

# **Analyzing Consumers' Value of "Product of USA" Labeling Claims**

**OMB No. 0583-NEW**

## **Supporting Statement**

### **B. Statistical Methods**

#### **B.1. Respondent Universe and Sampling Methods**

For this study, the target population is the U.S. general population of adults (18 years or older) who are primarily responsible for the grocery shopping in their household<sup>1</sup> and have purchased beef or pork in the last 6 months. To administer the web-based survey/experiment, RTI will subcontract with Ipsos's KnowledgePanel (<https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel>), a probability-based panel that is designed to be nationally representative of the U.S. adult population. This representation is achieved through address-based sampling (ABS), where every U.S. adult with an address (including those who do not have a landline phone number) has an equal probability of being selected for participation on the panel. Selected panelists without Internet access are provided with free Internet access and a tablet computer, if needed. The KnowledgePanel has some limitations that should be considered when interpreting survey results. The low recruitment rate for panel participation, panel attrition, and nonresponse among panelists selected to complete a particular survey may lead to a very low overall response rate (less than 10%), which may result in nonresponse bias if nonrespondents are systematically different from respondents (Tourangeau et al., 2013). Other potential limitations include sampling and coverage issues, nonresponse from breakoffs (i.e., not completing the survey), and measurement error (Tourangeau et al., 2013). This study will calculate response rates using American Association

---

<sup>1</sup> "Primarily responsible" means the individual does all, most, or about half of the grocery shopping in the household.

for Public Opinion Research response rate formulas (American Association for Public Opinion Research, 2016). The study procedures include a nonresponse bias analysis as described in Section B.3.

## **Study Population**

The sampling frame for the web-based survey/experiment is the U.S. general population of adults (18 years or older) who are members of the KnowledgePanel and speak English or Spanish (the survey will be translated into Spanish by TransPerfect, a professional translation company, certified to both ISO 9001:2015 and ISO 17100:2015) and who self-reported on their panel profile survey that they do all, most, or about half of the grocery shopping. From this sampling frame, panelists selected for the survey will be rescreened regarding their grocery shopping status to confirm their eligibility and also screened to determine if they have purchased beef or pork products from a grocery store, a butcher, or online within the past 6 months. At least 300 respondents will be Spanish-speaking people who will complete the Spanish version of the survey.

## **Panel Description**

**Panel Recruitment.** Since its inception in 1999, the KnowledgePanel has recruited participants based on industry standards for probability-based general population surveys. In the past, the panel relied on random-digit dialing (RDD) for recruitment. Currently recruitment is primarily through ABS. The ABS methodology is a random sample of addresses from the U.S. Postal Service's Delivery Sequence File. A residential household with at least one adult who is 18 years of age or older is considered an "eligible household." Individuals residing at randomly sampled addresses are invited to join the KnowledgePanel through a series of mailings (in English and Spanish); nonresponders are phoned when a telephone number can be matched to the sampled address. Non-internet households are provided a web-enabled tablet and

free internet access. Historical recruitment rates for participation in the panel are approximately 15 to 20%.

The KnowledgePanel's probability-based recruitment was originally based exclusively on a national RDD frame. In April 2009, in response to the growing number of cell phone-only households that are outside of the RDD frame, Ipsos migrated to using an ABS frame for selecting panel members. Most recently, approximately 10% of panel members were recruited through RDD methodology, while 90% were recruited using an ABS methodology. As previously noted, for both ABS and RDD recruitment, households without an internet connection were provided with a web-enabled device and free internet service. After initially accepting the invitation to join the panel, participants are asked to complete a short demographic survey (the initial profile survey); answers to these questions allow efficient panel sampling and weighting for surveys. Completion of the profile survey allows participants to become panel members. These procedures were established for the RDD-recruited panel members and continued with ABS-recruited panel members. Respondents sampled from the RDD and ABS frames are provided the same privacy terms and confidentiality protections.

ABS involves probability-based sampling of addresses from the U.S. Postal Service's Delivery Sequence File (DSF). The key advantage of the ABS sample frame is that it allows sampling of almost all U.S. households and improves population coverage—an estimated 97% of households are “covered” in sampling nomenclature. Regardless of household telephone status, those households can be reached and contacted through postal mail. The stratification plan for the ABS design shifts and evolves with time and currently leverages a combination of geographic oversampling and demographic oversampling, for example, slightly oversampling rural households and those likely to have a young adult. The stratification relies on ancillary

information that has been appended to the DSF by sample providers that includes a mix of commercial databases and Census data. Ipsos regularly reviews the stratification methodology, adjusting it as needed based on panel needs and differential response and attrition rates.

Randomly sampled addresses are invited to join the KnowledgePanel through a series of mailings, including an initial invitation letter, a reminder postcard, and a subsequent follow-up letter. Approximately 40% of the physical addresses selected for the sample can be matched to a corresponding valid telephone number. About 5 weeks after the initial mailing, telephone refusal-conversion calls are made to households for whom a telephone number was matched to the sampled address. Invited households can join the panel by: (1) completing and mailing back a paper form in a postage-paid envelope, (2) calling a toll-free hotline phone number maintained by Ipsos, or (3) going to a designated Ipsos website and completing the recruitment form at the website.

**Panel Management.** On average, panel members are invited to complete one survey per week and complete three to four surveys per month. Typical survey durations are 10 to 15 minutes per survey. Panelists are proactively withdrawn from the panel after nonresponse to numerous consecutive survey invitations.

## **Respondent Selection Methods**

There are two aspects of sampling for the KnowledgePanel: one occurs at the panel recruitment stage and the other occurs at the survey sampling stage.

**Panel Recruitment Stage.** Once panel members are recruited and profiled by completing the Core Profile Survey, they become eligible for selection for client surveys. Typically, specific survey samples are based on an equal probability selection method (EPSEM) for general population surveys. Customized stratified random sampling based on “profile” data can also be implemented as required by the study design, which is the case for this study in which profile

data on grocery shopping for the household are being used for prescreening—that is, members are drawn from a subsample of the panel. With this approach, all subsequent survey samples drawn that week are selected so that the resulting sample remains representative of the population distributions.

For selection of general population samples from the KnowledgePanel, a patented methodology has been developed such that samples from the panel behave as EPSEM samples. That is, the general population samples have proportional representation similar to the general population from which they are selected as if they were selected under an EPSEM design. Briefly, this methodology starts by weighting the pool of active members to the geodemographic benchmarks secured from a combination of the U.S. Census Bureau’s American Community Survey (ACS) and the March 2022 supplement of the U.S. Census Bureau’s Current Population Survey (CPS) along several dimensions. Typically, the geodemographic dimensions used for weighting the entire KnowledgePanel include the following dimensions, with additional nesting of dimensions as well:

- Gender (male/female)
- Age (18–29, 30–44, 45–59, and 60+)
- Race/Hispanic ethnicity (White/non-Hispanic, Black/non-Hispanic, other/non-Hispanic, 2+ races/non-Hispanic, Hispanic)
- Education (less than high school, high school, some college, bachelor and beyond)
- Census Region (Northeast, Midwest, South, West)
- Household income (under \$10 K, \$10K to <\$25 K, \$25K to <\$50 K, \$50K to <\$75 K, \$75K to <\$100K, \$100K to <\$150K, and \$150K+)

- Home ownership status (own, rent/other)
- Household size (1, 2, 3, 4+)
- Metropolitan Area (yes, no)
- Hispanic origin (Mexican, Puerto Rican, Cuban, other, non-Hispanic)
- Language dominance (non-Hispanic and English dominant, bilingual, and Spanish dominant Hispanic) when the survey is administered in both English and Spanish, which is the case for this study.

**Survey Sampling Stage.** Using the resulting weights as measures of size, a probability-proportional-to-size (PPS) procedure is used to select study specific samples. It is the application of this PPS methodology with the imposed size measures that produces demographically balanced and representative samples that behave as EPSEM. Moreover, in instances where a study design requires any form of oversampling of certain subgroups, such departures from an EPSEM design are accounted for by adjusting the design weights in reference to the Census benchmarks for the population of interest.

To determine the number of people that are to be sampled, the target respondent size of 4,400<sup>2</sup> was inflated by the cooperation, eligibility, and contact rates. Table B-1 shows the initial sample selected (starting sample size), expected contact rate, expected eligibility rate, the expected cooperation rate (targeted respondent sample size), and the expected design effect<sup>3</sup> (the effective sample size). The effective sample size is the equivalent to the simple random sampling sample size accounting for the design effect and is the sample size used in the power calculation.

---

<sup>2</sup> FSIS specified a sample size of 4,400 surveys based on the funding that was available for conducting the survey.

<sup>3</sup> The design effect is used to show the effect of the complex survey design on the variance of an estimator. It is the estimate of the variance of an estimator from the complex survey design divided by the estimate of the variance of an estimator for a simple random sample of the same size.

**Table B-1. Sample Size Table for the Web-Based Survey/Experiment**

| <b>Adjustment</b>                | <b>Adjustment Rate<sup>a</sup></b> | <b>Total Sample Size</b> | <b>Sample Description</b>     |
|----------------------------------|------------------------------------|--------------------------|-------------------------------|
| Selected sample                  |                                    | 9,778                    | Starting sample size          |
| Contact                          | 0.990                              | 9,681                    |                               |
| Eligibility                      | 0.909                              | 8,800                    |                               |
| Cooperation                      | 0.500                              | 4,400                    | Target respondent sample size |
| Design effect (2.0) <sup>b</sup> | 0.500                              | 2,200                    | Effective sample size         |

Note: Numbers may not be exact due to rounding.

<sup>a</sup>The adjustment rates were provided by Ipsos and are based on their prior experience with similar surveys.

<sup>b</sup>Ipsos indicated that for similar surveys the design effect is 2.0.

Ipsos sends randomly selected panel members a study participation invitation via email. The email includes a short description of the study and provides a unique link to the survey. Selected participants can also log-on to their password-protected panel home page to access the survey. For this study, approximately 9,778 English- and Spanish-speaking panel members will be sent an email invitation to the survey.

Although there is no explicit stratification for the survey, once the panel members are eligible and agree to participate in the survey, they will be randomized into 24 randomization groups. The 24 randomization groups will be created by cross-classifying four treatment conditions for the limited time exposure (LTE) experiment and six analytic groups (i.e., survey versions) for the discrete choice experiment (DCE) experiment as shown in Table B-2, with about 183 respondents in each randomization group. The treatment conditions for the LTE experiment and survey versions for the DCE are described below in Section B.2.

**Table B-2. Randomization Groups for the Web-Based Survey/Experiment**

| <b>Randomization Group</b> | <b>LTE Condition</b> | <b>DCE Survey Version</b> | <b>Number of Respondents</b> |
|----------------------------|----------------------|---------------------------|------------------------------|
| 1                          | 1                    | 1                         | 183                          |
| 2                          | 1                    | 2                         | 183                          |
| 3                          | 1                    | 3                         | 183                          |
| 4                          | 1                    | 4                         | 183                          |
| 5                          | 1                    | 5                         | 183                          |
| 6                          | 1                    | 6                         | 183                          |
| 7                          | 2                    | 1                         | 183                          |
| 8                          | 2                    | 2                         | 183                          |
| 9                          | 2                    | 3                         | 183                          |
| 10                         | 2                    | 4                         | 183                          |
| 11                         | 2                    | 5                         | 183                          |
| 12                         | 2                    | 6                         | 183                          |
| 13                         | 3                    | 1                         | 183                          |
| 14                         | 3                    | 2                         | 183                          |
| 15                         | 3                    | 3                         | 183                          |
| 16                         | 3                    | 4                         | 183                          |
| 17                         | 3                    | 5                         | 183                          |
| 18                         | 3                    | 6                         | 183                          |
| 19                         | 4                    | 1                         | 183                          |
| 20                         | 4                    | 2                         | 183                          |
| 21                         | 4                    | 3                         | 183                          |
| 22                         | 4                    | 4                         | 183                          |
| 23                         | 4                    | 5                         | 183                          |
| 24                         | 4                    | 6                         | 183                          |



The randomization for the LTE is for the four experimental treatment conditions in the LTE, and the randomization for the DCE is for the six survey versions. These two sets of groups will be cross-classified to ensure all six DCE groups are approximately equally represented in each of the four LTE groups. This eliminates the possibility that the DCE groups will be confounded with the LTE groups; however, analyses will be conducted post-data collection to confirm that the LTE group assignment did not adversely affect the DCE analysis.

## **B.2. Procedures for the Collection of Information**

As described in Part A, Section A.2, the web-based survey/experiment will comprise three components. For the first component, respondents will complete an LTE task to determine whether consumers notice the “Product of USA” labeling claim (i.e., to indicate saliency). For the second component, respondents will answer survey questions to address (1) their understanding of the current “Product of USA” labeling claim as it relates to product country of origin (e.g., born, raised, slaughtered, processed) and (2) their understanding of the meaning of other U.S. Department of Agriculture (USDA) labeling such as “USDA Choice” or the USDA mark of inspection, as related to product country of origin. For the third component, respondents will complete a DCE to measure their intrinsic value (willingness to pay [WTP]) for products bearing the “Product of USA” labeling claim for the current definition and potential revised definitions (e.g., the meat is from an animal that was both slaughtered and processed in the United States).

This section first provides information on the statistical methodology and sample selection for the web-based survey/experiment and describes the overall study procedures for the web-based survey/experiment. Next, for each of the three components of the web-based survey/experiment (i.e., LTE experiment for measuring saliency, survey questions for assessing knowledge, and DCE for measuring WTP), the collection of information, the estimation

procedures, and the degree of accuracy required for the study are described. There are no unusual problems requiring specialized sampling procedures and no periodic data collection cycles will be used.

## **Study Procedures**

To administer the web-based survey/experiment, Ipsos will send panelists selected for this study an email invitation to invite them to participate in the study (Appendix B). Once selected panelists click on the survey link, they will be provided information on informed consent and asked if they would like to proceed with the study (see Appendix A for the survey instrument; the first two screens provide information on informed consent). If panelists decline, they will be categorized as nonrespondents. If panelists accept, they will be asked several questions to determine eligibility (as noted in Section B.1). Panelists not eligible to complete the survey will be categorized as ineligible. Panelists who are deemed eligible will be randomly assigned to one of 24 study conditions as described in Section B.1 and will proceed with the survey. The survey will be available in English and Spanish and will take an average of 20 minutes to complete.

Before the administration of the full-scale study, Ipsos will conduct a pilot with a sample of 30 English- and Spanish-speaking panel members to ensure the programming logic is working correctly. Approximately 83 panel members will be sent an email invitation to complete the pilot. The same sampling and recruiting methods for the full-scale study will be used for the pilot. If changes to the survey instrument are required based on the pilot study, FSIS will submit a revised version of the instrument for review and approval before launching the full-scale study.

For the full-scale study, Ipsos will send up to two automatic email reminders to nonresponding panelists during data collection (see Appendix C).

## **Statistical Methodology Weighting**

Once the study sample has been selected and fielded, and the survey data edited and made final, the survey data will be weighted. The weighting process starts by computing base weights to address any departure from an EPSEM design. To minimize potential bias, the design weights will be adjusted for any survey nonresponse and for any under- or overcoverage imposed by the study-specific sample design. Depending on the specific target population for a given study, geodemographic distributions for the corresponding population will be obtained from the CPS, ACS, or in certain instances from the weighted KnowledgePanel profile data. For this study, the weighted KnowledgePanel profile data will be used for panel members who are primarily responsible for the grocery shopping in their household (i.e., the individual does all, most, or about half of the grocery shopping in their household) and adjusted for eligibility based on responses to the screening questions (primarily responsible for grocery shopping and have purchased beef or pork in the last 6 months). For weighting adjustments, an iterative proportional fitting (raking) procedure will be used to adjust design weights to produce final weights that will be aligned with respect to all study benchmark distributions simultaneously. In the final step, calculated weights will be examined to identify and, if necessary, trim outliers at the extreme upper and lower tails of the weight distribution. The resulting weights will be then scaled to the sum of the total sample size of all eligible respondents.

Standard weighting dimensions will be as follows, though adjustments will be made as needed based on the target population of interest (i.e., grocery shoppers):

- Gender (male and female) by age (18–29, 30–44, 45–59, 60+)
- Race-Ethnicity (White/non-Hispanic, Black/non-Hispanic, other/non-Hispanic, Hispanic, 2+ races/non-Hispanic)

- Census region (Northeast, Midwest, South, West) by metropolitan status (metro, nonmetro)
- Education (less than high school, high school, some college, bachelor’s degree or higher)
- Household income (under \$25K, \$25–\$49,999, \$50K–\$74,999, \$75K–\$99,999, \$100K–\$149,999, \$150K and over)

Because the study includes Spanish survey takers, there will be an additional adjustment for language proficiency:

- Language proficiency (English-proficient Hispanic, bilingual Hispanic, Spanish-proficient Hispanic, non-Hispanic)

## **Component 1: LTE Experiment for Measuring Saliency**

### *Collection of Information*

We will use methods from signal detection theory to measure saliency—the ability of a stimulus to attract attention in a complex field—for the “Product of USA” labeling claim on meat products. Signal detection is a branch of psychophysiology that examines the ability of a subject to discriminate visual or auditory stimuli that contain information (i.e., signal) from stimuli that do not contain information (i.e., noise) (MacMillan, 2002). The methodology typically involves exposing subjects to a stimulus and asking them to recall whether specific items were present or not. Subjects are typically exposed to the stimuli for a limited amount of time; thus, the approach is called LTE. For this study, respondents will be exposed to one of four randomly assigned mock packages for a meat product (ground beef) for 20 seconds and then respondents will be asked to answer a series of questions. Ground beef was selected as the meat product because it is a meat product often purchased by consumers of meat. Three of the

packages (treatment conditions) will bear the “Product of USA” labeling claim (the packages will be the same with the exception of the format of the “Product of USA” labeling claim), and one package will resemble the other three packages except for not having the claim (i.e., the control condition).

FSIS does not regulate the format or location of the “Product of USA” labeling claim; thus, a wide variety of label formats can be seen in the marketplace. Although the labeling claim may be displayed on the back of the package, this study will only examine saliency for the front of the package because the extent to which consumers may turn the package over to look at the back of the package is not known, and it would make the task too burdensome for respondents to consider the front and the back of the package.

To determine the format (e.g., size, color, use of icon) and placement (e.g., top right corner) for the three products bearing the “Product of USA” claim, a sample of 202 ground beef products bearing the “Product of USA” labeling claim that were randomly selected from the Label Insight database (<https://www.labelinsight.com/>), a proprietary data source with product attribute meta data, was reviewed. First, the products were coded to determine whether the “Product of USA” claim was on the front or back of the package and whether the packaging was tray/vacuum packed or a chub. Among the 202 products, 155 (83%) were packaged in a tray/vacuum packed. Among the 155 products packaged in a tray/vacuum packed, 99 had some type of the “Product of USA” claim on the front. About half of these claims (51%) had a similar version of “Product of USA,” but there were some other popular versions of the claim such as “Born, raised, and harvested in the USA” (26%) and “100% American” (12%). These 99 products were further coded for the following characteristics: placement of “Product of USA” claim (e.g., upper left, lower right), presence of flag or USA/state shape icon, use of contrast

color for text/icon, size of text/icon relative to product name, use of special formatting (e.g., border), and whether text for the claim is stand-alone or included in a list of other claims). The coded data were analyzed to identify the three most common formats of the “Product of USA” type claim among the sampled products. Among the 99 products, 42% had a flag or a USA-/state-shaped icon accompanying the claim, 30% had the claim formatted in a border, and the remaining labels did not have either of these formats. The majority of claims (52%) were of medium size relative to the product name, 29% were smaller, and 19% were larger. The placement of the claim on the product varied across all products and by how the claim was formatted. Based on the results of the analysis, the three labels bearing the “Product of USA” labeling claim (i.e., treatment labels) will be formatted as follows:

- (1) Claim is accompanied by flag/USA-shaped icon, is located in lower-left corner of package, is medium sized relative to product name, and printed in contrasting color.
- (2) Claim is formatted within a border, is located in upper-right corner of package, is medium-sized relative to product name, and printed in contrasting color.
- (3) Claim is stand-alone, is located in center right of package, is medium-sized relative to product name, and printed in contrasting color.

Selected panelists who meet the screening criteria and agree to participate will be randomly assigned to one of the four conditions for the LTE as previously described.

Respondents will receive the following instructions: “For the next question, assume you are at the grocery store, butcher shop, or shopping online and you are going to buy a package of ground beef. On the next screen, we are going to show you a package of ground beef. You will see the package for about 20 seconds. Carefully review the information on the product package because we are going to ask you a few questions about what you saw.”

Respondents will first complete an unaided recall task: “Please list everything you remember seeing on the food package. Please type each thing you remember seeing, such as words, pictures, and symbols, on a separate row. For pictures or symbols, please provide a description of what you saw.” The responses to the open-text question will be coded to indicate whether respondents recalled seeing the “Product of USA” claim (yes/no) and, if yes, the order in which it was listed (e.g., 1, 2, 3, 4).

Next, respondents will complete a cued recognition task. This task comprises eight dichotomous yes/no questions in which respondents will be asked whether they remember seeing certain words, pictures, or symbols on the package: “Now we are going to ask you if you remember seeing different words, pictures, or symbols on the product package. Only click YES if you are sure you saw the word, picture, or symbol; otherwise, click NO.” Four of the questions will ask about items that were on the package (including the “Product of USA” claim for the three treatment conditions) and four of the questions will ask about items not on the package (including the “Product of USA” claim for the control condition). The order of the eight questions will be randomized. For the three conditions with the “Product of USA” claim, the question about whether they saw the claim on the package will be considered a hit if they answer “yes.” For the control condition without the “Product of USA” claim, the question on whether they saw the claim on the package will be considered a miss or false alarm if they answer “yes.”

To practice the LTE task, respondents will first complete an example task for a mock chicken tender product (unaided recall and four cued recognition questions).

### *Estimation Procedures*

The primary analysis for the research question about saliency will investigate the comparison of four independent proportions of respondents who correctly recalled seeing the

“Product of USA” labeling claim on a package of ground beef (when viewing the front of the package for a limited time). The four independent proportions will be from three treatment conditions, each with a different format for the “Product of USA” labeling claim, and a control condition, which will not have the “Product of USA” labeling claim. The null hypothesis is that the three treatment condition proportions are the same as the control condition proportion. The alternative hypothesis is that at least one of the treatment condition proportions is different from the control proportion.

To answer this question, the proportions from the four independent samples estimated using the coded responses to the unaided recall task will be compared using a similar approach to the chi-squared test described in Fleiss, Levin, and Paik (2003, pp. 187–192) implemented using SUDAAN<sup>®</sup> (Research Triangle Institute, 2012) to account for the complex survey design and differential weighting. Secondary analysis using logistic regression in SUDAAN (Research Triangle Institute, 2012) may be used to determine the relationship among treatment and control conditions and the importance of other possible independent variables related to saliency (e.g., gender, race, ethnicity, or income). To provide additional information on saliency, the relative order in which respondents listed the “Product of USA” labeling claim (e.g., first, second ... last) will also be examined, assuming that items listed near the top of the list have more saliency than those listed farther down the list.

To provide qualitative information on saliency, the saliency of each of the four conditions (three treatment and one control) will be estimated based on the  $d'$  score (Bylinskii et al., 2017). The  $d'$  score is calculated from the responses to the set of eight dichotomous yes/no questions. As previously noted, four questions presented information that was on the package (including the “Product of USA” claim for the three treatment conditions); an affirmative (yes) response to each



will be referred to as a hit. Four additional questions presented information that was not on the package (including the “Product of USA” claim for the control condition); an affirmative (yes) response to each will be referred to as a false alarm. The order in which the eight questions (four hits and four false alarms) were shown will be randomized. First, the number of hits and false alarms will be summed separately. The respondents’ hit rate and false-alarm rate will determine  $d'$  using the following formula:

$$d' = H - score - F - score$$

where

*H-score* is the number of correct hits.

*F-score* is the number of incorrect hits (i.e., false alarms).

Applying this formula results in a  $d'$  score with a range of  $-4$  to  $+4$ . The condition with the highest  $d'$  score indicates that respondents were more likely to notice the “Product of USA” claim when it is formatted as shown on the treatment package.

#### *Degree of Accuracy Required for the Study*

To ensure that the study has adequate power to detect a difference, if one exists, a power analysis was conducted for the first research question, the unaided recall task, based on the estimation procedures described in the previous section. A range of null proportions (0.00 to 0.10 by 0.01) and a range of differences from the null proportion (0.00 to 0.05 by 0.01) were investigated to determine the power for each of these combinations. The range of null proportions represent a reasonable baseline for the control condition; that is, the proportion of respondents who said that they saw the “Product of USA” label, when it was not on the package. We expect this proportion to be small. The range of differences represent the higher proportions, baseline control condition proportion plus the difference, that have to be observed for at least one

treatment condition to be statistically significantly higher than the baseline control condition proportion.

The test statistic for the comparison of four proportions is the observed chi-squared value, and the critical value for the test is the theoretical chi-squared value based on 3 degrees of freedom and an alpha of 0.05. The power is calculated as a function of the critical value, degrees of freedom, and observed chi-squared value.

Given an initial sample of 4,400 completed surveys<sup>4</sup> and four conditions, each condition will have a sample size of 1,100 completed surveys. Because the power calculations are based on simple random samples, the condition sample size of 1,100 was divided by the design effect of 2.0<sup>5</sup> for the proposed design to determine the effective sample size of 550 per condition.

Figure B-1 shows the power curves for null hypotheses from 0.00 to 0.10 for the comparison of four independent proportions. The horizontal axis shows the differences for at least one treatment condition from the null hypothesis proportion. The vertical axis is the power with a black horizontal line at power equal to 0.8, which is the minimum power the test should achieve.

Of the analyzed null hypotheses, the worst-case scenario shown in Figure B-1, is equal to 0.10. That is, 10% of the respondents in the control condition recall seeing the “Product of USA” labeling claim when it was not present on the package. To achieve power approximately equal to 0.8, at least one of the treatment conditions need to have only a difference of 5 percentage points from the null hypotheses; that is, 15% of the respondents recall seeing the “Product of USA” labeling claim on the package. More realistic null hypotheses of 0.03 or less require differences of approximately 0.03 or less. That is, smaller proportions of respondents in the control condition

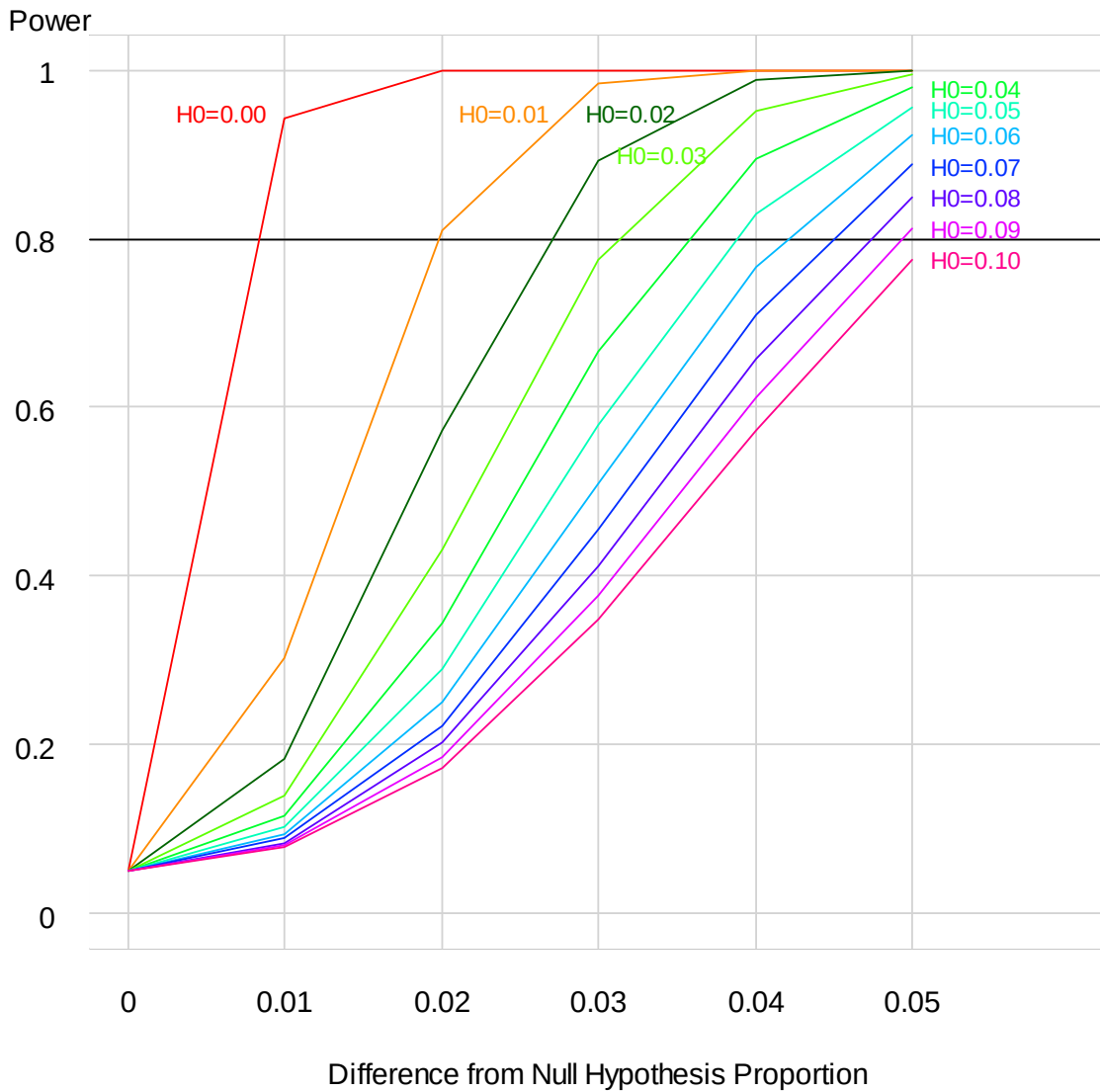
---

<sup>4</sup> FSIS specified a sample size of 4,400 respondents based on the funding that was available for conducting the survey.

<sup>5</sup> Ipsos indicated that for similar surveys the design effect is 2.0.

who misreport seeing the “Product of USA” labeling claim on the package when it was not present, are a more reasonable expectation than 10%.

**Figure B-1. Comparison of Four Independent Proportions for LTE (Condition Effective Sample Size = 550)**



## Component 2: Survey Questions for Assessing Knowledge

### *Collection of Information*

The following survey question will be asked to determine if respondents can correctly identify the current definition for the “Product of USA” labeling claim from a list of responses:

To your knowledge, what does the “**Product of USA**” labeling claim on meat products mean?

For the answer choices below, a meat product “**processed in the USA**” means the meat was packaged in the USA or cut/ground (for example, into pork chops or hamburger) and then packaged in the USA.

1. The product must be made from animals born, raised, and slaughtered and the meat then processed in the USA.
2. The product must be made from animals raised and slaughtered and the meat then processed in the USA; the animals can be born in another country.
3. The product must be made from animals slaughtered and the meat then processed in the USA; the animals can be born and raised in another country.
4. The product must be processed in the USA; the animals can come from another country.
5. Not sure/don’t know

Response option number 4 is the correct definition. The survey will include a distractor question that asks about the definition for the “Natural” claim. The order of the two questions and the response options for each question will be randomized.

FSIS is also interested in knowing if respondents have the misperception that quality claims from the USDA Agriculture Marketing Service (i.e., “USDA Choice”) and the USDA Seal of Inspection indicate that the meat product is a product of the United States. To collect information to address this issue, two different questions will be asked. The order of the two questions and the response options for each question will be randomized. The questions are shown below:

To your knowledge, what does “**USDA Choice**” on beef products mean? *Select all that apply.*

1. The beef was evaluated (graded) and is considered high-quality beef for tenderness, juiciness, and flavor. *[Correct response]*
2. The cows used to produce the beef were treated humanely from birth to slaughter on farms that provide suitable living conditions that meet the animals’ needs.
3. The beef does not contain any bacteria (e.g., *Salmonella*) that can cause foodborne illness.
4. The beef is a product of the USA.
5. Not sure/don’t know

Please look at this symbol.



To your knowledge, what does this symbol on meat products mean? *Select all that apply.*

1. The meat was produced under federal inspection of the U.S. Department of Agriculture (USDA). *[Correct response]*
2. The animals used to produce the meat were treated humanely from birth to slaughter on farms that provide suitable living conditions that meet the animals’ needs.
3. The meat does not contain any bacteria (e.g., *Salmonella*) that can cause foodborne illness.
4. The meat is a product of the USA.
5. Not sure/don’t know

### *Estimation Procedures*

The primary analysis for the second research question about knowledge is related to specific survey questions and investigates the proportion of all respondents who correctly identify the current definition for the “Product of USA” labeling claim on meat products. The null hypothesis is that the proportion is equal to 0. The alternative hypothesis is that the proportion is greater than 0.

To answer this question, the proportion and the standard error of the proportion will be estimated using SUDAAN (Research Triangle Institute, 2012) to account for the complex survey design and differential weighting. Secondary analysis using logistic regression in SUDAAN

(Research Triangle Institute, 2012) may be used to determine the relationship among treatment and control conditions and the importance of other possible independent variables related to knowledge.

#### *Degree of Accuracy Required for the Study*

Because the null hypothesis is that the proportion is equal to 0, if at least one respondent correctly identifies the current definition for the “Product of USA” labeling claim on meat products then the null hypothesis should be rejected and the alternative hypothesis accepted. Because there is uncertainty associated with taking a sample, the practical minimum is about three correct responses out of 2,200 total responses. Even in the worst-case scenario where respondents are randomly guessing the response with equal probability, the number of “correct” responses would be well above three. Given almost any knowledge at all of the definition, the number of correct responses should be well above three or four. Consequently, any power calculations involving a null hypothesis of 0 for a binary variable will show a power of 1 for any reasonable alternative estimates for the proportion.

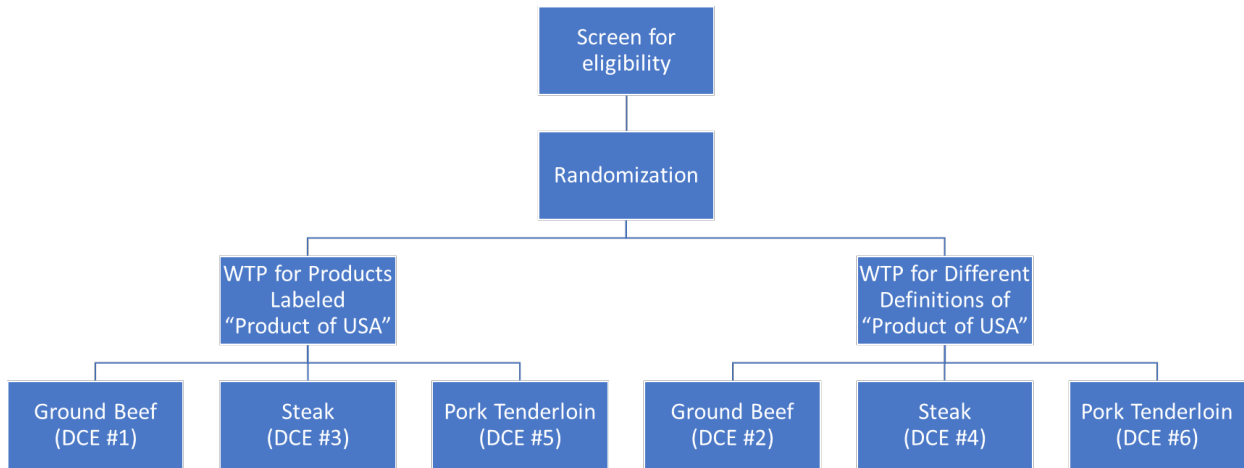
### **Component 3: DCE for Measuring WTP**

#### *Collection of Information*

The goal of this component of the study is to estimate how much respondents are willing to pay for meat products bearing the “Product of USA” labeling claim. In this experiment, respondents will be randomly assigned to one of six versions of the survey. Each survey will differ based on two label conditions (whether “Product of USA” is defined or not) and three meat product conditions (ground beef, NY strip steak, and pork tenderloin). This process is illustrated in Figure B-2. Eligible respondents will be randomly assigned to one of the six versions of the survey. For the DCE, respondents will be asked to complete a series of choice tasks where they

must choose between different hypothetical products that vary according to a set of attributes. This section provides detail on the study design, describes the attributes used in the choice experiment, and describes how the DCE questionnaire will appear to survey respondents.

**Figure B-2. Respondent Assignment to Survey Version**



### *Study Design*

This study will use six versions of a DCE experiment to estimate two different measures of how much respondents are willing to pay for products labeled “Product of USA”: (1) an estimate of how much respondents are willing to pay for a meat product labeled “Product of USA” when no definition of the label is provided and (2) how much respondents are willing to pay for meat products with differing definitions of “Product of USA.”

- **WTP for Products Labeled “Product of USA”:** Three versions of the survey (Versions 1, 3, and 5) will present the respondent with hypothetical products that may include a “Product of USA” labeling claim with no detailed description of the meaning of the labeling claim. Not including a definition will simulate the way most consumers will likely engage with a “Product of USA” labeling claim. Specifically,

when consumers are grocery shopping, they do not have access to educational material on the regulatory definition of “Product of USA.”

- **WTP for Different Definitions of “Product of USA”:** Three other versions of the survey (Versions 2, 4, and 6) will present the respondent with hypothetical products that will include a “Product of USA” labeling claim, but the respondent will be provided with additional information on how this label is defined for each product (current definition and three potential revised definitions that vary based on the production stages that take place in the United States). The value of this approach is it allows FSIS to determine which definition of “Product of USA” provides the average consumer with the greatest value. The definitions for a single meat product, ground beef, are shown below. Similar definitions will be used for steak and pork tenderloin by changing the type of meat and species as appropriate. Definition 1 is the current definition for “Product of USA.”

- **Definition 1 (Def 1):** The ground beef was processed in the USA, meaning it was packaged in the USA or ground and then packaged in the USA.
- **Definition 2 (Def 2):** The ground beef was made from cattle that were slaughtered and the meat then processed all within the USA.
- **Definition 3 (Def 3):** The ground beef was made from cattle that were raised and slaughtered and the meat then processed all within the USA.
- **Definition 4 (Def 4):** The ground beef was made from cattle that were born, raised, and slaughtered and the meat then processed all within the USA.



For each approach to estimating WTP, the study considers three types of meat products: 1) ground beef, 2) NY strip steak, and 3) pork tenderloin. Although the current “Product of USA” definition applies to all meat and poultry products, the study uses beef and pork products because these are the products that are most likely to be directly affected by changes to “Product of USA” labeling regulations. In addition, the study considers high-value beef products (i.e., steak) and low-value beef products (i.e., ground beef) because it seems possible that WTP for products produced in the United States may differ across these products.

#### *Attributes and Levels Used in DCE*

In the DCE component of each version of the survey, respondents will be asked 9 or 10<sup>6</sup> choice questions (in addition to a practice question) where they must choose between different hypothetical products. These products will be composed of a number of distinct characteristics or “attributes.” Some of these attributes (i.e., fixed attributes) will be constant across different hypothetical products such as the type of product under consideration. However, other attributes (i.e., randomized attributes) will be varied across questions to create 2 hypothetical meat products (e.g., two packages of ground beef) for each of the 9 or 10 choice tasks. Specifically, two types of randomized attributes are included in this study:

**Price:** This attribute will be included in the DCE because it is required to estimate the marginal utility of income, which is needed to estimate WTP (as discussed in more detail below). The levels used to describe this attribute for each meat product were selected using price data collected from USDA’s national weekly retail activity report (USDA, n.d.). Upon receiving Office of Management and Budget (OMB) approval, these prices will be updated using the same data source and methodology to ensure

---

<sup>6</sup> As described in the “DCE Questionnaire” section below, Versions 1, 3, and 5 will have 10 choice questions and Versions 2, 4, and 6 will have 9 choice questions.

the most up to date information is used. Each product will have three levels: the lowest observed price/pound, the national weighted average price/pound, and the highest observed price/pound.

**Labeling Claims:** For the purposes of this study, the most important labeling claim to include on each meat product is the “Product of USA” labeling claim described above. However, there are three reasons why we included other labeling claims when describing the hypothetical meat products. First, we wanted to avoid single-cue bias, where a product’s country of origin on a consumer’s perceptions and choices has a larger effect when they are told nothing else about the product. This bias has been observed in multiple Country of Origin Labeling studies (e.g., Peterson & Joilbert, 1995). The second reason to include other labeling claims on the hypothetical meat products is to measure how much consumers are willing to pay for products labeled as “Product of USA” relative to other attributes that consumers value. This information will help provide context for the WTP results. Lastly, by including other labeling claims on the meat products, we hope to make the choice tasks more realistic because consumers would have to make trade-offs between these attributes when choosing meat products in the real world. Therefore, in addition to “Product of USA,” the DCE includes other labeling claims consumers may consider when purchasing meat as product attributes. These labeling claims were chosen by reviewing which labeling claims are frequently included by manufacturers on these types of meat products using the Label Insight database. Each of the labeling claim attributes has two levels: (1) present on the label (yes) or (2) not present on the label (no).

Table B-3 presents the attributes and levels for ground beef and steak, and Table B-4 presents the attributes and levels for pork tenderloin.

**Table B-3. Attribute Table for Ground Beef and Steak Versions of the DCE**

| <b>Attribute</b>                         | <b>DCE #1</b>                        | <b>DCE #2</b>                        | <b>DCE #3</b>               | <b>DCE #4</b>                    |
|--|--------------------------------------|--------------------------------------|-----------------------------|----------------------------------|
| Product type                             | Ground beef<br>(85% lean/15%<br>fat) | Ground beef<br>(85% lean/15%<br>fat) | NY strip steak<br>(Choice)  | NY strip steak<br>(Choice)       |
| Price/pound                              | \$2.99<br>\$4.99<br>\$5.69           | \$2.99<br>\$4.99<br>\$5.69           | \$8.19<br>\$9.99<br>\$11.49 | \$8.19<br>\$9.99<br>\$11.49      |
| “Product of<br>USA”/location<br>produced | Yes<br>No                            | Def 1<br>Def 2<br>Def 3<br>Def 4     | Yes<br>No                   | Def 1<br>Def 2<br>Def 3<br>Def 4 |
| Grass fed                                | Yes<br>No                            | Yes<br>No                            | Yes<br>No                   | Yes<br>No                        |
| Free from<br>antibiotics                 | Yes<br>No                            | Yes<br>No                            | Yes<br>No                   | Yes<br>No                        |

**Table B-4. Attribute Table for Pork Tenderloin Versions of the DCE**

| <b>Attribute</b>                      | <b>DCE #5</b>              | <b>DCE #6</b>                    |
|---------------------------------------|----------------------------|----------------------------------|
| Product type                          | Pork tenderloin            | Pork tenderloin                  |
| Price/pound                           | \$2.79<br>\$3.99<br>\$5.49 | \$2.79<br>\$3.99<br>\$5.49       |
| “Product of USA”/Location<br>Produced | Yes<br>No                  | Def 1<br>Def 2<br>Def 3<br>Def 4 |
| Free from added hormones              | Yes<br>No                  | Yes<br>No                        |

(continued)

**Table B-4. Attribute Table for Pork Tenderloin Versions of the DCE (continued)**

| Attribute | DCE #5 | DCE #6 |
|-----------|--------|--------|
| Lean      | Yes    | Yes    |
|           | No     | No     |

By varying price and labeling claims according to an experimental design, one can see how a respondent's purchase decisions change when these attributes are changed. This variation can be used to quantify respondent preferences using statistical methods as described below.

The way attribute levels are varied into different combinations to create hypothetical products is called the experimental design. In most cases, the number of possible combinations is too large to ask respondents to evaluate all possibilities. However, if participant preferences meet some very basic assumptions, robust statistical results can be obtained from a fractional factorial design implemented in far fewer tasks. Using Sawtooth Software, experimental designs will be created for each of the six versions of the DCE with consideration for the following: (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; 2013).

#### *DCE Questionnaire*

At the beginning of the DCE portion of the survey, respondents will be presented with a description of the hypothetical choice tasks they will be asked to complete (customized to each of the six survey versions). This description will include details on the context of the hypothetical choice task (i.e., they are asked to imagine shopping for the product in a grocery store, at a butcher, or online). This description will also include plain language descriptions of all

product attributes that are sufficient to provide respondents with a basic understanding of each attribute. The exception is that a definition will not be provided for “Product of USA” for the survey versions examining WTP for products labeled “Product of USA.”

Next, a series of statements will instruct respondents to read the survey carefully and will describe the potential consequences of misleading survey results if the choice questions are not answered truthfully. Statements like these, referred to as “cheap talk,” are best practice in stated preference surveys and have been shown to reduce hypothetical bias in DCE responses (Tonsor et al., 2011).

After reading the “cheap talk” statement, respondents will be asked to complete a simplified choice task where they must choose between two meat products. The simplified choice task will begin with a verbal description of the attributes both products have in common:

“To start, consider Product A and Product B. Please assume they are the brand that you usually buy. Both products are packages of USDA-inspected 85% lean/15% fat ground beef sold by the pound and have the same weight and expiration (sell-by) date. The products are the same except for the features shown on the next screen. Please carefully consider each product.”

Next, the attributes that differ between the two meat products are described verbally as shown below:

| <b>Product Features</b> | <b>Product A</b> | <b>Product B</b> |
|-------------------------|------------------|------------------|
| Grass fed               | No               | No               |
| Free from antibiotics   | Yes              | Yes              |
| Product of USA          | Yes              | Yes              |
| Price/pound             | \$2.99           | \$5.69           |

Given these two options, which package of ground beef would you buy?

1. Product A
2. Product B
3. Neither

This format for presenting different meat product attributes is similar to the format used in Loureiro and Umberger (2007). Note that in this first, simplified choice task only the price attribute will be varied and all other attributes will be constant between Product A and Product B. This simplification not only makes the first choice question easier to answer, but also implies that there is a “correct” answer. Under the usual assumptions of consumer rationality, a respondent should prefer the less expensive version of two identical products. If a respondent chooses the more expensive product, the respondent is informed that the two products were identical other than price and is encouraged to read the survey carefully. The purpose of this question is to provide respondents with a warm-up question that can help them better understand the DCE part of the survey. In addition, this warm-up question can help identify whether a significant number of respondents had difficulties understanding the types of questions being asked. If it can be demonstrated that most respondents understood simplified survey questions, this can help support the internal validity of the survey (Johnson et al., 2019). The data from the pilot study will be analyzed to calculate the percentage of respondents who preferred the more expensive option. If a high number of percentages prefer the more expensive option, refinements will be made to improve understanding of the choice questions.

After completing the first simplified choice task, all respondents will be asked nine random choice questions where all attributes will vary according to an experimental design (as described above). For the survey versions in which respondents are asked to choose between a “Product of USA” labeling claim with no detailed description of the meaning of the labeling claim (Versions 1, 3, 5), respondents will complete an additional choice question, called a fixed choice question, in which all the attributes are the same except one product has the “Product of USA” labeling claim and one does not have the claim. The responses to this question can be used to directly assess whether respondents prefer products labeled “Product of USA” when all other attributes of the product are held constant.

#### *Estimation Procedures*

The data collected from the DCE component of the study will be used to test the following sets of hypotheses:

##### *Hypothesis 1—Willingness to Pay for Products Labeled “Product of USA” versus no Claim*

- $H_0$  = The difference between the amount respondents are willing to pay for meat products bearing the “Product of USA” labeling claim versus no claim is not statistically different from 0.

$H_1$  = The difference between the amount respondents are willing to pay for meat products bearing the “Product of USA” labeling claim versus no claim is statistically different from 0.

##### *Hypothesis 2—Willingness to Pay for Different Definitions of “Product of USA”*

- $H_0$  = The difference between the amount respondents are willing to pay for meat products bearing different definitions of the “Product of USA” labeling claim is not statistically different from 0.

$H_1$  = The difference between the amount respondents are willing to pay for meat products bearing more stringent definitions (i.e., more production stages take place in the United States) of the “Product of USA” labeling claim is statistically different from 0.

Using regression analysis, random utility models will be estimated to investigate each of these hypotheses. Such regression models assume that the probability of choosing one meat product is a function of the attribute levels in the profile. These models will follow best practices discussed in Hauber et al. (2016) and use the mixed logit estimator, which allows to quantify differences in preferences across survey respondents.

### *Testing Hypothesis 1*

Following Hauber et al. (2016) the simplified regression model for testing Hypothesis 1 for each product is formally stated as follows.

$$Pr(Choice_i) = \alpha_0 + \alpha_1 Price_i + \alpha_2 PUSA_i + \sum_{j=3}^k \alpha_j X_{ji} + \varepsilon_i$$

In this model, *Choice* is a dichotomous indicator for whether a survey respondent chose meat product *i* or not, *Price* is a continuous variable representing the price of meat product *i*<sup>7</sup>, and *PUSA* is a dichotomous indicator for whether meat product *i* includes a “Product of USA” label (where the variable equals 1 if the product is labeled “Product of USA” and equals 0 if not). The model also includes  $X_j$ , which represents dichotomous indicators for each other meat product attributes (e.g., whether the product has a “Grass Fed” label or not) and a random error term ( $\varepsilon$ ).

The parameters of this regression model have an intuitive interpretation. Specifically,  $\alpha_1$  is a measure of the utility the average respondent receives for each additional dollar increase in price. One would expect that this parameter is negative. In fact, the absolute value of this

---

<sup>7</sup> The model assumes that price enters the random utility model linearly. However, other functional forms will be explored when conducting the analysis to determine if they are more appropriate.



parameter is the marginal utility of income. Similarly,  $a_2$  measures the utility the average respondent receives from a meat product with the “Product of USA” labeling claim. Following Train et al. (2009) and assuming consumers are rational, the WTP for a product that has the “Product of USA” label can be estimated as follows:

$$WTP_1 = \frac{\alpha_2}{|\alpha_1|}$$

This estimate of WTP will be used to test Hypothesis 1 using a Wald Test to see whether WTP is significantly different from 0.

### *Testing Hypothesis 2*

Again, following Hauber et al. (2016), the simplified regression model for testing Hypothesis 2 for each product is formally stated as follows.

$$Pr(Choice_i) = \beta_0 + \beta_1 Price_i + \beta_2 \delta 2_i + \beta_3 \delta 3_i + \beta_4 \delta 4_i + \sum_{j=5}^k \beta_j X_{ji} + \varepsilon_i$$

In this model, *Choice* is a dichotomous indicator for whether a survey respondent chose meat product *i* or not, *Price* is a continuous variable representing the price of meat product *i*, 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. The current definition (Def1) is excluded because it is the reference value. The model also includes  $X_j$ , which represents categorical variables for other meat product attributes (e.g., whether the product includes a “Grass Fed” label or not) and a random error term ( $\varepsilon$ ).

It is important to note that the variables for the “Product of USA” definition will be effects coded. Effects-coding is similar to dummy-coding except that they take on a value of -1 when the attribute is at the “base” level (instead of zero). The benefit of using effects-coding in this context is the coefficient for the excluded definition (Def1) can be recovered as the negative

sum of the coefficients on the non-omitted levels of that attribute. Therefore, effects-coding yields a unique coefficient for each variable definition. The differences between effects and dummy-coding are discussed in more detail in Hauber et al. (2016).

As before, how much WTP differs by each definition will be estimated as follows:

$$WTP_{i1} = \frac{-(\beta_2 + \beta_3 + \beta_4)}{|\beta_1|}$$

$$WTP_{i2} = \frac{\beta_2}{|\beta_1|}$$

$$WTP_{i3} = \frac{\beta_3}{|\beta_1|}$$

$$WTP_{i4} = \frac{\beta_4}{|\beta_1|}$$

The difference between each WTP estimate will also be calculated to formally test Hypothesis 2 and to assess whether respondents are willing to pay more for meat products bearing more stringent definitions (i.e., more production stages take place in the United States) of the “Product of USA” labeling claim. Specifically, a Wald Test will be used to see whether the difference between each WTP is significantly different from 0.

### *Internal Validity Testing*

As part of the analysis, a series of internal validity tests will be conducted to assess the quality of data collected during the DCE. Internal validity tests check to see whether survey respondents were logical and consistent when answering DCE questions. Specifically, the following three internal validity tests will be conducted:

- **Dominant-option validation test:** As noted above, the first question in all six versions of the DCE is a choice question in which both meat products are identical except one is less expensive than the other. If a respondent understands the choice

task, they should obviously prefer the less expensive product. The greater the number of respondents that choose the less expensive option, the more confidence one has that the majority of respondents understood the questions being asked and provided well-considered answers. We will calculate how many respondents fail this internal validity test (i.e., choose the unambiguously worse option) and compare that number with other estimates in the literature. For example, in a review of 55 health-related DCEs, Johnson et al. (2019) found 21 studies that included this type of validity test. They found that the median failure rate across these 21 studies was 7% (i.e., 7% of respondents in each survey chose the unambiguously worse option).

- **Attribute dominance (noncompensatory preferences):** As Johnson et al. (2019) noted, choice experiments assume that respondents have compensatory preferences. This means that respondents should be willing to accept a reduction in one desirable attribute in return for a sufficiently large compensating increase in another desirable attribute. One can test whether respondents have noncompensatory preferences by looking at the respondent's answers to each DCE question and seeing whether they always chose the alternative with the better level of one attribute. We will estimate the percentage of respondents who exhibit noncompensatory preferences in our sample and compare that percentage with other estimates in the literature. For example, in a review of 55 health-related DCEs, Johnson et al. (2019) found that in the median study 20% of respondents exhibited noncompensatory preferences.
- **Straight-lining:** In the DCE component of this survey, respondents will answer nine questions where they must choose between two meat products (Product A and Product B). As Johnson et al. (2019) noted, the probability that the most preferred

option will always appear in the same position (i.e., Product A is always the most preferred or Product B is always the most preferred) in this many pairwise comparisons is less than 1%. As a result, if a respondent always chooses Product A or Product B as the most preferred option, this is evidence that they are not answering the DCE questions carefully. This behavior is referred to as “straight-lining.” We will estimate the percentage of respondents who exhibit straight-lining behavior and compare that estimate with other estimates in the literature. For example, in a review of 55 health-related DCEs, Johnson et al. (2019) found that in the median study 2% of respondents exhibited straight-lining behavior.

#### *Additional Analysis for External Validity*

As noted above, the survey asks respondents to make choices between different hypothetical products. Although the study design takes measures to minimize hypothetical bias (i.e., including cheap talk), the possibility of this bias influencing results still exists. To explore this possibility and obtain further external validity of our results, FSIS will conduct additional analyses using scanner data. For example, FSIS plans to use Label Insight and IRI retail scanner data to develop hedonic models that estimate the price premium of the “Product of USA” label claim on FSIS regulated products, such as ground beef.

#### *Degree of Accuracy Required for the Study*

As discussed above, respondents will be randomly assigned to one of the six survey versions. Ideally, the sample size would be based on a detailed power analysis, but traditional power calculations do not directly translate to preference surveys like DCEs because power is determined by a number of factors that are unknown before the study is conducted. When a

parametric approach for determining sample size is not feasible, simple rules of thumb are used.

One common rule of thumb for DCEs is known as Orme's rule:  $n \geq \frac{1,000c}{ta}$ :

- $n$  = number of participants
- $t$  = number of choice tasks per participant
- $a$  = number of alternatives per questions (not including the "neither" alternative)
- $c$  = the largest number of levels for any one attribute

In the context of this experiment, respondents will complete a minimum of nine choice tasks ( $t = 9$ ), with two alternatives per question ( $a = 2$ ) and a maximum of four levels for a single attribute ( $c = 4$ ). This suggests a minimum sample of 222 respondents per survey version.

However, it is important to note that this is strictly a bare minimum sample size. Orme (2019) recommends that DCEs conducted for research purposes should typically include a minimum of 300 respondents and that a larger number of respondents is preferable. Orme's recommendation of using a minimum of 300 respondents is consistent with the results of simulation studies conducted by Johnson et al. (2013), which found that the precision of DCE model estimates tend to flatten out after sample sizes exceed 300 for a given level of measurement error. However, Johnson et al. (2013) noted that measurement error can differ significantly across studies, making larger sample sizes preferable. With a sample size of 4,400 and six versions of the survey, the number of respondents per version will be approximately 733, which will allow for the estimation of models that yield reliable inferences.

### **B.3. Methods to Maximize Response Rate and Deal with Nonresponse**

RTI will maximize response rates for the web-based survey/experiment using a variety of approaches, monitor response rates throughout the data collection period, and conduct a nonresponse bias analysis. Ipsos will develop survey weights for use in the analysis.

#### **Procedures to Maximize Response Rate**

The approach for the web-based survey/experiment seeks to maximize response rates and minimize nonresponse. Specifically, the survey instrument is at 7.1 grade reading level (as measured by the Flesch-Kincaid Grade Level Readability Test) with simple instructions. Survey methodologists at RTI reviewed the survey instrument using their Question Appraisal System (QAS), a structured, standardized instrument review methodology that evaluates questions relative to the tasks they require of respondents, specifically with how respondents understand and respond. It also documents the question features that are likely to lead to response error such as comprehension, task definition, information retrieval, judgment, and response generation. Additionally, RTI conducted cognitive interviews (described in Section B.4) in April 2022 to evaluate participants' understanding and interpretation of the survey questions, to assess participant burden, and to enhance administration.

Based on experience conducting 20-minute online surveys with general population samples (i.e., adults 18 years or older), Ipsos estimates that about 50% of the eligible panelists will complete the survey. Ipsos will send up to two automatic email reminders to nonresponding panelists during the course of data collection (see Appendix C).

In addition, to encourage participation, the email invitations and reminders will state the study purpose, identify USDA as the study sponsor (see Appendix B for the email invitation and Appendix C for email reminders), and provide an email address and toll-free number for the RTI

team lead for panelists to obtain additional information about the study or verify the authenticity of the study.

#### *Monitoring Response Rates During Data Collection*

Throughout the course of the data collection, Ipsos will provide RTI daily status reports with the number of completes, ineligible, and nonrespondents. RTI will carefully review the daily reports to identify and, if necessary, work with Ipsos to address any concerns.

#### *Post-data Collection Weighting*

The post-data collection weighting is described in the section Statistical Methodology Weighting.

#### *Nonresponse Bias Analysis*

One advantage of using the KnowledgePanel for the survey sampling frame is that it provides a large set of possible variables to use in the nonresponse bias analysis. The best variables to use in the weight adjustment and the imputation processes are variables that are related to the response propensity and the values of key analytic variables. Consequently, response propensity models will be used to identify variables that should be incorporated in the weight adjustment models for each step in the weight adjustment process. In addition, prediction models will be developed for key analytic variables to identify variables that may be useful across all the steps in the weighting adjustment process. For this study, the contact indicator, cooperation indicator, and many of the key variables are binary variables. SUDAAN's logistic regression procedure LOGISTIC or RLOGIST (Research Triangle Institute, 2012, pp. 641-765) will be used for both the response propensity and prediction models to account for the complex survey design and differential weighting.

#### **B.4. Tests of Procedures or Methods to be Undertaken**

RTI conducted cognitive interviews with eight adults (five with English speaking adults and three with Spanish speaking adults) in April 2022 to test the survey instrument. The purpose of the cognitive interviews was to identify any survey questions and response items that were confusing or difficult for respondents to understand. The cognitive interviews were also used to measure respondent burden. Based on the findings from the cognitive interviews, several questions that were deemed unnecessary for the analysis were deleted to reduce the length of the survey. Additionally, the DCE was simplified by reducing the number of choice questions and removing some of the attribute levels and revisions were made to the instructions to facilitate understanding.

To ensure that the programming logic, sample distribution and fulfillment, and data compilation are functioning correctly, RTI will work with Ipsos to conduct a pilot with 30 randomly selected panelists. Data collection for the pilot will not commence until OMB approval is obtained. If changes are made to the survey instrument based on the pilot findings, FSIS will submit a revised survey instrument to OMB for review and approval before starting data collection.

#### **B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data**

Sheryl Cates is the RTI Project Director and will manage the study. Darryl Creel of RTI will oversee the sampling and analysis for the LTE experiment on saliency and the survey questions on knowledge. Dr. Dallas Wood of RTI developed the DCE and will conduct the analysis to estimate WTP. Andrew Pugliese, an FSIS employee, provided feedback on the study design and analysis procedures and is providing agency oversight of the study.



## References

- American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). AAPOR.  
[https://www.aapor.org/AAPOR\\_Main/media/publications/Standard-Definitions20169theditionfinal.pdf](https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf)
- Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2017). What do different evaluation metrics tell us about saliency models? <https://arxiv.org/pdf/1604.03605.pdf>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. (3rd ed.). John Wiley & Sons, Inc.
- Hauber, A. B., MarcosGonzález, J., Groothuis-Oudshoorn, C. G. M., Prior, T., Marshall, D. A., Cunningham, C., IJzerman, M. J., & Bridges, J. F. P. (2016). statistical methods for the analysis of discrete choice experiments: A report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health*, 129(4), 300–315.  
<https://www.sciencedirect.com/science/article/pii/S1098301516302911>
- Johnson, F., Kanninen, B., Bingham, M., & Özdemir, S. (2007). Experimental design for stated choice studies. In B. Kanninen (Ed.), *Valuing Environmental Amenities Using Stated Choice Studies* (pp. 159–202). Dordrecht, Netherlands: Springer.
- Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments. Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*, 16, 3–13.  
[https://www.valueinhealthjournal.com/article/S1098-3015\(12\)04162-9/pdf](https://www.valueinhealthjournal.com/article/S1098-3015(12)04162-9/pdf)

- Johnson, F. R., Yang J. C., & Reed, S. D. (2019). The internal validity of discrete choice experiment data: A testing tool for quantitative assessments. *Value in Health*, 22(2), 157–160.
- Loureiro, M. L., & Umberger, W. J. (2007). A choice experiment model for beef: What US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability. *Food Policy*, 32(4), 496–514.
- MacMillan, N. (2002). Signal detection theory. In J. Wixted (Ed.), *Steven's handbook of experimental psychology*. (3rd ed., pp. 43–89). John Wiley and Sons.
- Orme, B. (2019). *Getting started with conjoint analysis: Strategies for product design and pricing research* (4th ed.). Research Publishers LLC.
- Peterson, R. A., & Jolibert, A. J. (1995). A meta-analysis of country-of-origin effects. *Journal of International Business Studies*, 26(4), 883–900.
- Research Triangle Institute. (2012). *SUDAAN language manual, Volumes 1 and 2, Release 11*. Research Triangle Institute.
- Tonsor, G. T., & Shupp, R. S. (2011). Cheap talk scripts and online choice experiments: “looking beyond the mean.” *American Journal of Agricultural Economics*, 93(4), 1015–1031.
- Tourangeau, F., Conrad, F. G., & Couper, M. P. (2013). *The science of web surveys*. Oxford University Press.
- Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge University Press.

U.S. Department of Agriculture, Agricultural Marketing Service. (n.d.). Livestock, poultry, and grain market news national weekly retail activity reports.

<https://www.ams.usda.gov/market-news/retail>