Food Safety Education Post –Wave Tracking Survey

Comments from

Matt Gregg

Statistician

USDA/National Agricultural Statistics Service

202-720-3388

General – Comments

I reviewed the documents FSIS provided for the Food Safety Inspection Service provided for the Food Safety Education Post-Wave Tracking Survey. I did not use track changes in any of the documents provided all comments are included here. Bold headers identify the section the comments apply to.

Several places in the documents have been marked to change the age range from 20-40 to 20-45. Was this just a typo or is this a change from the benchmark survey? If it is a change from the benchmark survey, i.e. the benchmark survey was targeting adults between 20 and 40 and the post-wave survey will target adults between ages 20 and 45, then the results from the benchmark survey will not be directly comparable to the post-wave survey. During fielding of the benchmark survey in 2011, we expanded the age range from 20-40 to 20-45 because the incidence of 20-45 was coming in lower than predicted. The postwave survey age range is 20-45. Age distributions within this 20-45 range in the benchmark and the postwave will be comparable.

Part A

There are several references to the abbreviation JWT but the full name of JWT is not included anywhere. The full name of JWT should be included the first time it is used. We made the change as recommended. JWT stands for J. Walter Thompson.

I was confused by the description of the key measures on page 3. First, I found the wording of "Top Box Answer Choice" and "Top 2 Box Answer Choice" to be awkward. When referring to Top Box/Top 2 Box we use "Answer Choice" and "Answer Score" interchangeably. I was unsure how the comparisons will work. Will the pre and post survey comparisons be between the proportion of respondents that selected the top choice and then a second comparsion between the proportion that selected the top 2 choices? Yes Are the top 2 choices always the preferred choices for measuring increases in desired behavior or level of importance? For each item, we will look at top box scores *and* top 2 box scores from prewave to postwave, noting whether there was an increase, decrease, or no statistically significant change. -For example on question 14, subquestion 8 asks how important it is to "Wash meat or chicken before cooking". According to the FSIS website you should wash fruits and vegetables but not meat, poultry or eggs. For this particular question a statistically significant *increase* in the perceived

level of importance of this step would not be the desired outcome. Consider rewording parts of this section to reflect that statistically significant decreases are important on the not necessary steps on question 14. You are correct that for certain items we are more interested to see if there was a decrease (or it stayed flat) rather than an increase. The report will make this clear, and also note that we are most interested in the four food safety behaviors mentioned in the campaign advertising.

On page 6, under Questions of a Sensitive Nature, it is noted that there are no questions of a sensitive nature. However, question 33 asks about household income. Although the question asks which range the household income falls in and not the exact dollar figure, this may be considered sensitive information. The instructions on the questionnaire indicate that a popup option for the CATI instrument will answer the question of "Why we ask this question?" and option i. is "Prefer not to state". This is acknowledgement that people may not want to provide this information due to its sensitive nature. Justification for this question would be that the data will be summarized at the income level and this is the only source for that information. Good point. We amended page 6.

PART B – General

I'm unclear on some of the details of the sample design. For each group is the total sample size set at 3600 from which 600 respondents are expected? Yes, the total (initial) sample (3,600) size was determined as a function of incidence/contact rates. Working telephone numbers will be re-contacted a minimum of 3 times during a normal field cycle. If the yield is insufficient then additional sample will be applied on an equitable census distribution.-

Page 1 of Section B (English and Spanish) have been amended.

Or will households continue to be contacted until there have been 600 responses? The former may yield far fewer than 600 responses while the latter could result in a total number of contacts far more than 3600. I think it would be helpful to clarify how this will work.

More details on the assumption of the power analysis would be helpful. What alpha level was used? Is this based on the difference between two proportions? If so what was N for the benchmark survey? The alpha use for the power analysis was .05. There was no provision for the benchmarking survey (which is customary), instead theoretical proportions were applied based on N=600 and the difference in proportions were evaluated.

I have a few comments on Figures 1 and 2. First, do the x-axis values fall between the tick marks? It seems so on Figure 2 but I'm not sure on Figure 1. I would recommend aligning tickmarks and the values on the x-axis. Second, set apart the legend from the x-axis label and possibly provide a label for the legend. Please refer to figures 1 & 2 below. In both cases, the x-axis values were meant to align with the point values plotted along the curves. Hopefully, this is more pronounced with the addition of the gridlines. Page 3 of Section B (English and Spanish) have been amended.
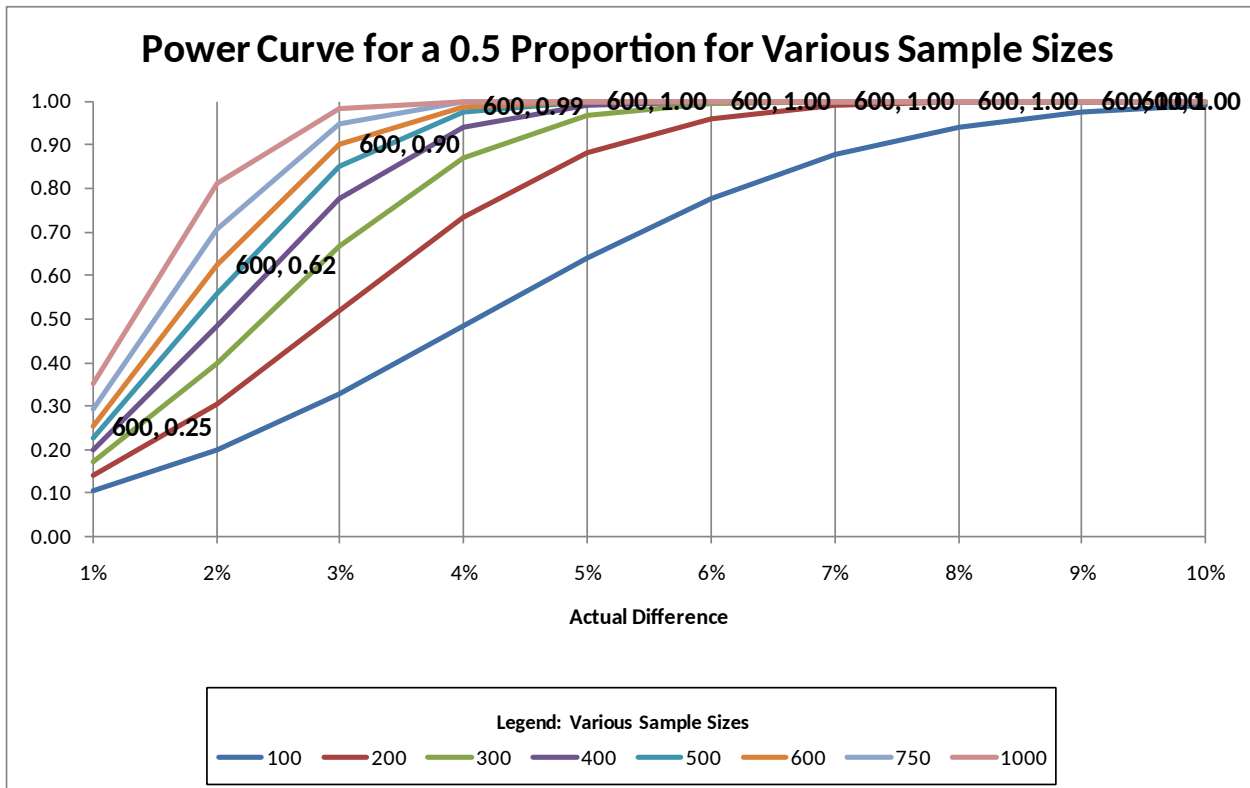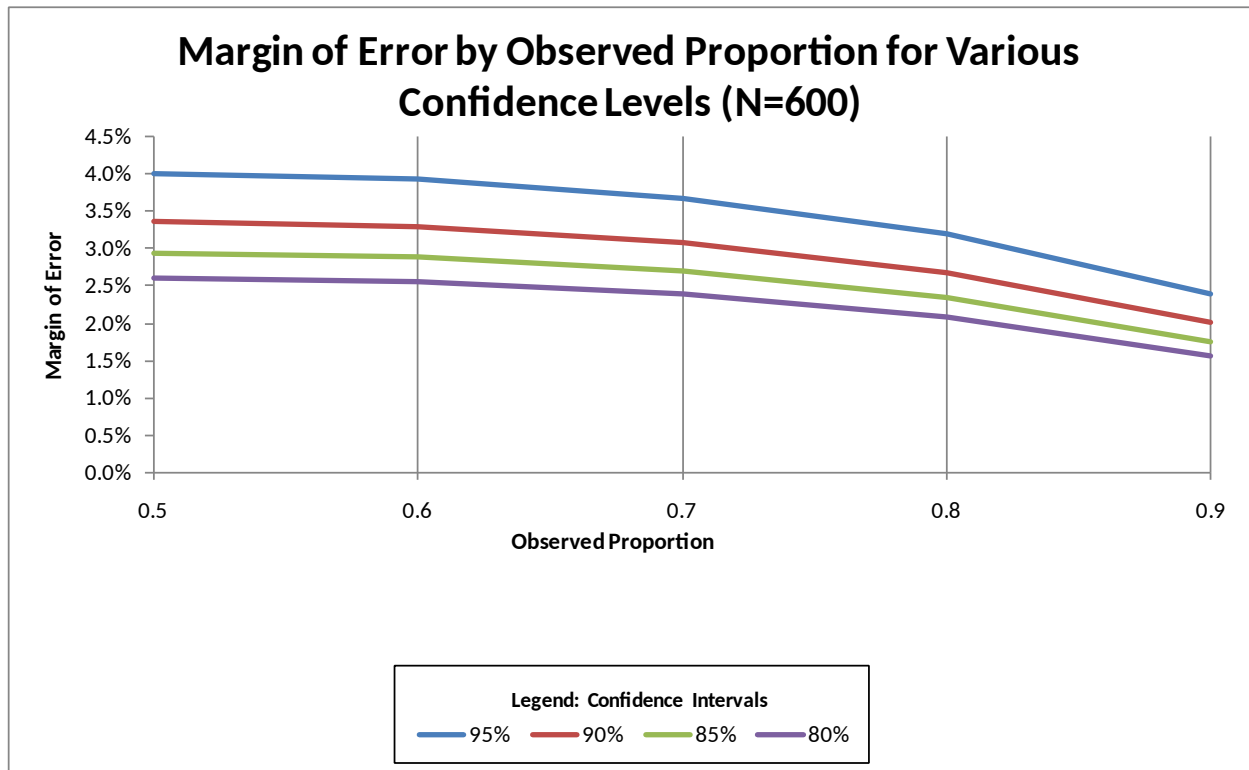
Figure 1

Power Curve for a 0.5 Proportion for Various Sample Sizes

Figure 2

Margin of Error by Observed Proportion for Various Confidence Levels (N=600)

The response to Question 3 was well done. Efforts to limit non-response are covered and potential sources of bias, particularly the issue of contacting only landline households was also well addressed.

Part B - English Survey

Statistical methodology for stratification and sample selection

The first sentence under the Geographic Stratification section says, "A sample frame will be developed using a stratified random sampling procedure that will proportionately divide the U.S. population into sampling units of geographic subpopulations". Generally speaking a sampling frame is constructed with the needs of the sample in mind and then the sample is drawn from the sampling frame incorporating the sampling specifications. A sampling frame is not developed from a sample. I would reword this paragraph. Good point. Changed "sampling frame" to "sampling approach".

This is my understanding of the sampling procedure based on the description in this section. If what I've written is incorrect, then first off I apologize but second, I would suggest rewriting portions of this section in order to provide a more detailed description of the procedure. I interpreted the sampling frame to be all households in the U.S. From there the specs are applied where the sample is allocated based on proportion of households in each of the nine census regions. Then a set of zip codes is systematically sampled within each region. Lastly, an RDD sample of landline phones is selected in

proportion to the number of households within each selected zip code. You are correct in your assumptions; however, OMB reviewers did not require this level of detail in the application for the benchmark survey, and we expect this original description to suffice, without causing unnecessary confusion.

Part B - Spanish Survey

Statistical methodology for stratification and sample selection

I have a similar critique of the wording that starts the Geographic Targeting section, "A sample frame will be developed using a sampling of the 30 DMAs with highest Hispanic household concentration." The sample will consist of households selected from the 30 DMAs but the sample frame will be all households within the 30 DMAs. I would reword this sentence.

The primary sampling unit will be the households in zip codes within the 30 DMAs that have at least 1% Hispanic population. The households that pass the surname targeting criteria will be the basis for the RDD sampling. Again, if I haven't properly stated the procedure then you might want to consider adding more details or rewording certain parts. Correctly defined, but we feel that the current description is sufficient to OMB requirements.

At the end of the Hispanic Acculturation Screener there is a reference to acculturation points. The purpose of the acculturation points should be addressed in Part B.

Questionnaries - General

In the materials sent from FSIS, there was a Hispanic Acculturation Screener (HAS) and a General Market Questionnaire (GMQ). How do these two fit together? If a Hispanic respondent qualifies for the survey based on the screener then what happens? The Hispanic acculturation screener is applied to an independent target acquisition sample (dedicated to Hispanic households) where the respondent must participate as part of the overall questionnaire and is subsequently classified based on the total number of points attained. Is there a Hispanic version of the General Market questionnaire or do they simply get the General Market version? Yes, there is a separate version of the General Market questionnaire which has been translated into Spanish for Spanish language households.

The respondent criteria box at the top of the HAS and GMQ lists a mix of 40/60% male/female as well as a percentage mix of race/ethnicity. Are these percentage breakdowns what is expected from the 600 respondents or are these specific targets that will be monitored during data collection? In other words if during data collection you hit 240 male respondents (40% of 600) on the English survey but only have 550 total respondents will males respondents be screened out for the remainder of data collection? Yes. Further, I don't understand how the race/ethnicity mix can only include 16% Hispanics when at least 50% of the respondents between the two surveys will be Hispanic. The English and Hispanic surveys are two separate surveys. One of the surveys is expected to accrue 16% Hispanic respondents, and the other survey 100% Hispanic respondents.

Be very specific in the instructions on how to ask for an adult to come to the phone when a child answers. You don't want enumerators asking "Can I speak to your mommy?" everytime as this will bias results. At the same time if a 7 year old answers the phone asking "Can I please speak to the head of household or caregiver?" will be confusing (although they will likely just hand the phone to the closest adult). There will be an [INTERVIEWER NOTE] added to ensure asking in a non-gendered format; additionally, quotas will be reviewed at the end of each field day.

There are several instances where the screening questions on the HAS and GMQ are not worded the same. Wording should be consistent across all questionnaire versions. Table 1 shows a list of questions that are asking for the same information but worded differently in the HAS and GMQ.

Table 1

HAS Question Number	GMQ Question Number	Description / Comment
S3	2	Respondents age
S4a	3	Care for any children 4-12 years of age
S4b	4	Capacity of care
S5	None	Age category of children in their care. Not on GMQ although it seems to be an important question
S6	7	Race/ethnicity. In GMQ repeat the wording from question 5

Hispanic Acculturation Screener

There is a typo on S3. The second line in the instructions after the question says "41 or above [THANK & TERMINATE]". The 41 should be 46.

What is the purpose of the acculturation points? This is not addressed anywhere other than a note at the end of the questionnaire that says that there will be a mix of acculturation levels. Is the screener completely separate from the main questionnaire meaning that a respondent completes the screener and then may be recontacted later based on their acculturation points? Or at the end of the HAS will the acculturation points be assigned by the CATI program and CATI will decide whether the interview continues? How will this work? I think the acculturation points need to be addressed in the Part B responses for the Spanish survey.

The acculturation point system allows the sub grouping of Hispanic respondents into three (3) acculturation levels: Low (Spanish dominant), Medium (Bilingual), High (Assimilated). This is a simple algorithm based on: (1) Language Preference, (2) Generation & Percent of Life in the U.S., (3) English Language Capability, and (4) Spanish Programming/Media Consumption. This aids in both reporting and analysis and ensures U.S. Census representation.

In the case of the Hispanic questionnaire, the acculturation screener is designed as an integral part. As such, the respondent will go through the entire survey and be classified via the CATI program during the

process. They will continue the survey independent of the acculturation points accumulated. This part is not intended to be broken-up for re-contact.