

Appendix J
Analysis Plan

J. Analysis Plan

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 J.1), stakeholder experiences (Section J.2), implementation processes (Section J.3), costs of HIP (Section J.4), and the feasibility of nationwide expansion (Section J.5).

J.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, and
- 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 surveys. 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).

J.1.1 Description of Household and Participant Characteristics

To provide a context, we will first describe the population participating in the HIP and the non-HIP group. 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 group 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 the head of household, number of adults employed, and participation in other assistance programs.

J.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 J.1.5 discusses, it is good statistical practice to identify the main “confirmatory” outcome in advance.¹ 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 J.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.

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,

¹ This practice avoids the well-known statistical problem of multiple comparisons. For each hypothesis test, a conventional approach allows a 5% 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).

- Consumption of all target fruits and vegetables, target fruits alone, target vegetables alone as a percentage of the MyPyramid/DGA recommendations (based on “usual” intake and recommendations by age and sex)
- Meeting the MyPyramid/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)

Exhibit J.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

- 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/or 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% 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 J.1, bottom panel).

J.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, 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:

$$(1) \quad y = \beta_0 + \beta_1 \text{HIP} + \beta_2 \text{ControlVars} + u .$$

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.

We will use this linear regression model even for binary outcomes (Angrist, 2001; Angrist and Pischke, 2009). This linear probability model (LPM) for binary outcomes consistently estimates the mean impact on the probability of the outcome of interest. As a sensitivity analysis for binary outcomes, we also will report some estimates using the logit model in place of linear regression.

J.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.²

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. Note that we have a “rich list sample” (see the discussion in Martorell, Klerman, and 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.

² Randomization does not guarantee homoscedasticity, and the linear probability model on binary outcomes induces heteroscedasticity.

J.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 J.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.

J.1.6 Subgroup Analyses

To maximize statistical power, subgroup analyses will be done on average consumption from Round 2 and Round 3.³ 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 B).⁴

³ Averaging across the two rounds will increase statistical power, which is particularly important in subgroup analyses.

⁴ We will use two approaches to determine 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.

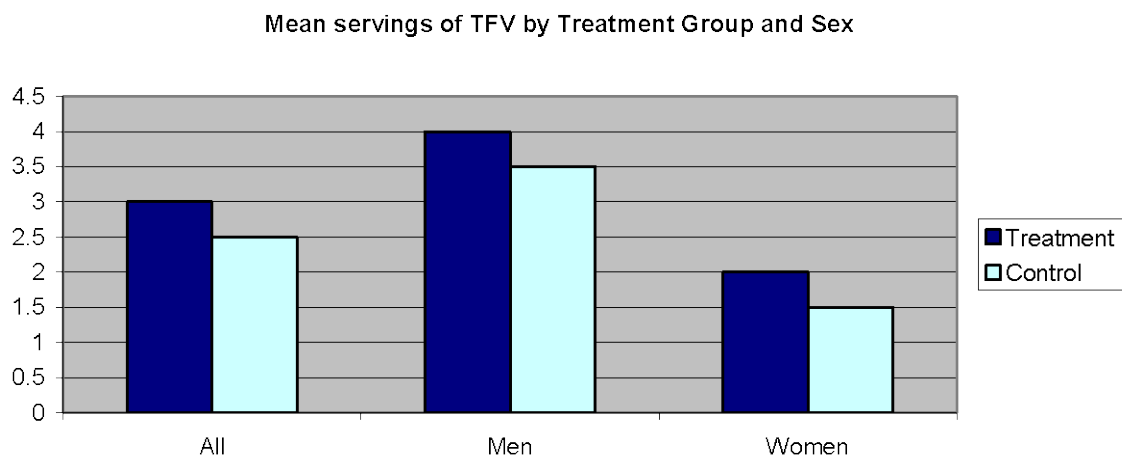
- Subgroups defined by Round 1 attitudes and barriers to fruit and vegetable consumption. The Round 1 survey includes questions about enjoying fruits and vegetables, exposure to nutrition education, the belief that fruits and vegetables are healthy, enjoyment of trying new foods, lack of ability to store 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 outcome, 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 J.1.2: MTFV intake, all fruit and vegetable intake, fruit intake, and vegetable intake (from the 24-hour recall instrument), and the five most common fruits and vegetables from the baseline fruit and vegetable screener 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 be added to show confidence intervals. Stars will indicate 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 will test for differences in impact magnitude between subgroups. See Exhibit J.2 for an example of the layout.

Exhibit J.2: Sample Bar Chart



J.1.7 Estimation of Usual Intakes

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 secondary 24HR will be selected at random, 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.

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.

J.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.⁵

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. Sub-group 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.

J.2 Stakeholder Experiences

This analysis addresses how HIP affects the experiences and satisfaction for several stakeholder groups: participants themselves (Sections J.2.1 and J.2.2), retailers of different types (Section J.2.3), and other stakeholders (Section J.2.4), including Massachusetts DTA (the HIP grantee), the local SNAP agency, EBT vendor/TPPs, and community organizations.

⁵ 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.

J.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 surveys (described in this subsection) and the focus groups (Section J.2.2).⁶

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

The first 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?

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 J.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 second 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 SNAP 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.

J.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

⁶ In addition, the retailer surveys provide some information about retailer perceptions of SNAP participant satisfaction with the retail experience of HIP (see Section 7.2.3).

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, and
- 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.

J.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, 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 surveys, 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.⁷

- 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 retailers⁸

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

⁷ 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.

⁸ For the maps, store locations will be geocoded.

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. We expect that the number of retailers that decline to participate 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 small samples. Response rates to the retailer survey will be presented as both counts and 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 J.2. Survey responses will be tabulated by HIP/non-HIP status and store type.

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 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.

The analysis of bottom-line business impacts 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.

In addition, 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.⁹

The design of the retailer survey offers the opportunity for an exploratory analysis of changes in food prices. The two rounds of the survey 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 and modest sample size ($n = 60$ retailers per round) 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.

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 retailer 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 to 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 J.2.1). Retailer perspectives on customer use of incentives will include: level of understanding

⁹ 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.

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.

J.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 J.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.

J.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 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 especially 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.

J.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, which in turn draws its source data from DTA reports and time sheets, contractor data, the retailer surveys, 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, 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)

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 J.3 shows a tabulation of costs by cost category.

Exhibit J.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 five years.

For a two-variable version of this analysis, Exhibit J.4 shows the cross-tabulation of stakeholder type and cost category.

Exhibit J.4: HIP Grant Expenditures by Stakeholder Type and Cost Category

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

ACS							
Retailers							
TPPs							
CBOs							
Total							

Note: combined costs assume that one-time costs are amortized over five 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 five-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?

J.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 measure how much it would cost to expand a program such as HIP nationally.

The quantitative component will use cost data classified according to the detailed cost frequency dimension (introduced in Section J.4). A key point is that this 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-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 employ two component factors.

Factor (A). The cost data, classified by detailed cost frequency, 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.

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 type¹⁰, and
- B4 Number of SNAP cases in the U.S.

We will multiply 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 J.5 provides a table shell for the results.

Exhibit J.5: Simulated Costs of Nationwide Expansion for a Healthy Incentive Program

	National Costs (By Cost Frequency Category)						Total
	State	County	Large Retailer	Medium Retailer	Small Retailer	SNAP Case	
One-time Costs							
Monthly Ongoing Costs							
Monthly Combined Costs							

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

The quantitative analysis will necessarily make certain assumptions: that Massachusetts costs reflect typical State costs, that Hampden County costs represent typical county 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 assess the robustness of the cost estimates using alternate assumptions. For each of the cost factors by detailed cost frequency type (factors A1 through A5 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. 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 J.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.

¹⁰ 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.

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 TFV), to FNS and to stakeholders? How does this ratio compare with other known interventions and to the costs associated with inadequate TFV consumption?
- 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.