
4. Date cost was most recently updated
5. Team member who entered cost into database
6. Information source (a categorical variable identifying the original provider of cost information)
7. Time period to which the cost applies
8. Stakeholder (a categorical variable identifying the unit to which the cost applies).
9. Stakeholder type (State headquarters, local office, retailer, CBO, TPP, ACS, Nova Dia)
10. Broad cost frequency type (one-time implementation costs or ongoing operational costs)
11. Detailed cost frequency type (once for the State, once per large retailer, medium retailer, small retailer, once per TPP, once county-wide, once per SNAP case, ongoing monthly costs for the State, ongoing monthly costs for the county, ongoing monthly costs per large, medium or small retailer, ongoing monthly costs per SNAP case)
12. Functional activity:
    a. Develop, design, test and operate payment processes
    b. Household recruiting and customer service
    c. Retailer recruiting and relations
    d. Community relations
    e. Training (clients, retailers, CBOs, etc.)
    f. General and administration
    g. Evaluation support
13. Cost category (salaries, wages and employee benefits, contractors, fees, supplies, other direct costs, and indirect/overhead)
14. Source of funds (HIP grant, other Federal funds, non-Federal funds, or private funds)
15. Non-duplication field. This field will indicate whether the cost duplicates any other cost entries in the database (as when two key informant interviews provided information about the same unit during the same time period).
16. Notes and documentation to help in later interpretation of the cost. This includes whether the cost was one-time only, or whether it would be incurred again in future implementations.

The cost database will be used to generate tables for the cost analysis (Section 7.4) and the feasibility of expansion (Section 7.5).

6.7 Creating Data Files for FNS and Public Use

At the completion of the data cleaning process for each round of the collection, the Abt team will develop SAS datasets for subsequent analysis. Direct personal identifiers will be maintained in separate databases. Full names, addresses and telephone numbers will be removed from the response and study management data before response cases are made available to data editing staff. Tracing information and alternate contact information collected in the Round 1 and Round 2 surveys will be removed from the datasets as well.

We will construct restricted-use data files and public-use data files that contain all data collected during the HIP study. Ideally, both types of files will exclude all direct personal identifiers such as respondent name, address, and date of birth and only include randomly assigned identifiers. However, we will review with FNS the requirements for the restricted-use data files. The public-use files will be created from the restricted-use files and will undergo the process described below to ensure that the data do not pose a risk of individual identification. In creating the public-use files that include participant data, we will comply with all relevant Federal statues regarding personally identifiable data. In addition to removing all personally identifiable information (PII), we will also apply a three-step approach for protecting the confidentiality of study participants:

Remove all individual identifiers and contact information. The files will exclude any identifying information used by the study team for tracking sample members, such as name, address, or telephone number.

Check for identifying demographic information. To guard against the possibility of identifying study participants from demographic information reported in the files, we will check that key demographic variables such as age, gender, race/ethnicity, and education level have at least five sample members in each response category. Any category with fewer than five sample members will be flagged for recoding. We will conduct similar checks for other potentially identifying variables, as well as combinations of demographic variables—for example, gender crossed by age and race/ethnicity.

Recode all identifying variables. If our checks uncover any variables with fewer than five sample members per response category, we will recode the variables to create larger categories. For example, we may recode narrow one-year age categories into broader five-year groups. We may also bottom- or top-code variables such as family income. For some variables, these recoding procedures may not be sufficient to protect the confidentiality of study participants, in which case we will either drop variables from the files or back code the information into numeric variables that better protect respondent confidentiality.

We will review these assumptions with FNS to ensure that the data files reflect USDA and IRB standards for restricted-use and public-use files and make changes requested by the COR or TWG. Data-sharing agreements with the State SNAP agency and the EBT contractor may impose further restrictions on the data elements that can be shared with FNS. In particular, there may be concern that retailer or transaction data could be linked with administrative data available to FNS through its data systems, leading to the potential for deductive disclosure. If these concerns are written into the agreement, we will also suppress or recode additional information such as time of EBT transactions.

Abt and Westat will create draft and final versions of restricted-use files containing raw and recoded variables for the interim report (November 6, 2012) and the final report (September 17, 2013). These deliverables will include files from SNAP case records, the participant survey, the retailer survey, EBT transactions, EBT and cost reports, and qualitative data (e.g. information collected from the key informant survey) in coded form. In addition, analysis files created by linking these data will be included (such as the EBT transaction analysis file). Restricted use files will be encrypted for transmission; passwords will be sent separately to FNS by a different method.

Abt and Westat will prepare a full set of corresponding documentation for the restricted-use data. While we will review with FNS the exact documentation features desired, we anticipate delivering, at a minimum:

- A user's guide that will provide a brief overview of the study design and a brief description of the data collection procedures, sample sizes, response rates, treatment or imputation of missing data, and data editing procedures. We will also provide descriptions of SAS programs used for variable creation and analysis.
- Description of the data files that includes the relationships of the analysis database and the structure of data files. This document will include data set name, number of records, SAS version, and a summary description.
- A codebook that includes information on each variable in each data file. The variable information will include variable name, variable label, variable length, variable type, variable format, allowed values, subsample for which collected, missing values, and value label. In some cases, we will also provide frequencies and/or means for the data items. If a variable in a survey data file is a response to a specific question, we will also include the question number, specific round and the wording of the question.

We will use software tools to automate data and documentation tasks which would otherwise be labor-intensive, time-consuming, and prone to human error. Among them is a tool that generates frequencies and means while automatically collapsing frequency counts for data value ranges and excluding missing data codes from means. Abt and Westat have successfully and efficiently prepared hundreds of documentation files.

We will create draft public-use files after the creation of the restricted use files (9/17/2013) and a final version after FNS review (December 31, 2013). The public-use file will contain data from all three rounds of the HIP evaluation. Since we expect to capture the same measures in more than one round of the data collection, SAS variable names will identify the round for all raw and analysis variables. This procedure will simplify analysis specifications, user documentation, and the creation of longitudinal files. Full documentation will also be provided with the public-use data.

7. Analysis

This section discusses the analyses required to respond to the evaluation's five research objectives. The analyses presented below are organized along five lines: HIP impact on SNAP participants (Section 7.1), stakeholder experiences (Section 7.2), implementation processes (Section 7.3), costs of HIP (Section 7.4), and the feasibility of nationwide expansion (Section 7.5).

7.1 HIP Impact on SNAP Participants

The goal of HIP is to change the food intake of SNAP participants; more fruits and vegetables, less of other foods. This section discusses how we will estimate HIP's impact on food intake. Specifically, the impact analysis will address:

- Impacts of HIP on fruit and vegetable consumption and other dietary intake measures by SNAP participants
- Factors that influence how HIP impacted participants, including regression analysis and subgroup comparisons based on demographic characteristics, Round 1 attitudes and barriers

Although it is not an experimental impact, the analysis of HIP participants will also address:

- Households' use of their HIP incentives

Longer-term outcomes that will not be addressed, because they are too expensive to study and beyond the scope of this evaluation, include HIP impacts on body weight and chronic disease.

Most analyses will use information from the 24-hour recall and household data collection instruments for HIP and non-HIP cases in the participant survey. Analyses of incentive claiming—the receipt of HIP incentives earned through purchases of TFVs—will use EBT data for the full HIP treatment group (rather than only the smaller sample that will be interviewed).

7.1.1 Description of Household and Participant Characteristics

To provide context, we will first describe the population participating in the HIP and the non-HIP groups. This background analysis will: (a) describe the household and respondent characteristics of SNAP participants in the pilot site, (b) compare these statistics to available measures for State and national SNAP populations; and (c) verify that randomization was correctly implemented.

Specifically, we will use both State SNAP administrative files and the evaluation's Round 1 (baseline) survey of the HIP and non-HIP groups to generate descriptive tabulations of the characteristics of the HIP population. Characteristics to be tabulated will include household composition, demographic characteristics of the Round 1 survey respondent and head of household, number of adults employed, and participation in other assistance programs.

7.1.2 Main Impact Analysis

The single most important goal of the HIP evaluation is to estimate the causal impact of HIP on fruit and vegetable consumption. As Section 7.1.5 discusses, it is good statistical practice to identify the main "confirmatory" outcome in advance.[22] For the HIP evaluation, we take the confirmatory outcome to be the HIP/non-HIP difference in the modified target fruit and vegetable (MTFV) intake, averaged over Rounds 2 and 3 of the participant survey, with regression adjustment for selected control variables. Exhibit 7.1 provides a table shell for the confirmatory outcome and several exploratory outcomes.

This definition of the confirmatory outcome distinguishes between the Targeted Fruits and Vegetables (TFV) that are eligible for the financial incentive and the Modified Targeted Fruits and Vegetables (MTFV) that can be measured using the 24-hour recall instrument.

- TFVs, eligible for the financial incentive, are the same foods that are eligible for WIC fruit and vegetable vouchers. These foods include fresh, canned, frozen, and dried fruits and vegetables without added sugars, fats, oils. Fruits may not have added salt; vegetables may be regular or lower sodium. Fruit juices and white potatoes are excluded, but yams and sweet potatoes are included. The class of foods eligible for HIP also excludes food-away-from-home and hot food served ready to eat.
- MTFV is identical to TFV except that it does not incorporate the restriction against added sugars, fats, oils, and salt. We make this modification because the 24-hour recall instrument cannot always identify whether such ingredients were included in a purchased product or added later as part of a recipe. A disadvantage of using MTFV is that the financial incentives in HIP presumably reduce consumption of sugary, fatty, oily, and salty fruits and vegetables that do not count toward the incentive bonus, even as they increase consumption of TFVs. The advantage of using MTFV is that it allows us to use the AMPM 24-hour recall instrument specified in the RFP. Modifying this instrument and then validating the modifications were deemed beyond the scope of this study.

In addition to the primary outcome, the evaluation will provide exploratory evidence of impacts on other outcomes, in three broad classes.

[22] This practice avoids the well-known statistical problem of multiple comparisons. For each hypothesis test, a conventional approach allows a 5 percent chance of incorrectly rejecting the null hypothesis (a Type I error) and concluding that an impact has occurred where none has. If the analyst runs multiple hypothesis tests using this same conventional approach, the risk of Type I error accumulates. Hence, it is recommended to identify a single confirmatory outcome ahead of time (Schochet, 2008).

Exhibit 7.1: Estimated Effect of HIP on Fruit and Vegetable Outcomes

Fruit And Vegetable Measure	HIP Group Mean	Non-HIP Group Mean	Estimated Impact Of HIP
	mean (se)		
Average Across Round 2 and Round 3			
Modified target fruits and vegetables (MTFV)			
Other fruits and vegetables (cups)			
All fruits and vegetables (cups)			
Target fruits (cups)			
Other fruits (cups)			
All fruits (cups)			
Target vegetables (cups)			
… <list continues>			
Round 2			
Modified target fruits and vegetables (MTFV)			
Other fruits and vegetables (cups)			
All fruits and vegetables (cups)			
Target fruits (cups)			
Other fruits (cups)			
All fruits (cups)			
Target vegetables (cups)			
… <list continues>			
Round 3			
Modified target fruits and vegetables (MTFV)			
Other fruits and vegetables (cups)			
All fruits and vegetables (cups)			
Target fruits (cups)			
Other fruits (cups)			
All fruits (cups)			
Target vegetables (cups)			
… <list continues>			
Change from Round 2 to Round 3			
Modified target fruits and vegetables (MTFV)			
Other fruits and vegetables (cups)			
All fruits and vegetables (cups)			
Target fruits (cups)			
Other fruits (cups)			
All fruits (cups)			
Target vegetables (cups)			
… <list continues>			

Note: Means are weighted and regression adjusted. Standard errors are corrected for complex survey design
** Statistically significant difference, $p < 0.01$
* Statistically significant difference, $p < 0.05$

The first class of exploratory outcomes will be based on survey responses to the 24-hour recall instrument:

- Cups of all fruits and vegetables (whether from retailers or restaurants), fruits alone, vegetables alone
- Cups of all target fruits and vegetables, target fruits alone, target vegetables alone
- Meeting the DGA recommendations for fruits and vegetables, for fruits alone, for vegetables alone, and for subgroups of vegetables
- Cups and percentage of appropriate recommendations for other food aggregates, including grains, meat and beans, dairy, and discretionary foods
- Total food energy (kilocalories) and percentage of food energy from the food groups listed above
- Cups of fruits and vegetables by form of preservation (fresh, frozen, canned, and dried)
- Cups of fruits and vegetables by subcategory (dark green vegetables, citrus fruits)
- Cups of selected leading fruits and vegetables with highest frequency of intake (apples, tomatoes)
- Quantities of other ingredients in foods with fruits and vegetables (salt, sugar)
- Other secondary nutrition measures that are expected to be related to fruit and vegetable intake, including Healthy Eating Index (HEI) scores, fiber, beta carotene, vitamin A, and vitamin C
- Threshold measures of consuming fruits and/and vegetables during the 24-hour reference period (none, up to 1 cup, 1 cup or more)
- Consuming any fruits and vegetables by form of preservation (fresh, frozen, canned, and dried)
- Consuming any fruits and vegetables by subcategory (dark green vegetables, citrus fruits)
- Variety of fruits and vegetables consumed.

The second class of exploratory outcomes will be based on the Fruit and Vegetable Screener (FVS). From this source, impacts on the frequency and amount of consumption in the past 30 days will be computed for selected foods, including:

- 100 percent juice
- Fruit
- Salad
- Fried potatoes
- Other potatoes
- Beans
- Other vegetables.

The third class of exploratory outcomes will be based on survey responses by the sample person or the household's primary shopper to questions about other outcomes, perceptions, and experiences:

- Expenditures on food
- Expenditures on fruit and vegetables
- Purchasing more fruits and vegetables than previously (self-assessment)
- Trying new fruits and vegetables
- Finding fruits and vegetables to be affordable.

In addition to cross-sectional comparisons of the HIP and non-HIP samples, we will estimate the change in consumption between the Round 2 and Round 3 interviews. To do so, we will use the longitudinal sample of those who complete both of these interviews. The outcome of interest is change in consumption in the HIP group relative to change in consumption in the non-HIP group (see Exhibit 7.1, bottom panel).

7.1.3 Multivariate Models

As noted in the previous section, our primary estimates of HIP's impact will be regression adjusted rather than a simple difference in means of HIP versus non-HIP. Regression adjustment improves the comparability of the HIP and non-HIP groups and increases the precision of our estimators. We will use a model of the form:

(1) $y = \beta_0 + \beta_1 HIP + \beta_2 ControlVars + u$.

where y is an outcome of interest, HIP is a binary variable that identifies the treatment group, and *ControlVars* is a vector of characteristics measured as of the Round 1 (baseline) survey or at baseline from administrative data. *ControlVars* will include individual demographics (own age, own gender) household demographics (number of adults, number of children, age of oldest adult), income and earnings, size of SNAP benefit, and measures related to baseline consumption according to the Fruit and Vegetable Screener. These control variables were selected as likely to be strong predictors of the confirmatory outcome. The coefficient on the HIP indicator gives the treatment effect.

For binary outcomes, we will use the corresponding logistic regression model:

(2) $\ln[\text{odds}(y=1)] = \beta_0 + \beta_1 HIP + \beta_2 ControlVars$,

where odds(y=1) indicates the odds that the binary outcome happened, and ln is the natural logarithm operator. As a second functional form for binary outcomes, we will use the linear probability model (LPM) in addition to logistic regression.

7.1.4 Survey Weights, Standard Errors, and Confidence Intervals

All analyses of survey data will use person-level or household-level survey weights depending on the unit of analysis, so that sample estimates provide unbiased estimates of the corresponding population statistics (for those who consent and complete the Round 1 survey) in the pilot site. Standard errors

and confidence intervals will correctly take account of the survey design, using the replicate sampling weights. All standard error estimates will be robust to heteroscedasticity.[23]

Our field methods—in particular, large incentives and intensive tracking—are designed to achieve the maximum possible response rate (we expect to achieve greater than an 80 percent response rate). Nonetheless, some non-response is inevitable, and such non-response might affect HIP and non-HIP groups differentially.

We will follow conventional approaches in dealing with this potential threat to internal validity, modeling non-response and constructing weights to force respondents to align with the unattrited sample in terms of observed characteristics (see Section 6.4). Note that we have a "rich list sample" (see the discussion in Martorell, Klerman, & Loughran, 2010)—from the SNAP administrative data we know a considerable amount about the population, including non-respondents. That information includes demographics (e.g., gender, age, race/ethnicity, family structure, household composition), and also income and earnings, history of SNAP receipt (duration and benefit amount), and some information about usual fruit and vegetable consumption (from the Fruit and Vegetable Screener). We will proceed by modeling response as a function of the rich list information.

7.1.5 Multiple Comparisons

As noted earlier, having a large number of hypothesis tests creates a danger of finding "false positives," seemingly significant impacts when in fact the true impact of HIP is zero. To address this multiple-comparisons problem, we have specified one "confirmatory" outcome: the HIP/non-HIP difference in average MTFV intake, based on data from rounds 2 and 3 of the participant survey, using regression adjustment for control variables (see Section 7.1.2).

If the main confirmatory HIP/non-HIP difference is statistically significant at the 5 percent level, we will use the conventional approach to testing HIP/non-HIP differences for all outcomes and subgroups. In presenting results, we will describe analyses other than the main confirmatory outcome as "exploratory," pointing out that occasional "significant" differences could appear simply due to sampling variation in multiple hypothesis tests.

If the main confirmatory HIP/non-HIP difference is statistically insignificant at the 5 percent level, we will still use the conventional approach to testing differences for all outcomes and subgroups, but the accompanying discussion will warn that seemingly significant differences for particular outcomes and subgroups could be spurious. As before, the discussion will describe the analysis of these other outcomes and subgroups as exploratory. The Executive Summary and other summary documents will simply report that the HIP evaluation found no significant impact on the main outcome and not mention any of the exploratory results.

[23] Randomization does not guarantee homoscedasticity, and the linear probability model on binary outcomes induces heteroscedasticity.

7.1.6 Subgroup Analyses

To maximize statistical power, subgroup analyses will be done on average consumption from Round 2 and Round 3.[24] The subgroup analyses are exploratory. Our subgroup analyses will consider three sets of subgroups:

- Demographic subgroups based on age (youth [16-17], adults [18-59], elderly [60+]); gender; and employment status (employed, unemployed and looking for work, not in labor force, and not looking for work).
- Subgroups that are more likely or less likely to receive the maximum HIP incentive for which they are eligible. A household may reach its maximum possible incentive either by spending all SNAP benefits on TFVs or by reaching the $60 monthly cap on incentives. According to economic theory, households that reach the maximum possible incentive could in principle respond differently from other households (see Appendix C).[25]
- Subgroups defined by Round 1 attitudes and barriers to fruit and vegetable consumption. The Round 1 survey includes questions about enjoying fruits and vegetables, exposure to nutrition education, the belief that fruits and vegetables are healthy, enjoyment of trying new foods, lack of ability to store fresh fruits and vegetables, and transportation difficulty in accessing retailers with fruits and vegetables. Separate scales/scores will be computed for preferences and barriers; these will incorporate all relevant survey questions on preferences and barriers to purchasing and consuming fruits and vegetables at Round 1. Rounds 2 and 3 respondent preferences and barriers to fruit and vegetable consumption will be examined as secondary outcomes, whereby we will examine if HIP influenced preferences and barriers.

Each of these subgroups is defined using only Round 1 characteristics, so there is no endogenous selection.

The subgroup analyses will consider the following outcomes from the longer list in Section 7.1.2: MTFV intake, all fruit and vegetable intake, fruit intake, and vegetable intake (from the 24-hour recall instrument).

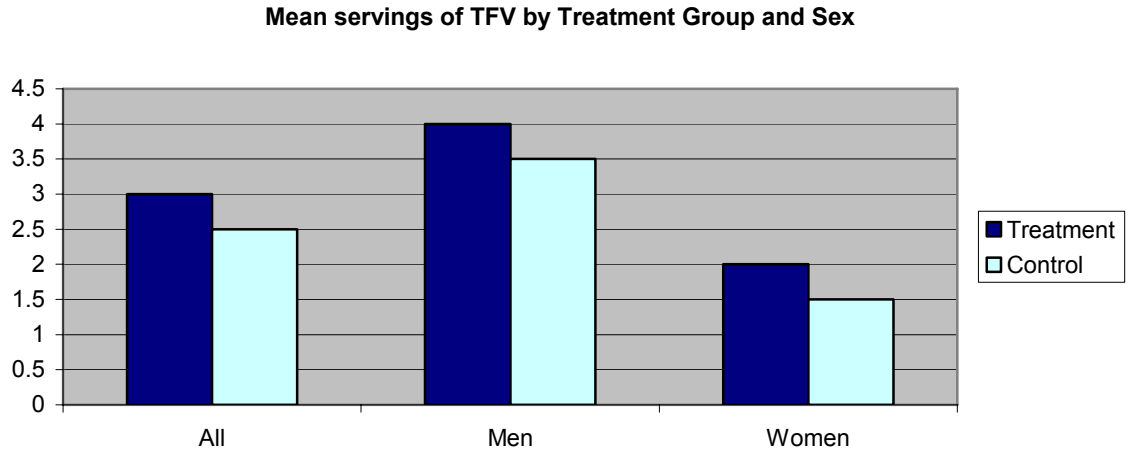
For selected outcomes and subgroups, the Final Report will display vertical bar charts showing the main results. These charts will use a simple design and a consistent color scheme. Adjacent bars will compare outcomes for the HIP and non-HIP groups. Whiskers on each bar will indicate confidence intervals and stars will note statistical significance of the HIP/non-HIP difference. Pairs of vertical bars will distinguish results for the HIP (treatment) and non-HIP (control) samples by subgroups. We

[24] Averaging across the two rounds will increase statistical power, which is particularly important in subgroup analyses.

[25] We will use two approaches to identify households that are "likely to receive the maximum HIP incentive for which they are eligible": (a) based on their MTFV expenditures in Round 1, and (b) using a model of HIP incentives received, estimated using EBT records for the HIP households that were not included in the survey sample. Under both approaches, this subgroup analysis will measure HIP/non-HIP differences in MTFV intake for households that were more or less likely to receive the maximum HIP incentive for which they are eligible.

will test for differences in impact magnitude between subgroups. See Exhibit 7.2 for an example of the layout.

Exhibit 7.2: Sample Bar Chart



7.1.7 Estimation of Usual Intakes

As explained in Section 3.1, there will be two 24-hour dietary recalls in each round (Rounds 2 and 3) in a 10 percent subsample of the population. Based on within-respondent variances estimated using this subsample, we will be able to obtain estimates of the usual intake of foods, food groups, and nutrients and calculate HIP's impact on these outcomes. It is also of interest to describe the proportion of the participants in the HIP and non-HIP groups who consumed less than a particular number of cups of fruits and vegetables and other foods and nutrients.

We will answer these questions using a statistical method developed by researchers at the National Cancer Institute (NCI), the USDA, and elsewhere, termed the NCI method (Tooze et al., 2006). The NCI method is a unified framework for estimation of usual intake of nutrients and foods. In the remainder of this section, we provide additional background on the rationale for estimating impacts on usual intake rather than single-day intake, and describe specific procedures to be employed in calculating usual intake using the NCI method.

A single 24-hour dietary recall measures consumption at one point in time. Individual consumption varies substantially from day to day. Intake estimates calculated based on a single day of recall data do not therefore accurately represent long-term average intake for that individual, also known as "usual intake." Although single-day recall estimates the mean without bias, when substantial within-person variation is included in the estimate of usual intake, the estimates of the percentage of a group below or above a certain cutoff are biased.

The mean of multiple days of intake for an individual is a better measure of usual intake than a single day; however, it is often not practical to collect more than one day of intake in each round on the entire sample. Doing so would either dramatically increase the cost of the study or—to offset the added cost—reduce the sample size and power to detect differences in intake. Therefore, researchers

have developed statistical methods for estimating usual intake of foods and nutrients for samples in which only a subset of respondents report a second day of recall data.

In order to estimate usual intakes, replicate observations are needed for at least a subset of individuals in the sample. We will collect a second day of intake data for a 10 percent subsample, or 75 individuals in the HIP group and 75 in the non-HIP group in Rounds 2 and 3.

The 10 percent sample who complete a second 24HR will be selected at random at each round, so their ratio of between-person to within-person variation may be used in the estimation of usual intake. However, we will conduct preliminary tabulations to describe any differences between the full sample and the 10 percent sample for which a second dietary recall is collected. In the unlikely event that there are statistically significant differences, we will: (a) revisit the random sampling algorithm to confirm that no errors were made, and (b) include a tabulation of the differences in the final report and warn the reader to treat estimates of usual intake with caution.

Statistical Methods to Estimate Usual Intake

Dodd et al. (2006) provide a useful overview of recent statistical methods for estimating usual intakes of nutrients and foods. Though details vary across methods, in general these statistical approaches share a common analytic framework. Our analysis will proceed using a method developed by staff at the National Cancer Institute (NCI) in collaboration with staff at the USDA Center for Nutrition Policy and Promotion (Tooze et al., 2006). This method appears to be becoming the de facto standard. Of particular relevance to the HIP evaluation, this "NCI method" offers significant advantages for foods not consumed every day by a significant proportion of individuals. Such foods include fruits and vegetables, the primary outcome of interest for this study.

The NCI method models usual intake as the product of the probability of consumption on a given day and the average amount consumed conditional on having consumed a positive amount. Specifically, it takes into account reported zero-consumption days and reported consumption-day amounts that are positively skewed, and distinguishes between within-person and between-person variation in consumption.

Finally, the NCI method allows for efficient estimation of usual intake for subgroups. Instead of stratifying the sample by subpopulation and estimating usual intake separately for each subgroup, covariates defining subgroups are included in the NCI model, such that (difficult to estimate) variance components may be estimated simultaneously for the full sample, and only covariate values differ. For subgroups of respondents comprising a relatively small proportion of the full sample, the efficiency gains from this capability are likely to be relatively substantial.

7.1.8 Analysis of the Incentives Received

If HIP is to have an impact on TFV consumption, it seems necessary (though not necessarily sufficient) that HIP households spend SNAP benefits on TFVs and receive the corresponding incentives. Information about SNAP spending on TFVs and HIP incentives received will be available

in the administrative data for the entire HIP group, not merely for the much smaller HIP survey sample.[26]

From daily EBT transactions data, we will obtain SNAP benefits issued, regular SNAP purchases, HIP purchase transaction amounts, and incentives received. We will conduct analyses of how much incentive households received and the percentage of households that reached the maximum incentive for which they are eligible.

From the periodic extracts of SNAP administrative data, we will obtain demographic characteristics, employment, and income. Subgroup analyses will show how the amount of incentive received, and the percentage receiving the maximum for which they are eligible, differed by demographic type, employment status, income category, and SNAP benefit level.

7.2 Stakeholder Experiences

This analysis addresses how HIP affects the experiences and satisfaction of several stakeholder groups: participants themselves (Sections 7.2.1 and 7.2.2), retailers of different types (Section 7.2.3), and other stakeholders (Section 7.2.4), including Massachusetts DTA (the HIP grantee), the local SNAP agency, EBT vendor/TPPs, and community organizations.

7.2.1 Experiences and Satisfaction as Measured in the Participant Survey Data

The analysis of the first group of stakeholders, participants themselves, will combine data from the participant survey (described in this subsection) and the focus groups (Section 7.2.2).[27]

Analysis of experiences and satisfaction from the participant survey will have three parts.

The first part will assess whether HIP participants changed their shopping patterns in ways that are relevant for fruit and vegetable purchase:

- Did HIP affect the choice of store type where SNAP participants shop?
- Did HIP affect the frequency of shopping, in number of times per week or per month?
- Did HIP affect special efforts to shop for fruits and vegetables?
- Did HIP affect the proportion of respondents who report that they find it hard to eat fruits and vegetables because they are hard to find where the respondents shop?

The second part will address questions about experiences and satisfaction with SNAP:

- Did HIP affect whether SNAP participants had difficulty using the SNAP benefits?
- Did HIP affect whether SNAP participants perceived barriers to the purchase of fruits and vegetables?

[26] However, it is not possible to infer total TFV purchase amounts from the incentive receives, because households may also purchase TFVs using their own cash income.

[27] In addition, the retailer survey provides some information about retailer perceptions of SNAP participant satisfaction with the retail experience of HIP (see Section 7.2.3).

If the HIP participants perceive fewer barriers, that could be interpreted as a favorable or satisfactory experience with HIP. This analysis will have the same design as the experimental impact analysis of the main outcomes (see earlier Section 7.1). It will compute regression-adjusted HIP/non-HIP differences in outcomes from the participant survey related to overall views of the SNAP program and its impact on dietary quality. These outcomes include: where the participant shopped, whether the participant had problems with the EBT card, whether the participant used a SNAP help hotline, and whether the participant perceived cost or transportation or other barriers to purchase of fruits and vegetables.

The third part of the participant survey analysis will determine whether HIP participants themselves report favorable experiences and satisfaction with HIP. This analysis will address experiences and views that are relevant only for the primary HIP sample. This part will tabulate responses to several questions:

- Had the respondent heard about HIP?
- How well was the HIP program explained?
- How easy was it to keep track of which fruits and vegetables are eligible?
- Was the DTA hotline called about HIP problems?
- How did the respondent keep track of HIP incentive amounts used and remaining?
- What was the overall level of satisfaction with the new fruit and vegetable incentive?

To complement the main impact analyses, this part will describe whether HIP participants perceived an increase in their FV intake, and, if so, whether the change was due to the financial incentive or changing attitudes toward and exposure to fruits and vegetables. The analysis will integrate results from the survey with discussion of insights from the focus groups.

7.2.2 Participant Focus Groups

The methodological steps for focus group data involve creating an analytic framework from which to evaluate the data, and indexing the data based on specific research questions. The variables of interest will be based on the focus group questions and the data from the focus group transcripts.

Analytic Framework

We will use a combination of two different frameworks for the data analysis: constant comparative and key concepts. The objective of the constant comparative method is to identify patterns or trends in the data and discover the relationships between ideas or concepts (Krueger & Casey, 2009). The second method, key concepts, helps to identify factors that are of key and moderate importance to the participants and document how participants view a topic.

Indexing Data

The goal is to bring together all extracts of data that are pertinent to a particular theme, topic, or hypothesis (Coffey & Atkinson, 1996). Using the focus group's original moderator as the analyst facilitates this process, since he or she is already familiar with the content of the transcripts. Indexing involves the analyst reading and rereading the text and assigning index codes to the text based on the proposed analytic framework. Initially, the index codes tend to be broad and subgroups form under

these broader groupings. Westat staff will use NVivo, a commercial qualitative data analysis package that can retrieve all text for a particular code.

Analysis
Using the coded data, we will report how participants describe:

- The user-friendliness of the HIP processes
- Expectations of families and level to which HIP met expectations
- Changes in willingness to purchase fruits and vegetables in general
- Changes in willingness to purchase new fruits or vegetables
- New one-time and multi-time purchases of fruits and vegetables since inception of pilot
- Unexpected effects of HIP on eating habits
- Other unexpected outcomes

The descriptive analysis will be developed once the focus group topics have been finalized. The analyses will be organized by the protocol used during the focus groups, the composition of the groups, and their responses. Westat will prepare a draft descriptive report summarizing key themes for each round.

7.2.3 Retailer Impact and Satisfaction Analysis

The retailer impact analysis will address the research questions about the impact of HIP on retailers:

- Did the pilot affect business for participating retailers and non-participating retailers?
- What changes in checkout procedures, stocking of products (including foods stocked, replenishing and ordering practices, and effects on particular products or product types), or other retailer practices were reported as occurring as a consequence of HIP?
- What challenges were observed with checkout procedures during on-site retailer visits?
- What perceptions did retailers have about customer use of the incentives?
- How satisfied were retailers with their ability to implement and participate in the pilot?

For this analysis, we will link and analyze the retailer database (from FNS and DTA), the retailer survey, EBT transaction data, and on-site retailer visits. Transaction data will be used to compute HIP and non-HIP SNAP redemption totals for individual stores and groups of stores. The EBT vendor's retailer ID will be used to link survey and transaction data with retailer characteristics from the retailer database, including store type, chain/independent, type of EBT terminals used (integrated POS, stand-alone EBT, no terminal/paper voucher), HIP participation, method of identifying HIP-eligible items (automated or manual), and location.

Retailer Population Description and HIP Participation

The analysis will begin with a description of the SNAP retailer survey sample and population in the site using the retailer database and transaction data:[28]

- Number of retailers, percentage of all monthly SNAP redemptions in Hampden County, and average monthly redemptions per store, by store type, for the survey sample and the population
- Number of retail chains, percentage of all monthly SNAP redemptions in Hampden County, and average monthly redemptions per chain, by type of retail chain, for the survey sample and the population
- Density of retailers per square mile (overall and for the most common store types)
- Maps showing the distribution of SNAP retailers[29]

Second, we will analyze patterns of retailer participation in HIP, again using the population data from FNS and DTA and the EBT transaction data for both surveyed retailers and the population. Specifically we will tabulate the following statistics for participating and non-participating retailers:

- Number and percentage of SNAP retailers, overall and by store type
- Number and percentage of SNAP retail chains
- Total amount, average amount, and percentage of all SNAP redemptions in Hampden County, overall and by store type
- Total number, average number, and percentage of all SNAP transactions in Hampden County, overall and by store type

General retailer population statistics and comparisons of participating/non-participating retailers will be computed at baseline (before the first round of the retailer survey), for the third month of HIP operations (early implementation) and at the time of the second retailer survey (9 months after HIP goes live).

For retailers that decline to participate in HIP and those that leave HIP, survey data will be used to analyze the reasons for their decisions. The sample of non-participating retailers will be small, and thus these tabulations will simply present counts of retailers giving various reasons for declining or ceasing to participate (cost, lack of interest, lack of information, difficulty of separating eligible items, reconciliation problems, etc.). Use of weights and subgroup analysis will not be appropriate for this small sample. Response rates for the retailer survey will be presented as both counts and

---

[28] In these tabulations, the retailer database and transactions data provide the authoritative description of the retail environment in Hampden County. Redemption totals will be based on the transaction data for the HIP and non-HIP households identified at the time of initial random assignment, and will not include redemptions by newer SNAP households in Hampden County or by out-of-county SNAP households. The retailer survey data are included here merely to assess the representativeness of the survey sample.

[29] For the maps, store locations will be geocoded.

percentages. We will also use the administrative data to tabulate the number and characteristics of the full population of retailers that decline to participate.

Impacts of HIP on SNAP Retailers

The analysis of HIP impacts will address the following questions:

- Did the pilot affect business for participating retailers and non-participating retailers?
- What changes in checkout procedures, stocking of products, or other retailer practices were reported as occurring as a consequence of HIP?

This analysis will rely primarily on the two rounds of data from the retailer survey on the topics listed in Exhibit 4.2. Survey responses will be tabulated by HIP/non-HIP status and store type. The analysis will use weights as described in Section 6.5.

The question of impacts of HIP on retailers' business is particularly critical to the long-run feasibility of HIP. Retailers will face implementation costs to set up processes for HIP and operating costs at the checkout counter and the "back office" (reconciliation). While some implementation costs may be reimbursed with HIP grant funds (provided by FNS to Massachusetts DTA as the HIP grantee), retailers are likely to have uncompensated costs for implementation and operations. In some cases they may also incur costs to stock TFV (more inventory, coolers, etc.). On the other hand, they may experience increases in purchases of TFV, and retailers with good selection of TFV (i.e., supermarkets, superstores, and other stores that increase their TFV inventory) may see a general increase in SNAP redemptions as participants do more of their shopping in locations where they can benefit most from the incentive.

While the transaction data analysis described below will address changes in SNAP redemptions from the retailer perspective, it will not address the bottom-line impact of HIP, for which costs must be taken into account. For the question of bottom-line business impacts, the analysis will rely primarily on survey responses from retailers. While self-reported assessment of the financial impact of HIP will be subject to reporting error, the alternative would be to collect sensitive and burdensome financial data and assume a pre/post analysis framework to calculate causal impacts.

Using transaction data, we will estimate the impact of HIP on SNAP redemptions for each retailer in Hampden County and for retailers by store type and participation status. For each retailer, the impact of HIP on redemptions is the difference between their actual redemptions and what they would have received in the absence of HIP. The difference between what HIP and non-HIP households spend in a given store allows us to estimate this from the experiment, since the behavior of HIP households shows us the average amount *all* households would have spent at the store with HIP, while the behavior of non-HIP households shows us the average amount *all* households would have spent at the store absent HIP. We will scale up this difference up in two steps: first, multiplying by the number of HIP households to estimate the total impact on the store at the scale of the demonstration, and second, multiplying by the total number of households in Hampden County to project the impact of county-wide implementation. In the analysis, we will sum the total impact across retailers in each relevant

subgroup (by store type and participating/non-participating). This method will allow us to estimate the magnitude of the HIP impact on redemptions in each retail subgroup.[30]

The design of the retailer survey offers the opportunity for an exploratory analysis of changes in food prices. The two rounds of the survey (described in Section 3.3) will collect prices for a market basket of commonly purchased target fruits and vegetables (TFV), such as apples, bananas, carrots, and tomatoes. Round 1 will provide the baseline before the HIP incentive becomes available. Round 3 (so named because it corresponds in time with Round 3 of the participant survey) will provide the post-intervention data. The statistical significance of the change in prices across rounds will be tested, accounting for the survey weights and stratified sampling design. A significant change in prices across rounds may reflect the influence of HIP. It will not reflect seasonal variation, because the two rounds of the retailer survey take place at the same time of year (September–October 2011 and September–October 2012). We acknowledge that the simple pre/post design will preclude a definitive finding on price effects. In particular, the pre/post comparison may be confounded by changes in the composition of the retailer population and secular trends in prices. We will use changes in prices among supermarkets and superstores as an indicator of secular trends, since these stores are highly unlikely to change their prices in response to HIP. We will also investigate applicable price indexes.

Retailer Perceptions and Satisfaction with HIP

We will analyze survey responses from HIP participating retailers and from retailers that have dropped out of the pilot to address the following questions:

- What perceptions did retailers have about customer use of the incentives?
- How satisfied were retailers with their ability to implement and participate in the pilot?

Retailer satisfaction data will be critical to understanding both the observed patterns of participation and the broader feasibility of HIP (would retailers continue participating if HIP were permanent? To what extent to retailers consider the benefit in sales to be worth the cost in preparations, training, reconciliation, and trouble-shooting?). Concerns of retailers that drop out will be especially important in identifying ways that HIP implementation could be improved. An important determinant of retailer satisfaction, incidence of transaction problems, will be analyzed using EBT system reports or transaction data as well, to supplement the perceptual data from the retailer survey.

The retailer perceptions section of the retailer survey instrument also provides some information for the analysis of SNAP participant experience and satisfaction (discussed earlier in Section 7.2.1). Retailer perspectives on customer use of incentives will include: level of understanding by customers on how HIP works and their responsibilities; common errors by customers and problems in transactions; and perceptions of changes in SNAP customer purchases.

[30] This approach is preferable to comparing total redemptions for participating and non-participating retailers, because total redemptions include purchases by households that were not on SNAP when households were flagged. As a result, the total redemptions reflect a population of non-HIP households that are not comparable to the HIP households, and the differences between the populations could confound the comparison.

7.2.4 Other Stakeholders' Impacts and Satisfaction

Qualitative data from the key informant interviews with the State and local SNAP staff, EBT vendor staff, third-party EBT processors, and community organizations will provide both high-level indicators of satisfaction with HIP and supporting details on the reasons for reported satisfaction. For the SNAP staff and community organizations, this analysis will look at satisfaction with both the intended results of HIP—encouraging healthier eating by SNAP participants—and the side effects or unintended impacts on operations. These will be weighed against the impacts on staff workloads and organizations' abilities to accomplish their other goals. (Cost impacts will be analyzed separately as discussed in Section 7.4.) For the EBT vendor and third party processors, the focus will be narrower: what parts of the implementation and operations worked well, what parts did not, and was the overall experience positive or negative. Data from these interviews will also provide alternate perspectives on how participants interacted with HIP and insights into the ways that these interactions may have contributed to participant impact findings.

7.3 Implementation Processes

Qualitative data from the key informant interviews will provide valuable insight into State and local agency operations in the implementation of HIP. The analysis topics represent the major areas of activity to implement and operate HIP:

- Designing, developing, and implementing payment processes—identifying HIP-eligible purchases, posting incentive credits to participants' accounts, settlement of HIP transactions, and resolving problems such as reversals and refunds
- Recruiting and retaining SNAP households selected for the primary HIP sample, providing information and training, responding to concerns and problems
- Recruiting SNAP retailers to participate in HIP, providing information and training, responding to concerns and problems, and managing retailer relations
- Local SNAP agency preparations for and involvement in HIP—establishing procedures, staffing HIP-related functions, providing information to participants and other parties, responding to participant and community needs, adapting other aspects of operations to accommodate HIP
- Community partners' preparations for and involvement in HIP—tasks depending on roles, but similar to those of the local SNAP agency
- Coordination among cooperating organizations
- Coordination with the evaluation

The analysis will begin with compiling narratives of HIP implementation and operations: who did each task, what successes and challenges they encountered, and how they adapted to changes over time. These narratives will combine information from the various stakeholders who were involved with each process, using multiple perspectives to provide a comprehensive picture. Local-level perspectives will be particularly important as a check on statements from State-level respondents; local administrators and community groups may have their own biases, but they are often especially

aware of operational realities. We will identify insightful quotes from the interviews to highlight respondents' perspectives. From this narrative, we will draw the lessons of the demonstration, especially the unanticipated consequences of events and decisions. This phase of analysis will inform the interpretation of the impacts on participants and retailers.

7.4 Costs

The evaluation of HIP depends on costs as well as benefits. In the long term, an important consideration is ongoing costs, including the actual cost of the fruit and vegetable incentive plus any additional operational costs to SNAP agencies, retailers, and contractors. In the shorter term, policy-makers will also have to consider one-time implementation costs, such as the development of new program administration systems and procedures.

The cost analysis will estimate costs of the pilot, disaggregated in several dimensions described below (such as stakeholder type and functional activity). It will draw heavily on the master cost database (Section 6.6), which in turn draws its source data from DTA reports and time sheets, contractor data, the retailer survey, and key informant interviews.

The cost analysis will consider both SNAP benefit outlays and administrative costs. Settlement reports from the EBT vendor will be used to determine the benefit outlays for HIP incentives. Administrative costs will include costs incurred by DTA (headquarters and local levels) and its contractors (the EBT processor and the EBT project management contractor), and also the retailer and TPP costs reimbursed by DTA. The administrative cost analysis will also consider uncompensated costs incurred by community partners, retailers, contractors, and third party processors.

Costs will be subdivided on several dimensions:

- Stakeholder type (FNS, State, county, local office, large retailer, medium retailer, small retailer, CBO, TPP, ACS)
- Broad cost frequency type (one-time implementation costs or ongoing operational costs)
- Detailed cost frequency type (once per State, once per county, once per retailer, once per SNAP case, ongoing monthly costs per State, ongoing monthly costs per county[31], ongoing monthly costs per retailer, ongoing monthly costs per SNAP case)
- Functional activity (develop, design, test and operate payment processes; household recruiting and customer service; retailer recruiting and relations; community relations; training (clients, retailers, CBOs, etc.); general and administration; and evaluation support)
- Cost category (salaries and wages, employee benefits, contractors, supplies, other direct costs, and indirect/overhead)
- Source of funds (HIP grant, other Federal funds, non-Federal funds, or private funds)

[31] Interview data may indicate that the local office rather than the county is the most appropriate unit for this analysis.

Summary tabulations will distinguish one-time costs from ongoing costs. Tables will use a standard format with three tiers, showing: (a) one-time costs (b), monthly ongoing costs, and (c) monthly combined costs, assuming that the one-time costs are amortized over 5 years. A 5-year period is commonly used in budget forecasting for Federal programs. For example, Exhibit 7.3 shows a tabulation of costs by cost category.

Exhibit 7.3: HIP Grant Expenditures by Cost Category

	Cost Category						
	Salaries and Wages	**Employee Benefits**	**Contractors**	**Supplies**	**Other Direct Costs**	**Indirect/ Overhead**	**Total**
One-time Costs							
Monthly Ongoing Costs							
Monthly Combined Costs							

Note: combined costs assume that one-time costs are amortized over 5 years. Other categories of costs that represent a significant share of the total will be broken out where data are available.

For a two-variable version of this analysis, Exhibit 7.4 shows the cross-tabulation of stakeholder type and cost category.

Exhibit 7.4: HIP Grant Expenditures by Stakeholder Type and Cost Category

	Cost Category						
Stakeholder Type	**Salaries and Wages**	**Employee Benefits**	**Contractors**	**Supplies**	**Other Direct Costs**	**Indirect/ Overhead**	**Total**
One-Time Costs							
DTA							
Local DTA Offices							
ACS							
Novo Dia Group							
Retailers							
TPPs							
CBOs							
Total							
Monthly Ongoing Costs							
DTA							
Local DTA Offices							
ACS							
Retailers							
TPPs							
CBOs							
Total							
Monthly Combined Costs							

Stakeholder Type	Cost Category						
	Salaries and Wages	Employee Benefits	Contractors	Supplies	Other Direct Costs	Indirect/ Overhead	Total
DTA							
Local DTA Offices							
ACS							
Retailers							
TPPs							
CBOs							
Total							

Note: combined costs assume that one-time costs are amortized over 5 years.

In a similar manner, we will construct tabulations for the several subdivisions of costs described above. The narrative accompanying the cost tabulations will address an array of questions, including the following:

- What costs were imposed on each type of stakeholder?
- Over a 5-year period, what would start-up costs be as a proportion of total costs?
- What proportion of total costs for the pilot were reimbursed and unreimbursed?
- What functional activities cost the most?

7.5 Feasibility of Expansion

The HIP evaluation will provide important information as policy-makers consider whether and how to expand the pilot. The analysis of the feasibility of expansion will have qualitative and quantitative components. Building on the implementation analysis (Objective 3), the qualitative components will discuss each hurdle that was encountered during the pilot, assessing lessons learned and the possibility of overcoming the hurdle in the future. Building on the cost analysis, the quantitative components will estimate how much it would cost to expand a program such as HIP nationally.

The quantitative component will use cost data classified according to the dimensions identified in Section 7.4, including detailed cost frequency, functional activity, cost category, and source of funds.[32] The cost frequency dimension classifies costs according to their frequency, not according to the party that actually pays the costs. Here are two examples:

- "One-time costs per SNAP case" include the average per-case costs of mailing a new EBT card sleeve and fielding help-line calls from new HIP participants. These costs are paid by the State agency.
- "Ongoing monthly costs per SNAP case" include the monthly per-case costs of the incentive itself (paid by the State agency using grant funds from FNS) plus the average monthly per-

[32] The cost data from the pilot will also identify the stakeholder type. However, for the estimation of costs of HIP expansion, this dimension will probably be redundant given the differentiation by functional activity and detailed cost frequency.

case costs specifically for processing HIP transactions (regardless of whether they are reimbursed or absorbed by retailers and the EBT processor).

Our estimates of the costs of national expansion will employ two sets of factors.

Factor (A). The cost data, classified by detailed cost frequency. At the highest level, the costs will be organized and numbered as follows:

- A1(a) One-time costs per State, and A1(b) ongoing monthly costs per State
- A2(a) One-time costs per county, and A2(b) ongoing costs per county
- A3(a) One-time costs per retailer by type, and A3(b) ongoing costs per retailer by type
- A4(a) One-time costs per SNAP case, and A4(b) ongoing costs per SNAP case

As shown below in Exhibit 7.5, each of these categories includes one or more functional activities. The underlying cost estimation model will also differentiate costs by cost category and source of funds.

Factor (B). The number of national units corresponding to each stakeholder type, organized and numbered as follows:

- B1 Number of States
- B2 Number of counties in the U.S.
- B3 Number of retailers in the U.S. by type[33]
- B4 Number of SNAP cases in the U.S.

We will multiple the average cost estimates in each detailed cost frequency category by the corresponding number of national units. For example, A1(a), the one-time costs per State in Massachusetts will be multiplied by B1, the 50 States plus the District of Columbia. Exhibit 7.5 provides a table shell for presenting the results broken down by two dimensions: cost frequency and functional activity. Additional tables will be produced as needed using the other dimensions in the cost model.

[33] The appropriate typology of retailers for this analysis may be based on ownership (chain vs independent), size, or type of processor. This will be determined during the key informant interviews. It may be necessary to use the distribution of stores by FNS store type to approximate the more relevant distribution (e.g., estimating the percentage of EBT-only stores in each store type as a way of estimating the total number of such stores).

Exhibit 7.5: Simulated Costs of Nationwide Expansion for a Healthy Incentive Program, by Cost Frequency and Functional Activity

Timing and Functional Activity	National Costs (By Cost Frequency Category)						
	State	County	Large Retailer	Medium Retailer	Small Retailer	SNAP Case	Total
One-Time Costs:							
System design, development and testing							
Household recruiting							
Retailer recruiting							
Community relations							
Training							
General/administrative							
Monthly Ongoing Costs:							
System operations							
Household customer service							
Retailer customer service							
Community relations							
General/administrative							
Monthly Combined Costs							

Note: combined costs assume that one-time costs are amortized over 5 years.

This approach recognizes that larger and more populous States (which generally have more counties, retailers, and SNAP cases) are likely to have greater implementation costs. The quantitative analysis will necessarily make certain assumptions: that Massachusetts one-time State-level costs reflect typical one-time State-level costs, and so forth. These assumptions will be only approximately correct. The need to make such assumptions is a consequence of the decision to use a strong random assignment research design in a single county. We will tabulate our estimated total implementation costs by State.

We will assess the robustness of the cost estimates using alternate assumptions. For each of the cost factors by detailed cost frequency type (Factors A1 through A4 in the preceding discussion), we will establish alternate low and high estimates for comparison with our best estimates. For some factors, we will be able to choose the low and high estimates by making alternate assumptions about which detailed costs from the master cost database should be included, or by using data on the variability of these costs (e.g., the variation in system design, development, and testing costs among large retailers). For other factors, the only approach is to use expert judgment in picking the low and high estimates. For these expert judgments we will use internal input from members of the Abt team and also from FNS. The analyses of total national costs (as in Exhibit 7.5) will be re-estimated in a parallel format for multiple combinations of the low and high alternate estimates for each factor. We will report the extent to which uncertainty about each cost factor contributes to uncertainty about the total national cost estimate.

To look at the question of expansion from perspectives besides costs, we will address several further feasibility issues:

- Is HIP technically feasible for all of the partners in the EBT system: State, EBT vendor, retailers, and their processors?
- Does HIP produce increased intakes of fruits and vegetables? How large is the effect?
- What is the cost per unit of benefit (e.g., per cup of fruits and vegetables), to FNS and to stakeholders? How does this ratio compare with other known interventions?
- Are there indications in the findings of how the feasibility and cost-effectiveness of HIP can be improved?

Discussions of possible improvements will draw primarily on the insights and suggestions of informants at all levels. In addition, the researchers will draw on their own experience and expertise, particularly that of the EBT experts who have assisted with the implementation of a wide variety of innovative SNAP payment systems reforms.

8. Reporting

The Abt team will produce four reports and two briefings on the HIP evaluation. Due dates for each report are noted in Exhibit 8.1 along with the due dates for the briefing materials. As stipulated in the evaluation contract, the dates for each revised report assume that we receive FNS's comments on the preceding draft report within 2 weeks of submitting a report, allowing Abt 2 weeks to respond to comments and suggestions and produce the next version of the report. Our project quality advisors will review each version of the report before it is sent to FNS. For the revised and final versions, they will make certain that all of the comments received from FNS have been properly dealt with in the report.

The four reports (and associated FNS briefings) address the five study objectives (see Exhibit 1.1).

- The Implementation Report describes the process of implementing HIP (Objective 3). It will utilize information obtained in the first round of stakeholder data collection activities.
- The Interim Report evaluates the impact of HIP on participant fruit and vegetable consumption and other key measures of dietary intake, and assesses stakeholder experiences and satisfaction, using results from early data collection (Objectives 1, 2, and 4). As part of the report preparation review process, we will hold a meeting with the TWG after the first draft of the report; we will also conduct a briefing for FNS staff after the second draft report.
- The Final Report addresses all five objectives of the HIP evaluation, using results from early and later data collection combined (Objectives 1–5). We will conduct a briefing for FNS staff after the second draft report to discuss findings and solicit input for final revisions.
- The Summary Report, prepared at the end of the project, provides an overview of major evaluation findings in non-technical language.

Reports will include an executive summary which will use graphics to present the main results, the main text, and technical appendices, as appropriate.

Exhibit 8.1: Reports and Briefings

Report/Briefing	Due Date		Objectives Addressed
Implementation Report	Draft Revised Final	1/10/12 2/7/12 3/6/12	3
Interim Report	Draft Revised #1 Revised #2 Final	6/26/12 8/21/12 10/9/12 11/6/12	1, 2, 4
Interim Report Briefing	Draft Materials Final Materials Interim Briefing	8/21/12 9/11/12 9/18/12	1, 2, 4
Final Report	Draft Revised #1 Revised #2 Final	5/7/13 6/25/13 8/20/13 9/17/13	1, 2, 3, 4, 5
Final Report Briefing	Draft Materials Final Materials Briefing	6/25/13 7/16/13 7/23/13	1, 2, 3, 4, 5
Summary Report	Draft Revised Final	10/15/13 11/12/13 12/10/13	1, 2, 3, 4, 5

8.1 Implementation Report

This report addresses **Analysis 3 (Section 7.3)** and uses data from the **Retailer Survey**, **Retailer On-Site Observations**, and **Key Informant Interviews**.

Work on the HIP Implementation Report will begin after Round 1 of the retailer survey is completed and Round 1 of the SNAP staff and community partners and EBT vendor staff and processor interviews have been completed (November 2011). The focus of this report will be to describe the processes involved in the implementation and operation of HIP from the stakeholders' perspective. The report will include a description of the roles and responsibilities of key stakeholders other than participants, challenges encountered in the implementation and operation of HIP, and the resolution of those challenges. The lead staff will prepare the report with input from other relevant team members. Three versions of the Implementation Report will be produced: first draft, revised draft and Final Implementation Report (see Exhibit 8.1 for dates).

Each version of the Implementation Report (first, second, and final draft) will include the following components:

- An executive summary

- A technical report that addresses:
  - Payment processes
  - Recruiting and training SNAP households
  - Recruiting and training SNAP retailers
  - Local SNAP agency preparations for and involvement in HIP
  - Community partners' preparations for and involvement in HIP
  - Coordination among stakeholders
  - References

8.2 Interim Report

This report addresses **Analysis 1 (Section 7.1)** and **Analysis 2 (Section 7.2)** and uses data from **Participant Round 1 Survey** and **Round 2 Survey**, the **Retailer Survey**, and **Retailer On-Site Observations.**

Work on the Interim Report, describing early pilot impacts, will begin after the completion of analysis of data for Rounds 1 and 2 of the evaluation (May 2012). The confirmatory outcome is the average of MTFV in Rounds 2 and 3, and this cannot be computed until the Round 3 data become available. Thus, all impact analyses in this Interim Report will be exploratory.

We will prepare a detailed report outline to be submitted to FNS on May 22, 2012. The draft report outline will include table shells, sample charts and graphs, and report covers, to be discussed with FNS prior to submission of the first draft. Four versions of the Interim Report will be prepared for the HIP evaluation: the first, second, third and final versions. The first draft will be submitted on June 26, 2012 and the final version on November 6, 2012. After the second draft has been submitted, the Abt evaluation team (project director and key analysis staff) will conduct a half-day briefing for FNS on or about September 18, 2012. Each revision of the Interim Report will incorporate feedback from FNS. In addition, TWG feedback and comments from FNS received from the briefing will be incorporated into the third draft and the final Interim Report.

Each version of the Interim Report (first, second, third and final draft) will include the following components:

- An executive summary
- A technical report that addresses:
  - An introduction with a description of the pilot background
  - Study issues, objectives and research questions
  - Evaluation methods
  - The impact of HIP on fruit and vegetable consumption and other key measures of dietary intake, based on Rounds 1 and 2 of the participant survey
  - Stakeholder experiences and satisfaction, based on Rounds 1 and 2 of the participant survey, the first round of the retailer survey, and key informant interviews
  - References

- Appendices with all study data collection instruments, the detailed sampling plan, and a technical description of all analysis procedures.

We will prepare two versions of the Interim Report briefing materials (draft and final) prior to the briefing. The draft materials for the briefing to FNS will be prepared and submitted concurrently with the second draft of the Interim Report. The purpose of the briefing will be to discuss and explore the analyses and the findings thus far with FNS and members of the TWG, to determine what adjustments or additional analyses may be necessary or desired for the remainder of the evaluation and to get feedback on further revisions needed to the Interim Report. The final draft of the interim briefing materials will be prepared after receiving comments from FNS, assuming that comments will be received within 1.5 weeks of the first draft submission. Both versions of the briefing materials will include:

- A copy of the Interim Report (second draft)
- PowerPoint slides with graphs, visual displays and bulleted highlights
- Handouts of the slides (when final)

8.3 Final Report

This report addresses all of **Analyses 1** through **5** and uses data from **Participant Round 1**, **Round 2**, and **Round 3 Surveys**, **Participant Focus Groups, EBT Transactions Data**, the **Retailer Survey**, **Retailer On-Site Observations,** and **Key Informant Interviews**.

Work on the Final Report, describing the pilot and impacts, will begin after the completion of analysis of Round 3 data. This report will include the central evaluation finding, impact on the confirmatory outcome of modified target fruit and vegetable intake average over Rounds 2 and 3. It will also provide information on program costs for the first time. We will prepare a detailed report outline to be submitted to FNS on January 29, 2013. The draft outline will include table shells, sample charts and graphs, and report covers, to be discussed with FNS prior to submission of the first draft. Four versions of the Final Report will be prepared for the HIP evaluation: the first, second, third and final versions. TWG members will be asked to comment on the first draft and their comments will be incorporated into the second draft. After the second draft has been submitted, the Abt evaluation team (project director and key analysis staff) will conduct a half-day briefing for FNS on or about July 23, 2013. Each revision of the Final Report will incorporate feedback from FNS. In addition, TWG feedback and comments received from the briefing will be incorporated into the third and final drafts.

Each version of the Final Report (first, second, third and final draft) will include the following components:

- An executive summary
- A technical report that addresses:
  - An introduction with a description of the pilot background
  - Study issues, objectives and research questions
  - Evaluation methods

- The impact of HIP on fruit and vegetable consumption and other key measures of dietary intake.
- Stakeholder experiences and satisfaction.
- The process of implementation, based on the earlier Implementation Report.
- The costs of introducing and operating HIP.
- The feasibility of expansion to a national program, based on the Implementation Report and the cost analysis.
- References

- Appendices with all study data collection instruments, the detailed sampling plan, and a technical description of all analysis procedures.

Two versions of the Final Report briefing materials (draft and final) will be prepared prior to the final briefing. The draft materials will be prepared and submitted concurrently with the second draft of the Final Report. The purpose of the briefing will be to discuss and explore the analyses and findings with FNS, allowing us to fine-tune the interpretation of the results for the third and final drafts of the report. The Final version of the final briefing materials will be prepared after receiving comments from FNS, assuming that comments will be received within 1.5 weeks of the draft submission. Both versions of the briefing materials will include:

- A copy of the Final Report (second draft)
- PowerPoint slides with graphs, visual displays and bulleted highlights
- Handouts of the slides (when final)

8.4 Summary Report

After FNS acceptance of the Final Report, the Abt team will prepare a short (20 pages or fewer), free-standing Summary Report providing an overview of the study and the major evaluation findings, written in non-technical language. The free-standing summary will be professionally designed and formatted, full-color and glossy, including a professionally designed full-color cover with photographs, and will use colors in graphical displays. Three versions of the Summary Report will be produced: first, second and final versions, with rounds of FNS review and comment in between. The first draft will be submitted on October 15, 2013 and the final version on December 10, 2013.

Upon acceptance of the final version, we will submit an unbound camera-ready copy and print files on a CD in a pdf file.

Appendix A: HIP Incentive Cap

The HIP design includes a fixed monthly cap on the amount of the incentive payments per household. The size of this statutory cap will influence the pilot's impact and cost.

If the statutory cap were set too low, it could limit the impact of HIP on households' expenditures for target fruits and vegetables. (HIP participants who are not capped face a price for fruits and vegetables that is reduced by 30 percent, whereas participants who are capped face the original price of fruits and vegetables.) However, if the statutory cap were set too high, it could encourage fraud and cause the total incentive payments to exceed the budgeted amount of $2 million.

Balancing these two considerations, an analysis by Abt Associates recommended—and FNS selected—a cap of $60 per month per household.

Participants reach an effective limit or "cap" on their incentive if they use their SNAP benefits entirely for TFVs, or if they reach the fixed "statutory cap" determined by FNS. Based on an analysis of three simulation models, Abt recommended a statutory incentive cap in the range of $50 to $80 per SNAP household per month. Within this range, the best choice depended which of the three simulations one preferred. Specifically based on the second of three models, we suggested a $60 cap. The second model predicts that a $60 cap would constrain the behavior of only 1 percent of HIP participants, without exhausting the budget for incentive payments. However, based on the two other models, one could make a good argument for a slightly higher or lower cap.

The Models

For a range of potential incentive caps, from $30 to $80, we estimated (a) the percentage of households that would be capped and (b) the total Healthy Incentive Pilot (HIP) incentive payment that would be disbursed over the 12 months of the demonstration. Our simulations used three alternative models of consumer purchasing behavior.

Model 1. This model uses an Engel function approach, similar to a model presented at the first Technical Working Group (TWG) meeting. An Engel function shows how food spending responds to changing household resources. This model assumes: (a) participants spend 20 percent of total resources on food,[34] (b) participants use 12 percent of food spending for fruits and vegetables in the absence of HIP (Stewart & Blisard, 2008), (c) the own-price elasticity for fruits and vegetables is –0.85 (at the upper end of the 95 percent confidence interval of the elasticities presented in Andreyeva, Long, & Brownell, 2010). This model generates the lowest estimates of TFV spending, and hence the lowest estimates of the budget required for incentive payments.

[34] This is an approximation, appropriate for low-income households. Based on Consumer Expenditure Survey data, at-home food as a share of income is higher than 20 percent for households with annual income below $20,000 and lower than 20 percent for households with annual income above $20,000 (Frazao et al., 2007). We assume participants will use all of their SNAP benefits on food if their benefits exceed their desired (inframarginal) food spending level.

Model 2. This model assumes participants seek to purchase 53 percent of the Thrifty Food Plan (TFP) recommendation for fruits and vegetables. This assumption is based on Blisard and Stewart's (2006) estimate of average fruit and vegetable spending as a fraction of TFP. This model generates intermediate estimates of TFV spending and the budget required for incentive payments.

Model 3. This model assumes participants seek to purchase 80 percent of the TFP recommendation for fruits and vegetables. Although there is no particular literature to justify this assumption, we wanted one model to consider more conservatively what would happen if the HIP impact were quite large, generating higher spending on fruits and vegetables than has been seen in the previous literature.[35] Model 3 generates the highest estimates of TFV spending and the budget required for incentive payments.

Data

To describe the distribution of caseload characteristics (including earnings and SNAP benefits), we used administrative records from Hampden County for October 2010. Each record includes the household composition (including age and gender), monthly SNAP benefit, and monthly cash income.

We summed the TFP recommendation for each individual, based on the individual's age and gender, to compute each household's TFP recommendation for all food and for fruits and vegetables. For households with a particular household size, the TFP recommendation for fruits and vegetables varies because of changes in the age and gender composition of households. We used the mean TFP recommendations for fruits and vegetables in descriptive statistics below, and we used a percentage of this recommendation in our results for Models 2 and 3.

Results

Descriptive statistics are reported in Table 1. There were 52,650 SNAP cases in Hampden County in October, 2010. Mean income per month was $849 and mean monthly SNAP benefit was $260.

Table 1. Descriptive Statistics, Hampden County

Cases	52,650
Mean income ($/month)	849.46
Mean benefits ($/month)	260.48
Mean household size	2.89
Mean TFP, all food ($/month)	420.68
Mean of 53% of TFP, fruits/veg. ($/month)	68.52
Mean TFP, fruits/veg. ($/month)	129.30

[35] Model 3's assumption that 80 percent of the TFP recommendation would be spend on fruits and vegetables is quite high. If the starting spending level without HIP were 53 percent of the TFP recommendation, and the own-price elasticity were -0.3, then the new spending level would be 66 percent of the TFP recommendation. We chose 80 percent simply as a round number, more conservative than this assumption based on existing spending patterns.

In advance of our simulations, a preliminary tabulation of TFP recommendations for different household sizes is reported in Table 2. It shows the incentive payment that would correspond to fruit and vegetable purchases equal to the TFP fruit and vegetable recommendation for each household size. For example, for a household with 6 persons, the mean monthly TFP recommendation for fruits and vegetables is rounded to $135. The incentive payment corresponding to fruit and vegetable purchases at this level is $40. (equal to 30% * $135). If the monthly incentive cap were below $40, the household would be unable to earn the incentive on all of its fruit and vegetable spending.

Table 2. Mean TFP recommendations for fruit and vegetables, by household size

		Thrifty Food Plan ($/month)		
Household Size	Frequency (%)	Total	Fruit/Veg	30% of Fruit/Veg
1	37.4	188	43	13
2	15.3	307	50	15
3	14.7	436	48	14
4	12.7	562	63	19
5	8.5	675	94	28
6	5.0	819	135	40
7	2.9	911	276	83
8	1.6	1045	316	95

Source: Administrative records for Hampden County SNAP, October, 2010.

The main results for our simulations are reported in Table 3. The rows show incentive caps from $30 to $80. The three broad groups of columns show the three models. Within each model, the columns show: (a) percentage of participants who reach the statutory cap, (b) mean incentive payment per household, and (c) total incentive payments for HIP in thousands of dollars.

Table 3. Simulated Incentive Payments Under Six Statutory Caps and Three Models of Expenditure

	(1) Engel Function Simulation			(2) 53% TFP Simulation			(3) 80% TFP Simulation		
		Incentive Paid			Incentive Paid			Incentive Paid	
Statutory Cap	% Capped	per case month ($)	total (1000 $)	% Capped	per case month ($)	total (1000 $)	% Capped	per case month ($)	total (1000 $)
30	0.2%	14.20	1054	19.7%	17.78	1320	40.1%	23.18	1721
40	0.0%	14.45	1072	7.7%	19.04	1413	24.4%	25.38	1884
50	0.0%	14.49	1076	2.5%	19.51	1448	13.7%	27.27	2025
60	0.0%	14.50	1076	1.0%	19.68	1461	7.3%	28.28	2100
70	0.0%	14.50	1076	0.4%	19.75	1466	3.4%	28.80	2138
80	0.0%	14.50	1076	0.2%	19.78	1468	1.7%	29.03	2155
Assumed number of case months:			74232						

Based on Model 1, few participants would be capped. Only $1 million of the $2 million budget for incentive payments would be exhausted irrespective of any of the statutory cap amount.

Based on the intermediate Model 2, 19.7 percent of participant households would be capped with a statutory cap of $30, and 7.7 percent of participant households would be capped with a statutory cap of $40. We consider these percentages to be sufficiently high that we are concerned that these low statutory caps could constrain the overall impact of HIP. According to this model, the $2 million budget would not be exhausted under any of the incentive caps we studied up to $80. The main reason we do not consider caps even higher than $80 is concern about fraud, which was not addressed in the model.

Based on Model 3, 40 percent of participant households would be capped with a statutory cap of $30, and 24 percent of participant households would be capped with a statutory cap of $40. Again, these percentages are so high that we are concerned about limiting the overall impact of HIP. Yet, even under the conservative assumptions of Model 3, the $2 million budget would not be exhausted with incentive caps in this range. With higher incentive caps, fewer households are capped, but the total incentive payments in this conservative model exceed $2 million.

Limitations

There are several limitations to our analysis:

First, in Model 1, we used elasticity estimates and budget share estimates that did not distinguish target fruits and vegetables from all fruits and vegetables. In further work, one could use the Consumer Expenditure Survey (CEX) to measure Target Fruits and Vegetables (TFVs) rather than all fruits and vegetables.[36]

Second, in all three models, we assumed that each participant household followed the same behavior model, with different spending outcomes only because the households have different income, benefits and household sizes. It would be more realistic to consider the distribution of spending levels for households with a particular income, benefits, and household size. In further work, one could attempt to estimate more precisely the fraction of households that would be capped. In particular, one could extract information on the distribution of Engel fractions (i.e., the fraction of income spent on TFV) from the CEX.

Third, we did not formally address the potential for fraud, in which participants seek to purchase more fruits and vegetables than they need, either for resale or as part of a scheme involving an authorized retailer. Instead, because of concern about fraud, we simply ruled out in advance potential incentive caps over $80. In further work, one could make a more formal assumption about the fraction of the HIP sample that would engage in this type of fraud.

Despite these limitations, we are confident in the main conclusions below. For example, we doubt that further analysis would show either that a $60 cap will constrain a large fraction of households or that a $60 cap will exhaust the $2 million budget.

Conclusion

Within the range of $50 to $80, the best choice depends on which of the three models one prefers. To determine a single recommendation, we focused on Model 2, because it is more cautious than Model 1, and yet more realistic than Model 3. In the column of results for Model 2, we would be satisfied

[36] The initial (non-HIP) spending level for TFVs will be lower than the initial spending level for all fruits and vegetables. However, the price elasticity for TFVs may be higher than the price elasticity for all fruits and vegetables, because narrow category definitions allow more cross-category substitution than a broad category definition does. Overall, we do not know the direction of the change to expect by considering TFVs more specifically.

with having only 1 percent of participant households reach the statutory cap, a level that indicates a recommended cap of $60.

Appendix B: Evaluation Project Schedule

Project Activities for Tasks 3-13: Tasks, Deliverables, Due Dates
Start Date June 30, 2010; End Date December 31, 2013

		Period of Performance			
Task/ Subtask	**Description**	**Start Date**	**End Date**	**Deliverables**	**Due Date**
BASE CONTRACT					
3	Create a Technical Work Group	6/30/10	12/31/13		
3A	First TWG Meeting	8/1/10	10/19/10		Completed
3B	Second TWG Meeting	6/19/12	8/7/12		
				Draft Agenda and Materials	6/26/12
				Final Agenda and Materials	7/10/12
				Hold Second TWG Meeting	7/17/12
				Technical Memorandum—TWG Feedback on Interim Report	8/7/12
3C	Additional TWG Activities	5/15/13	12/31/13		
				Technical Memorandum—TWG Feedback on Final Report	6/4/13
				Ad-hoc Activity Summaries	As needed
4	Establish Evaluation Reporting Requirements and Technical Support for the Pilot	8/1/10	2/28/13		
4A	Set up Evaluation Reporting Requirements	8/1/10	12/14/10		Completed
4B	Provide Technical Support to Pilot for Their Evaluation Activities	11/2/10	2/28/13	*Provide technical support, as needed*	
5	Update Study Plan	10/19/10	3/15/11		
				Draft Updated Study Plan	Submitted
				Final Updated Study Plan	Submitted
6	Develop and Pretest Data Collection Instruments and Procedures	10/19/10	8/23/11		

Project Activities for Tasks 3-13: Tasks, Deliverables, Due Dates
Start Date June 30, 2010; End Date December 31, 2013

		Period of Performance			
Task/ Subtask	**Description**	**Start Date**	**End Date**	**Deliverables**	**Due Date**
				Draft Instruments	Submitted
				Revised Instruments	Submitted
				Instrument Testing Complete	Completed
				Technical Memorandum—Testing of Data Collection Instruments and Procedures	Submitted
				Final Instruments	Submitted
				OMB Revisions to Instruments/Procedures	8/23/11
7	Create OMB Clearance Package	11/16/10	8/23/11		
				Draft OMB package	Submitted
				Final OMB package	Submitted
8	Recruit and Train Data Collectors	3/15/11	9/15/12		
				Draft Data Collector Training Plan	Submitted
				Final Data Collector Training Plan	Submitted
				Training Summary Memorandum	9/15/11
9	Implement Evaluation Design	5/1/11	3/13/13		
				Technical Memorandum—Testing Sample Selection Procedures	7/26/11
				Initial Sample Draw Memorandum	9/27/11
				Sample Selection Complete	11/15/11
				Final Sample Draw Memorandum	11/29/11
10	Collect Data				
10A	Baseline Interviews	9/1/11	12/13/11		
				Round 1 Household Data Collection Complete	12/13/11
10B	Month 3-5 Interviews	2/6/12	5/1/12		
				Round 2 Household Data Collection Complete	5/1/12
10C	Month 9-11 Interviews	8/6/12	12/11/12		
				Round 3 Household Data Collection Complete	12/11/12
10D	Participant Focus Groups	2/20/12	11/6/12		
		2/20/12	3/27/12	Round 2 Focus Groups Complete	3/27/12

Project Activities for Tasks 3-13: Tasks, Deliverables, Due Dates
Start Date June 30, 2010; End Date December 31, 2013

		Period of Performance			
Task/ Subtask	**Description**	**Start Date**	**End Date**	**Deliverables**	**Due Date**
		10/1/12	11/6/12	Round 3 Focus Groups Complete	11/6/12
10E	SNAP Retailers	7/1/11	12/11/12		
				First Retailer Survey Complete	10/21/11
				Second Retailer Survey Complete	12/11/12
10F	SNAP Staff	10/15/11	12/11/12		
		10/15/11		Round 1 Interviews Complete	10/21/11
		4/10/12		Round 2 Interviews Complete	5/1/12
		11/10/12		Round 3 Interviews Complete	12/11/12
10G	EBT Vendor Staff and third Party Processors	10/15/11	2/2613		
		10/15/11		Round 1 Interviews Complete	10/21/11
		4/10/12		Round 2 Interviews Complete	5/1/12
		1/15/13		Final Interviews Complete	2/26/13
10H	EBT Transaction Data	7/1/11	2/26/13		
				Months 1-5 EBT Data Obtained	4/24/12
				Months 6-15 EBT Data Obtained	2/26/13
11	Creating Data Files				
	Interim Report	11/22/11	1/15/13		
				Data and Analysis Files for Interim Report	11/6/12
				Revisions to Files if necessary	1/15/13
	Final Report	12/11/12	12/31/13		
				Data and Analysis Files for Final Report	9/17/13
				Revisions to Files (if necessary)	12/31/13
	Public Use Files	7/17/13	12/31/13		
				Public Use Data and Analysis Files	9/17/13
				Revisions to Files (if necessary)	12/31/13
12	Analysis				

Project Activities for Tasks 3-13: Tasks, Deliverables, Due Dates
Start Date June 30, 2010; End Date December 31, 2013

		Period of Performance			
Task/ Subtask	**Description**	**Start Date**	**End Date**	**Deliverables**	**Due Date**
	Round 1 and 2				
		2/1/12	11/6/12	Round 2 Draft Tables (Interim Report)	5/22/12
	Round 3				
		12/11/12	9/17/13	Round 3 Draft Tables (Final Report)	3/26/13
13	Briefings and Reports				
13A	Implementation Report	11/1/11	3/6/12		
				Draft Report	1/10/12
				Revised Report	2/7/12
				Final Report	3/6/12
13B	Interim Report	4/1/12	11/6/12		
				Outline	5/22/12
				Draft Report	6/26/12
				Revised Report	8/21/12
				Draft Briefing Materials	8/21/12
				Final Briefing Materials	9/11/12
				Hold Interim Report Briefing	9/18/12
				Revised Report	10/9/12
				Final Report	11/6/12
13C	Final Report				
				Outline	1/29/13
				Draft Final Report	5/7/13
				Revised Final Report	6/25/13
				Draft Briefing Materials	6/25/13
				Final Briefing Materials	7/16/13
				Hold Final Report Briefing	7/23/13
				Revised Report	8/20/13
				Final Report	9/17/13
13D	Summary Report				

Project Activities for Tasks 3-13: Tasks, Deliverables, Due Dates
Start Date June 30, 2010; End Date December 31, 2013

		Period of Performance			
Task/ Subtask	**Description**	**Start Date**	**End Date**	**Deliverables**	**Due Date**
				Draft Report	10/15/13
				Revised Report	11/12/13
				Final Report	12/10/13

Appendix C: Economic Theory

Economic theory and past research give just a rough indication of the type of effects one might see. HIP may influence fruit and vegetable consumption through (a) a budget effect and (b) a marketing effect. The budget effect arises because HIP provides additional resources in the form of SNAP benefits. The marketing effect arises because HIP carries an explicit and implicit message favoring fruits and vegetables, which may influence participants to consume more fruits and vegetables.

The budget effect, in turn, has two parts: (a1) a price effect and (a2) an income effect.

- (a1) The price effect arises on some purchases when HIP reduces the effective marginal price of TFVs relative to the price of other SNAP-eligible foods. For example, if a participant spends an additional $1 on TFVs, and HIP gives her 30 cents back on her EBT card, the real price of the food from her perspective is only 70 cents, equivalent to a 30 percent price discount.
- (a2) The income effect arises because HIP increases the total amount of SNAP benefits available to the household. As with any increase in SNAP benefits, consumers may spend some of the additional resources on TFVs.

Income is expected to have a small marginal effect, while a price change might in principle be expected to have a larger effect (Guthrie et al., 2007). Mean estimates of the elasticity of fruit and vegetable consumption with respect to price are 0.70 and 0.58, respectively (Andreyeva, Long, & Brownell, 2010). This implies that a 30 percent price discount could lead to a 21 percent increase in fruit consumption and a 17.4 percent increase in vegetable consumption, which are considered large consumption increases.

HIP participants only face the discounted price on marginal additional purchases of TFVs if they have not yet reached the maximum TFV amount on which they can earn the incentive. Participants can reach this maximum either by using all of their SNAP benefits on TFVs or by reaching the $60 monthly cap on incentive payments. For a participant who reaches the maximum, additional marginal spending on TFVs using his or her own cash income does not earn the incentive, so this additional spending is not subject to the price effect.

Overall, because of gaps in the existing research literature, it is difficult to forecast the likely magnitude of changes in TFV consumption based on theory and past research alone. The HIP evaluation will provide new evidence on this issue.

References

Andreyeva, T., Long, M. W., & Brownell, K. D. 2010. The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food. *American Journal of Public Health* 100: 216-222

Beatty, P. C., & Willis, G. B. 2007. The practice of cognitive interviewing. *Public Opinion Quarterly* 71(2): 287-311.

Blisard, W.N., Stewart, H. 2006. *How low-income households allocate their food budget relative to the cost of the Thrifty Food Plan.* Alexandria, VA: U.S. Department of Agriculture, Economic Research Service.

Cody, S., Castner, L. A., Mabli, J., & Sykes, J. 2007. *Dynamics of Food Stamp Program participation, 2001-2003.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Research, Nutrition, and Analysis.

Coffey A., & Atkinson, P. 1996. *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage Publications.

Cole, N. and Fox, M.K. 2008. *Diet Quality of Americans by Food Stamp Participation Status: Data from the National Health and Nutrition Examination Survey, 1999-2004.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Research, Nutrition and Analysis.

Dodd, K. W., et al. 2006. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *Journal of the American Dietetic Association* 106(10): 1640–1650.

Dong, D., & Lin, B. H. 2009. *Fruit and vegetable consumption by low-income Americans: Would a price reduction make a difference?* (ERS Report 70). Washington, DC: U.S. Department of Agriculture, Economic Research Service. Available at: http://www.ers.usda.gov/publications/err70/err70.pdf

Fraker T. 1990. *The effects of food stamps on food consumption: A review of the literature.* Washington, DC: U.S. Department of Agriculture, Food and Nutrition Service.

Frazão, E., Andrews, M. S., Smallwood, D., & Prell, M.A. 2007. Food spending patterns of low-income households will increasing purchasing power result in healthier food choices? (Economic Information Bulletin Number 29-4). Alexandria, VA: U.S. Department of Agriculture, Economic Research Service.

Groves, R. M., Presser, S., & Dipko, S. 2004. The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly* 68(1): 2–31.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. 2004. *Survey methodology*. New York: Wiley.

Guthrie, J., Lin, B., Ver Ploeg, M., & Frazao, E. 2007. *Can food stamps do more to improve food choices? (*Economic Information Bulletin No. 29-1). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Institute of Medicine. 2000. *Dietary Reference Intakes: Applications in dietary assessment.* Washington, DC: National Academies Press.

Li, T., Tooze, J., Bethel, J., & McNutt, S. 2007. *Food Attitudes and Behavior (FAB) Study analysis results for pilot data.* Bethesda, MD: National Cancer Institute.

Mabli, J., Gleason, P., & Hall, J. 2009. Final memorandum for HIP power analysis. Contract Number: AG-3198-D-08-0090/GS-10F-0050L. Memo number 014. Princeton, NJ: Mathematica Policy Research.

Martorell, F., Klerman, J.A. , & Loughran, D. 2010. Assessment of and adjustment for nonresponse bias in earnings data. Presented at the December 2010 meeting of the Abt Associates Journal Authors Support Group, Cambridge, MA.

Ministry of Health. 2003. *NZ food NZ children: Key results of the 2002 National Children's Nutrition Survey*. Wellington: Ministry of Health.

Nusser, S.M., et al. 1996. A semi-parametric transformation approach to estimating usual nutrient intake distributions. *Journal of the American Statistical Association* 91: 1440-1449.

Rust, K.F., & Rao, J.N.K. 1996. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 5: 283-310.

Schochet, P. Z. 2008. Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics* 33(1): 62-87.

Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N. Tanur, J., & Tourangeau, R. 1999. *Cognition and survey research*. New York: Wiley.

Stewart, H., & Blisard, N. 2008. Are lower income households willing and able to budget for fruits and vegetables? (ERS Report 54). Alexandria, VA: U.S. Department of Agriculture, Economic Research Service.

Tooze, J. A., et al. 2006. A new statistical method for estimating the usual intake of episodically consumed foods with applications to their distribution. *Journal of the American Dietetic Association* 106(10): 1575-1587.

Tourangeau, R., Rips, L. J., & Rasinski, K. 2000. *The psychology of survey response.* New York: Cambridge University Press.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. 2010. *Dietary guidelines for Americans, 2010*, 7th ed. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. 2010. *Healthy people 2010,* 2nd ed. Washington, DC: U.S. Government Printing Office.

Willis, G. B. 2005. *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, CA: Sage.