Appendix G2

National Agricultural Statistical Service (NASS) Comments

Brian Richards
USDA / NASS
April 21, 2011

My Comments: ICR Package / Healthy Incentives Pilot Evaluation

I. Introduction:

On April 18, 2011, I received a request to review an evaluation which is referred to as the Healthy Incentives Pilot (HIP) Evaluation. In the sections below, I provide my comments associated with this evaluation. Being a mathematical statistician, my primary focus relating to these surveys is how the population (from which we will sample) is formed, how the sampling will be carried out, etc. So these related comments appear in section II. below. But I have offered a few other comments, aside from those related to statistical methodology; and these appear in section III. below.

II. My Comments Relating to the Statistical Methodology:

Part B of this overall package that I was forwarded does an excellent job, overall, of explaining pertinent details associated with the statistical methodology. In the very first paragraph of section B.1.1, it is explained that "… we will randomly assign 7,500 of the 53,000 recipients to the treatment group and the remainder to the control group." Although this is explained in later sections of Part B, it might be helpful to clarify, at this introductory stage, that the 53,000 recipients are actually households, as opposed to individual persons. So one can conceptualize, at this first stage of sampling, that the sampling units are households.

I compliment the authors / creators on the diagrams contained in Part B. More specifically, the diagrams displayed in Exhibit B1.1 and Exhibit B1.2 do a very good job of depicting the important details of the overall sample design (of the part of the evaluation involving the SNAP participants).

At different locations in Part B, there are explanations relating to stratification, stratification variables, blocking variables, sorting, and non-response adjustment cells / classes. For some of these explanations, the author may consider clarifying certain details. I am uncertain how closely OMB might examine such methodological details, and if some of the explanations might raise questions from OMB. For example, in the very first paragraph of Part B (page 22), it is stated that "We will then stratify each group by person-level characteristics such as age, race/ethnicity, gender to ensure that they are balanced." If I understand all of this properly, although reference is made to stratifying, we are not actually creating strata which will be part of our sample design associated with the SNAP participants. That is, for the SNAP participants' portion of our evaluation, we are **not** forming

strata, and then sampling by stratum. But rather, the person-level characteristics referenced above (in the context of strata) are variables that are actually blocking variables or sorting variables, which are explained at the top of page 25. That is, we will sort our list of participating SNAP households using a pre-determined algorithm (potentially using some or all of these person-level characteristics), and then we will systematically sample every "n-th" household. Then, at our second stage of sampling, and in a similar fashion, we will systematically sample every "n-th" individual in our new sorted list. The stratification which is referenced early on perhaps relates to post-stratification, which would be used to carry out a non-response adjustment during the estimation phase. This is referenced on pages 39 and 40. I am assuming that the stratification variables referenced earlier will be used (some of them, anyway) to form non-response adjustment cells or classes. So for the different variables referenced in different sections of part B, some of them will be used **both** to sort our frame in the sampling phase **and** to post-stratify in the estimation phase. And some of the variables might be used as part of this earlier phase, but not the later phase … or vice versa.

I also wanted to point out some apparent contradictions across different parts of the explanations in section B.1.2.; more specifically, with the sampling explanations associated with the SNAP retailers. The explanatory text and table (Exhibit B1.3) on pages 30 and 31 lay out details pertaining to the planned sample design of the SNAP retailers. The second paragraph on page 30 explains that "We will stratify stores participating in HIP by store type." This paragraph also explains that "With an expected response rate of 80 percent, this sample will yield 60 completed surveys, with at least 10 per stratum (except for farmers' markets, where we will survey all participating markets)." However, when I review some of the information in the table at the top of page 31 (Exhibit B1.3), some of the counts appear questionable. When I examine the row of this table relating to "Farmers Markets", the 4th column displays that the number of expected participants is 4. When I move over two more columns to the right, to the column displaying the participants' sample, I see a count of 2 for Farmers Markets. Shouldn't this count of 2 actually be 4? That is, don't we want to sample **all** of the expected participants in the Farmers Markets stratum? This is based on the explanation on page 30, referenced above.

I also had some questions on the row in this table relating to "supermarkets". The column associated with population for this stratum displays a count of 19; and the column associated with expected participants also displays a count of 19. So from this, we project that all 19 retailers in the supermarket stratum will choose to participate (in the HIP evaluation). When I go over to the right-most column in this row, however, I see a count of 4 for the non-participating sample. Shouldn't this count be 0? If we

project that all supermarkets will participate, there would be no non-participating supermarkets to sample from.

In different parts of the overall package, I paid attention to discussions and explanations of the **minimum detectable difference.** This quantitative measure seems like it has been considered and studied very rigorously. It seems like a very important aspect of this overall evaluation. That is, how likely will we be able to detect a plausible HIP / non-HIP difference in the main outcome? More specifically, will the monetary incentive make a significant difference in the amount of targeted fruits and vegetables consumed among SNAP participants? This aspect of the evaluation seems to be very well addressed; so I have no further comments.

III. My Comments Relating to Other Aspects of the Evaluation Proposal:

Table A.12.1 in Part A (Supporting Statement) displays respondent burden and cost estimate details of interest. I understand that in some separate correspondence between David Hancock and Rachelle Ragland-Greene, that David pointed out a slightly different way to display this information. This is necessary because of the need to identify the burden for good responses and for non-responses separately. I assume that this information was adequately explained, so I will not comment further on this.

My last set of comments clearly is outside of the “arena” of statistical methodology, but I will offer these comments anyway. Perhaps they will be useful in some way. These specific comments relate to Appendix D – “Focus Group Materials”. As I understand the plans for the focus groups, there will be about 10 participants in each group (per the information provided on page 28 of Part B). And there will be a total of six groups – three groups in round 2 and three groups in round 3. In the middle of page 9 of Appendix D, it is stated that the focus group meetings should take about an hour. However, as I read through the prospective questions to be asked of the focus group participants, it seems like there are many good questions that will be asked. The questions begin at the bottom of page 9 and stretch over to the bottom of page 13. Some additional questions are displayed at the top of page 14. All of the questions listed in these 4+ pages seem like relevant questions, and they may elicit much discussion among the 10 participants. So I am wondering if one hour seems unrealistic (that is, too little time) in planning how long these focus group meetings will take.