Technical Memorandum: TWG Meeting on HIP Evaluation Analysis

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

April 2013

Prepared for:
Danielle Berman
US Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

Submitted by:
Abt Associates
55 Wheeler Street
Cambridge, MA 02138

Table of Contents

Summary Memorandum

Introduction

The Healthy Incentive Pilot (HIP) investigates the impact of making fruits and vegetables more affordable for SNAP participants by offering financial incentives at the point of sale. The HIP Evaluation uses a random assignment research design to measure the impact of the pilot intervention on fruit and vegetable intakes and other outcomes.

The evaluation plan included formation of a Technical Working Group (TWG) to provide external advice and input on the evaluation. The first meeting of the TWG, held on October 5, 2012, provided feedback on the evaluation design. The second meeting of the TWG was held on March 13, 2013 with the objective of discussing the HIP Interim Report and plans for Final Report analyses.

The TWG is composed of five outside experts in relevant areas of research:

- Tom Baranowski, Children's Nutrition Research Center, Baylor College of Medicine;
- Simone French, University of Minnesota;
- Joel Gittelsohn, Bloomberg School of Public Health, Johns Hopkins University;
- David Just, Cornell University; and
- Diane Whitmore Schanzenbach, Northwestern University.

All five TWG members attended the meeting in person. In addition, Laura Tiehen, USDA Economic Research Service, attended the meeting as a federal government expert. FNS staff attending the meeting were: Danielle Berman, Kelly Kinnison, Anita Singh, and Kathryn Law.

This Technical Memorandum summarizes the discussion at the meeting, focusing on comments raised by TWG members and the ensuring conversation. We organize the document according to the five sessions of the meeting: (1) Highlights of TWG implementation; (2) Overview of interim analysis; (3) Group discussion of issues raised by the interim analysis; (4) Participant impact analysis; and (5) HIP purchases and incentive earnings. The meeting agenda and presentation slides are included as Appendix A and Appendix B, respectively. Appendix C provides a technical discussion of mediator analyses (referenced in Section 4, below).

Session 1: Highlights of HIP Implementation

In the first session, Susan Bartlett presented on the highlights of the implementation of HIP, including challenges faced in implementation and early operations. TWG members had several questions and comments:

- ***Technical systems issues***. In the early months of operations, two technical systems issues affected incentives earned for some HIP participants. How were the issues discovered? And what does the fact that HIP participants noticed (or did not notice) the problems indicate about their awareness and understanding of HIP?
    - **Action step:** elaborate, as appropriate, in reports.
- ***Retailer participation in HIP.*** How extensive was coverage of all retailers and of farmers markets? (Clarified that all farmers markets in Hampden County participated, representing 15 percent of HIP participating stores.)

- **Action step:** Interim Report will clarify that retailer coverage was better than the number of participating stores suggests, because over half (57 percent) SNAP purchases in Hampden County are made at HIP participating retailers. In the Interim Report, we will add the limited information currently available about the percent of Hampden County SNAP purchases redeemed in HIP participating retailers. The Final Report will provide additional information on this topic.

Session 2: Overview of Interim Analysis

In the second session, Lauren Olsho provided an overview of the analysis covered in the Interim Report. Two points were clarified:

- Our survey does not include the full sample of those who were randomly assigned to HIP. We do have EBT data for all participants, but survey costs did not allow the evaluation to survey all participants assigned to HIP.
- For some analyses in the Final Report, we plan to link EBT data and survey data.

In addition, questions and discussion centered around the definition and coding of the evaluation's confirmatory outcome measure, modified targeted fruits and vegetables (MTFV), and more broadly how dietary recall data were collected and coded. The issues raised and our planned responses are presented in the next section.[1]

Defining Modified Targeted Fruits and Vegetables (MTFV)

The primary objective of the HIP evaluation is to determine the effectiveness of HIP in increasing intake of those targeted fruits and vegetables (TFV); i.e., those fruits and vegetables eligible to earn the HIP incentive payment. However, since the incentive is applied at the point of purchase, while the dietary recall interview asks about foods as consumed (rather than as purchased), it was not feasible to construct an exact measure of TFV. Instead, the Interim Report described two proxy measures, lower-bound and upper-bound modified targeted fruits and vegetables (MTFV) intake, which appeared to be the best approximation of true TFV intake that could be constructed using available data. In addition, the Interim Report designated the lower-bound intake measure ("lower-bound MTFV intake") as the study's confirmatory outcome.

TWG members raised several questions and issues about the two MTFV proxies:

- ***Measure Construction.*** TWG members probed about which food items are included in the "upper bound" versus "lower bound" MTFV intake measures. Several specific issues were discussed:
  - Treatment of mixed foods. Mixed foods are defined for our purposes as foods that contain both fruit and/or vegetable ingredients as well other non-fruit, non-vegetable ingredients. For mixed foods with TFV ingredients recorded in the recall interview database as a single food code, we cannot ascertain whether the food was purchased

[1] Since the TWG meeting, we have obtained additional information about our ability to reliably create outcome measures for cup-equivalents of fruits and vegetables by form of preservation. Only 45 percent of vegetables and 36 percent of fruits can be classified by form of preservation. Thus, we cannot create these outcome measures.

in already-prepared form, in which case the included fruits and vegetables would not have earned the HIP incentive payment, or whether it was prepared from individual fruit and vegetable ingredients, in which case the included fruits and vegetables would have earned the HIP incentive payment if purchased from participating retailers. Upper-bound MTFV includes intake from mixed foods, while lower-bound MTFV excludes intake from mixed foods. The upper-bound measure thus likely captures some fruit and vegetable intake not actually from TFV, while the lower-bound measure likely excludes some TFV intake. It is in this sense that the two proxy measures represent upper and lower bounds on true TFV intake.

- Example – If reported as a single food item, "green salad with salad dressing," "chili con carne," and "broccoli with butter" would each be included in upper-bound MTFV but not lower-bound MTFV under our current coding scheme. (More detail provided below.)
- **Action step**: We will clarify discussion of measure construction in the revised Interim Report, including a diagram indicating various components of lower- and upper-bound MTFV (the briefing slides included such a diagram). In addition, we will add an appendix with further details in the Final Report.

- Coding depends on how foods are reported. In practice, whether mixed foods are represented by a single food code or by multiple codes depends on whether the ingredients are reported together or separately. If a mixed food was prepared at home, respondents are asked as part of the interview protocol to list individual ingredient components. If they are able to do so, then ingredients are represented by multiple food codes in the dataset, allowing us to explicitly identify whether individual fruit and vegetable ingredients are TFV. In this situation, TFV intake from the mixed food would be included in both upper- and lower-bound MTFV. On the other hand, if the respondent cannot identify individual ingredients, either because the food was not prepared at home or for any other reason, the food will be represented by a single default food code, and fruit and vegetable intake from that food would be included in upper-bound MTFV, but not lower-bound MTFV.
  - Example – a stir fry with vegetables made at home. If respondents can report the ingredients, vegetable intake from the stir fry would be included in both the lower-bound and upper-bound measures. If respondents cannot report ingredients, vegetable intake from the stir fry would be reported only in the upper-bound measure.
  - Example – a chicken pot pie with broccoli, carrots, and potatoes purchased in prepared form. In this case, the pot pie would be reported using a single default food code, and thus would be included only in the upper-bound measure.
- Is lower-bound MTFV the "preferred" measure? The Abt team clarified that the reason for selecting lower-bound MTFV over upper-bound MTFV as the confirmatory outcome measure was because it is considered a more conservative measure, in the sense that it does not include intake from mixed foods that may not have qualified to earn the HIP incentive.

  - Further refinement of MTFV may be possible. Westat staff attending the meeting who have been involved in the food coding process noted that they may be able to separate out fruit and vegetable intake for some foods currently classified as "mixed foods," e.g. broccoli with butter, where it can be inferred that the other ingredients were added during the preparation phase.
    - **Action step**: Westat will work on refining coding of the MTFV measures in accordance with the above. We will use the revised coding in the analyses for the Final Report. This implies that the Round 2 estimates in the Final Report may diverge from those in the Interim Report.
- ***Terminology.*** One TWG member disliked the characterization of the measures as "upper and lower bounds" due to potential confusion about these terms as often used in the context of sensitivity analyses.
  - **Action step:** Though we plan to retain the current terminology, we will clarify use of the terms in the text of the report.
- ***Fatigue.*** There was concern about a fatigue factor in reporting consumption. As respondents get further into the survey would they be more likely to report 'mixed foods' because they don't feel like reporting the individual ingredients? Some studies involving multiple days of 24-hour dietary recall interviews do show a systematic decrease in the number of foods reported in later interviews.
  - **Action step**: We will examine whether there is any evidence of this in our data, by counting the number of foods reported and any differences between the first interview and the second interview for the 10 percent subsample as well as across the two rounds of the AMPM survey.
- ***Rationale.*** There was a question about the rationale for distinguishing between salad bought with salad dressing versus a salad prepared at home with salad dressing added. The original rationale is to distinguish between what qualifies for HIP (individual ingredients of salad made at home) and what doesn't (prepared salad with dressing included). However, we do provide results for other fruit and vegetable aggregates (e.g. total fruit and vegetable intake), that show influences on intake more broadly defined.

Session 3: Group Discussion of Issues Raised by the Interim Analysis

In the third session, Lauren Olsho moderated a discussion during which the TWG identified issues raised by the interim analysis. Issues raised included:

- ***The limited awareness of HIP***, including possible explanations for the limited awareness, and how awareness and understanding may impact consumption and potentially mediate the impact of HIP on consumption. We also clarified that focus groups also indicated confusion about participating retailers and qualifying foods.
  - **Action step:** Will use conduct mediator analysis using this concept. (See the review of Session 4 below for more discussion of this issue.)
- ***The analysis of "any" versus "total cups" of consumption.*** The impact of greater alcohol consumption among HIP participants raised the issue of whether the impact would still occur if alcohol consumption was dichotomized into any versus none, which led to the issue of

"any" versus "total cups" of TFV. This analysis would examine the notion of whether more people are consuming TFVs.

  - **Action step:** We will perform exploratory analyses of several binary outcomes in the Final Report, including any/no alcohol consumption and (as already specified in the Updated Study Plan) any/no lower-bound MTFV intake.

- ***Halo effects***, such as the impact on fruit juice, which does not qualify for HIP. Are there other food items we can isolate to see if there is a halo effect?
  - **Action step:** We will perform exploratory analyses of other non-qualifying fruits and vegetables and of other mixed foods (for which we cannot definitively establish whether or not they qualify).
- ***Communication about HIP when rolling out widespread implementation***, and how this would differ from the pilot given the constraints of communication due to random assignment. Would widespread implementation improve understanding?
  - **Action step:** Final Report analysis that examines potential nationwide expansion of HIP will consider this issue.
- ***Low retailer participation***, reasons for low participation, how this might differ with a permanent program, and potential differential impacts on shoppers in participating versus non-participating stores.
  - **Action step:** Final Report analysis that examines potential nationwide expansion of HIP will consider this issue. Also planned retailer analysis will examine whether HIP households changed SNAP purchases to HIP-participating retailers.
- ***How meaningful the interim findings were***, in terms of the overall impact on health and the impact on high-risk groups. There appeared to be disagreement among TWG members about whether the estimated impacts were large enough to be substantively important. One comment suggests considering the trade-off between program impact and cost (i.e., if the program costs are higher, this implies greater incentive use and greater impact; if the impact is low, however, costs will also be low).

Finally, there were several questions/issues raised during the discussion that we subsequently clarified in the revised version of the Interim Report, or noted (in annotated comments submitted with the revised report) that the issues would be addressed in the Final Report:

- Whether ***study participants who went off SNAP were retained in the sample***.
- Why ***spending was higher in November*** compared to the other months.
- ***Shopping behaviors***, including spending in participating stores, and why we would expect any changes in shopping behavior.
- Why the ***average caloric intake*** was around 200 calories lower than the average American caloric intake.

Session 4: Participant Impact Analysis

The afternoon began with a presentation and discussion on participant impact analyses led by Jacob Klerman. The session focused on four topics:

- Confirmatory outcome for the evaluation;
- Longitudinal analysis;
- Moderator analysis;
- Mediator analysis.

These topics, including decisions reached, are discussed in separate sections below.

In addition, throughout the participant impact analysis discussion session, FNS emphasized the importance to them of "telling a story".

- **Action Steps:**
  - We will present the findings in the Final Report in a way that is clearer to a non-research audience by using a better narrative to describe what we have done and to describe the findings. Additional graphical presentations will also be used. Particular attention to these issues of exposition will be given in the Executive Summary, the introduction, and the conclusion.
  - We will also discuss with FNS the language to use in describing treatment-control differences, in particular framing the observed differences as changes caused by HIP. We will also augment the methods section, explaining the connection between random assignment research design and the ability to give a cause and effect interpretation to the key HIP/non-HIP outcome differences. This discussion will explain the technical circumstances in which is it statistically valid to make interpretation sentences that include active and dynamic verbs, such as "All else equal, when a person joins HIP, his or her daily targeted fruit and vegetable intake increases on average by 0.2 cups."

Confirmatory Outcome Measure

We discussed our proposed confirmatory outcome of lower-bound MTFV intake, pooled across the Round 2 and Round 3 data. Pooling (sometimes called "stacking") the data makes the most efficient use of the available data. It involves creating an analysis file with one record per 24-hour recall. Most survey respondents will have two 24-hour dietary recalls, one at Round 2 and the other at Round 3. The 10 percent of respondents (at both Rounds 2 and 3) who completed a second dietary recall interview will have an additional 24-hour dietary recall record. The resulting file is then analyzed, "clustering" on individual; i.e., allowing for correlation between 24-hour recalls of a given individual, but assuming independence across individuals.

- **Action step:** After the discussion, the TWG concurred with the earlier decision as to confirmatory outcome; as well as the decision to implement that concept for Round 3 by stacking the data.

Longitudinal Analysis

Our approach to longitudinal analysis is to compute four estimates for each outcome:

1. *Pooled estimate*—this single estimate from the Round 2 and Round 3 dietary intake recalls is the most powerful test of the impact of HIP.
2. *Round 2 impact estimate.*
3. *Round 3 impact estimate*.
4. *Round 2/Round 3 change estimate*—explores whether maturation of the program (and participation in the program) changed the program's impact over time.

In most cases, the discussion in the Final Report will focus on the pooled results. In addition, when—based on the test for R2/R3 change—there is clear evidence of a difference in impact between Rounds 2 and 3, we will discuss this evidence for a difference and the specific estimates for Round 2 and for Round 3. As of now, it is our expectation that statistically significant R2/R3 changes will be rare (incentive earned data does not suggest a major change over time; power to detect changes is weak).

The presentation in the Final Report will include graphics as a way to clearly illustrate the story. For example, we anticipate including a line graph that plots the three rounds of outcomes based on the fruit and vegetable screener, separately for the HIP and non-HIP groups. Depending on the findings, we might present some outcomes as bar charts (HIP vs. non-HIP), with separate sets of bars for the pooled estimate, Round 2 estimate, and Round 3 estimate.

There was also an interest in conventional panel data models that explore change in behavior over time for an individual and how that change over time varies with treatment status (i.e., receipt of HIP). During the discussion, Abt noted that in many cases these models are formally identical to the Round 2 vs. Round 3 models. The Round 2 vs. Round 3 models first compute treatment/control differences and then compute Round 2/Round 3 differences of the treatment/control differences. In the random assignment setting, the panel data models first compute Round 2/Round 3 differences and then compute the treatment/control difference of the Round 2/Round 3 differences. In simple cases, the result is algebraically identical, regardless of the method (i.e., which difference is taken first). In most complicated models, the results are not necessarily algebraically identical, but they are conceptually similar. Nevertheless, some readers are likely to find it easier to think of the analysis one way; others are likely to find it easier to think about the analysis the other way.

- **Action step:** We will report (in an appendix) limited models in which we first compute Round 2/Round 3 differences and then take the Treatment/Control difference, assuming FNS thinks this presentation would be valuable.

Moderator Analysis

Abt expressed concern about the large number of analyses implied by doing all possible subgroup analyses for all outcomes. Abt's background materials proposed limiting both the outcomes considered and the subgroups considered.

Specifically, Abt propose to conduct subgroup analyses for four intake outcomes:

1. Lower Bound-Modified Target Fruits and Vegetables (LB-MTFV)
2. LB-MT Fruits
3. LB-MT Vegetables
4. All Fruits and Vegetables (including LB-MTFV plus fruits and vegetables from mixed foods, white potatoes, 100% fruit juices, and fruits and vegetables obtained from restaurants and other non-store locations).

TWG members agreed that these were the most important intake outcomes and would be the only ones included for moderator analysis of intake. Inasmuch as some subgroups appear to be consistently and strongly statistically significant, we will explore those subgroups for a broader set of other outcomes, including attitudes and preferences.

The background materials for the meeting (and the briefing slides) included a long list of subgroups. The discussion suggested the following changes to that list:

- Drop household income subgroups;
- Include White, Black, and Hispanic subgroups (as opposed to White/Non-White);
- Add disability status (from case record);
- Retain all suggested subgroups on the likely amount of benefit as well as attitudes, barriers, and family food environment; and
- Drop subgroups related to shopping patterns.

There was also interest in subgroups on geographic area and distance from the grocery store. However, current plans do not include geocoding individual respondents, so these subgroups are not feasible.

There was also discussion of continuous versus binary subgroups, and attendees ultimately agreed on the possibility of examining some continuous moderators, including the scales and/or SNAP benefit amount.

- **Action step:** The final list of subgroups includes the following *demographic and economic* subgroups:

    - Gender (male/female)
    - Age (16-40/41+)
    - Primary shopper employment status (working/not working)
    - Household composition (children in HH/no children in HH)
    - Household monthly SNAP benefit amount ($200 or less/over $200)
    - WIC participation (yes/no)
    - Race/ethnicity (white/black/Hispanic)
    - Disabled (yes/no)

    To test variation based on the *likely amount of HIP incentive received*, we will test four binary subgroups based on a respondent's predicted HIP incentive[2] and fruit and vegetable consumption on the fruit and vegetable screener:

    - Any predicted incentive/no predicted incentive
    - High predicted incentive/low predicted incentive
    - Any fruit and vegetable consumption on screener/no fruit and vegetable consumption on screener

[2] To estimate predicted incentive amounts, we will first estimate a regression model with the HIP incentive amount as the dependent variable, and baseline respondent and household characteristics as explanatory covariates. Estimated coefficients from this regression will be used to construct a predicted incentive amount for each sampled household, and the sample will be stratified based on this predicted value.

- High fruit and vegetable consumption on screener/low fruit and vegetable consumption on screener

We will also examine four subgroups based on attitudes, barriers, and family food environment: [3]

- Positive attitudes towards food, fruits, and vegetables (high/low)
- Barriers to eating fruits and vegetables (high/low)
- Barriers to grocery shopping (high/low)
- Fruits and vegetables available in the home (more/less frequently)

Mediator Analysis

The TWG meeting included a fruitful discussion of mediator analysis. As with the longitudinal analysis discussed earlier in this section, this discussion had multiple objectives. In part, the discussion was motivated by interest in a particular family of regression models for analyzing mediators. In part, it was motivated by interest in improving the narrative account of key results pertaining to intermediate variables.

In the discussion , TWG members agreed that discussions of mediator analysis should plainly state that giving mediator analysis a causal interpretation requires very strong assumptions about omitted variables, assumptions that are not justified by the random assignment framework of the broader evaluation. Nevertheless, augmenting the treatment/control analyses and the moderator analyses with some limited single mediator analyses—probably in an appendix and with appropriate caveats—seems appropriate. Appendix C provides a detailed discussion of mediator analysis, particularly the assumptions required to make causal inferences.

- **Action step:** After the group discussion, individual discussions with FNS staff, Diane Schanzenbach, and Tom Baranowski, and further discussion among the Abt analysis team on the pros and cons of mediator analyses in the context of a random assignment design, we plan to conduct a **limited** amount of single mediator analyses, using the Baron and Kenney (1986) method. These mediator analyses will be included in an appendix of the final report. Suggestions for single mediators to include were:
  - Food spending
  - Barriers
  - Awareness/Understanding of HIP
  - Attitudes
  - Receiving messages about fruits and vegetables

[3] These subgroups are based on an indexes, not single items.

Session 5: HIP Purchases and Incentive Earning

The final session, led by Parke Wilde, included a presentation and discussion of analyses on HIP purchases and incentive earning. Discussion focused on two topics: HIP purchases in participating retailers and zero-incentive earners.

HIP Purchases in Participating Retailers

Using data for both the HIP and non-HIP groups, we will estimate the impact of HIP on mean monthly purchases of HIP-eligible foods in supermarkets/superstores (the two principal IECR-equipped participating retailer types). In these stores, we can observe treatment/control differences in purchases. We will also break down overall HIP purchases into intermediate steps, showing in greater detail how any observed HIP/non-HIP differences arose. First, we will estimate experimental HIP/non-HIP differences in EBT transactions in participating and non-participating retailers. Then, we will examine HIP/non-HIP differences in purchase of TFVs.

- **Action step:** TWG members agreed that if a participant's decision to shop at a participating versus non-participating store is not responsive to HIP then the decision to purchase TFVs versus other items can be interpreted as an impact estimate. Otherwise these can just be discussed as descriptive outcomes.

Zero-incentive Earners

In the Final Report, we will conduct additional analyses of participants who purchased few or no TFV's (called zero-TFV purchasers). Wilde discussed (a) a two-group approach in which we compare food intake outcomes for HIP and non-HIP participants who were predicted to be likely zero-TFV purchasers based on independent explanatory variables and (b) a three-group approach in which we compare food intake outcomes for actual non-HIP participants, HIP participants with zero-TFV purchasers, and HIP participants who were positive-TFV purchasers..

- **Action step**: Due to concerns about confounding variables in the three-group approach, the TWG meeting participants advised the two-group approach for examining the impact of zero-TFV purchasers. In descriptive analyses only, we will also plan to compare those in the treatment group who do and do not use HIP.

FNS and TWG members were also interested in:

- Those who maximized the incentive versus those who did not use the incentive all. However, people who maximize (hit the cap) are very rare – a very low percentage either received $60 or used all of their SNAP benefit. We could consider dividing people with positive incentives into low, medium, and high incentive groups for comparison purposes.
- How learning has happened. The Interim Report examines selected outcomes over time. We will consider other variables for analyses of month to month changes.

Appendix A: Agenda

Healthy Incentives Pilot (HIP) Evaluation
Technical Work Group (TWG) Agenda
March 13, 2013

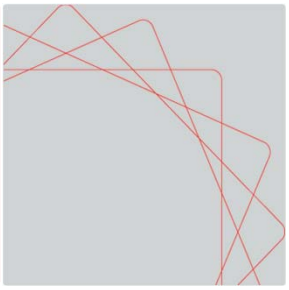
9:30-10:00	Opening Remarks • Introductions • Overview of Meeting	Danielle Berman and Susan Bartlett
10:00-10:30	Highlights of HIP Implementation	Susan Bartlett
10:30-11:00	Overview of Interim Analysis	Lauren Olsho
11:00-12:00	Issues Raised by Interim Analysis	Group Discussion
12:00-1:00	Lunch	
1:00-2:45	Participant Impact Analysis • Next Steps: pooling across waves; additional outcomes; longitudinal analysis; additional subgroup analyses; mediator analysis	Jacob Klerman
2:45-3:00	Break	
3:00-4:30	HIP Purchases and Incentive Earning • Next Steps: understanding low incentive earning; additional subgroup analyses; analysis of households with no incentive earnings	Parke Wilde
4:30-5:00	Next Steps and Closing	Susan Bartlett

Appendix B: PowerPoint Slides

Healthy Incentives Pilot (HIP) Evaluation

Technical Work Group Meeting

March 13, 2013



HIGHLIGHTS OF HIP IMPLEMENTATION

Highlights of HIP Implementation

- Background and objectives of HIP
- HIP intervention
- HIP implementation
- Challenges of implementation

HIP Background and Objectives

- Investigates impact of making fruits and vegetables more affordable to SNAP participants
- Authorized under 2008 Farm Bill
- Determine if financial incentives at point of sale increase consumption of fruits, vegetables, and other healthful foods

HIP Intervention

- Financial incentive
  - 30% incentive on SNAP purchases of Target Fruits and Vegetables (TFV)
  - Incentive in the form of additional SNAP benefits
- Target fruits and vegetables
  - Same as for WIC fruit and vegetable vouchers
  - Excludes juice, white potatoes, mature legumes
  - No added sugar, salt, oils (with some exceptions)

HIP Intervention

- Limits on TFV spending eligible for incentive
  - Must be SNAP spending (not cash or WIC voucher)
  - Must be spent in retailers participating in HIP
  - Subject to a cap
- Two mechanisms for affecting outcomes
  - Vigorous financial incentive
  - Implicit or explicit fruit and vegetable promotion

HIP Pilot Site

- Hampden County, Massachusetts
- County has lowest median income and highest poverty rate in State
- Western MA has highest obesity rates and related chronic illnesses
- 7,500 SNAP households randomly assigned to participate in HIP

Overview of HIP Implementation & Operations

- HIP planning stage: December 2009-November 2011
- Roll-out over 3 months: November 1, 2011-January 1, 2012
- 2,500 households began HIP each month
- Households earned HIP incentives for 12 months
- Pilot close-out occurred over 3 months: November 1, 2012-January 1, 2013

HIP Implementation

- Complicated undertaking that involved:
  - Assembling team
  - Designing and implementing EBT and SNAP eligibility system changes
  - Recruiting retailers to participate in HIP
  - Developing participant notification and training materials
- Implementation posed many challenges
- Pilot began on the intended schedule

Implementation Challenges

- Technical systems issues
- Retailer participation
- Participant understanding of HIP

Technical Systems Issues

- HIP transaction processing code inadvertently removed from 11 stores for 45 day period during December-January 2011
  - 1,140 households affected
  - Incentives subsequently credited
- Databases of HIP-eligible items not accurately maintained by 2 chain retailers in early 2012
- Some receipts not accurately reflecting HIP purchases/incentives earned

Retailer Participation

- Recruitment was challenging both for chain retailers and independent retailers
- Approximately 120 stores participated in HIP
  - 40% large supermarkets
  - 30% convenience stores
  - 15% small grocery stores
  - 15% farmers markets
- One major supermarket chain did not participate
  - Result: 40% of supermarkets participated in HIP

Participant Understanding

- DTA spent considerable time and effort developing notifications and training materials
- Notifications
  - Household received 3 notices at beginning of pilot
  - Several notifications during pilot—updated list of retailers, farmers markets, simple brochure
- Attendance at trainings low—less than 2% of HIP participants attended
- Feedback from participants suggest some confusion about pilot

Overview of HIP Interim Analysis

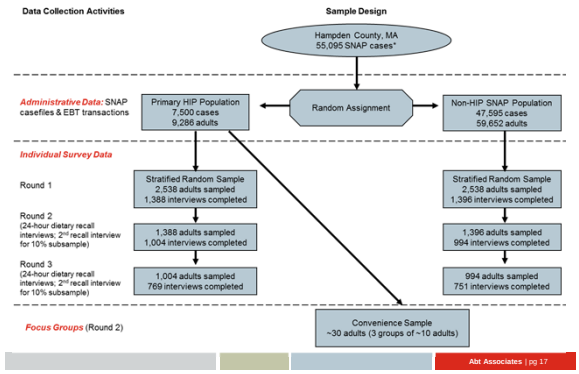
Research Objectives

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

Research Objectives

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

Evaluation Design Overview



Targeted Fruits & Vegetables

Any variety of fresh whole or cut fruit without added sugars.[a]

Any variety of fresh whole or cut vegetable, except white potatoes, without added sugars, fats, or oils (orange yams and sweet potatoes are allowed). [a]

Any variety of canned[b] fruits (must conform to FDA standard of identity (21 CFR Part 145); including applesauce, juice pack or water pack without added sugars, fats, oils, or salt (i.e. sodium). Any variety of frozen fruits without added sugars.[c]

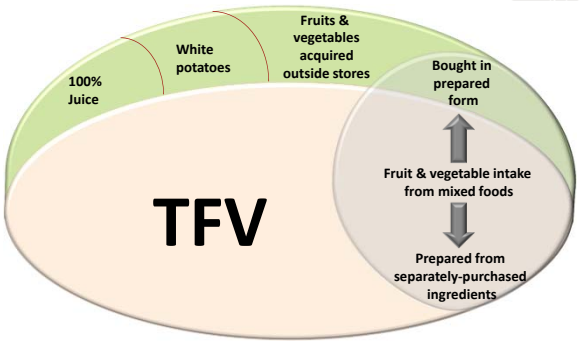
Any variety of canned[b] or frozen vegetables (must conform to FDA standard of identity (21 CFR Part 155)) except white potatoes (orange yams and sweet potatoes are allowed); without added sugars, fats, or oils. May be regular or lower in sodium.[c]

Any type of dried fruits or dried vegetable without added sugars, fats, oils, or salt (i.e., sodium).[a]

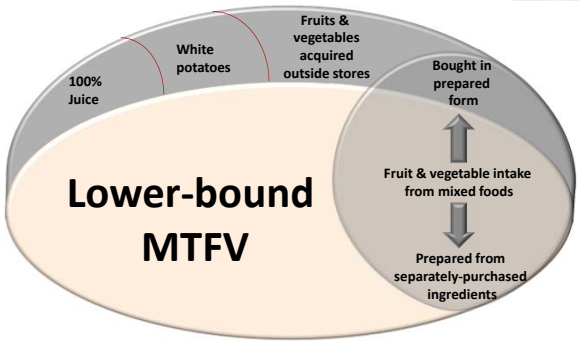
[a]Herbs or spices; edible blossoms and flowers, e.g., squash blossoms (broccoli, cauliflower and artichokes are allowed); creamed or sauced vegetables; vegetable-grain (pasta or rice) mixtures; fruit-nut mixtures; breaded vegetables; fruits and vegetables for purchase on salad bars; peanuts; ornamental and decorative fruits and vegetables such as chili peppers on a string; garlic on a string; gourds; painted pumpkins; fruit baskets and party vegetable trays; and items such as blueberry muffins and other baked goods are not authorized. Mature legumes (dry beans and peas) and juices are not authorized.
[b]"Canned" refers to processed food items in cans or other shelf-stable containers, e.g., jars, pouches. Home canned fruits and vegetables, such as those sold at Farmers' Markets, are not allowable.
[c]Excludes white potatoes; catsup or other condiments; pickled vegetables, olives; soups; juices; and fruit leathers and fruit roll-ups.

Source: Reproduced from FNS Request for Application, Supplemental Nutrition Assistance Program (SNAP), Healthy Incentives Pilot (HIP), CFDA #:10.580, Figure 1.

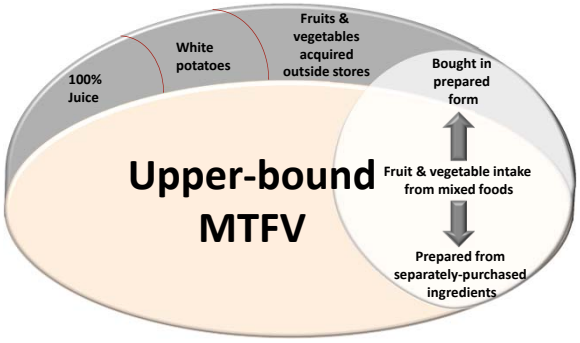
TFV and Total Fruits and Vegetables



Measuring TFV Intake



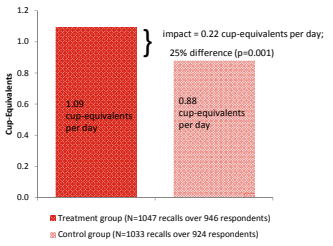
Measuring TFV Intake



Two TFV Proxy Measures

- Establish upper and lower bounds on actual TFV intake
- Both measures exclude
  - white potatoes
  - 100% fruit juice
  - fruit and vegetable intake from non-store-acquired foods
- Difference is exclusion (lower-bound MTFV) or inclusion (upper-bound MTFV) of fruit and vegetable intake from mixed foods
- We take lower-bound MTFV as our confirmatory outcome

Impacts on Lower-Bound MTFV



Other Fruit & Vegetable Impacts

	Mean		Impact			
	Treat.	Control	T-C	[S.E.]	{t-stat}	(P-value)
Lower-bound MTFV	**1.093**	**0.878**	**0.215**	**[0.064]**	**{3.373}**	**(0.001)*** **
Plus MTFV from mixed foods	0.425	0.431	-0.006	[0.032]	{-0.192}	(0.848)
Upper-bound MTFV	**1.518**	**1.309**	**0.209**	**[0.072]**	**{2.926}**	**(0.003)*** **
Plus additional components:						
100% fruit juice	0.594	0.477	0.117	[0.051]	{2.269}	(0.023)**
White potatoes	0.361	0.351	0.011	[0.032]	{0.332}	(0.740)
Other fruits & vegetables acquired outside stores	0.257	0.251	0.006	[0.029]	{0.206}	(0.837)
All fruits and vegetables	**2.731**	**2.388**	**0.343**	**[0.101]**	**{3.395}**	**(0.001)*** **

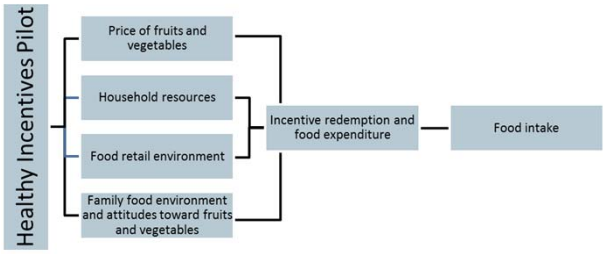
Fruit and Vegetable Groups

- Statistically significant impacts on intake of both fruit and vegetables considered alone (Exhibit 6.3)
- Intake higher in HIP group for both MyPyramid fruit subgroups…
  - Citrus, melons, & berries
  - Other fruits (e.g. apples, pears, bananas, grapes, peaches)
- …but only a subset of vegetable subgroups.
  - Dark green vegetables
  - Tomatoes
  - Other vegetables (e.g. celery, cucumbers, mushrooms, green beans, onions, asparagus)
  - NOT starchy or orange vegetables

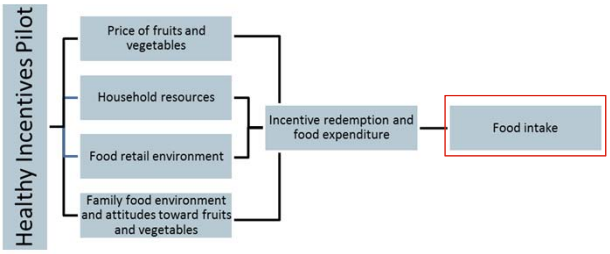
Other Intake

- Interim Report Exhibits 6.6, 6.7, 6.8
- HIP group had higher intake of
  - Refined and total grains
  - Alcohol
  - Vitamin C
- No difference detected in
  - Other MyPyramid food groups
  - Total energy
  - Other nutrients (Vitamin A, fiber, beta carotene)

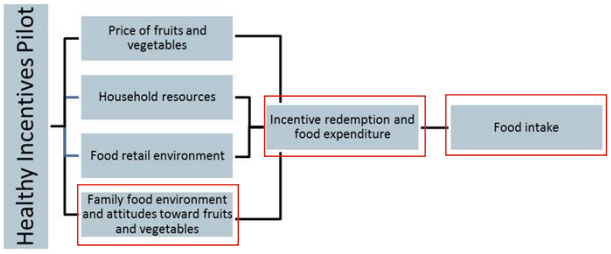
Logic Model



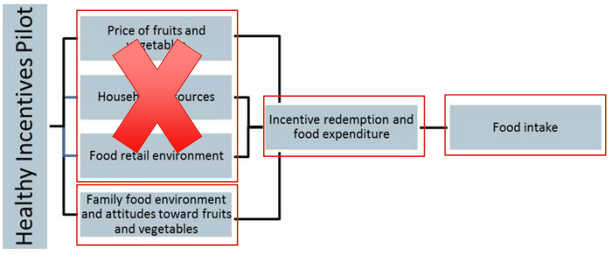
Logic Model



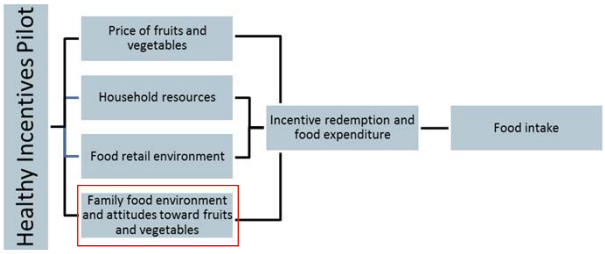
Logic Model



Logic Model



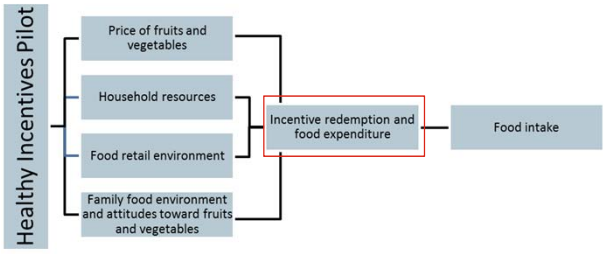
Logic Model



Attitudes & Family Food Environment

- HIP group more likely to report exposure to messages about fruits and vegetables in past month (Exhibit 4.1)
- More likely to have fruits and vegetables available at home (Exhibit 4.2)
- No consistent impacts on
  - Food preferences & beliefs (Exhibit 4.3)
  - Perceived barriers to fruit & vegetable consumption (Exhibit 4.4)

Logic Model



HIP Take-Up

- In each month, about two thirds of the HIP group made at least one HIP purchase
- Households making HIP purchases differed from other households (Exhibit 5.2):
  - Larger SNAP benefit
  - More likely to have a child in the household
  - Younger household head

Total HIP Spending

- Roughly one-third of households had zero HIP purchases
- Average monthly HIP purchases ranged from $10-$12 after initial roll-out period
  - corresponding to $3-$4 incentive earned
- Note this presents a puzzle: take-up rates and total TFV spending imply lower increases in intake than we actually observed

Monthly HIP Purchases

Self-Assessed Changes in Shopping Patterns

- 70% of HIP respondents agreed that HIP made fruits and vegetables more affordable (Exhibit 5.7)
- 60% agreed that they bought more and a greater variety of fruits and vegetables due to HIP (Exhibit 5.7)
- Only about a quarter reported that they changed their shopping location due to HIP (Exhibit 5.8)
- Note these measures are non-experimental (HIP respondent self-reports only)

Impacts on Shopping Patterns & Spending

- No consistent impacts on shopping patterns:
  - Usual grocery store location type (Exhibit 5.9) or reasons for choosing that location (Exhibit 5.10)
  - Grocery shopping frequency (Exhibit 5.11)
  - Perceived barriers to grocery shopping (Exhibit 5.11)
- Higher self-reported monthly spending on fruits and vegetables in HIP group ($7.39) (Exhibit 5.12)

Subgroups

- Subgroup analyses (Exhibit 6.5) intended to describe potential variation in impacts:
  - Respondent gender (males vs. females),
  - Respondent age group (age 16-40 years vs. 41+ years),
  - Primary shopper employment status (employed full- or part-time vs. not employed),
  - SNAP benefit size ($200 or less vs. over $200), and
  - Household composition (households with children vs. households with no children)
- No differences in impacts across subgroups detected (borderline difference by primary shopper employment status)
- Note analysis was not explicitly powered for subgroups

Conclusions

- Higher fruit and vegetable consumption in HIP group (25% higher intake of lower-bound MTFV)
- Broadly consistent with expected increase given price elasticities from the literature
- Conundrum in explaining increase in intake given relatively low HIP take-up and spending
- Differences in attitudes and self-reported spending provide evidence on possible mechanisms
- Subgroup analyses provide evidence on heterogeneity in impacts

Discussion

- Final part of morning session is reserved for general TWG member comments/reactions
- Note that afternoon sessions will cover two key issues raised in written TWG comments:
  - Mediator/moderator analyses
  - Spending/intake conundrum

Impact Analysis Issues

Goal for this Session

- Motivate and describe our proposed approach and some technical issues in specifying models for the Impact Analysis
- Better understand issues, motivation and specific models suggested by FNS and TWG members

Specific Discussion Topics

- Main (Pooled) Analysis
  - Handling multiple comparisons/ fruit and vegetable summary measures
  - Longitudinal analysis
- Moderator Analysis
  - Which subgroups? For which outcomes?
  - Continuous moderators
- Mediator analysis

Two Definitions

- ***Moderator Analysis***: A moderator is a variable that influences the strength of a relationship between two other variables
  - Usually suggests a subgroup analysis (or an interaction of the moderator variable with the random assignment dummy variable)
- ***Mediator Analysis***: A mediator is a variable that explains the relationship between the two other variables
  - Usually suggests a relationship between the intervention (in our case, HIP), an intermediate outcome (in our case, attitudes or incentives earned), and the ultimate outcome of interest (in our case, TFV Intake)

Summary of the Argument

	Pooled	Moderators	Mediators
Research Questions			
Methods			
Assumptions			
Comments			

Pooled Analysis

- ***Research Question***: What is the impact of HIP on TFV/Targeted Fruits and Vegetables?
  - This RQ corresponds to the policy choice: *Fund/do not fund the program*
- ***Methods***: RA/Random assignment and then T/C (treatment/control) comparisons
  - In practice, regression adjusted (not needed for unbiasedness/consistency; improves precision)
  - Regressors include baseline measures, including components of EATS FV Screener
- ***Assumptions***: Minimal
  - (properly implemented) Random assignment guarantees that T and C are balanced on unobservables, so T/C differences estimate the impact of the program
  - i.e., outcomes w/HIP relative to outcomes w/o HIP

Multiple Comparisons

- Multiple Comparisons
  - Conventional statistical tests give proper p-values taking each test separately
  - In practice, we run hundreds of tests
  - Even if there were no significant impacts, looking over all tests, by chance, some tests would appear significant
- We address multiple comparisons by pre-specifying one primary and confirmatory outcome
  - For Interim Report: LB-MTFV, in R2 data (including 10% "usual intake sample")
  - For Final Report: LB-MTFV, in pooled R2 and R3 data (including 10% "usual intake sample")
- We then determine the success of the program based on the estimated impact for this confirmatory outcome

FV Concepts

	Mean		Impact			
	Treat.	Control	T-C	[S.E.]	{t-stat}	(P-value)
Lower-bound MTFV	**1.093**	**0.878**	**0.215**	**[0.064]**	**{3.373}**	**(0.001)*** **
Plus MTFV from mixed foods	0.425	0.431	-0.006	[0.032]	{-0.192}	(0.848)
Upper-bound MTFV	**1.518**	**1.309**	**0.209**	**[0.072]**	**{2.926}**	**(0.003)*** **
Plus additional components:						
100% fruit juice	0.594	0.477	0.117	[0.051]	{2.269}	(0.023)**
White potatoes	0.361	0.351	0.011	[0.032]	{0.332}	(0.740)
Other fruits & vegetables acquired outside stores	0.257	0.251	0.006	[0.029]	{0.206}	(0.837)
All fruits and vegetables	**2.731**	**2.388**	**0.343**	**[0.101]**	**{3.395}**	**(0.001)*** **

Multiple Comparisons, Again

- We report—and discuss—results for all outcomes
  - For all outcomes, we report "uncorrected" standard errors"
- This appears to be the standard approach in random assignment analyses
  - We have taken this approach in other FNS analyses (FFVP/Fresh Fruit and Vegetable Program, SEBTC/Summer EBTC for Children)
- Choice does not appear to be consequential
  - Results for LB-MTFV, UB-MTFV, all F&V are qualitatively similar (t=3.33, 2.94, 3.39)

FV Concepts

	Mean		Impact			
	Treat.	Control	T-C	[S.E.]	{t-stat}	(P-value)
Lower-bound MTFV	**1.093**	**0.878**	**0.215**	**[0.064]**	**{3.373}**	**(0.001)*** **
Plus MTFV from mixed foods	0.425	0.431	-0.006	[0.032]	{-0.192}	(0.848)
Upper-bound MTFV	**1.518**	**1.309**	**0.209**	**[0.072]**	**{2.926}**	**(0.003)*** **
Plus additional components:						
100% fruit juice	0.594	0.477	0.117	[0.051]	{2.269}	(0.023)**
White potatoes	0.361	0.351	0.011	[0.032]	{0.332}	(0.740)
Other fruits & vegetables acquired outside stores	0.257	0.251	0.006	[0.029]	{0.206}	(0.837)
All fruits and vegetables	**2.731**	**2.388**	**0.343**	**[0.101]**	**{3.395}**	**(0.001)*** **

Changing this Decision

- It's late to revisit this decision
  - We and you have already seen the R2 outcomes
  - R2 outcomes are a strong signal of pooled R2/R3 outcomes
  - So, we could not argue that any new choice was "pre-specified"
- We could still revisit how we pool R2 and R3 data
  - We originally proposed averaging
  - We now believe that superior approaches exist
  - Instead, we propose "stacking" all 24-hour recalls and "clustering"
  - This is how we handled the 10% "usual intake sample" in the Interim Report analyses

Longitudinal Analysis

- Current Approach:
  - Regression with baseline measures (including baseline measures of The EATS battery) as regressors
  - This is usually considered the most efficient way to use baseline measures of outcomes in random assignment analyses
  - *Plus*, estimating and testing for change in impact R2 to R3
- Alternatives:
  - Construct R3-R2 and model it in terms of treatment
  - Where we have R1 measures; pool R1/R2/R3, include individual fixed effects, and dummy variables for treatment (and perhaps wave)
  - Some form of "growth curve" analysis
- Questions:
  - What is the research question?
  - What is the proposed analytic approach?

Summary of the Argument

	Pooled	Moderators	Mediators
Research Questions	What is overall impact of HIP?		
Methods	T/C comparisons, regression adjusted		
Assumptions	Minimal		
Comments	Random assignment eliminates concerns about unobservables		

Moderator Analysis

- ***Research Question***: How does the impact of HIP vary with observable characteristics? (e.g., SG/subgroups)
  - This RQ corresponds to the policy choice: *Fund/do not fund the program for a SG*
  - Moderator analysis may also be informative about "mechanism" (e.g., impact on high baseline FV vs. impact on low baseline FV)
- ***Methods***: RA and then T/C comparisons w/in SG
  - In practice, only discuss results if reject equality of impacts across subgroups
- ***Assumptions***: Minimal, as long as SG are defined at baseline (so not affected by intervention)

Specifying Subgroup Analyses I

- For three complementary reasons, we prefer to limit subgroup analyses
- First, otherwise we will "cut down a lot of trees"
  - Our current specification has 79 outcomes and 22 subgroups; i.e., outcomes for 5,214 outcomes = 79 outcomes x (pooled + SG-A + SG-B)
- Second, this raises multiple comparison problems
  - Even if there were no differential impacts, some of these outcomes would appear to be different from each other
  - One (partial) approach to this is to pre-specify the subgroups of primary interest

Specifying Subgroup Analyses II

- Third, our power to estimate differential subgroup impacts is not strong
  - Impact on LB-MTFV has a large t-statistic of 3.4
  - Second wave of data, doubles sample size, but observations are correlated
  - (best guess) Pooling two waves will cut standard error by 25%; i.e., t-statistic of ~5
  - Best case (even split between subgroups) standard error of estimate of difference will be: 0.10 (vs. 0.05 for pooled analysis)
  - So can only detect a differential impact of 0.19 cup equivalents (i.e., 85% of main impact); which seems unlikely

Subgroups in Interim Report

- We considered 5 subgroups
  - Only one test of equality (for employment status) was borderline significant: p=0.057
  - In that case, impact was nearly zero in one subgroup
  - Others were clearly insignificant: p=0.718, 0.551, 0.815, 0.516
- Simply adding more data
  - Would push p=0.057 to 0.018
  - But other four will not become significant

We expect to find few significant differential subgroup effects. We would be happy to be shown wrong…

Subgroup Analysis for …

1. Lower Bound-Modified Target Fruits and Vegetables (LB-MTFV)
2. LB-MTFruits
3. LB-MTVegetables
4. All Fruits and Vegetables (including LB-MTFV plus fruits and vegetables from mixed foods, white potatoes, 100% fruit juices, and fruits and vegetables obtained from restaurants and other non-store locations).

Are there other variables we should add? Why?

Subgroups to be Considered

- Our current plan is to consider 21 subgroup dimensions/42 subgroups
  - Those in the *Interim Report* (5/10)
  - Demographic Subgroups (4/8)
  - Likely Amount of HIP Benefit Received (4/8)
  - Attitudes, Barriers, Family Food Environment (based on indexes; 4/8)
  - Shopping Patterns (4/8)
- This seems like a long list
  - What is the argument for each of these subgroups?
  - Can we drop some?
  - Are there some we should add?

Subgroups to be Considered I

- **As in the *Interim Report*** (5/10)**:** Gender (male/female), Age (16-40/41+), Primary shopper employment status (working/not working), Household composition (children in HH/no children in HH), Household monthly SNAP benefit amount ($200 or less/over $200)
- **Demographic Subgroups** (4/8): WIC participation (yes/no), Household income ($787 or less/$788 or more; i.e., split at the median), Household income adjusted for household size (above/below sample median), Race/ethnicity (white/non-white)

Subgroups to be Considered II

- **Likely Amount of HIP Benefit Received** (4/8)**:** Any predicted incentive /no incentiverebate, High predicted incentive/low predicted incentive, Any FV consumption on screener/no FV consumption on screener, High FV consumption on screener/low FV consumption on screener
- **Attitudes, Barriers, Family Food Environment** (based on indexes; 4/8): Positive attitudes towards food, fruits, and vegetables (high/low), Barriers to eating fruits and vegetables (high/low), Barriers to grocery shopping (high/low), FV available in the home (more/less frequently)

Subgroups to be Considered III

- **Shopping Patterns** (4/8)**:** Choose usual shopping place because it's close to home (yes/no), Choose usual shopping place because of affordable prices (yes/no), Choose usual shopping place because of variety of products (yes/no), Choose usual shopping place because of better or fresher produce (yes/no)

Can we drop any of these?
Are there other subgroups we should add? Why?

Other Moderator Analyses

- In Interim Report: All moderator analyses were based on binary subgroups
- For continuous variables, we dichotomized approximately at median
- Questions:
  - Should we interact the continuous moderator with Treatment?
  - How should we report the results?

Summary of the Argument

	Pooled	Moderators	Mediators
Research Questions	What is overall impact of HIP?	How does impact vary with (pre-treatment) characteristics?	
Methods	T/C comparisons, regression adjusted	T/C comparisons w/in groups (and tests for equality)	
Assumptions	Minimal	Minimal	
Comments	Random assignment eliminates concerns about unobservables	Sometimes subgroup analysis is informative about mechanism	

How HIP Works and for Whom

- i.e., we proposed to address how HIP works and for whom, through a combination of:
  - Subgroup analyses, and
  - T/C analyses of intermediate outcomes
    <see list on next slide>

Impact on Intermediate Outcomes

- Interim Report includes models of the impact of random assignment status on several intermediate variables:
  - Awareness of HIP (E3.4), Problems w/EBT Card (E3.8)
  - Exposure to Nutritional Education (E4.1), Food Environment (E4.2), Food Preferences and Beliefs (E4.3), Perceived Barriers (E4.4)
  - Usual Grocery Story Type (E5.9), Reasons for (E5.10), Shopping Behaviors and Barriers (E5.11)
  - Expenditures (E5.12)
- Are there others that we should model?

How HIP Works and for Whom

- i.e., we proposed to address how HIP works and for whom, through a combination of:
  - Subgroup analyses, and
  - T/C analyses of intermediate outcomes
  - Analyses w/in Treatment group of the correlates of HIP incentive receipt <see next session>
- Several comments from FNS and TWG members suggest mediator analysis

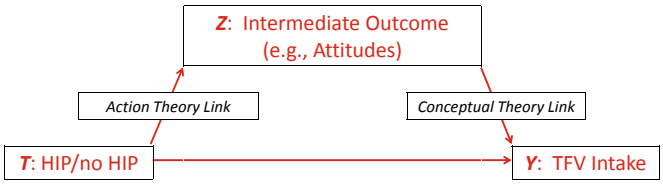
We would like to devote the balance of this session to discussion of these issues

We Would Like to Understand

- What is the corresponding research question?
  - What does it mean?
  - Why/how is that information useful?
- What specific methods/models would you suggest?
- What are the implicit assumptions required for those models?
  - Are those assumptions reasonable in this context? Why?
- Perhaps an example will help …

Example: FV Attitudes

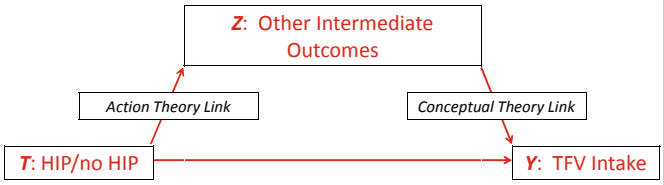
- ***Research Question***: *Does HIP work through lower prices or through changing attitudes/knowledge of FVs?*



- What does this question mean? Why is the answer useful?
- What model would we estimate?

Other Examples

- ***Research Question***: *Does HIP work through lower prices or some other pathway?*



- What does this question mean? Why is the answer useful?
- What model would we estimate?

Questions for Discussion

- What is the corresponding research question?
  - What does it mean?
  - Why/how is that information useful?
- What specific methods/models would you suggest?
- What are the implicit assumptions required for those models?
  - Are those assumptions reasonable in this context? Why?

Summary of the Argument

	Pooled	Moderators	Mediators
Research Questions	What is overall impact of HIP?	How does impact vary with (pre-treatment) characteristics?	What are the mechanisms through which HIP impacts participants?
Methods	T/C comparisons, regression adjusted	T/C comparisons w/in groups (and tests for equality)	Path Analysis/ Structural Equation Modeling
Assumptions	Minimal	Minimal	No unobservables or exclusions
Comments	Random assignment eliminates concerns about unobservables	Sometimes subgroup analysis is informative about mechanism	These assumptions are not supported by random assignment

BACK-UP SLIDES

… Strong Assumptions

- A model for mediation will have the following form:

$$Y = \theta_0 + T\theta_T + Z\theta_Z + X\theta_X + v$$

- i.e., we want to estimate
  - θ_Z giving the effect of HIP "through *Z*", and
  - θ_T gives the "residual effect" of HIP not "through *Z*"
- The problem is omitted variables bias
  - To estimate the quantities of interest, we need that (conditional on the observed and included *X*s) there are no unobservables that both affect *Z* and directly affect *Y*
  - When this assumption is violated , θ_Z does not give the effect of change *Z*
- That assumption seems (to us) implausible
  - If that assumption would be valid, you would not need random assignment
  - Just run the program in some places and compare (regression adjusted) outcomes

Alternative Assumption

- There is some variable that affects *Z*, but does not directly affect *Y*
  - Under those conditions, one can apply instrumental variables/two-stage least squares/structural equation models
- But no such variables are immediately apparent

SHOPPING, SPENDING, AND INCENTIVES EARNED

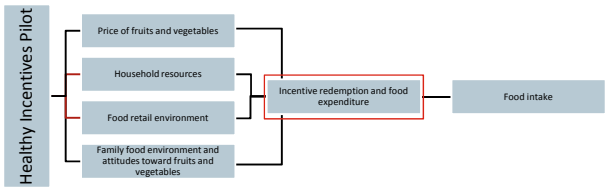
Outline

- Overview and motivation
- Data
  - EBT
  - Survey
- Analysis
  - SNAP Redemptions by Retailer Type
  - HIP Purchases in Participating Retailers
  - Smaller Analyses

Outline

- **Overview and motivation**
- Data
  - EBT
  - Survey
- Analysis
  - SNAP Redemptions by Retailer Type
  - HIP Purchases in Participating Retailers
  - Smaller Analyses

Logic Model



A Puzzle from the Interim Report

- Lower-bound MTFV intake for HIP participants: 1.09 cup-equivalents per person per day.
- HIP purchases: $12.90 per household per month.
- Rough reconciliation (from Chapter 7): The HIP purchases might supply approximately 0.5 cup-equivalents per day.
- Where did HIP participants acquire the remainder of their TFV intake?

Ways of Acquiring TFVs

- HIP purchases from participating retailers
- TFV purchases from non-participating retailers
- Cash resources for TFV purchases from any retailers
- Other (WIC, food pantries, school meals)

Outline

- Overview and motivation
- **Data**
  - **EBT**
  - **Survey**
- Analysis
  - SNAP Redemptions by Retailer Type
  - HIP Purchases in Participating Retailers
  - Smaller Analyses

Data: EBT

- Population
  - 6,000 HIP participant households
  - 40,000 non-HIP participant households
- Variables
  - HIP purchases (purchases of HIP-eligible TFVs)
  - HIP incentives earned

Household-Month Analysis File

- Monthly SNAP spending by retailer type
- Monthly SNAP benefit amount
- Household size
- HIP identifier
- Survey sample identifier

Data: Retailer Types

	Participating retailers		Non-participating retailers
	IECR	Non-IECR	
Supermarket/Superstore	1		6
Grocery and specialty		3	7
Convenience	2	4	8
Other		5	9

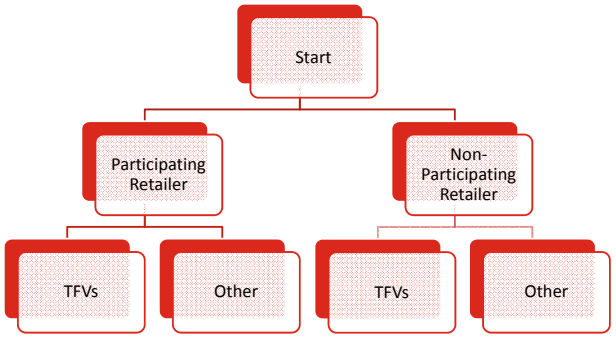
Data: Survey

- Outcomes from shopper survey (rounds 2 and 3)
  - Self-reported shopping behavior changes
  - Self-reported food expenditure

Outline

- Overview and motivation
- Data
  - EBT
  - Survey
- **Analysis**
  - SNAP Redemptions by Retailer Type
  - HIP Purchases in Participating Retailers
  - Smaller Analyses

A HIP Participant's Decision Tree



Outline

- Overview and motivation
- Data
  - EBT
  - Survey
- Analysis
  - **SNAP Redemptions by Retailer Type**
  - HIP Purchases in Participating Retailers
  - Smaller Analyses

Redemptions by Retailer Type

- Design
  - Impact estimates from HIP/non-HIP comparison
- Outcomes: SNAP redemptions
  - $ per household per month
  - % of all SNAP redemptions
- Retailer types
  - All 4 retailer types
  - Participating and non-participating retailers

Outline

- Overview and motivation
- Data
  - EBT
  - Survey
- Analysis
  - SNAP Redemptions by Retailer Type
  - **HIP Purchases in Participating Retailers**
  - Smaller Analyses

HIP Purchases (1)

- Design
  - Descriptive statistics for HIP participants
- Outcomes: HIP purchases
  - $ per household per month
  - % of all SNAP redemptions
- Retailer types
  - All 4 retailer types
  - Participating retailers, both IECR and non-IECR

HIP Purchases (2)

- Design
  - Descriptive comparison of HIP and non-HIP participants
  - Interpretation is conditional on prior shopping decision (may approximate impact estimates if shopping effects are negligible)
- Outcomes: eligible TFV purchases
  - $ per household per month
  - % of all SNAP redemptions
- Retailer types
  - 2 retailer types combined
  - IECR-equipped participating retailers

Outline

- Overview and Motivation
- Data
  - EBT
  - Survey
- Analysis
  - SNAP Redemptions by Retailer Type
  - HIP Purchases in Participating Retailers
  - **Smaller Analyses**

Zero-Incentive Earners

A. Do they predominantly shop in non-participating retailers?
   - Geographic barriers
   - Retailer preferences combined with low responsiveness to HIP incentive

B. Do they have no HIP purchases even when shopping in participating retailers?
   - Low preferences for fruits and vegetables
   - Limited understanding of HIP

Impact of Zero-Incentive Use

- Two-group approach. Uses a regression-based prediction of zero-incentive use for HIP and non-HIP participants.
  - HIP participants (with low predicted incentive use)
  - non-HIP participants (with low predicted incentive use)
- Three-group approach. Descriptive comparison.
  - HIP participants (with zero-incentive use)
  - HIP participants (with positive incentive use)
  - non-HIP participants

Farmers Markets

- Review of implementation
- For **HIP participants**, monthly HIP purchases and incentives earned
- For **HIP and non-HIP participants**, impact estimates of SNAP redemptions in farmers markets
- For **HIP and non-HIP participants**, descriptive estimates of TFV and non-TFV purchases in farmers markets

Analysis: Subgroups

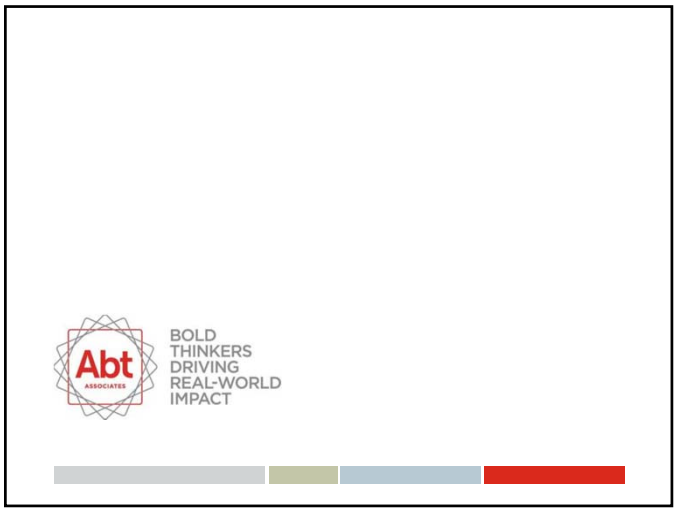
- Demographics (sex and age race/ethnicity for household head, presence of children)
- Household monthly SNAP benefit amount ($200+)
- Household income ($788+)
- Household income adjusted for household size (above/below sample median)
- Geographic location within Hampden County (Springfield/Chicopee and Holyoke/other)

Analysis: Subgroups (Survey)

- Attitudes
- Barriers
- Family food environment
- Shopping patterns

Discussion Questions

- If choice of retailer is not responsive to HIP, are we justified in describing HIP/non-HIP comparisons of TFV purchases in IECR-equipped participating retailers as nearly equivalent to impact estimates?
- What is your view of the two-group and three-group approaches to the connection between zero-incentive use and food intake outcomes?
- What other analyses do you suggest?


Abt
ASSOCIATES
BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT

Appendix C: Mediator Analysis

Our analytic approach in the HIP Interim and Final Reports is designed to take full advantage of the random assignment evaluation design, while recognizing the sample size limitations. Key elements in our approach include:

- Main analyses based on regression-adjusted differences between the HIP and non-HIP groups. This approach allows us to more precisely estimate the impact of HIP.[4]
- Exploring how the impact of HIP on TFV intake varies with observable subgroups (i.e., moderator analyses); and
- Estimating the impact of HIP on intermediate outcomes (what Cerin et al, 2009, call the "Action Theory Link").

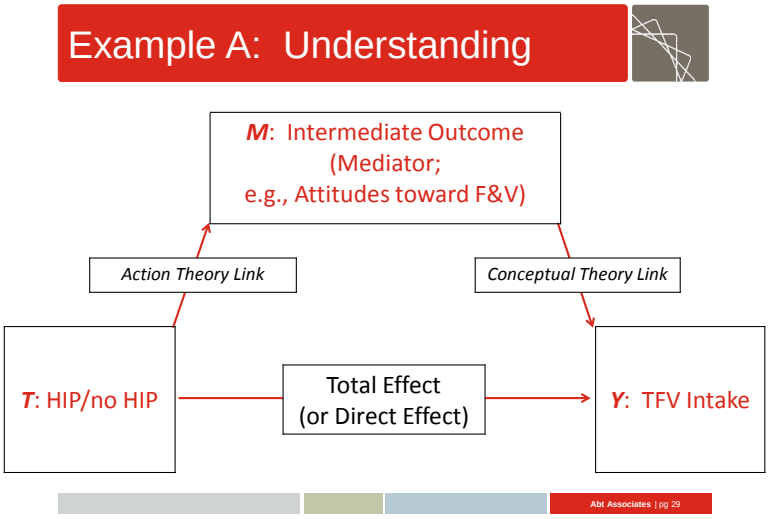
FNS as well as some TWG members have expressed interest in the analytic approach referred to as "mediator analysis" to further explore how HIP might affect fruit and vegetable intake. This appendix discusses this approach and considers how it might be used in the HIP analysis.

The first section of this appendix describes and motivates the statistical models—single mediator path analysis models in the spirit of Baron and Kenney (1986). The second section critiques the models, considering several questions. First, in what sense are the estimates useful or informative? Second, what assumptions are required to give these estimates a causal interpretation? And third, how do those assumptions relate to the random assignment approach used in the HIP evaluation? And, how reasonable are those assumptions?

Single Mediator Models

The following figure presents a depiction of a single mediator models.

[4] As discussed in the Interim Report (Section 2.6), covariates include stratification/blocking variables used in sampling, demographic characteristics of respondents, baseline fruit and vegetable consumption derived from the fruit and vegetable consumption screener, and baseline outcome variables (if available).



The analysis presented in the Interim Report and planned for the Final Report estimates the "total effect" of HIP (*T*/Treatment) on the ultimate outcomes of interest (*Y*/TFV Intake).

From a theoretical perspective, it is sometimes insightful to conceptualize this total effect as operating through two distinct pathways:

1. *The Mediated Effect* through *M*; e.g., HIP affects attitudes towards fruits and vegetables; attitudes in turn affect TFV intake.
2. *The Direct Effect*; HIP affects TFV intake directly, without changing attitudes. For HIP, this might be the impact through the incentive's lowering the effective price of TFV.

In equations, we might express these ideas as follows. First, we express the total effect as:

(1) $Y_i = \alpha_0 + \alpha_1 T_i + \alpha_3 X_i + u_i^{\alpha}$

Second, we express the Action Theory Link as:

(2) $M_i = \beta_0 + \beta_1 T_i + \beta_3 X_i + u_i^{\beta}$

Third, we express the Conceptual Theory Link as:

(3) $Y_i = \gamma_0 + \gamma_2 M_i + \gamma_3 X_i + u_i^{\gamma}$

Fourth, we express the Mediator Model as:

(4) $Y_i = \delta_0 + \delta_1 T_i + \delta_2 M_i + \delta_3 X_i + u_i^{\delta}$

Where each of the equations includes other variables measured at baseline *X*, and a regression residual, *u* (with appropriate superscript).

In terms of this formulation, we can explain mediation as follows: Equation 2 implies that the intervention (*T*, in our case HIP) increases the mediator by β_1. In addition, Equation 3 implies that each additional unit increase in the mediator, *M* (in our example attitudes), increases the outcome by

γ_2. Thus, in Equation 4, the mediating effect of *T* on *Y* can be expressed as $\beta_1\delta_2$. By subtraction, we can compute the direct effect (i.e., the total effect, net of the mediating effect), δ_1, as $\alpha_1 - \beta_1\delta_2$.[5]

Finally, note that all four of these expressions are written in terms of observable variables and in linear form. It follows that they can be estimated by linear regression. Furthermore, we can create standard errors and confidence intervals for as $\beta_1\delta_2$ by the delta method (Cramer, 1946; Oehlert, 1992; Greene, 2003). The next section discusses whether it is appropriate to given the resulting parameter estimates a causal interpretation.

Critiquing These Models

The previous section has specified the models to be estimated; this section considers their usefulness. Specifically, this section considers three questions. First, if these estimates had a causal interpretation, how could we use them for advancing theory and policy? Second, what assumptions are required to give these estimates a causal interpretation? Three, how valid are these assumptions? If these assumptions are not valid and these results do not have a causal interpretation, the analysis has limited utility in helping us to understand how HIP impacts fruit and vegetable intake.

With respect to theory, knowledge of "how HIP works" might be useful for designing future interventions. If a mediation analysis shows HIP works through attitudes, rather than through prices, that might suggest designing a program that emphasized attitudinal change, perhaps with a smaller incentive. Conversely, if changes through attitude appear unimportant, then we might retain (or perhaps expand) the incentive. In either case, future experimental work would be needed to establish an impact of any other program design.

Crucially, this use of mediation models for theory requires a causal interpretation of the regression parameters. In particular, in order for the effect of HIP through the mediator to be expressible as $\beta_1\delta_2$, two conditions must be satisfied. First, it must be true that as β_1 gives the causal effect of HIP on the mediator; where by "causal effect", we mean the change in the mediator with an exogenous (i.e., outside the system) change in HIP status. Second, it must be true that as δ_2 gives the causal effect of the mediator on TFV intake; where by "causal effect", we mean the change in TFV intake with an exogenous (i.e., outside the system) change in the mediator variable.

It is not clear that both these conditions are satisfied. In a best linear predictor sense, regression gives the best fit between the included covariates and the dependent variable. However, standard analysis for linear regression suggests that only when there are no unobservable variables correlated with the included variables will regression estimate the causal effect (e.g., Greene, 1993)

---

[5] This step is crucial for the discussion of identification in the next section. If one is willing to assume that all of the effect of treatment works through the mediator, then other modeling strategies with weaker identification conditions are worthy of consideration (Sobel, 2008; Angrist, Imbens, and Rubin, 1996; see the discdussion in Imai, Keele, and Yamamato, 2010, p. 58). However, that assumption seems extremely unattractive for HIP. It seems reasonable to assume that the primary pathway through which HIP affects TFV intake is the effect on net price; i.e., not through the mediators considered in this appendix.

For β_1, the conditions required for a causal interpretation are minimal. The key regressor is HIP treatment status. HIP treatment status is randomly assigned. As long as the other regressors are measured before random assignment (as they are), no omitted variable can be correlated with HIP treatment status. Thus, we can conclude that β_1 gives the causal effect of HIP on the mediator.

With respect to as δ_2, the assumptions seem more problematic. The value of the mediator is determined by (randomly specified) intervention status, but also by anything else that affects the mediator. Anything that affects the mediator directly may also affect TFV intake directly. To some extent observed (and included in the regression model) covariates (i.e., X) control for such factors that affect the mediator and TFV directly. However, a causal interpretation would require that those omitted variables have no (or at least only small) direct impact on the outcome. Given how little we observe, that condition seems unlikely. A key unobservable would be underlying taste for TFV, in the absence of HIP.

Similarly, measurement error in the mediator will induce problems. Classical measurement error biases coefficients towards zero. In addition, it induces omitted variable bias. Even if an effect works totally through a mediator, if the mediator is measured with error, then omitted variables will matter.

These issues have been widely noted in the literature on path analysis. The discussion in MacKinnon's (2008, pp. 365-366) standard textbook on mediation is careful:

> *The [Rubin Causal Model] demonstrates the problems in the interpretation of the relation between* M *and* Y *in mediation models, at least in part because this relation is not randomized but is self-selected in most applications. The main benefit of all these detailed causal approaches is the causal consideration of the limitations and strengths of different types of evidence of causal inference. … At a minimum the causal inference approaches force researchers to consider the assumptions under which mediation is investigated. For the most part, the sensitivity of the estimates to violation of assumptions is not generally known[.]*

More recently, in discussing the identification results in their paper (which are themselves slightly weaker than those required for Baron-Kenney like models), Imai, Keele, and Yamamoto (2010, p. 56) explain the required assumption as being that the mediator "is effectively randomly assigned" given the treatment and the observed covariates. About this assumption, they state bluntly (p. 51):

> *Like many identification assumptions, the proposed assumption may be too strong for the typical situation in which causal mediation is employed. For example, in experiments where the treatment is randomized but the mediator is not, the ignorability of the treatment assignment holds but the ignorability of the mediator may not.*

Later they explain why ignorability may not hold (p. 61):

> *First, there may exist unmeasured pre-treatment covariates that confound the relationship between the mediator and the outcome. Second, there may exist observed or unobserved post-treatment confounders. These possibilities, along with other obstacles encouraged in applied research, have led some scholars to warn against the abuse of mediation analysis (e.g., Green, Ha and Bullock, 2010). Indeed, as we formally show below, the data generating process has no information about the credibility of the sequential ignorability assumption.*

What the final sentence means is that the required assumptions are untestable; they are pure assumptions.

Green, Ha, and Bullock (2009) and Bullock, Green, and Ha (2010) also express strong skepticism about the plausibility of the assumptions required for mediator models and therefore for the utility of the resulting estimates. The abstract of Green, Ha, and Bullock (2009) is blunt:

> *Recent years have seen growing enthusiasm for regression models that purport to establish claims about mediation. Despite their growing popularity, these regression models rest on naïve assumptions. The point of this essay is to puncture the widely held view that it is a relatively simple matter to establish the mechanism by which causality is transmitted. This means puncturing the faith that has been placed in commonly used statistical methods of establishing mediation.*

Similarly, the abstract to Bullock, Green, and H (2010) is skeptical with regard to mediation in context of experimental designs, such as HIP:

> *[E]xperiments cannot overcome certain threats to inference that arise chiefly or exclusively in the context of mediation analysis – threats that have received little attention in psychology…. Our conclusion is that inference about mediators is far more difficult than previous research suggests, and best tackled by an experimental research program that is specifically designed to address the challenges of mediation analysis.*

We note that HIP was not "specifically designed to address the challenges of mediation analysis."

This leaves us with a philosophical issue in evaluation. These models—and implicitly assumptions of this form—are common in the academic literature. These models are the standard part of the curriculum in many doctoral programs. One would certainly not want to argue that these models are without value.

However, these models are inconsistent with the spirit of this experimental evaluation. The power of random assignment is that only minimal assumptions are required. In particular, random assignment does not require assumptions about omitted variables or measurement error. It does require assumptions about proper randomization and loss to follow-up, but those assumptions will be required no matter what the analytic approach. Other details of the analytic procedure are specified deliberately to minimize the required assumptions. Thus, for example, standard errors are computed without an assumption of normality; we use linear regression models, rather than logistic regression models, so as not to assume the extreme value distribution; and we cluster the standard errors, rather than attempting to model the correlation structure. Alternative assumptions would yield more precise estimates; but the spirit of random assignment is to make as few assumptions as possible. Thus the limited assumptions of the random assignment design are not sufficient to satisfy the assumptions required for mediation models.

However, within the constraints of the design and data collected, as discussed in the body of this memorandum, we plan to conduct and report the results of a limited number of single mediator models in an appendix to the Final Report in order to explore potential pathways through which HIP may be linked to consumption.

References

Angrist, J.D., Imbens, G.W., & Rubin, D.B. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444-455.

Baron, R.M., & Kenny, D.A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology*, 5(6), 1173-1182. http://www.public.asu.edu/~davidpm/classes/psy536/Baron.pdf

Bullock, J.G., Green, D.P., & Ha, S.E 2010. Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550-558. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1543925

Cerin, E., Barnett, A., & Baranowski, T. 2009. Testing theories of dietary behavior change in youth using the mediating variable model with intervention programs. Journal of Nutrition Education and Behavior, 41(5), 309-318. doi: 10.1016/j.jneb.2009.03.129.

Cramér, H. 1946. *Mathematical Models of Statistics*. Princeton: Princeton University Press, p. 353.

Green, D.P., Ha, S.E., & Bullock, J.G. 2010. Enough already about 'black box' experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Science*, 628, 200-208. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1544416

Greene, W. H. 2003. *Econometric Analysis*, 5th ed. New York: Prentice-Hall, pp. 913f.

Hoyle, R. H., & Kenny, D. A. 1999. Statistical power and tests of mediation. In R. H. Hoyle (ed.), *Statistical Strategies for Small Sample Research.* Newbury Park: Sage.

Imai, K., Keele, L., & Yamamoto, T. 2010. Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 51-71.

Judd, C.M. & Kenny, D.A. (1981). Process Analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5(5), 602-619.

MacKinnon, D.P. & Dwyer, J.H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144-158.

MacKinnon, D.P. (2008). Introduction to statistical mediation analysis. Mahwah, NJ: Erlbaum.

MacKinnon, D.P., Fairchild, A.J., & Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614.

Oehlert, G.W. (1992), A note on the delta method. *The American Statistician*, 46(1), 27-29.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290–312.