memorandum

Healthy Incentives Pilot Evaluation

Date February 14, 2014 (updated August 20, 2014)

To Danielle Berman,

From Susan Bartlett

Subject Technical Memorandum: Technical Work Group (TWG) Feedback on HIP Final Evaluation Report

The purpose of this memorandum is to report on the comments received from TWG members on the draft HIP final evaluation report.

The HIP evaluation plans called for assembling for a Technical Working Group (TWG) to provide external advice and input on the evaluation research design, early results, and the final report. The TWG includes 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
- Diane Whitmore Schanzenbach, Northwestern University

The TWG met twice with the HIP evaluation team and FNS, the first time to review the research design and the second time to review the interim results as well as the analytic plans for the final evaluation report.

We sent TWG members the second draft of the final evaluation report on November 5, 2013 and asked that they review the draft report and provide comments. Several TWG members provided comments in the draft report; others prepared memoranda/emails with their comments. Attachment A provides the comments received from TWG members.

The comments we received were helpful and indicated areas where clarification and expansion of the text were needed. No suggestions for additional analyses (that were possible with the available data) were made. However, based on a question from Tom Baranowski, we conducted principal components analysis and factor analysis to further justify the attitudes scale. As noted above, at the second TWG meeting, members provided input on analyses that should be included in the final report. The Abt evaluation team incorporated these suggestions into the final report, and thus the final draft report reflected TWG input concerning appropriate analyses.

For the most part, we incorporated TWG comments and suggestions into the revised report. We did not necessarily respond in the place in which the comment was made. For example, some comments made in the executive summary were more appropriately addressed later in the report.

Some comments, particularly those that we did not incorporate into the revised draft, require explanation as to why we did not make the requested change. We have provided these comments and our response to them in Attachment B.

Attachment A: TWG Comments on Draft Final Report

Tom Baranowski

Thank you for allowing me to review and comment on your excellent and certainly comprehensive report. My comments are organized by section. Many of my early questions were answered later, so I deleted them. The generic big issues are discussed next.

General Big Comments:
The amount of information in this report is mind numbing. After a few pages, it is hard to keep things straight. The current approach to reporting analyses is a list-wise sequential analysis, as if from a list of many issues, which tend to blend together. Stating specific hypotheses and combining all methods to address a hypothesis might be a way of simplifying the presentation into palatable, digestible mouthfuls?

Since the changes in TFV benefits per month were small, it is not surprising the availability of HIP didn't result in more change in grocery selection and other variables. This suggests two research questions are key: Why was use of the benefit so low? Why did so many not use the benefit? Can we find explanations for non-participation? Was it just lack of knowledge? What could be done to further minimize non-participation?

It is a shame you don't have impact on children. Should you also conduct an analysis on the increase in spending on the total TFV divided by the number of people in the household (who presumably should benefit equally from the changes)?

Executive Summary:
You report only differences that are statistically significant at $p<0.05$. Given the large numbers of tests reported, shouldn't a more conservative approach be taken?

How did the available HIP $ benefit earned by participants change over the course of the year?

Exhibit E5.2: it would be helpful to show confidence intervals on the mean values.

The word fruit is both singular and plural, and does not need an "s" at the end to denote plural.

It would be helpful for the report to spend a paragraph talking about the meaningfulness for health of an increase of 0.23 servings, by citing relevant articles.

How likely would it be for the FV intake to increase if the incentive were increased beyond 0.3/$? Is there a point of diminishing returns?

An increase of 0.23 is roughly half a serving. Half a serving increase is roughly what most of the child FV change interventions attained by other intervention procedures. Is ½ a serving really just some community induced measurement error? Or is there some reason we can't exceed very modest increases?

Will a public data set become available that graduate students and others can be assigned for analysis? Should a section of this report identify issues of high priority to USDA that could be investigated in this data set, but Abt ran out of time and money to do?

1. Introduction

1.1 Pilot Background

It would be helpful to at least briefly present the economics rationale of price elasticity underlying this intervention.

In light of the Wilde et al (2000) findings, it would be helpful to separate the report of FV intake of participants and non-participants eligible for receiving HIP benefits.

1.2 Research Objectives

--

1.3 Theory and Conceptual Model

Exhibit 1.1 is very helpful. Did you measure all the variables here? If not, can you highlight the ones you did?

It would be helpful if the authors could state the primary hypotheses that were specified prior to data collection, and use them to highlight the methods and findings at all steps of reporting.

1.4 Previous Research

This is an interesting section.

Is the $0.30/$1 a lower price or an additional/earned incentive? Would the findings have been expected to be different from an across the board equivalent lower price?

You may be interested in a recent report: MR Richards & JL Sindelar, Rewarding healthy food choices in SNAP: Behavioral economic applications. The Millbank Quarterly. 2013; 91(2):395-412. It would be informative for you to compare your approach (which is similar to their first proposal) to all of theirs (which are based on Behavioral Economics principles).

The lack of HIP participants transferring their SNAP FV purchases to stores offering HIP benefits appears to fly in the face of your hypothesis. What does that tell us? Can the authors estimate the level of missed benefits by not shopping at a new store?

2. Evaluation Design

Exhibit 2.1 – helpful.

Research review procedures on RCTs require the method of random selection be clearly stated both at the step of trt-ctl assignment, and at the step of interview selection. How were the participants notified of group assignment? Were they notified before or after baseline assessment?

Highest household educational attainment is a common indicator of SES, but not included in your descriptive statistics. Was this collected?

3. HIP Implementations and Operations

No comment. Whatever can go wrong, will.

4. Retailer Experiences

--

5. HIP Participant Experiences

Exhibit 5.9: Is this the 30% of money spent on TFV or is this the 130% on TFV?

Exhibit 5.13: It would be helpful to have a table that compared demographic characteristics of the HIP households with and without HIP purchases. While it is the 30%'s right not to, it would be interesting to know who they are, and have some idea what might be needed to get them to participate. Comparisons of interview data between participants and non-participants might be helpful?

6. Effects on Expenditures and Shopping Behaviors

Many of the findings in this report are gems. It is a shame that most scientists in this area will never see these gems since lengthy reports such as this do not usually get read by scientists in this area. I hope Abt will take steps to publish research articles on these findings.

7. Impacts on Attitudes toward FV and Family Food Environment

Do the analyses of scales in this chapter co-vary out the baseline values?

With the very large samples, and the large numbers of tests, $p<0.01$ or even less should be the criterion for significance? Indicating $p<0.1$ as marginally significant is inappropriate.

8. Impact on FV Consumption

This is an important chapter.

9. Costs of Pilot and Feasibility of Nationwide Expansion

This is not an area of my expertise.

9.1. Data Description

--

9.2. Pilot Costs

pp175-9: missing/did not print

Exhibit 9.7: First and second row totals are substantially different (factor of 2), yet the column entries appear to be the same?

9.4. Feasibility of HIP Expansion

--

10. Conclusions

10.1. Review of Key Results

These are important findings. Hopefully, these findings will be transformed into several journal articles to maximize their accessibility to interested academic and policy audiences. The integrated economic, attitudinal & ecological mediation analysis should be given wider exposure (even given its flaws/limitations).

10.2. Limitations

--

10.3. Comparing HIP Impacts on Spending and Intake

--

Appendix A: Economic Theory

This is not my strength.

A.1. A Basic Elasticity Analysis

--

A.2. Incorporating Income Effects

--

Appendix B: Random Sampling

--

Appendix C: Weighting Method

--

Appendix D: Participant Data Preparation

Nice level of detail for a very complicated process.

Were any recalls double coded to estimate inter-coder reliability?

p 245, line 7: typo – "of" should be "or"

D.2. Coding of Other Participant Survey Data: The 'C's alphas for positive attitudes were low. Did you do a principal components analysis to assess if all the items fit on the same dimension?

Appendix E: Analytic Methods

--

Appendix F: Supplemental Tables

--

Appendix I. Mediator Analysis

--

Simone French

Overall Summary

This is an interesting, well-organized, dense and detailed report of the pilot program HIP. It was difficult to review because there was so much detail. Overall, it was very well written and presented complicated information as clearly as can be expected. Readers of the report will be impressed with the pilot and how well it was conducted and evaluated, given its scale, complexity and the number of components required to be implemented for the program to be evaluated.

Below are only a few comments about sections or findings this reviewer had difficulty understanding as presented. It would also be informative to highlight the types of data that should be collected in subsequent efforts that will both clarify the lack of understanding about how and why the program worked or didn't work, and strategies suggested to address the observed limitations in the pilot study.

Specific Comments

HIP can only be evaluated for 50% of all SNAP purchases.

HIP led households to increase their expenditures on targeted fruits and vegetables in participating stores. Nonparticipating stores accounted for half of all SNAP purchases by sample members. The EBT data allow us to compare the TFV purchases of HIP and non-HIP households in supermarkets that participated in HIP.

This seems like a significant limitation. The potential effect of this on interpretation of pilot results may have been explained, but I did not see it.

What are recommendations for increasing SNAP participant understanding and awareness, and use, of the program? Communication and understanding of HIP by SNAP participants seems to have been a major important issue.

"While 60 percent of HIP participants reported that they had heard about HIP when asked in the Round 2 survey (4-6 months after implementation), the other 40 percent reported that they had not heard about HIP. Forty percent of Round 2 respondents and 25 percent of Round 3 respondents also reported that it was hard or somewhat hard (or they didn't know) to understand how HIP worked"

I can't understand the following two paragraphs. They seem contradictory to me. Exhibit 6.10 is the table

"Not surprisingly, monthly TFV purchases in participating supermarkets and superstores were much higher for households that primarily shopped in participating stores during the pre-implementation period. Among HIP households, pre-HIP participating store shoppers purchased an average of $14.58 of TFVs each month, compared $9.20 for pre-HIP non-participating store shoppers (Exhibit 6.10).

However, somewhat surprisingly, the HIP impact was not larger for households that primarily shopped in participating stores prior to HIP. For these households, the HIP impact on TFV purchases in participating supermarkets was $1.24. For households that primarily shopped in non-participating stores prior to HIP, the impact was $1.70. We had hypothesized that the HIP impact would be greater for those households that even before implementation were already shopping in retailers who later ended up participating in the pilot, but this hypothesis did not turn out to be correct. This may suggest

that some households who had previously shopped mainly in non-participating stores shifted some of their TFV purchases to participating stores."

The text below is interesting and does it contradict the idea of an economic/pricing effect for the HIP impact on TFV intake?

"It is noteworthy that we see no significant difference in agreement to the statement, "I don't eat fruits and vegetables as much as I like to because they cost too much," as this suggests that the HIP price rebate was not strongly affecting participants' perceptions of costs. Taken as a whole, these findings provide little evidence that HIP affected perceived barriers to eating fruits and vegetables."

"Given HIP's financial incentive, one might have expected to find that HIP reduced the problem of cost as a reported barrier to fruit and vegetable consumption. However, there was no difference between HIP and non-HIP participants in the probability of agreeing with a statement about not eating fruits and vegetables as much as they would like, due to the cost. Perhaps HIP participants did not reflect upon the price adjustment provided by HIP in answering this question. Alternatively, perhaps even a 30 percent rebate was not sufficient to eliminate cost as a perceived barrier. We note that when asked directly whether HIP made fruits and vegetables more affordable, 72 percent of HIP respondents replied in the affirmative (see Exhibit 6.6)."

Conclusions/Summary Section

Since there was communication confusion, both about the HIP and about the foods that qualified, (e.g., juice increased), what is recommended to communicate more clearly to get broader and more intense, and more correct, program participation among SNAP participants?

"HIP's impact on mean daily intake of all fruit and vegetables was 0.31 cup-equivalents, somewhat higher than the 0.23 cup-equivalent impact on intake of TFVs alone. The bulk of the difference in impacts between TFV and total fruit and vegetable intake was attributable to higher intake of 100% fruit juice among HIP participants. The promotion of TFVs through HIP may have encouraged consumption of juices, because of their fruit content, even though juices did not earn the incentive. Or, alternatively, HIP participants may have been confused about which items qualified to earn the incentive, mistakenly believing that fruit juice qualified."

"To reflect the Federal and State budget environment that would face a broader roll out of a HIP-like program, *HIP did not include dedicated funding for social marketing or nutrition promotion*."

This seems like a mistake prospectively and even more so in retrospect. Investment is needed in development of effective communication and participation strategies for the target audience, since participation and awareness was low.

Joel Gittelson

Comment #	Page	Paragraph	Comment
1	2	Data Collection	A table laying out the different data collection methods and when performed would be helpful.
2	3	Consumption of Targeted Fruits and Vegetables (TFVs)	Suggest bolding or italitization of key findings
3	4	Expenditures on Fruits and Vegetables	Is it possible to also show this as a proportion of total spending?
4	5	Expenditures on Fruits and Vegetables	First part is good news for stores. Latter part is less good news for stores. Later I suggest adding a recommendations section at the end of the report. One of these recommendations would be to have separate feedback to stores and other food retailers.
5	5	Knowledge, Attitudes, and Behaviors	I would think low understanding of HIP would mean that those assigned to the HIP group would not use the benefit. Can you report on that association?
6	5	Knowledge, Attitudes, and Behaviors	What is the association between positive attitudes and HIP use among the HIP group?
7	5	Knowledge, Attitudes, and Behaviors	I would report on the percentages here
8	6	Effects of HIP on Retailers	This is an important case study. Could this be written up as an appendix? It represents a potential barrier to expansion of the program, and having the case study will enhance transferability of your findings.
9	6	Effects of HIP on Retailers	In terms of amounts and/or variety?
10	6	Effects of HIP on Retailers	Recommend retrospective qualitative study to explore this. Many convenience stores do substantial SNAP business.
11	7	Effects of HIP on Retailers	Would add definitions as footnotes to the exhibit above.
12	7	HIP Implementation, Costs, and Feasibility of Expansion	Recommendation is to do much more of this for a proposed full scale implementation. Would suggest specifically what should be done. I think promotional materials that reflect how the program works would be helpful.
13	8	HIP Implementation, Costs, and Feasibility of Expansion	I would reference where these could be found
14	8	Conclusions	Would refer here to a recommendations section later in the report.
15	11	The Healthy Incentives Pilot	How representative compared to the rest of the USA?
16	11	The Healthy Incentives Pilot	More setting description would help with transferability
17	12	1.2 Research Objectives	Some sort of table or figure that presents the methods used would be helpful

Comment #	Page	Paragraph	Comment
18	13	1.3 Theory and Conceptual Model	Where did this model come from? Is it based on the social ecological model? Maybe draws from Karen Glanz's model of the nutrition environment?
19	13	1.3 Theory and Conceptual Model	I don't think increased food consumption is the desired outcome! I think you mean increased FV consumption.
20	15	Theoretical Issues	So it will be importantly to clearly state later the reasons you feel that only a 0.23 cup increase was seen when you expected a 0.5 cup increase. Perhaps in the conclusions section where you describe the response to objective 1.
21	17	Interventions to Change Food Prices	Implies that greater attention needs to be paid to the differences between fruits and vegetables than is presented in the report.
22	27	Participant Survey Instruments	I would describe a bit more here – including how you were able to estimate portion size. I assume you gave them the companion food model booklet? I would mention that. We want to enhance the credibility of your findings.
23	30	Participating Retailer Surveys	I would provide a table with these definitions in an appendix.
24	32	Store Observations	What instrument was used for this? What data were collected?
25	32	2.5 Stakeholder Interviews	These interviews are not mentioned anywhere else in the report, nor are any findings reported as far as I can tell.
26	33	2.5 Stakeholder Interviews	Were the interviews transcribed, coded? Incentives given?
27	33	2.6 Participant Focus Groups	Need to describe the analysis a bit more. Were the focus groups transcribed? Coded? How analyzed?
28	34	2.6 Participant Focus Groups	Is there insight in the focus groups as to why there was a 0.23 c increase as opposed to the 0.5 c increase anticipated?
29	49	3. HIP Implementation and Operations	Chapter would benefit from IDI and observational data.
30	49	3.1 Overview of Implementation and Operations	I am curious about why this is not framed in process evaluation terms – i.e., reach, dose, fidelity of implementation? This would certainly be a more typical form of presentation of implementation.
31	52	Retailer Recruitment	This is essentially REACH
32	53	Retailer Recruitment	Can you expand slightly on the basis of your qualitative work?
33	53	Retailer Recruitment	Again, this is a hugely important case study. I suggest that it be added as an appendix, or at least mention the plan to develop this case study.
34	54	Retailer Recruitment	Report percentages here
35	56	Mailings	All of this could be reported in terms of reach, dose, fidelity…
36	57	Training Sessions	Why did they keep going if average attendance was <1 person? Is there a lesson learned from this about what NOT to do?
37	59	Participant Understanding of HIP	So this is something that SHOULD be done as part of a full-scale role-out of HIP. I would make some recommendations of what this component would look like.

Comment #	Page	Paragraph	Comment
38	60	3.3 Discussion	This makes it sound like the obviously poorly planned training session was the fault of the HIP participants.
39	61	3.3 Discussion	I would use the qualitative work to help confirm or converge on these themes.
40	63	4. Retailer Experiences	Can you also include some of the in-depth interview information?
41	65	Activities to Promote and Sell Fruits and Vegetables	Did you collect data on number of varieties, shelf space for produce, etc? This could be relevant.
42	65	Activities to Promote and Sell Fruits and Vegetables	Can the observation data also be put in table format, or perhaps included in the table above?
43	65	Activities to Promote and Sell Fruits and Vegetables	Can you look at the relationship between FV promotional activities (created via some sort of score at the store level), and sales of FV? This would provide evidence for a future recommendation to enhance these activities as part of a national roll-out.
44	81	4.4 Discussion	I think the perceived increase in FV sales should be mentioned. Perceptions are very important and would certainly drive future participation in the program.
45	83	5.1 Awareness and Understanding of HIP	Nothing from the in-depth interviews?
46	84	5.1 Awareness and Understanding of HIP	Some of this could be described as “dose received”
47	89	5.1 Awareness and Understanding of HIP	I think there are implications for this in terms of your recommendations on how to promote an expanded HIP program in the future. Maybe a FAQ about HIP would be in order?
48	92	5.2 Experiences and Satisfaction	Is there a training problem for store workers? This could be another area for future work.
49	97	5.3 HIP Purchases and Earning HIP Incentives	Is there a holiday spending effect as well?
50	101	SNAP and HIP Purchases by Retailer Type	I think you are forgetting the monthly cycle in patterning of use of SNAP benefits. When you get the big bolus of money once a month – a common practice is to go out and do the “big shop” – most likely at a supermarket or superstore – where its worth the investment to pay extra for transportation. So people may be making the big shop and buying a lot of “luxury” foods like fresh FV this one time a month.
51	101	SNAP and HIP Purchases by Retailer Type	Did your observations look at this?
52	103	SNAP and HIP Purchases by Retailer Type	How is this related to store type?
53	103	SNAP and HIP Purchases by Retailer Type	Are these more educated participants?

Comment #	Page	Paragraph	Comment
54	104	SNAP and HIP Purchases by Participant Subgroup	Does this imply that people who already eat a lot of FV are using the benefit, versus people who normally don't eat very much? This would be an important secondary finding.
55	106	5.4 Discussion	Should also go in recommendations section
56	107	5.4 Discussion	Do IDIs shed some light on this?
57	111	HIP Impact on TFV Purchases	Can you break apart F from V purchases here?
58	129	7. Impacts on Attitudes toward Fruits and Vegetables and Family Food Environment	What about focus group findings?
59	137	7.6 Discussion	What is supposed to go here?
60	146	8.2 Comparing HIP Impacts on Fruit and Vegetable Spending and Intake	Wouldn't it be a higher bound, if children's consumption not factored in?
61	169	9. Costs of Pilot and Feasibility of Nationwide Expansion	Some of the topics I expected to see in this section were acceptability (economic, cultural), operability, and perceived sustainability. I saw a big emphasis on economic acceptability mainly, and not much on the other aspects that would be part of feasibility.
62	186	9.4 Feasibility of HIP Expansion	It seems like the electronic record-keeping involved is a major operability problem for many of the stores.

David Just

Comment #	Page	Paragraph	Comment
1	57	Training Sessions	140 trainings with only 100 participants? How many had none? I understand voluntary training might not sound too appealing. Nonetheless, this may need some explanation. This may need some explanation.
2	65	Activities to Promote and Sell Fruits and Vegetables	This sentence should clarify whether this restocking frequency etc. is specifically restocking of HIP qualified goods.
3	103	SNAP and HIP Purchases by Retailer Type	How does this compare to other research. I believe one study found close to a third of children in low income households don't consume a vegetable in a week. This suggests a sixth don't buy any fruit or veg (other than 100% juice, potatoes or already prepared in other foods) in an average month. Is there any way to determine what percent of those not availing themselves of HIP didn't know about HIP? How much of this could be that they only used their own cash?
4	107	5.4 Discussion	May be helpful to remind the reader what size effect was expected.
5	134	7.5 Family Food Environment	Perhaps this reflects an opportunity to make the program more effective by encouraging placing fruits or veg on the counter where they will be eaten more often?
6	137	7.6 Discussion	?
7	145	White Potatoes, Legumes, and 100% Fruit Juice	Or it could be a compliment to other foods which increased consumption (perhaps eating oranges gives you a taste for orange juice?).
8	158	8.7 Round 2/Round 3 Differences	This happens at the same time awareness of the program was increasing and understanding of the program is increasing. I can understand how understanding could decrease fruit juice consumption. I have a hard time not believing awareness in general wouldn't be increasing fruit juice consumption. Hard to make too much of this, but it may be worth mentioning.
9	160	Other Intake	I think this is a very fair way to deal with this result. My own suspicion is that the incentive can drive alcohol expenditures among those who already consume (or are prone to consume) large quantities of alcohol. That said, this seems to be a really waffley result and should rightly be dismissed as chance.
10	201	10.3 Comparing HIP Impacts on Spending and Intake	Though that would be inconsistent with increasing TFV purchases in the first place.
11	201	10.3 Comparing HIP Impacts on Spending and Intake	That in itself is a pretty astounding conclusion.

Diane Schanzenbach

General

I have two "big-picture" comments:

1) There is a substantial difference between the self-reported spending and the observed spending from EBT. Is there a way to narrow the gap between the $12 or so people are observed to spend and the $75 or so th`ey say they spend? Is this difference the result of different definitions? Can you reconcile $75/month on fruits and vegetables with a broader definition? This difference is large, and I fear it is evidence against (in the broader literature) asking people what they spend instead of observing it directly. Is there a way to better reconcile the spending amounts? If not, that is important to know, too.

2) Throughout the process, I have objected to the nomenclature of “upper bound” and “lower bound” because those are usually considered to be statistical terms indicating bounds on effect sizes. In this report, you are instead using those to refer to the upper and lower bounds of TFV consumption. Furthermore, as I describe below, I do not think that these even represent upper and lower bounds on TFV consumption. Can you provide a response to me why you all decided to reject my strong objection to the nomenclature of "upper bound" and "lower bound"? I really think these terms are misleading.

Chapters 2-3

Overall, I think this reads extremely well and was executed in an excellent manner. I am glad to see such good work done on such an important topic.

Here are a few minor comments:

P 25: “most HH had 3 months of experience with a fully operational pilot “ at the time of the survey. Is this at least 3 months, or 3 exactly? It reads like 3 months exactly and I don't think that is correct.

Table 2.7: Why are there more HH homeless than “other” (b/c private, public + other sum to 1.0)?

Quoting from P 58: “A fourth technical issue was discovered as HIP was being implemented in the four IECRs who joined HIP on October 1, 2012. All of the new IECRs used the same TPP, one that was also being used by two of the existing HIP chain retailers. It was the same TPP that had been involved in the December 2011 problem. This time, the HIP month-to-date (MTD) total did not appear on HIP clients’ receipts. However, this issue involved only the receipts and did not affect the actual HIP incentives earned. The receipts were accurate at the start of the pilot, but since no HIP participants called the toll-free line to report the incorrect receipts, the problem went undetected until September 2012. The problem was not corrected by the end of the pilot.” Are these dates correct? The chronology doesn't make sense to me.

Chapter 8

Re: TFV intakes.

These aren’t upper and lower bounds on “true” TFV *intakes*, because it only includes purchased foods not bought in prepared forms. Let’s say that HIP encourages people to consume vegetables and people discover that they have a taste for them. Now they consume them more than they used to – e.g. they buy

the beef stew with extra vegetables, they order veggies as a side when they eat out, etc. Even though those weren't directly targeted by the incentive, they may be an important spillover effect. (I see that you acknowledge this on page 144). I don't think you want to use language that minimizes the possibility of this effect. (Or the opposite – it could be that this intervention shifts the veggies that people eat to the incentivized set and then they reduce their intake from other sources. I hope this doesn't happen though!)

As a result, I suggest "narrow-definition" and "broader-definition", or something like this, as the terms. I strongly argue that upper-bound is not appropriate since there are still TFV's that are potentially part of the dietary intake that this measure does not include. That makes it inaccurate to call it an upper-bound!

And I renew my concern about the term "bound" in here at all, as in the context of a randomized experiment most readers who are trained in program evaluation will think of this being an impact bound instead of an intake bound.

I think you over-state your preference for the narrow definition of TFV consumption. Note that that definition is so narrow – i.e. no added fats – that it excludes eating a salad with salad dressing (!!!). When we think about the purpose of this demonstration project, it is to see if a financial incentive can encourage Americans to eat more healthy foods. I just don't think it makes sense to argue that those vegetables somehow don't "count" if we sprinkle some olive oil and vinegar on them! I know that we discussed this at the original TWG meeting and we agreed with your conceptual approach, but that was before we all understood how difficult it would be to measure intake in practice.

On page 150, you state about the EATS measure: "These measures are less precise than 24-hour dietary recall measures, so in general we would expect to detect fewer impacts". Does this necessarily follow? My fear is that it is easier to falsely report on the EATS (and tell you what you want to hear). I agree that the results line up nicely with the dietary recall numbers, though.

What is the SNAP benefit > vs < $200 supposed to tell us? Is that this is of interest to the USDA for some reason? Benefit level is a combination of family size and income from other sources. This $200 cutoff is a weird combination of variation across those areas. I don't think we learn anything from this (and it is not surprising to me that there is no difference across this cut of the data!)

It surprises me that there is no difference for WIC participants, but there is one for children in the house. Do you have any insights into the difference?

I found the results on the rest of consumption to be very intriguing (and wish you had more statistical power to detect impacts! Wouldn't it be interesting if this intervention caused people to substitute TFV for empty carbs? It does not look like we can statistically reject that.)

Chapter 9

I think this is such an interesting chapter and really important to think about how to push the policy forward.

Is it worth making the following point explicitly: in the pilot, the infrastructure costs were much larger than the incentive costs. Going nationwide, the infrastructure costs will be relatively much smaller,

because the heavy lifting has already been done to get the cash register programs ready, and the incentive costs will be larger (because it would be nationwide). You make this point perhaps too subtly in the current draft.

The CBO always seems to do projections based on households, not individuals. I followed that convention and got a smaller projection for the cost of expanding across the nation. Do you want to justify why you use individuals instead, or show both? (Fine with me to leave as is, too, I just wanted to raise this.)

I wonder if you want to blow up nationwide estimates at the 95% CI of the mean, too? I think it might be helpful to state.

This might be convention, but I think it reads weird that you are talking about 53 state agencies.

I have wondered a lot about why on earth that one retailer refused to participate. My instincts are that if this were to be expanded nationwide, it would be preferable to make participation in HIP mandatory for stores that accept SNAP. I am glad you talk about this in the chapter. One point of confusion on my part is that I thought the depth of stocking requirements meant that everyone who accepted SNAP would also have HIP-eligible items – is that not correct? This might merit a brief discussion in this chapter.

Attachment B: TWG Comments Requiring Additional Clarification and Abt Responses

Page #	Comment	Abt Response
Executive Summary		
Baranowski	You report only differences that are statistically significant at p<0.05. Given the large numbers of tests reported, shouldn't a more conservative approach be taken? (Also similar comment regarding Chapter 7.)	No, we chose a single confirmatory outcome prior to analysis (as discussed in Chapter 2) as the way to deal with the multiple comparisons issue. Added note in executive summary; full discussion is in Chapter 2 and Appendix E.
Baranowski	How likely would it be for the FV intake to increase if the incentive were increased beyond 0.3/$? Is there a point of diminishing returns?	We have no direct evidence on this. To answer this question would have required a different design and much larger samples. However, note that the marketing argument might imply that a smaller benefit would have similar results. We suggest this as a future research question.
Gittelsohn, p. 5	J7: I would report on the percentages here	Analysis in body of report converts Likert categories to numbers and this is what is reported in chapter and executive summary. Appendix does present percentages from logistic regressions for interested readers.
Gittelsohn, p. 8	J8: This [supermarket chain] is an important case study. Could this be written up as an appendix? It represents a potential barrier to expansion of the program, and having the case study will enhance transferability of your findings. [Also J33, p. 53]	We have not added any detail on [supermarket chain]. While retailer recruitment is an issue for a temporary pilot, it is not clear (particularly given [supermarket chain] reasons for not participating in HIP) that it would be an issue in wider implementation.
Chapter 1		
Gittelsohn, p. 11	JG16: More setting description would help with transferability	We have not added any additional description given that this language was approved some time ago. If FNS desires we can add more detail on the county.
Gittelsohn, p. 13	JG18: Where did this model come from? Is it based on the social ecological model? Maybe draws from Karen Glanz's model of the nutrition environment?	This model is guided by the research questions and the data sources used in the HIP evaluation.
Chapter 2	No outstanding issues	
Chapter 3		
Gittelsohn	Several comments suggest reframing the discussion to more explicitly use terminology he states are process evaluation terms (e.g. reach, dose, fidelity.	Current discussion, much of which is taken from the Early Implementation report addresses these issues. We do not see that the final report would be improved by such a reframing.
Chapter 4		
Gittelsohn, p. 63	J40: Can you also include some of the in-depth interview information?	We did not interview any respondents who could discuss the types of store information included in this chapter.
Gittelsohn, p. 65	J42: Can the observation data also be put in table format, or perhaps included in the table above?	We added some more findings from the observational data, including numbers of stores, when informative. We also noted that the observations were consistent with survey findings. Given that we only had 10 observations, we did not feel that including the information in tables was appropriate.

Page #	Comment	Abt Response
Gittelsohn, p. 65	J43: Can you look at the relationship between FV promotional activities (created via some sort of score at the store level), and sales of FV? This would provide evidence for a future recommendation to enhance these activities as part of a national roll-out.	We do not have good data on store sales of fruits and vegetables (only respondents opinion about whether sales increased, decreased, or stayed the same) to do this type of analysis.
Chapter 5		
Gittelson, p. 103	J53: Are these more educated participants?	The only available education data are for the sampled respondent. TFV purchases are made by the primary shopper, so this analysis would not be very useful.
Chapter 6	No outstanding issues	
Chapter 7	No outstanding issues	
Chapter 8		
Gittelsohn, p. 146	J60: Wouldn't it be a higher bound, if children's consumption not factored in?	No, lower bound is correct. Our base intake estimates from the recall data are expressed in terms of average daily consumption per adult per day, since we did not sample children. We convert from an adult-level measure to a household-level measure by multiplying the adult-level measure times the number of adults per household (1.49, per Exhibit 2.8). We could alternatively have multiplied by the total number of individuals per household (2.34, also as reported in Exhibit 2.8), including children (2.34 - 1.49 = 0.85), thereby resulting in a higher consumption estimate in per-household terms. However, this would require the implicit assumption that children in the household consumed the same amount, on average, as adults, which seems unlikely – in fact, we might expect that children consume less, on average, than adults. If, for example, children in the household consumed half as much as adults, then household-level consumption would be [1.49*(estimated adult-level consumption)] plus [0.85*(½)*(estimated adult-level consumption)]. By making the more extreme assumption that children consume no fruits and vegetables at all, we eliminate the second term of that calculation entirely, thereby obtaining a lower bound as stated. We have added a footnote to attempt to clarify this point.
Chapter 9	No outstanding issues	
Chapter 10	No outstanding issues	

Page #	Comment	Abt Response
Appendix D		
Baranowski, Section D.2	D.2. Coding of Other Participant Survey Data: The 'C's alphas for positive attitudes were low. Did you do a principal components analysis to assess if all the items fit on the same dimension?	These 6 items were designed to go together when the survey was developed, and the alphas are above .70 for Rounds 2 and 3. The alpha for Round 1 is also close to .70. Removing items with lower correlations with the total did not increase the alpha suggesting that the items go best together. A principal components analysis on the covariance matrix also indicates that the items load onto one factor. While other methods of principal components and factor analysis indicate a second factor with an eigenvalue slightly over 1, alphas of the scale split into two subscales demonstrate poor reliability. Thus we felt it is best to keep all 6 items together.