Abt Associates Inc.

Cambridge, MA
Bethesda, MD
Durham, NC
Atlanta, GA

Technical Work Group (TWG) Meeting: Evaluation Design and Study Plans

Healthy Incentives Pilot (HIP) Evaluation

Task Order AG-3198-D-10-0044

February 16, 2011

Prepared for
Kelly E. Kinnison, PhD
Social Science Research Analyst
Office of Research and Analysis
USDA Food and Nutrition Service
3101 Park Center Drive Rm 1014
Alexandria, VA 22302

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138-1168

Contents

Introduction

The Healthy Incentives Pilot (HIP) is a project of USDA's Food and Nutrition Service (FNS) to promote intake of fruits and vegetable among participants in the Supplemental Nutrition Assistance Program (SNAP). The HIP evaluation will use a random assignment research design to measure the impact of the pilot intervention on fruit and vegetable intakes and other outcomes. The Massachusetts Department of Transitional Assistance (DTA) received the grant to operate the pilot in Hampden County, Massachusetts for 14 months, beginning in late 2011 and concluding in early 2013.

The HIP Evaluation plans call for a Technical Working Group (TWG) to provide external advice and input on the evaluation research design, early results, the final report, and to provide other consultation services as needed. The TWG includes five outside experts in relevant areas of research. In addition, subject matter experts from Federal government agencies are participating in the TWG review activities.

The five external TWG members are (see Appendix A for biographical sketches):

- 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
- Diane Whitmore Schanzenbach, Northwestern University

The three federal government experts are:

- Margaret Andrews, USDA Economic Research Service
- Sue Krebs-Smith, National Cancer Institute
- Alanna Moshfegh, USDA Agricultural Research Service[1]

First TWG Meeting

The TWG met for the first time on October 5, 2010 to discuss the HIP evaluation's research design, data collection plan, and analysis plan. This memorandum describes the topics of the first TWG meeting, summarizes the key feedback received from TWG members, and provides decisions and rationale for how input from the TWG will be used to improve data collection and analysis.

This memorandum is organized into four sections, corresponding to the four principal sessions of the first TWG meeting:

1. Background and theory, including estimates of plausible possible effect sizes;
2. Participant data collection;
3. Participant impact analysis; and
4. Stakeholder and electronic benefit transfer (EBT) data collection and analysis.

Within each section, this memorandum is organized around the specific questions and issues posed to the TWG in the advance materials, as well as several issues that arose during the TWG meeting.

[1] Alanna Moshfegh was unable to attend the first TWG meeting, and designated a representative from the Agricultural Research Service to attend the meeting.

1. Background and Theory

1. How strong is the evidence in the literature that a 30 percent price reduction should be expected to increase total fruit and vegetable consumption by ¼ cup or more? Is the HIP intervention sufficiently strong that it has a reasonable chance of affecting behavior?

The design is expected to be sufficiently powerful to detect a plausible HIP/non-HIP difference in the main outcome. Specifically, as discussed below, the minimum detectable effect (MDE) for one interview is 0.164 cups targeted fruits and vegetables; pooling across the two waves (as we propose to do), the MDE about 25 percent smaller. Thus, we will be able to easily detect a ¼ cup impact.

As to whether a ¼ cup impact is plausible in HIP, we considered the implications of two different approaches. In each approach, fairly strong assumptions are required. The existing research literature on price responses addressed spending outcomes rather than daily food intake outcomes, so one must either apply spending elasticities to intake outcomes, or convert estimates of baseline food intake into estimates of baseline food spending.

Approach 1. Simulation Presented to the TWG

In this simple simulation, we estimated the possible increase in spending that one might anticipate based on the change HIP makes to a SNAP household's budget constraint.

Average daily intake of targeted fruits and vegetables (TFVs)[2] is estimated to be 1.32 cups, so ¼ cup represents an increase of 19 percent (Mabli, Gleason, and Hall, 2009). Although food intake and food spending are distinct outcomes, a goal of the simulation was to assess whether a spending increase of 19 percent is theoretically plausible.

Our simulation addressed only the financial incentive aspects of HIP, not any implicit marketing aspects. Based on current estimates of price elasticities for fruit and vegetable spending (Dong and Lin, 2009; Andreyeva et al., 2010), with a statutory cap on the incentive in the range of $16 to $129 of rebate per month per household, the simple simulation suggested fruit and vegetable spending could increase 10 percent to 23 percent.[3] Thus, based on this simulation, a 19 percent increase is toward the higher end

[2] Fruits and vegetables targeted for HIP are those allowed by Federal regulations for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Fruit and Vegetable Voucher. These can be summarized as follows: (a) Most fruits and vegetables are included (mature legumes and juices are not authorized). (b) Included fruits and vegetables can be in any of the following forms: fresh, canned, frozen, and dried. (c) In general, eligible fruits and vegetables must not have any added sugars, fats, oils, or salt. (d) White potatoes are excluded, but yams and sweet potatoes are allowed.

[3] The simulation assumed that participants spend 20% of resources on food. Before HIP, 12% of food spending is for fruits and vegetables (Stewart and Blisard, 2008). Due to lack of existing data on food spending for TFVs, the simulation did not seek more precisely to distinguish TFVs from all fruits and vegetables. A low elasticity scenario assumed an own-price elasticity of -0.3; a high elasticity scenario assumed an own-price elasticity of -0.7. The low statutory cap is $5 per week per person, which is equivalent to approximately $16 of rebate per month per household; the high statutory cap is $40 per week per person, which is equivalent to approximately $129 of rebate per household per month.

of the plausible range for impacts that one might anticipate based on the financial incentive component alone.

Approach 2. Additional Estimates Produced for the Updated Study Plan

As an additional check, we revisited this question while drafting the Updated Study Plan. The second approach used a different method for linking the existing literature on spending elasticities to the task of estimating a change in daily food intake (in cups). Whereas the first approach (above) converted a change in food intake into a percentage change in food spending, the second approach applied elasticity estimates from the food spending literature to generate food intake estimates.

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 intake[4] is 2.0 cups (Dong and Lin, 2009; see also Blisard, Stewart, and Joliffe, 2004). Mean estimates of the elasticity of fruit and vegetable consumption with respect to price are 0.70 and 0.58, respectively (Andreyeva, et al., 2010). Together, these figures suggest an impact of HIP on fruit and vegetables of about 0.384 cups.[5] An impact of this magnitude is estimated to be detectable using this study's research design.

Of course, real-world results may differ greatly from the simulations. Due to data limitations, the simulations could not distinguish TFVs from other fruits and vegetables in some respects: the fruit and vegetable spending shares (Steward and Blisard, 2008) and price elasticities in the existing literature apply to all fruits and vegetables. Moreover, any marketing elements of HIP could enhance the pilot program's impact on fruit and vegetable spending and intakes (see question 2 below).

Input from TWG Members

In response to the analysis presented at the TWG meeting, some comments from TWG members agreed that the existing literature on financial incentives offers hope that HIP could have an effect on TFV intake as large as ¼ cup per person per day. French pointed out that the literature in agricultural economics and behavioral nutrition both indicate a large potential impact of price incentives. However, she noted that the individual intake outcome is different from the household level spending variable that the intervention most directly affects. Food spending (which is not closely measured in the HIP evaluation design) is a more proximate outcome, while food intake (which is a main focus of the HIP evaluation design) is a more distal outcome. Just said that the evidence from previous literature that the anticipated impact could be achieved is "extremely mixed," and felt that ¼ cup impact of fruit and vegetable impact would be at

[4] This estimate was computed from Dong and Lin (2009), Table 1: 2.0 cups = 1.43 Total Vegetables + 0.96 Total Fruits – 0.39 White Potatoes. As discussed in Section 6, "modified" targeted fruits and vegetables also exclude fruit juices, consumption away from home, and some fruits and vegetables consumed in combination with other foods.

[5] 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 upper end of reasonable anticipated effects. Gittelsohn suggested that, assuming there is good recognition and awareness by participants, the intervention could be effective.

Andrews noted that the prior literature is hindered by differing conventions about what foods are included in the food spending measure. HIP acts only on targeted fruits and vegetables. By contrast, the category for fruits and vegetables in much of the prior literature on food spending includes items that are not eligible TFVs under HIP. Comments from Andrews and Just suggested that more precisely distinguishing TFVs from all fruits and vegetables might lead to greater pessimism about the potential of HIP to achieve an increase of ¼ cup in TFV intake. On the other hand, in the oral presentation, Wilde pointed out that empirical elasticity estimates using a narrow category of good (such as TFVs) tend to be larger than elasticity estimates using a broad category of good (such as all fruits and vegetables), because there are more opportunities for consumers to substitute into or out of a good when the good is narrowly defined.

Brief Summary of Power Analysis

The main confirmatory outcome is the average treatment/control difference in targeted fruit and vegetable (TFV) intake based on Rounds 2 and 3 of the participant survey. The specific definition of intake in the main outcome is called "modified" targeted fruits and vegetables, as explained under Question #7 below. In a single round of the survey, the design is estimated to have a minimum detectable effect (MDE) of 0.164 cups targeted fruits and vegetables (TFV).[6] This MDE is below FNS's design goal of a quarter (0.250) cup. The MDE for the main confirmatory outcome based on two rounds of the survey will be even smaller.[7]

Conclusions

The impact of HIP on intake of Targeted Fruits and Vegetables is a genuinely open empirical research question, worthy of a pilot project with a strong research design. Two advance simulations, and input from the TWG, suggest that a ¼ cup increase in daily TFV is within the range that one might realistically expect from the financial incentive provided by HIP. Furthermore, HIP carries with it an implicit and explicit promotional message in favor of fruits and vegetables, which could increase the overall effect of the pilot beyond what was estimated in the simulations. Because the a ¼ cup effect is within the plausible range of likely effects, and the minimum detectable effect for the study is substantially smaller than a ¼ cup effect, the current design and sample size seem appropriate.

---

[6] This power calculation assumes 1,500 completed interviews, R-squared=12% (a conservative value based on Westat studies of the predictive power of the Fruit and Vegetable Screener), design effect=1.05 (for non-response), and conventional power parameters 1-α=80%; β=5%, and a one-sided test. A one-sided test is appropriate because we would treat a negative impact as equivalent to no impact. Source: personal communication with Janet Tooze; analysis from unpublished paper.

[7] How much smaller will depend on the correlation between TFV intake across the two waves. If the correlation was zero, the MDE would drop from 0.164 to 0.116. Given the high day-to-day variability in TFV consumption, it seems likely that the correlation will be low, so MDEs for the average impact across the two waves are likely to be about 0.140.

2. What features of how HIP is implemented are most likely to affect whether this potential effect is realized?

TWG members emphasized the importance of vigorous outreach to participants to promote awareness of HIP. This discussion covered whether HIP should focus on estimating the effect of a pure financial incentive, or instead whether nutrition education/promotion aspects should be encouraged; and how much promotion should be part of HIP. Several points of view were expressed:

- Isolate the impacts of a pure financial incentive as cleanly as possible.
- Promotional component must be economically and logistically feasible without large nutrition education budgets. Promotional efforts must be sufficiently low that they would be affordable, and the design of such promotional efforts must be sufficiently simple that they would be feasible for State agencies, if HIP is later rolled out nationwide.
- HIP necessarily marries a financial incentive with promotional aspects. An overly narrow focus on isolating the effect of a financial incentive in a 2-arm design (treatment vs. control) would fail to achieve the intended research goal, and it would reduce the probability that HIP would have an effect on TFV as large as ¼ cup per person per day.
- An option to the research contract would permit three research arms – a control group, a HIP financial incentive group, and a HIP plus nutrition education group. Distinguishing financial effects from marketing effects is only possible if this 3-arm design (treatment with nutritional education, treatment without nutritional education, control) is used.

Decision and rationale. The option to permit three research arms will not be exercised due to funding limitations and thus it will not be possible to test a third arm that includes both the financial incentive and nutrition education.

Instead, the financial incentive will include a low cost nutritional promotion component. Specifically, the following strategy will be used to promote fruits and vegetables among the HIP participants:

- DTA will incorporate language into the HIP "introduction" letter to SNAP recipients that highlights the importance of fruits and vegetables and explains why SNAP is providing an incentive for these foods. The text will be distinguished from the rest of the letter through use of a border, font changes, color, etc., such that recipients may be drawn to read it.
- There will not be a separate flier or document that provides messages or other promotion of fruits and vegetables.
- The sleeve produced for participants' EBT cards may have some appealing images of fruits and vegetables and may have a short promotion message.
- A reminder or prompt about the incentive will be sent to all HIP participants mid-way through the pilot. This reminder will consist of a new sleeve and a letter similar to the introduction letter, which will also include a paragraph that promotes the purchase and consumption of fruits and vegetables.

FNS, MA DTA, and Abt considered creating a separate HIP-specific flier, but decided against this for several reasons:

- Concerns about the ability of States to produce and distribute a flier as part of HIP if the program is rolled out nationwide;

- Lack of evidence that such a sheet would actually influence participants' purchases and;
- Reaching consensus and obtaining clearance from the *Dietary Guidelines* committee on the content is not feasible within the pilot schedule.

3. FNS is considering an incentive cap in the range of $30-$40 per month. What would you recommend as a lower bound on the incentive cap to ensure a sufficiently strong incentive?

TWG members agreed that it was important to set the incentive cap at the highest "safe" level. If the cap is set too low, it could limit the impact of HIP on households' expenditures for target fruits and vegetables, which could suppress the potential average impact of HIP. However, if the statutory cap is set too high, it could encourage fraud and cause the total incentive payments to exceed the budgeted amount of $2 million.

Members of the TWG raised several points:

- Concern that HIP would use a single household-level statutory cap that applies to all household types. The Abt team suggested that it is possible to choose a sensible statutory cap that is fixed and applies to all households.
- Cap estimates of $30-40 are based on assumptions that participants will use the full incentive up to the statutory cap. The Abt team suggested that many participant households will use far less than their entire maximum rebate. Because these households may have lower fruit and vegetable spending levels, money is freed up, so the fixed cap may be set somewhat higher without undue risk of exceeding the incentive budget.

It was agreed that Abt would undertake additional analysis to estimate the highest level for the statutory cap that can be afforded without exceeding the fixed budget for incentive payments. The estimates would use observed or assumed expenditure patterns, estimates of the sensitivity of expenditure to price, and various possible rebate caps.

Decision and rationale. The incentive cap will be set at $60 per month for all HIP households. This decision is based on analysis performed by Abt Associates and detailed in a memorandum (12/21/10). The analysis involved simulations using three alternative models of consumer purchasing behavior, representing varying levels of spending on fruits and vegetables. Using a $60 monthly cap and the most realistic assumptions of spending behavior, only 1 percent of households will have their behavior constrained by the cap (i.e. their fruit and vegetable spending eligible for the rebate will be limited by the cap). The households constrained by a $60 cap are those with 9 or more members whose SNAP benefit are greater than $200. This makes a clear case for a flat cap as opposed to one that is tied to household size or benefit level, given the operational difficulty and cost associated with a variable cap.

2. Participant Data Collection

The evaluation will collect data from pilot participants in three rounds—the first round prior to HIP implementation, round 2 at 3 months post-implementation and round 3 at 11 months post-implementation. The most important participant impacts are estimated from data on fruit and vegetable intake for adults based on a 24-hour dietary recall in the post-implementation surveys in round 2 and round 3.

4. We would like TWG's thoughts on the proportion of time allocated in the interview on each domain.

Most comments indicated that the time allocation is sound. The TWG's most substantial input on this topic concerned three issues, which are discussed in more detail in subsequent questions:

- fruit and vegetable expenditure measure (see Question 5);
- an additional 24-hour dietary recall in the pre-implementation survey in round 1 (see Question 8); and
- fruit and vegetable outcomes for children in the household (see Question 9).

Decision and rationale. Keep the time allocation in the participant survey instrument as proposed, maintaining a sharp focus on questions that achieve the project's most central evaluation goals and eliminating duplication in data collection. The instrument must be short to limit respondent burden and maximize response rates. To this end, we will evaluate how each question relates to our conceptual framework, while simultaneously retaining a sharp focus on a short list of questions that most strongly illuminate the overall HIP impact on intake of fruits and vegetables. We intend to eliminate duplication by asking questions on household characteristics and participation in nutrition assistance programs only in the baseline questionnaire.

5. To the extent possible, we will use survey questions that have been validated in other studies. What is your advice about the most appropriate sources for the following domains: (a) fruit and vegetable screener; (b) food security; and (c) food expenditures?

(a) Fruit and vegetable screener. Our proposal suggested using the NCI all-day fruit and vegetable screener in all three rounds. This will provide a consistent measure of fruit and vegetable consumption at baseline and allow for a comparison with Rounds 2 and 3. Several TWG members expressed support for using this screener; one member expressed misgivings about simple fruit and vegetable screeners, preferring to conduct additional 24-hour recalls (see Questions 8 and 9, below).

Decision and rationale. Based on the discussion, we will use the NCI all-day fruit and vegetable screener with some adaptations (updates in item wording to match current NHANES dietary screener, inclusion of a new item for intake of salsa, and deletion of the existing items on soup and vegetables consumed through mixed dishes).

(b) Food security. Our proposal initially suggested the short 6-item food security questionnaire, and we asked whether we should use a 12-month or 30-day reference period. The TWG suggested the 30-day reference period was fine for this application, and recommended the 10-item food security questionnaire. In subsequent discussion, Mark Nord at USDA/ERS indicated that the 6-item questionnaire would also suffice. The TWG also endorsed our proposal for a brief follow-up for households that reported running out of food, asking for further detail.

Decision and rationale. Based on further consideration of the issue and discussions with FNS, we decided to eliminate the food security questionnaire to reduce respondent burden. Food security is not a central outcome measure, nor is it useful for subgroup analyses. For example, we concluded it would be unsound to measure HIP/non-HIP differences separately for food secure and insecure SNAP participants, because food security is not an independent (exogenous) explanatory variable. In reviewing the instrument, we sought to eliminate questions that could not be used either as outcomes or as subgroup identifiers in this manner.

(c) Food expenditures. All TWG members agreed that it is important to measure fruit and vegetable spending if possible. These expenditures represent an important intermediate outcome. At the same time, the TWG clearly understood that there does not appear to be any good option for collecting detailed spending information. (Using a full consumer expenditure instrument, such as collecting UPC information through home scanner technology or collecting grocery receipts is out of the scope of the present study.)

NHANES and other major federal surveys ask broad questions about total spending in several types of food retail stores. Those broad questions do not ask about fruits and vegetables in particular. Several TWG members suggested that these questions work adequately, in their experience; others doubted that they could answer these questions themselves and questioned whether SNAP respondents would be able to do so.

Decision and rationale. While it is difficult to measure average household spending with self-reported survey questions, it is important that we do so. To better understand the difficulty with answering these questions, we evaluated four Consumer Expenditure Survey questions (one of which has been modified to collect expenditures on fruits and vegetables) in a cognitive testing setting. As a result of the testing, we changed the purchase recall window from one week to one month, which we found is more reflective of the way SNAP recipients purchase foods; and separated the first expenditure question into two questions, in order to separately ask about food purchases with SNAP cards and with other funding sources.

6. Focus groups

TWG members made several suggestions about the focus groups, which will be conducted twice during the post-implementation period.

- Focus group participants could be selected using random sampling rather than convenience sampling.
- Focus groups should include specific questions about (1) if information on grocery receipts concerning qualifying fruit and vegetable expenditures and incentives earned were used, and (2) how the pilot was perceived and understood.

Decision and rationale.

- We will use a convenience sample for the focus groups. The small sample size makes statistical generalization impossible, so there is little additional loss in selecting a convenience sample. Sampling for focus groups will aim to select a range of participant characteristics to capture the range of opinion and experiences with HIP. The discussion of focus groups in the report will clearly note the sampling approach used and will discuss generalizability.
- The recommended topics will be included in focus group discussions.

7. Definition of TFV based on 24-hour recall

Targeted fruits and vegetables (TFVs) exclude foods with added sugars, salt, and oils (see footnote 2). Additionally, SNAP eligible foods exclude food away from home and hot grocery store foods intended for immediate consumption. With 24-hour recall data it is not always possible to distinguish processed foods from foods that were prepared at home using a recipe. For example, it may be difficult to distinguish a vegetable recipe with oil added at home from a store-bought processed food with the same

composition. Yet, the vegetable purchased for home cooking would be eligible for HIP, while the store-bought processed food would not be eligible.

Options for defining the main TFV outcome variable for 24-hour recall analysis include:

(a) Define the variable to match actual HIP eligibility rules for food purchases as closely as possible (excluding restaurant food, and excluding foods with added sugars, salt, and oil);
(b) Define the variable to exclude food-away-from home, white potatoes, fruit juices, and mature legumes, but not excluding foods with added sugars, salt, and oil; or
(c) Define the variable only to exclude white potatoes, fruit juices, and mature legumes (but including other fruits and vegetables from restaurants).

Decision and rationale. For reasons of practical data processing, and to avoid emphasizing a nutritionally less relevant distinction between similar home-cooked and processed foods, the main outcome measure of fruit and vegetable consumption will be defined as (b), above. We will refer to this outcome as "modified target fruits and vegetables" (MTFV's).

8. Would you give up one of the post-implementation waves of 24-hour recalls to collect baseline 24-hour recalls? If yes, how would you propose using the baseline data?

The majority of TWG members would not give up one of the post-implementation rounds of 24-hour recalls, though one member would give up the first post-implementation round of dietary recall (round 2) if that sufficed to make the baseline dietary recall possible.

Use of baseline 24-hour recalls includes:

- Testing the equivalence of the treatment and control groups;
- Examining changes in food intake outcomes before and after implementation; and
- Testing for differential attrition.

Decision and rationale. The design will remain unchanged: collect 24-hour recalls at two points post-implementation and not during the pre-implementation period. This decision is based on several factors:

To be cost neutral, moving the dietary recall from Round 2 to baseline would require interviewing a smaller sample, since the baseline sample size is larger than the Round 2 sample.

The research strategy relies on random assignment to assure equivalence between the HIP and non-HIP arms of the study. Given random assignment, baseline data is not essential for a strong research design. We recognize that under some circumstances baseline data would improve precision; however, given the availability of a reliable FV screener, it does not appear that baseline data using a smaller sample size would increase precision of the main outcome.

Pre/post changes in the treatment group are secondary to treatment/control differences. Pre/post changes in the treatment group could be due merely to the seasonality of TFV consumption (perhaps induced by the seasonality of TFV availability and price).

- The analysis will examine, and, if needed, correct for differential attrition based on information in the case records and in the baseline survey.

9. Are there any alternatives to 24-hour recall data that you consider worth the investment for obtaining information about food consumption among children in the family? If yes, please describe.

If you do not see any worthwhile alternatives to 24-hour recalls, would you be willing to sacrifice power in adult analyses in order to get child observations?

The TWG expressed consistent interest in obtaining more information about impacts on children. Suggestions included:

- Include children older than 10 years in the sampling frame for the dietary recall;
- Administer brief parent-reported questions about child fruit and vegetable intake, perhaps just during the part of the day that the child is at home.
- Administer child dietary intake screeners such as those collected in ECLS-K or YRBS.

However, most TWG members would not sacrifice power in adult analyses to add questions about children.

Decision and rationale. Collection of outcome data for children is not feasible in this study due to fixed study resources and a lack of feasible instrumentation that could be implemented at a smaller (less costly) scale. Adding a parent questionnaire about child intake would not yield data of sufficient quality and lowering the age of eligibility for the 24-hour recall would have significant cost implications and thus not feasible within the current budget. Maintaining power in adult analyses is crucial to answering the central research questions.

10. What elements of the family food environment are most important to capture? How do you see them moderating the main impact?

TWG members suggested extensive research literature with relevant survey questions, particularly on home fruit and vegetable availability, fruit and vegetable pantry management, outcome expectancies for purchasing fruits and vegetables, fruit and vegetable shopping practices, and social support for purchasing fruits and vegetables.

Decision and rationale. The participant survey will include several questions on the family food environment. Measures include:

- Availability of fruits and vegetables at home;
- Availability of salty snacks, low-fat/fat-free milk, and soda/fruit-flavored drinks at home;
- Evening meals—eaten together as family; and cooked at home

At the same time, we have limited the number of questions about the family food environment to minimize respondent burden. In designing the survey instrument, priorities include questions that can be used as outcome variables or to measure differential HIP impacts by subgroups, as well as descriptive questions that are essential to characterize the SNAP population under study. Additional information about the family food environment will be collected from the focus groups.

3. Participant Impact Analysis

11. Please consider the protocol for addressing multiple comparisons and provide comments.

We will report a wealth of primary and secondary outcome results, for the full sample and for selected subgroups. In order to ensure that hypothesis tests and p-values are valid, it will be essential to correctly

account for the large number of outcomes and therefore the large number of hypothesis tests being conducted.

Decision and rationale. The protocol we will follow:

- Identifies a primary outcome or a single composite of primary outcomes in advance (i.e., before any data analysis). In the language of the multiple comparisons literature, we will treat this test as "confirmatory."
- If the HIP/non-HIP difference in the primary outcome is statistically significant, use conventional thresholds to report statistical significance of the components of any composite and secondary outcomes and subgroup analyses. In the language of the multiple comparisons literature, treat these tests as "exploratory" (and explain clearly what that means).
- If the HIP/non-HIP difference in the primary outcome or primary composite is not statistically significant, accompany all discussion of secondary outcomes with a clear caveat about the multiple comparisons issue.

12. The primary nutritional outcomes that we are proposing to use are: targeted fruits and vegetables (TFV) -- 24-hour recall and 30-day recall using the NCI Screener. Do you agree that these are the most appropriate? If not, what would you suggest?

The TWG members generally agreed with the identification of primary outcomes.

Decision and rationale. The primary outcome will be modified target fruits and vegetables (MTFVs) from 24-hour recall. This outcome measure will use both the round 2 and round 3 post-implementation surveys and be defined as the average of the two 24-hour recalls.

We considered using a composite outcome, incorporating an outcome measure from the fruit and vegetable screener in addition to the MTFV outcome from 24-hour recall. However, computations based on relevant data (personal communication with Janet Tooze, data presented in unpublished paper) suggest little gain in precision.

13. Secondary outcomes include: discretionary calories and total calories. Do you agree?

In addition to these secondary outcomes, TWG members suggested:

- Other foods and beverages, for example, sugar, snacks, and eating out.
- Measures of fruits and vegetables separately. Based on previous research, several TWG members suggested that interventions may have an easier time changing the targeted fruits and than the targeted vegetables.
- Intermediate constructs such as liking fruits and vegetables.

Decision and rationale. Based on TWG suggestions, questions about fruits and vegetables will be separated in several parts of the instrument. The participant survey will also include a small number of representative questions about attitudes toward fruits and vegetables, which will be analyzed as secondary outcomes.

14. What is your advice on the most important subgroup analyses that should be done? One thought would be to stratify on baseline (pre-HIP) TFV intake.

As priority subgroups, TWG members identified basic household composition and demographic categories—presence of children, female headship, family size, education, benefit level—as well as pre-HIP fruit and vegetable intake. One suggestion was to add residential location in food deserts versus other areas.

Decision and rationale. Although we cannot identify food deserts and non-deserts within the pilot site, we will explore using survey questions characterizing barriers to access as a basis for subgroup analysis. Other suggested categorical variables will be used in subgroup analyses.

15. Should we sample those no longer on SNAP?

Retaining those no longer on SNAP would entail costs and difficulties. The advantage of retaining them would be:

- permitting an intention-to-treat design, thereby ameliorating analytic difficulties introduced by potential differential sample exit between treatment and control groups;
- allowing analysis of what happens when the external reward is removed; and
- examining any unintended negative responses to the incentive.

Decision and rationale. The sample frame from which HIP households will be drawn consists of SNAP cases active as of the end of July 2011. Households will be excluded from the sample frame if they: live in long-term residential facilities (and turn their benefits over to the facility); receive only disaster SNAP benefits; and receive $0 benefits.

Rules concerning HIP eligibility include:

- Participation in HIP will be tied to the head of household;
- If a HIP household changes composition, HIP participation will be retained by the original head and any new or retained members of the household;
- If the head of a HIP household leaves SNAP and re-enters during the pilot, s/he will continue as a HIP participant; and
- HIP participants who move away from Hampden County will be retained as they will still be able to earn incentives shopping at retailers who roll-out HIP for all of Massachusetts.

SNAP participants sampled for the evaluation will be followed as long as they remain in their original household or become part of a household whose HIP status is the same as their original status. For example, a respondent from a HIP household will remain in the analytic sample as long as s/he is either in the same household or becomes part of a different HIP household.

SNAP households who are not on SNAP during data collection periods will not be interviewed. If such households return to SNAP during a subsequent data collection period, they will be interviewed.

16. Should we oversample those with large SNAP benefits as these households are likely to have larger impacts?

In discussing oversampling, the goal is not actually to cherry-pick households with bigger effects, but rather to increase our ability to measure distinct impacts on households with larger SNAP benefits while –

in weighted analyses – still representing the entire pilot site SNAP population in the main estimates. Even so, TWG members recommended against oversampling.

Decision and rationale. We will not oversample households with large SNAP benefits.

4. Stakeholder and EBT Data Collection and Analysis

As currently designed, we will conduct the first round of the retailer survey just prior to HIP implementation and the second round several months before the end of the demonstration. The timing of the first round allows collection of baseline data on food prices and minimizes the recall period for questions about decisions to participate.

17. Would it be preferable to conduct the first round of the retailer survey early in the implementation process? This would give us data on early implementation problems and experiences that could be included in the Interim Report.

TWG members recommended that pre-implementation baseline data on retailers is essential. They also acknowledged the potential value of having a retailer survey fairly early in the implementation process.

Decision and rationale. We will retain the original plan to collect one round of retailer data prior to implementation and the second round later in the implementation process. Cost considerations prohibit data collection at three points in time.

18. We need to keep the retailer survey relatively brief to assure an adequate response rate, especially for retailers surveyed twice. Are there topics that should be dropped or scaled back because of low priority or limitations of the study design?

The TWG emphasized the importance of the retailer survey to the project as a whole and did not suggest deleting any topics. One member suggested more extensive retailer data collection, including measuring transaction times, which may be an economically important factor influencing retailer support.

Decision and rationale. We will retain the self-administered survey with telephone follow-up. Self-administered surveys with telephone follow-up allow retailers to complete the survey when they have time.

While resources preclude in-person interviews with all sampled retailers, stakeholder interviews with integrated retailers, will include issues such as transaction times, corporate efforts to assure fidelity of implementation, and perceived success of these efforts. In addition, participant focus groups and stakeholder interviews, will explore stakeholder perceptions of how retailers implemented HIP.

19. The analysis plan includes two analyses using EBT transaction data. Are these the most useful analyses that can be done with these data? Are there other feasible and valid analyses that should be considered?

The discussion favored these two priorities for EBT analysis.

Decision and rationale. The priorities for EBT analysis include:

- Take-up of HIP incentive by participants; and
- Impact on retailer redemptions.

Sources Cited

Andreyeva, T., M. W Long, and K. D Brownell. "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, no. 2 (2010): 216.

Dong, D., and B. H Lin. "Fruit and vegetable consumption by low-income Americans: would a price reduction make a difference?." Washington, DC: USDA Economic Research Service, Economic Research Report 70 (2009).

Mabli, J., P. Gleason, and J. Hall. Final Memorandum for HIP Power Analysis. Washington, DC: Mathematica Policy Research, Inc. (2009).

Stewart, H., and N. Blisard. "Are lower income households willing and able to budget for fruits and vegetables?" Washington, DC: USDA Economic Research Service, Economic Research Report No. (ERR-54) 29 (2008).

Blisard, N., H. Stewart, and D. Jolliffe. "Low-income households' expenditures on fruits and vegetables." Washington, DC: USDA Economic Research Service, Agricultural Economic Report No. 833 (2004).

Appendix A: Healthy Incentives Pilot Evaluation Technical Working Group (TWG) Member Bios

Tom Baranowski, PhD

Tom Baranowski is a Professor in the Department of Pediatrics at Baylor College of Medicine. His research is directed toward understanding why children eat the foods and engage in the physical activities they do as well as designing and evaluating programs to help change these dietary and physical activity behaviors in community settings. He is principal investigator for several grants that incorporate experimental research design related to diet and physical activity. He is author or co-author of over 260 peer-reviewed articles, 10 non-peer reviewed articles, 20 book chapters, and two editions of a textbook on methods of evaluation for health promotion programs. Tom graduated from Princeton University in 1968 with a B.A. in Politics, and received an M.A. and Ph.D. in Social Psychology from the University of Kansas.

Simone French, PhD

Simone French is a Professor in the Division of Epidemiology and Community Health in the School of Public Health at the University of Minnesota. Dr. French has conducted research in the area of obesity prevention for the past 18 years, with a focus on social and environmental influences. Dr. French's research has evaluated interventions to promote healthful food choices in community settings such as worksites, schools, and households. These environmental interventions have included experimental studies in schools and worksites to evaluate the use of pricing incentives for healthful food choices such as fruits and vegetables. She is the Co-Director of the Obesity Prevention Center at the University of Minnesota and the author of over 100 publications and was a member of the Committee on Prevention of Obesity in Childhood and Youth held by the Institute of Medicine. Simone received her Ph.D. in Psychology from the University of Minnesota.

Joel Gittelsohn, PhD, MSc

Joel Gittelsohn is a Professor in the Department of International Health at Johns Hopkins Bloomberg School of Public Health who specializes in combining qualitative and quantitative approaches for the design, implementation and evaluation (including process evaluation) of health and nutrition intervention programs. Dr. Gittelsohn’s research investigates culture-based beliefs and behaviors regarding dietary patterns, and how these factors influence the success or failure of dietary and lifestyle modification strategies. He applies these methods and interventions for the prevention of obesity and diabetes among different indigenous and ethnic groups including African Americans, American Indians, First Nations, Inuit and Pacific Islanders. He is the author of over 122 peer-reviewed articles the areas of environmental and behavioral interventions in food stores, schools, churches and households. Joel received his Ph.D. in Medical Anthropology from the University of Connecticut.

David Just, PhD

David Just is an Economist and Associate Professor in the Department of Applied Economics and Management at Cornell University. His work focuses on applications of behavioral economics to agricultural markets, including risk response, information use and food consumption. Several of his studies have won wide recognition, including the *American Journal of Agricultural Economics* Outstanding Journal Article (2009, 2008 honorable mention) and being cited by *Discover Magazine* as one of the top science stories of 2006. He is an author on several reports from the Economic Research Service at the USDA regarding about how behavioral economics can be used to design incentives for healthy food consumption and its potential for improving diet quality among food assistance program recipients. David received in Ph.D. in Economics from the University of California, Berkeley.

Diane Whitmore Schanzenbach, PhD

Diane Whitmore Schanzenbach is an Associate Professor in the School of Education and Social Policy at Northwestern University and faculty research fellow at the National Bureau of Economic Research. Her research focuses on education policy, child health, and food demand analysis. Her work on SNAP has measured how households alter their consumption of food, leisure and other goods when they receive SNAP, and whether the benefits improve the health of recipients. She graduated magna cum laude from Wellesley College with a B.A. in economics and religion, and received a Ph.D. in Economics from Princeton University.