Consumer Research on the Safe Handling Instructions Label for Raw and Partially Cooked Meat and Poultry Products and Labeling Statements for Ready-to-Eat and Not-Ready-to-Eat Products

OMB No. 0583-NEW

Supporting Statement

B. Statistical Methods

B.1. Respondent Universe and Sampling Methods

Web-Based Experimental Study

RTI will subcontract with Lightspeed (www.lightspeedresearch.com), a provider of a non-probability-based, opt-in national online consumer research panel. Lightspeed will program the survey instrument and administer the data collection for the web-based experimental study. We are using nonprobability sampling because it is not feasible to draw a random probability-based sample of the entire population given time and cost considerations. The use of a non-probability sampling approach has some limitations, as described below, specifically that inferences cannot be made to the U.S. population. Probabilistic (i.e., random) sampling is important in surveillance studies, where data are collected on a single sample of participants and the sample estimate is assumed to be representative of the broader population. In such studies, probabilistic sampling is the only mechanism for controlling selection bias (i.e., influences on the estimate due to factors related to the people in the study and unrelated to the measured outcome). In comparison, experimental designs employ random assignment to experimental conditions to control selection bias. Random assignment uses a probability-based approach—each participant has a known and equal chance of being assigned to any of the experimental conditions—to allocate sample members. This known probability mechanism serves to distribute the potentially

confounding influences of selection equally across experimental conditions, and the method of difference serves to remove both measurable and unmeasurable confounding from the estimated effect (Rosenthal & Rosenow, 1991). Accordingly, the proposed experimental design with random assignment to one of the 27 experimental conditions will provide an unbiased assessment of which among the 27 labels are more likely to attract consumer attention.

The study population for the web-based experimental study is the U.S. general population of adults (18 and older) who are members of the Lightspeed panel. Although some subpopulations are at greater risk of contracting foodborne illness (e.g., young children, the elderly, pregnant women), the SHI label is intended to reach the general population with information on safe handling practices; thus, the study is not limited to at-risk populations and will include a diverse sample of participants who vary on specific demographic characteristics, as described in more detail below.

Opt-In Panel Description

Lightspeed's U.S. consumer panel consists of approximately 1.5 million adults. Lightspeed uses the following methods to recruit panelists: email, co-registration, e-newsletter campaign, traditional banner placements, social marketing, and internal and external affiliate networks. A double opt-in approach is used to enroll consumers interested in participating on the panel. When a respondent clicks on a link from a panel ad, s/he is directed to the panel registration survey. Each prospective panelist must provide demographic and household information, pass through some validation checks (e.g., verify postal address), and agree to the website Terms and Conditions and Privacy Policy. Those who pass the Lightspeed checks are sent an email to confirm their email address. After clicking on a link within the email, they

complete the double opt-in process and become panel members. Table B-1 shows the demographic composition of the panel and how it compares to 2014 Census data.

Table B-1. Comparison of Lightspeed Panel to Census Data

Category	2014 Census Data[a]	Lightspeed Panel Composition[b]
Gender		
Male	49%	30%
Female	50%	70%
Region		
Midwest	21%	22%
North East	18%	15%
South	38%	43%
West	24%	19%
Age		
15–24	17%	26%
25–34	17%	22%
35–44	16%	16%
45–54	17%	14%
55+	32%	22%
Income (categories used by Lightspeed)		
Less than $20,000[c]	13%	37%
$20,000–$39,000[d]	21%	27%
$40,000–$74,000[e]	31%	17%
$75,000–$99,000	12%	9%
$100,000–$150,000	13%	7%
Greater than $150,000	10%	3%

[a] Source: U.S. Census Bureau. (n.d.). 2010–2014 American Community Survey 5-year data profiles. Retrieved from https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2014/

[b] Source: Lightspeed. 2018. Global Panel Book: 2018.

[c] U.S. Census Bureau 2010–2014 American Community Survey 5-year data income categories are different from Lightspeed's categories. Equivalent of "Less than $20,000" is "Less than$15,000".

[d] U.S. Census Bureau 2010–2014 American Community Survey 5-year data income categories are different from Lightspeed's categories. Equivalent of "$20,000–$39,000" is "$15,000–$34,999".

[e] U.S. Census Bureau 2010–2014 American Community Survey 5-year data income categories are different from Lightspeed's categories. Equivalent of "$40,000–$74,000" is "$35,000–$74,999".

Respondent Selection Methods

Lightspeed sends randomly selected panel members a study participation invitation via email. The email includes a short description of the study and instructs respondents to log-on to their password-protected panel home page to access the survey. For this study, approximately 70,000 English- and Spanish-speaking panel members will be sent an email invitation to the survey (see Appendix A). Interested panelists will be screened to ensure they meet the following criteria (see Appendix B):

- are 18 years of age or older;
- cook raw meat or poultry at home at least once per week;
- have not cooked or worked professionally in a food preparation setting within the past 5 years; and
- have not taken any type of food safety training, such as ServSafe, in the past 5 years.

Inbound quotas based on the most current demographic data from the U.S. Census Bureau will be set to obtain a diverse sample across all respondents with respect to education, age, race, and English vs. Spanish speaking (see Table B-2). This approach will not use a probability-based, nationally representative sample in any statistical sense, but setting quotas will help ensure that the sample is diverse with respect to the demographic characteristics noted above. Previous research suggests that consumer safe handling practices vary based on certain

demographics such as education, age, race, and ethnicity (see, for example, Patil, Cates, & Morales, 2015; Kosa et al., 2019; Quinlan, 2013; and Redmond & Griffith, 2003).

Table B-2. Quotas for Inbound Sampling for Web-Based Experimental Study

Category	Quota based on Census Data[a]
Race	
White	74%
Non-White	26%
Speak Spanish at home	
No	87%
Yes	13%
Age	
18–34	28%
35–54	36%
55+	36%
Education	
Less than high school or high school diploma/GED (including vocational training)	42%
Some college (no degree) or associate or 2-year degree	29%
Bachelor's degree	18%
Graduate or professional degree	11%

[a] Source: U.S. Census Bureau. (n.d.). 2010–2014 American Community Survey 5-year data profiles. Retrieved from https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2014/

The study will include 3,600 participants with approximately 133 participants randomly assigned to view each of the 27 SHI labels created by fully crossing the three primary study features—label shape, safe handling instruction text, and safe handling icons—each of which will have three options. Appendix Q describes the process used to develop the revised SHI labels and presents the 27 SHI label options to be tested in the web-based experiment.

Before the administration of the full-scale study, Lightspeed will conduct a pretest with a sample of 100 English- and Spanish-speaking panel members to ensure the programming logic is working correctly. Approximately 1,700 panel members will be sent an email invitation to complete the pretest. The same sampling and recruiting methods for the full-scale study will be used for the pretest.

Limitations

The web-based experimental study does not use a probability-based sampling methodology because it would be cost-prohibitive. The main limitation to using opt-in panels instead of a probability-based sample as noted by the American Association for Public Opinion Research (AAPOR, n.d.) is that panel members are not representative of the general U.S. population. Although 89% of Americans have access to the Internet (Pew Research Center, 2015), people who do not have access differ systematically from those who do have access. People without access to the Internet tend to be older, less educated, and poorer than those with Internet access, and this difference can bias the results. Panel surveys often have very low response rates; as a result, the sample respondents may not be a representative subset of the panel members initially invited for the survey.

These limitations will be taken into consideration when interpreting the results of the web-based experiment. As previously noted, inferences cannot be made to the U.S. population; however, by using an experimental design with random assignment to treatment condition, we will be able to obtain an unbiased assessment of which among the 27 labels are more likely to attract consumer attention with a high degree of internal validity. Additionally, setting quotas with regard to education, age, and other demographic characteristics will ensure that our sample is diverse and includes people who may be less likely to have Internet access such as older adults

and less educated individuals. Setting a quota for education level also helps ensure that we include people who may have lower literacy compared with more educated people. Although there are some limitations to the study, many communication research studies regularly face similar constraints, and none of these limitations should threaten the internal validity of our results.

Behavior Change Study

The behavior change study employs an experimental design in which participants are randomly assigned to one of three treatment groups (that will be used to assess three alternative SHIs as identified by the web-based experimental study) or a control group (that will be used to assess the current SHI). We are using nonprobability sampling because for the behavior change study because it is not feasible to draw a random probability-based sample of the entire population given time and cost considerations. The use of a non-probability sampling approach has some limitations, as described below, specifically that inferences cannot be made to the U.S. population.

The data collection will be conducted in test kitchens, each similar in design and layout, in four locations across the country: (1) Wake, Orange, and/or Durham Counties, NC; (2) Brazos County, TX; (3) Yolo County, CA; and (4) Providence County, RI. The four locations were selected to provide geographic diversity with data collection in three of the four Census regions. The Census regions are groupings of states and the District of Columbia that subdivide the United States for the presentation of Census data and are often used in other studies to ensure geographic diversity. Additionally, areas of the United States that have a relatively large percentage of Hispanics were selected so that the percentage of Spanish-speaking participants in the study sample is equal to that of the U.S. population (13% of U.S. adults speak Spanish at

home).[1] The North Carolina counties (South Atlantic Division, South region) were selected because North Carolina State University (NCSU) has three test kitchens located in this area. The Texas (West South Central Division, South region) and California (West region) locations were selected because of the availability of suitable test kitchens and to reach Spanish-speaking individuals (25.1% Hispanic for Brazos County, TX, and 31.5% for Yolo County, CA). Providence County (Northeast) was selected to reach an urban population in the Northeast because of the availability of a suitable test kitchen and to reach Spanish-speaking individuals (22.1% Hispanic).[2]

The respondent universe is English- and Spanish-speaking adults living within driving distance (e.g., 30-minute drive) of each of the four locations. Because probability-based sampling is not being used, inferences cannot be made to the U.S. population; however, by using an experimental design that is powered to detect change in the desired outcome (the power analysis is described in Section B.2), we can assess which label most likely to encourage consumers' adherence to the safe handling instructions.

Respondent Selection Methods

Study participants will be recruited in the four locations using strategies that include social media outreach via Craigslist, Facebook ads, Google Ad Words, and Instagram (Appendix D); by sending emails to Expanded Food and Nutrition Education Program participants (for the North Carolina location only) (Appendix E); and through notices about the study that may also be posted in local grocery stores, food retailers, food banks, and other locations (Appendix F). If necessary, we may conduct additional outreach with hard-to-reach populations such as Hispanics, older adults, and high-school educated individuals (e.g., reaching

[1] U.S. Census Bureau. (n.d.). 2010–2014 American Community Survey 5-year data profiles. Retrieved from https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2014/

[2] Percentage of population that is Hispanic available at http://www.city-data.com/.

out to parents and guardians of the Juntos program in North Carolina [which helps Latinos have more success in middle and high school] and community groups who work with cooperative extension programs), similar to the approach FSIS has used for other observational studies.

Interested individuals can complete the screening questionnaire online or by telephone (see Appendix G). Eligible participants will be contacted by phone and invited to participate in the study and an appointment scheduled. Appointments will be scheduled during work hours, evenings, and weekends to allow for a broader participant pool.

These recruiting methods will be supplemented using outbound recruiting to recruit individuals with specific demographics that may be challenging to recruit using social media (e.g., individuals with a high school education or less and older adults), as necessary. For the outbound recruiting, the study team will work with a local market research company in each of the four locations. Each market research company will contact adults from their database and screen them for eligibility. The databases maintained by each of the local market research companies have been developed over time through advertisements, word-of-mouth referrals, outreach to various organizations and industries, and Internet list searches, among other means. The databases are constantly refreshed through natural attrition and additions. In addition to contact information, basic socioeconomic characteristics are collected through the registration process and updated periodically so that the market research firms can easily identify individuals with specific demographic characteristics (e.g., high school education).

RTI/NCSU has used similar recruiting methods for conducting observation studies for FSIS (OMB No. 0583-0169: *In-Home Food Safety Behaviors and Consumer Education: Annual Observational Study*, expiration 6/30/2020) and has enrolled study participants with diverse

demographic characteristics (e.g., for a recent observational study, 14% of participants were Hispanic or Latino, 25% were 50 years or older, and 24% had a high school diploma/GED).

Participants must meet the following criteria to be eligible to participate in the study:

- are 18 years of age or older;
- cook raw meat or poultry at home at least once per week;
- have not cooked or worked professionally in a food preparation setting within the past 5 years;
- have not taken any type of food safety training, such as ServSafe, in the past 5 years;
- have never had a seizure or have not been diagnosed with epilepsy (it is recommended that such individuals not take part in eye-tracking studies)
- do not wear corrective lenses that may interfere with the calibration of the eye-tracking device (i.e., progressive lenses, hard or semi- hard contact lenses).

Sample Size and Allocation

Based on the power analysis (described in Section B.2), a sample size of 480 is required for the behavior change study, with participants randomly assigned to one of the four study groups (120 per group). Of the 480 participants, data collection for 360 participants will take place in the North Carolina location, and 120 participants in each of the other three locations (40 per location). Most of the data collection will take place in North Carolina because of the greater availability of test kitchens, the logistics of conducting this type of data collection, and budgetary constraints. Because we are using an experimental design (i.e., random assignment to one of four study conditions) with the aim of estimating causal effects, our study will have a high degree of internal validity but may not be generalizable to the broader U.S. population. Within each

location, we will randomly assign participants to one of the four study groups so that the number in each group is similar for each location (i.e., 90 per group in North Carolina and 10 per group in the other locations). With random assignment, the distributions of demographic characteristics for the four study groups are expected to be similar. Table B-3 provides the sample allocation by location and study group.

Table B-3. Sample Allocation by Location and Study Group for the Behavior Change Study

Location	Control (Current SHI)	SHI Option 1	SHI Option 2	SHI Option 3	Total
North Carolina	90	90	90	90	360
Texas	10	10	10	10	40
California	10	10	10	10	40
Rhode Island	10	10	10	10	40
Total	120	120	120	120	480

As part of the screening process, information on participant and household demographic characteristics will be collected. The study will ensure a diverse sample of participants with respect to race, ethnicity, age, education level, and presence of children (0 to 17 years) in the household by using inbound quotas. As previously noted, research suggests that consumer safe handling practices vary based on certain demographics such as education, age, race, and ethnicity. The expected allocation for the full sample is shown in Table B-4. When randomly assigning participants to the four study groups, we will take necessary and reasonable steps to ensure that the demographic characteristics of each group are similar.

Response Rate

The percentage of recruited participants who complete the data collection for the behavior change study is expected to be 85% based on the contractor's experience with similar studies. Thus, the starting sample size for the behavior change study is 565, yielding 480 completed observations (565 x 0.85).

Table B-4. Sample Allocation by Demographic Characteristics for the Behavior Change Study

Characteristic	Response Categories	Percentage	Number
Race	White	68%	326
	Non-White	32%	154
Ethnicity	Not Hispanic or Latino	84%	403
	Hispanic or Latino	16%	77
Age	18–34	35%	168
	35–54	39%	187
	55+	26%	125
Highest level of education	Less than high school, high school diploma/GED, or vocational school	26%	125
	Some college (no degree) or associate or 2-year degree	40%	192
	Bachelor's degree	19%	91
	Graduate or professional degree	15%	72
Child 0–17 years in household	Yes	48%	230
	No	52%	250

Limitations

The behavior change study uses convenience sampling instead of probability-based sampling because it would be cost-prohibitive to select a random sample of the U.S. population and conduct in-person data collection in a large number of locations throughout the United

States. The convenience sampling approach relies primarily on social media recruiting coupled with targeted efforts to recruit hard-to-reach populations (e.g., sending emails, posting signs, and conducting outreach efforts) and, if necessary, outbound recruiting using lists from market research facilities. There are some limitations to convenience sampling instead of probability-based sampling. External validity is a concern because only certain types of people may choose to opt into the study (e.g., those interested in cooking), and they may be different from those who do not in ways that could potentially limit the generalizability of the results.

Also, users of social media differ from the U.S. population as a whole. In 2015, 67% of American adults reported using Facebook, and 72% of online American adults use Facebook. Facebook usage is lower among adults aged 65 or older compared with other age groups (e.g., 48% for people 65 years or older vs. 79% for people 30 to 49 years), as noted by the Pew Research Institute.[3] Usage is similar for different education levels and races/ethnicities. To reach older adults and people who are Hispanic or have a high school education or less (who are traditionally more difficult to recruit using social media), the study uses targeted efforts as previously noted to ensure that people from these demographic groups are included in the final sample.

These limitations will be taken into consideration when interpreting the results of the web-based experiment. As previously noted, inferences cannot be made to the U.S. population; however, by using an experimental design with random assignment to study conditions, we will be able to obtain an unbiased assessment of which among the four labels (three SHI label options vs. the current SHI label) performs best at encouraging participants' adherence to the recommended safe handling instructions for handwashing, using a food thermometer, keeping raw meat/poultry separate from RTE product, and cleaning/sanitizing kitchen surfaces and

[3] http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/

equipment. Additionally, by setting quotas with regard to education, age, and other demographic characteristics, we are including a diverse population in the study. Although there are some limitations to the behavior change study, many communication research studies regularly face similar constraints, and none of these limitations should threaten the internal validity of our results.

B.2. Procedures for the Collection of Information

This section provides an overview of the study procedures, provides information on the degree of accuracy required for the study, and discusses the estimation procedures for each study component. There are no unusual problems requiring specialized sampling procedures for either study component. A separate set of participants is being contacted for each study component and will only be contacted one time, so periodic data collection cycles are not applicable.

Web-Based Experimental Study

Study Procedures

Lightspeed will send panelists selected for this study email invitations to invite them to participate in the study (Appendix A). The email invitations will instruct each panelist to log-on to the Lightspeed website and enter their email address and password to access the link to the survey. 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 B for data collection instrument). 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 a study condition and will proceed with the survey. The survey will be available in English and Spanish and is estimated to take 20 minutes to complete.

Before the administration of the full-scale study, Lightspeed will conduct a pretest with a sample of 100 English- and Spanish-speaking panel members to ensure the programming logic is working correctly. Results of the pretest will also be used to determine the appropriate amount of time for the limited-time exposure task.

Degree of Accuracy Required for the Study

Statistical power (1−ß) provides a measure of the degree of confidence we have that we will not make a Type II error by accepting a false null hypothesis. In other words, 1−ß is our probability of accepting the alternate hypothesis when it is in fact true. By convention, we set the Type II error rate at 0.20 and view 0.80 as an acceptable level for statistical power. To calculate statistical power, we began with an assumed sample size of 3,600 and assign 133-134 individuals to each of 27 treatment groups, where participants in each treatment group are exposed to one of the 27 different SHI labels, generated by the 3x3x3 full factorial research design with factors that include safe handling instructions, safe handling icons, and label shape.

The main effect of the primary study outcome (label salience) among the 27 treatment groups was used to determine statistical power for the study. Label salience will be assessed using a measure of sensitivity (see Estimation Procedures for details) that gauges the participant's ability to accurately recall the stimuli (MacMillan, 2002). The null hypothesis is that mean label salience scores will not differ among the 27 groups.

To test this hypothesis, we applied a one-way analysis of variance. The effect size of this test is *f*, estimated as: $f = \frac{\sigma_m}{\sigma}$, where σ_m is the standard deviation of the mean of each group and σ is the within-group standard deviation (Cohen, 1988; Desu & Raghavarao, 1990). We note that the reader should not confuse the use of notation "*f*" reported here with the *F*-statistic in the

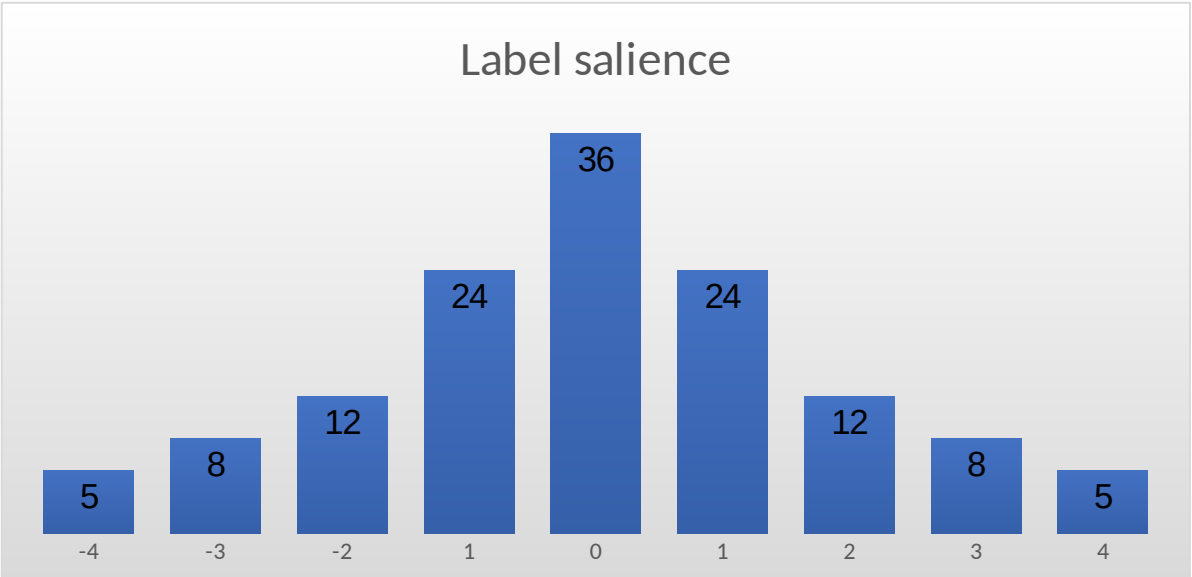
analysis of variance which is a ratio of between-group variances and within-group variance. We use *f* to signify the effect size, which is the ratio of the standard deviations described above, because this is how it is described in Cohen (1988).[4]

We begin with the assumption that among the 27 groups, 20 groups have an attention score mean of 0, five groups have a mean of 0.25, and two groups have a mean of 0.5. We chose these values because smaller values might not be meaningful and larger values might not be observed. With this distribution of means, we calculated the standard deviation of the group means (σ_m) as 0.16 and assumed that the standard deviation within the groups is equal. Accordingly, to achieve power of 80% with a Type I error rate of 5% the standard deviation within the groups must be 1.89 or less. We obtained this result using PASS software and confirmed the result through Monte Carlo simulations.

The study will have 80% power to detect an effect size of $f = \frac{\sigma_m}{\sigma} = \frac{0.16}{1.89} = 0.08$(this is a small effect size based on Cohen's classification) as long as it is reasonable to expect the within-group standard deviation of the label salience variable to be less than 1.89. Consider Figure 1, which shows a distribution of label salience scores that have a standard deviation of 1.89 within group; here, 8% of this distribution has the most extreme values (−4, 4) and 12.5% of this distribution has the next most extreme values (−3, −3). This distribution presents more spread (i.e., extreme scores) than one would expect in the planned experiment. Additionally, because the distribution is symmetrical around zero, it has the greatest chance to have a largest variance. Consequently, we anticipate that the within-group label salience variance will be less than 1.89; therefore, the study will be adequately powered.

[4] It is also described in the PASS software documentation. See page 550-5 in this document: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/PASS/One-Way_Analysis_of_Variance_F-Tests.pdf.

Figure 1. Distribution of Label Salience Outcome Scores with Standard Deviation of 1.89



Estimation Procedures

- Our purpose is to select the three visual label formats that perform best at attracting participants' attention to test in the behavior change study. We refer to the ability of a label to attract attention as salience and define salience as the degree to which a visual feature (target) predominates perceptual attention when presented in a complex environment. Selection will be accomplished by ranking the 27 labels from highest to lowest on the visual salience score.

In this experiment, we will infer salience from a set of items designed to assess the participant's ability to recall information presented in the target from information not presented in the target; we refer to the former as a "hit" and the latter as a "false alarm." The participant will be asked a series of dichotomous (yes/no) questions about information that may (or may not) have been presented in the visual target (see Appendix B for the survey instrument). The item set will be balanced to include the same number of hits and false alarms. For each potential hit, the participant receives one point (+1) for a correct answer (yes), and zero otherwise. For each false

alarm, the participant receives one point (+1) for each incorrect answer (yes), and zero otherwise. The number of hits and false alarms reported can be summarized as proportions and transformed to Z-scores so that each participant's hit rate and false alarm rate are realizations from a unit-normal distribution. These two pieces of information will be used to calculate the individual's ability to accurately differentiate elements that were present from those that were not (*d'*) using the following formula:

$$d' = z(H) - z(F)$$

where

z(H) is the Z-score corresponding to the proportion of hits.

z(F) is the Z-score corresponding to the proportion of false alarms.

This formula describes visual salience as the difference between true positive responses and false negative responses. Label formats with higher positive values indicate that participants were more attentive to the visual target (i.e., high salience). The three SHI labels with the highest visual salience scores will be selected for inclusion in the behavior change study.

The rationale statement associated with the SHI explains the importance of following safe handling instructions to prevent foodborne illness.[5] Survey participants will be presented with five variants of a newly proposed SHI rationale statement (provided in Appendix Q) and asked to rank (i.e., order) the statements in terms of how clearly each one communicates the dangers of foodborne illness. The three rationales with the highest ranking will be retained for developing the final three labels for testing in the behavior change study. Appendix R describes how the ranking data will be analyzed to select the top three rationales. Experts in risk communication

[5] The rationale shown on the 27 different SHI labels is one of the five options that will be ranked by participants and will be kept constant across the 27 labels.

will assign one of the three rationales to each of the three SHI labels with the highest visual salience score to create the final three labels for testing in the behavior change study.

Behavior Change Study

Study Procedures

Upon arrival to the test kitchen, participants will read and sign an informed consent form (see Appendix K). Next, each participant will be asked to watch a video on an iPad so that participants receive a consistent description of the study and what to expect during their participation in the study (see Appendix L). Afterward, participants will be directed to put on a Tobii Pro Glass 2 mobile eye-tracking unit. A trained data collector will adjust the unit for fit and comfort and calibrate the unit following manufacturer recommendations. To ensure precision, the participant will be given a set of two-dimensional test images and directed to examine each in sequence. This step will serve to establish a base rate, confirm the accuracy of the eye-tracking unit, and provide the participant with an opportunity to become accustomed to the wearable data collection device.

After the calibration of the eye-tracking unit, participants will be directed to the general locations of kitchen utensils and staples but will not be prompted to use any particular equipment; kitchen drawers and cabinets will also be labeled with their contents.

While under video observation, participants will be given recipes and ingredients, including two raw meat products bearing the assigned SHI label (control or one of the three treatment labels), and asked to prepare three dishes: (1) gluten-free pasta and meatballs using frozen, preformed raw meatballs; (2) gluten-free pasta and meatballs using raw ground beef; and

(3) a cherry tomato garnish.[6] Participants will also be instructed to clean up afterwards as they would do at home.

The general kitchen layout and setup of equipment will be consistent across test kitchens in the four locations. Video recording equipment will be set up to record meal preparation. Recording of meal preparation will begin as soon as the participant enters the test kitchen and will end after the participant leaves the test kitchen. Participants' cleaning and sanitizing of equipment and the kitchen environment before and after preparation will also be recorded. The meal preparation/observation portion of the study will take 50 to 80 minutes to complete.

Following the meal preparation/observation portion of the study, participants will be provided an opportunity to take a break. If participants choose to take a break, they will remove the eye-tracking device and then it will be refitted/calibrated upon their return. Participants will then be directed to examine each of six mock meat and poultry products (i.e., stimuli). The following products will be used as stimuli for the eye-tracking study and the in-depth interviews (IDI):

- NTRE ground beef patties in a Styrofoam container sealed with plastic wrap (with SHI)
- frozen NRTE ground beef patties in a box (with SHI)
- frozen RTE chicken nuggets/tenders with breading in a bag
- frozen NRTE chicken nuggets/tenders with breading in a bag (with SHI)
- frozen RTE chicken cordon bleu in a box
- frozen NRTE chicken cordon bleu in a box (with SHI)

[6] Participants will be told the dish is for an individual who is on a gluten-free diet to avoid eliciting socially desirable behavior.

Appendix N provides the script the data collector will use to administer the eye-tracking study. The script will direct participants' attention to each product, and participants will be asked to complete several tasks to determine which version of the SHI label (current or one of three alternative versions) is most often attended on a meat and poultry package and to assess whether participants can properly distinguish between RTE and NRTE products that appear to be ready to eat. The eye-tracking study will take up to 30 minutes to complete and will be recorded in audio and video formats.

Lastly, participants will take part in an IDI. Using a semi-structured interview guide (see Appendix N), the data collector will ask participants a series of questions about their views, opinions, and experiences during the meal preparation/observational study and questions to understand how participants determine whether a meat or poultry product is raw versus already cooked. The IDI will take up to 30 minutes to complete and will be recorded in audio and video formats.

Degree of Accuracy Required for the Study

Statistical power (1−ß) provides a measure of the degree of confidence we have that we will not make a Type II error by accepting a false null hypothesis. In other words, 1−ß is our probability of accepting the alternate hypothesis when it is in fact true. By convention, we set the Type II error rate at 0.20 and view 0.80 as an acceptable level for statistical power. To calculate statistical power, we began with an assumed sample size of 480 with 120 individuals randomly assigned to either a control condition (current SHI) or one of three experimental conditions (alternative SHI labels). A sample size of 480 was selected because it is feasible given the budgetary and logistical constraints of conducting in-person data collection. The purpose of the behavior change study is to evaluate the impact of alternative SHI labels on participants'

adherence to the instructions on the label using a measure that quantifies adherence to four safe handling instructions (for details see Estimation Procedures). The measure of behavioral adherence has a range of 0 to 4 points, where participants scoring 0 points demonstrated none of the recommended safe handling practices and participants scoring 4 points consistently demonstrated all recommended safe handling practices (see Appendix S for information on scoring). We anticipate the measure will produce a normal distribution of scores with a mean adherence score among the control group (current SHI) of 1.00 and a standard deviation of 1.30 (Chapman, 2019). Using these parameters as a starting point, we calculate statistical power for a range of small to medium treatment effects using PROC GLMPOWER (SAS Institute, 2013).

We assume observed differences between the control (current SHI) and treatment (alternative SHI) groups will range from 0.29 points to 0.65 points. These differences correspond to effect size estimates (i.e., Cohen's *d*) of 0.22, 0.37, and 0.50, which are in the small to medium range as defined by Cohen (1988) and Lipsey and Hurley (2009). Table B-5 shows the results of the statistical power of the omnibus test and condition-specific, post hoc comparisons. As the table indicates, we anticipate that the omnibus test will have 94% statistical power to determine that adherence scores for at least one of the alternative SHI labels will be significantly different from the current SHI label and that condition-specific tests will be sufficiently powered (80% or better) for differences that are 0.48 points or greater.

Table B-5. Sample Size Requirements for Different Observed Differences between the Control and Treatment Groups for the Behavior Change Study

Condition	Safe Food Handling Score	Test	Observed Difference Between Groups	Statistical Power
Control (current SHI)	1.00	Omnibus	0.29, 0.48, 0.65	0.94

Condition	Safe Food Handling Score	Test	Observed Difference Between Groups	Statistical Power
Treatment 1	1.29	Trx1 v. Control	0.29	0.40
Treatment 2	1.48	Trx2 v. Control	0.48	0.81
Treatment 3	1.65	Trx3 v. Control	0.65	0.97

Note: Trx = Treatment

Accordingly, the proposed sample size of 480 participants (120 per each of the four conditions) takes into consideration several important features of the study: the anticipated base rate for following the instructions on the SHI label among those attending to the instructions, the anticipated distributional characteristics of a continuous outcome, and input from food safety scientists. The study will employ the following strategies to maintain a constant and robust (i.e., 80% or better) level of statistical power for all hypothesis tests:

- Use reliable and validated data collection methods to control inter-subject variability.
- Employ common measures across all study conditions.
- Collect covariates related to the outcome measure that can be applied in multivariate modeling to improve statistical precision.
- Develop analysis strategies that are matched to the distributional assumptions of the outcome measures.

Estimation Procedures

Trained coders will watch the recorded videos for the meal preparation experiment and use an observation rubric (see Appendix M) to code participants' behaviors as following vs. not following the safe handling instructions on the label (handwashing, using a food thermometer,

keeping raw meat/poultry separate from RTE product, and cleaning/sanitizing kitchen surfaces and equipment). We will conduct statistical analyses comparing the label adherence scores among the four groups (i.e., current SHI label and three alternative SHI labels) to determine the most effective label. Appendix S provides additional information on the approach for calculating the label adherence scores.

The data from the eye-tracking study will be reviewed, coded, and processed. The primary outputs from the analysis are summarized below:

- Eye-tracking metrics for each area of interest (AOI), such as percentage of participants who visually notice the relevant AOIs. AOIs will be standardized across packages and will include the SHI labels (and sublabel components), competing instructions, and other labeling information.
- Time to first viewing: The time until each AOI is first noticed.
- Total viewing duration: The total time spent viewing each AOI.
- Number of viewings: The average number of times an AOI is viewed.
- Distribution of attention: The percentage of package viewing time spent on each AOI.
- Perceptual flow: The typical order in which each AOI on a package is seen.
- Eye-tracking heat maps or gaze plots of attention to SHI labels and other package information.

Appendix S provides additional information on the approach for analyzing the eye-tracking data and the statistical analysis procedures for measuring attention to the SHI label among the four groups (i.e., current SHI label and three alternative SHI labels).

Using the findings from the behavior change study, we will construct regression models to examine the association between attention to SHI labels (from the eye-tracking study) and the label adherence score (from the meal preparation experiment). These data will provide empirical evidence on the SHI label option that is most effective at encouraging consumers to follow recommended safe handling practices for raw and partially cooked meat and poultry products.

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

Web-Based Experimental Study

Based on experience conducting 20-minute online surveys with general population samples (i.e., adults 18 years or older), Lightspeed estimates that about 19% of the selected panelists will be eligible and complete the online survey. Lightspeed will send up to 3 three automatic email reminders to nonresponding panelists during the course of data collection. Panelists who do not complete the survey will be categorized as nonrespondents.

To maximize participation, we will conduct cognitive interviews and a pretest to help improve the understandability and usability of the questionnaire, reduce participant burden, and enhance administration.

In addition, to encourage participation, each email invitation and reminder will state the study purpose, identify USDA as the study sponsor (see Appendix A for email invitation and Appendix C for email reminders), and provide an email address and toll-free number (provided by the contractor) for panelists to obtain additional information about the study or verify the authenticity of the study.

Behavior Change Study

We estimate 1,695 prospective participants will complete the screening questionnaire by telephone or via a web link (Appendix G), and 33% of these individuals will be eligible and

subsequently contacted by phone to schedule an appointment. Of the 565 scheduled participants, we estimate that 480 (85%) will participate in the behavior change study.

To maximize the response rate, each participant will receive a cash honorarium of $100 and a small gift (food thermometer valued at $5.38 and magnet valued at $0.23) for their participation. In addition, we will send a confirmation email and up to two additional email reminders (see Appendix I) with directions and make a reminder call (see Appendix J) to recruited individuals before their scheduled appointment.

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

Web-Based Experimental Study

In July and September 2018, the contractor conducted cognitive interviews in Research Triangle Park, NC with nine target audience members (including people who speak Spanish) to determine if any survey questions or response items were confusing or difficult to understand. Based on the cognitive interview findings, we refined the programmed instrument. Specifically, we revised the distractor questions regarding the food packages to ask about images, words, and phrases that are more often found on food packages. We determined that the questions to collect information on which rationale statement (to include on the SHI label) would be most motivating to consumers were not being answered as intended, so we developed a different question to collect this information. To improve understanding and readability, we reformatted several questions and simplified the instructions for some questions. The cognitive interviews also confirmed the estimated burden of 20 minutes (the average time to complete the survey was 17 minutes).

To ensure that the programming logic, sample distribution and fulfillment, and data compilation are functioning correctly, Lightspeed will conduct a pretest with 100 randomly selected panelists. Data collection for the pretest will not commence until OMB approval is

obtained. As previously noted, results of the pretest will also be used to determine the appropriate amount of time for the limited-time exposure task. If changes are made to the survey instrument based on the pretest findings, a revised survey instrument will be submitted to OMB for approval before conducting the full-scale study.

Behavior Change Study

The contractor conducted a pilot study to test the methodology and instruments for the behavior change study in September 2018 with two people in a test kitchen. Based on the pilot, we made several revisions to the methodology and instruments. For the meal preparation study, we had originally considered having participants prepare a parley garnish but decided to change this to a cherry tomato garnish to ensure participants used a knife to prepare a ready-to-eat food. Additionally, we simplified the instructions and tasks for the eye-tracking study, in particular to make the script more conversational. We also revised some of the questions in the IDI script to avoid using a judgmental tone when referring to participants' adherence to recommended handling practices during the meal preparation study. The contractor will conduct additional pilots with test subjects before fielding the main study to assess and refine the training procedures for data collection staff.

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. Dr. Jonathan Blitstein of RTI developed the experimental design for the web-based experimental study and the eye-tracking study and will oversee the analysis for these studies. Dr. Benjamin Chapman of NCSU will manage the data collection for the behavior change study and oversee the analysis for the meal preparation/observational study. Christopher Bernstein, an FSIS employee, will review the study results.

References

AAPOR. (n.d.). Online panels. Oakbrook Terrace, IL: American Association for Public Opinion Research. Retrieved from https://www.aapor.org/Education-Resources/Election-Polling-Resources/Online-Panels.aspx

Cates, S. C., Kosa, K. M., & Muth, M. K. (2015, January). *Professional services to support requirements gathering sessions for Safe Food Handling Instructions: Strategic planning session.* Research Triangle Park, NC: RTI International.

Chapman, B. (2019, March 25). Personal communication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Desu, M. M., & Raghavarao, D. (1990). *Sample size methodology*. New York: Academic Press.

Kosa, K. M., Cates, S., Brophy, J. E., Godwin, S., Chambers, D., & Chambers, E. (2019). Older adults and parents of young children have different handling practices for raw poultry. *Journal of Food Protection*, 82(2), 200-206.

Lipsey, M. W., & Hurley, S. M. (2009). Design sensitivity: Statistical power for applied experimental research. In L. Bickman and D. Rog (Eds.), *The SAGE handbook of applied social research methods* (pp. 44–76). Los Angeles, CA: Sage.

MacMillan, N. (2002). Signal detection theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 4, pp. 43–89). New York, NY: John Wiley and Sons, Inc.

Patil, S., Cates, S., & Morales, R. (2005). Consumer food safety knowledge, practices, and demographic differences: Findings from a meta-analysis. *Journal of Food Protection*, *68*(9), 1884–1894.

Pew Research Center. (2015, September 22). *Coverage error in internet surveys—Who web-only surveys miss and how that affects results*. Washington, DC: Pew Research Center. Retrieved from http://www.pewresearch.org/methods/2015/09/22/coverage-error-in-internet-surveys/

Quinlan J.J. (2013). Foodborne illness incidence rates and food safety risks for populations of low socioeconomic status and minority race/ethnicity: a review of the literature. *International Journal of Environmental Research and Public Health*, *10*(8), 3634–52.

Redmond, E. C., & Griffith, C. J. (2003). Consumer food handling in the home: a review of food safety studies. *Journal of Food Protection,* 66, 130–161.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill, Inc.

SAS Institute Inc. (2013). *SAS/STAT® 13.1 user's guide.* Cary, NC: SAS Institute Inc.

Yang, S., Angulo, F. J., & Altekruse, S. F. (2000). Evaluation of safe food-handling instructions on raw meat and poultry products. *Journal of Food Protection, 63*(10), 1321–1325.