Appendix H:

Comments Received from ERS on the “Product of USA” Study

Review for Analyzing Consumers' Value of "Product of USA" Labeling Claims

OMB No. 0583-NEW

This study proposes three methods to investigate consumer perception and preference for "Product of USA" label, which FSIS is contemplating a revision of its definition. The label is an important for marketing of meat products. It entails considerable consequences to consumers, producers, and businesses. The potential real-world impact of this study cannot be understated. Thus, it is crucial that the inferences derived from the study can truly assist in the rulemaking process.

Overall, the LTE experiment and the DCE may be useful for FSIS purpose. However, the proposal leaves open some questions whether the proposed method may indeed accomplish its intended purpose, I have some comments that may be helpful.

LTE experiment

The LTE experiment is proposed to test the saliency of the "Product of USA" claim.

1. This deals more with how the structure of the proposal (section B: statistical method) rather than the LTE technicality. Four LTEs are proposed. I find it hard to follow even after reading the proposal twice. It might be helpful to describe *what* is being done first, only then to proceed to *why* it is being done. Perhaps some figures would help to illustrate.
2. The proposal goes into some details about different variations of "Product of USA" from page 11-12, where it notes that the use of flag, position, text size, verbiage, and etcetera are among the variables being considered. On page 13, three version of the labels were settled. An important assumption is made that the chosen labels can represent all the different variations that are being used in the market, which is rather strenuous. While it may not be practical to represent all the variations, why would the three chosen be sufficient? It needs a strong justification.
3. The LTE also involve an important assumption that the 20 seconds limit is justifiable. Is there any research to suggest the optimal exposure time?
4. Another important assumption is that the LTE reflects shoppers' behavior in store. I question if this is true as the respondents are specifically being asked to focus on the label for 20 seconds, whereas a shopper may only give it a mere glance in the store. This may pose a serious question on whether the LTE can really tell us about the saliency of "Product of USA." This is not to say that the LTE has no value at all, it may need other procedures to buttress its validity.

Survey Questions for Assessing Knowledge

5. Could not some basic Likert scale questions, such as "rate the importance of country of origin when you are buying meat" be used to supplement the LTE test? Another way to approach it could be some eye-tracking experiments; yet another way could be some heat map experiments; yet another is by measuring the "affect" of the label.

DCE for measuring WTP

6. Four definitions were provided in the second version of the DCE. While these definitions are correct in the technical sense, their negative connotations are well hidden. It is doubtful that the respondents, by looking at definition 1, could arrive the conclusion that the cattle were raise in Uruguay, processed in Uruguay, and further processed in USA, and be labeled as product of USA.
7. Assuming that the WTP for definition 1 is not statistically significant from the other definitions, I wonder what practical inference that can be made. Was it because consumers don't care if the cattle are from Uruguay? Or was it because they didn't know if the cattle are from Uruguay?
8. I find that the prices chosen in the DCE are rather low, especially in the context of 2022. Further, three levels may not be sufficient.
9. The price variation for steaks is rather too narrowed, it would be a deal to buy angus, grass-fed, choice grade NY strip at $11.49. (Note that the USDA choice logo may yet be another way consumers can confound the grade as origin label.)
10. Justification is needed for the rather numerous chosen attributes. You need to weigh between the complexity vs. realism of your CE. And ultimately, what do you want to accomplish with the CE.
11. Another assumption here is that the text-based CE can mimic the actual labels enough. Why not picture- based CE as it mimics the product more realistically?
12. Another point worth emphasizing here: the way which the definitions are presented in the choice sets is crucial. As there is no specific example attached, I can't fully comment on that.
13. Wouldn't free-range imply grass-fed? Then it might be confusing to have free-range and grain-fed product. I don't recall free-range of being prominently used to market steak.
14. "(1) the levels of an attribute occur with equal frequency so that each respondent sees most or all attribute levels, (2) the occurrences of any two levels of different attributes are uncorrelated, and (3) attribute levels that do not vary within a choice set are minimized. This approach is consistent with best practices for experimental design development in DCEs (Johnson et al., 2007)" – these principles are rather dated now. D-optimality is no longer the sole consideration. Bayesian design may be more useful. Also, certain interaction terms may warrant attention.
15. $WTP_1 = \alpha_2 / |\alpha_1|$ – the absolute value on alpha 1 assumes that respondents are all rational about prices.
16. "…and Def2–Def4 are effects-coded variables for whether the product used definition 2, definition 3, or definition 4 to define the "Product of USA" labeling claim." --Hu et al. (2022) AJAE argues for dummy coded variables, especially when interaction terms are considered.
17. Descriptions about the practical implications and inferences from the DCE are somewhat lacking. The intersections between the rulemaking and the DCE are not sufficiently described. I.E., if the hypothesis from DCE 1 is rejected, what is its implication to the rulemaking process, and vice versa? Similarly for other DCEs.