

Experimental Studies of Nutrition Symbols on Food Packages

0910-NEW

DRAFT SUPPORTING STATEMENT

B. Statistical Methods (used for collection of information employing statistical methods)

1. Respondent Universe and Sampling Methods

1a. Respondent universe

Study 1 - The respondent universe of the experimental study is members on the U.S. online consumer panel of Survey Sampling International (SSI). SSI recruits panelists by working with various Web sites directly as well as with data aggregators and Web media agencies. Consumers who are 18 or older can sign up, via a double opt-in process, to join SSI's panel. Currently, the established panel has about 1 million participants.

The respondent universe of the eye-tracking study is members of Eyetracking, Inc.'s Participant Database, which has been built up over the past 10 years by the company. Consumers who are 18 or older and reside in the San Diego, CA metro area voluntarily sign for the database via an advertisement or word of mouth. Currently, there are over 100,000 participants in the Database.

Study 2 - The respondent universe of this study is Synovate's online Consumer Opinion Panel ("ePanel"). U.S. Consumers who are 18 or older are invited to join the ePanel primarily through an affiliate marketing program. Select web sites, portals and Internet Service Providers partner with Synovate to promote ePanel membership through targeted email campaigns as well as placement of banner and pop-up advertisements. Consumers may also join ePanel through referrals from existing ePanel members and re-enlistment of former members. Currently, ePanel has over 2.5 million participants.

1b. Sampling methods

Study 1 and 2 (cognitive interviews) – Respondents in the cognitive interviews will be recruited from a commercial database of residents in the Washington, D.C. metropolitan area. We will recruit approximately 10 to 12 respondents to make sure at least 9 of them will show up for the interviews.

Study 1 (experimental study) - Using the quota sampling method, the experimental study will attempt to achieve a target sample size of 2,400. Quotas will be developed so that the distribution of the overall sample of completed interviews matches that of SSI's consumer panel in age, gender, education, and

ethnicity/race. The contractor will send email invitations to panelists (see Appendix D for invitation email) and monitor completed interviews until the quotas are achieved.

Study 1 (eye-tracking study) - a sample of 30 adult consumers will be drawn by the contractor using pre-determined quotas of gender, age, and education to insure a reasonable degree of diversity within the sample. The contractor will recruit respondents by sending email invitations to members on the Participant Database (see Appendix E for invitation email). Since this is a pilot study, the agency does not intend to generate nationally representative results or precise estimates of population parameters from the eye-tracking study.

Study 2 (experimental study) - The target sample size is 4,800. A quotas will be developed prior to sampling so that the overall sample of panelists who are sent invitations to participate in the study are reflective of the ePanel in gender, age, education, and race/ethnicity, i.e., outbound-balanced (see Appendix F for invitation email). The planned balancing categories are: (a) gender: female and male, (b) age: 18-34, 35-54, and 55+, (c) education: high-school graduate or less and one year or more college education, and (d) race/ethnicity: non-Hispanic white and other. For a given quota, Synovate will send email invitations to randomly selected panelists who qualify for the quota until the quota is filled and each of the predetermined experimental conditions is filled with the predetermined cell size.

The agency does not intend to generate nationally representative results or precise estimates of population parameters from the experimental studies. The studies will use a convenience samples rather than probability samples. Despite the attempt to match between the study's sample and the respondent universe in four demographic characteristics, matching is used solely to produce a sample with a reasonable degree of diversity in key demographic characteristics.

Rather, the strength of the experimental studies lies in its internal validity, on which meaningful estimates of differences across experimental conditions can be produced and generalized. As discussed in the following sections, the agency has taken commonly accepted measures to enhance internal validity of the study. Examples of these measures include random assignment of respondents and conditions, counterbalancing condition assignments within the sample, and use of comparison conditions and relevant covariates.

2. Procedures for the Collection of Information

2a. Studies 1 and 2 (cognitive interviews)

The contractors will use a telephone invitation (Appendix M) to recruit respondents. The recruitment will target for diversity in respondents' gender, age, race and education. They will also be required to meet other eligibility conditions

(see Appendix M). Eligible respondents will complete the draft questionnaires (Appendices G and I) on a personal computer by herself/himself. Then, a moderator will debrief the participant about how she/he interpreted certain questions and the process by which she selected her/his responses.

2b. Study 1 (experimental study)

Members of the contractor’s online consumer panel will be invited by email to an SSI Website to complete the study online (Table 3). Each respondent will first be presented experimental conditions (i.e., images of various mock food labels) and asked about their reactions to the conditions and then be asked about their product experience, knowledge and perceptions and other individual background information (see Appendix G for the full questionnaire). We estimate that it will take respondents about 20 minutes to complete the study.

Table 3. Structure of Study 1 - experimental study

Section	Topic
A (order of A and B will be randomized between respondents)	Which product in a pair of products would be bought; which product in the pair would be bought as a healthier product for the family and reason(s) (open-ended), judgment of the chosen “healthier” product in its (1) overall healthiness, (2) relationship with weight gain and risks of chronic illnesses, (3) contribution toward the overall diet, (4) implication on intake, (5) levels of calories and three of total fat, sodium, sugars, Vitamin A, Vitamin C, calcium, and fiber.
B	Judgment of a single product and its label, including (1) overall healthiness, (2) relationship with weight gain and risks of chronic illnesses, (3) contribution toward the overall diet, (4) implication on intake, (5) level of calories and three of total fat, sodium, sugars, Vitamin A, Vitamin C, calcium, and fiber in the product; (6) taste; (7) perceived helpfulness and trustworthiness of the symbol in communicating nutritional qualities of the product and in helping intake decisions.
C	Judgment of a pair of products of different categories, same type of symbol, and different nutritional qualities: which is healthier.
D	Awareness and use of symbols and who decides eligibility of products for a symbol (government, manufacturer, retailer, other)
E	Consumption and purchase experiences, purchase considerations, and label-reading practice related to the product categories that were asked in Sections A and B, respectively.
F	Prior knowledge and perceptions of the product categories that were asked in Sections A and B, respectively.
G	Dietary interests.
H	Motivation related to use of food labels
J	Health status and demographics

The experiment uses a 5 (labeling condition) x 3 (product category) x 2 (order of two product judgment tasks) fully factorial between-subjects design. The five types of labeling schemes are: Smart Choices Program, Guideline Daily Amounts, a variant of Multiple Traffic Light, NF only control, and no FOP information control. The three product categories are breakfast cereal, savory snack, and frozen entrée. The two orders of task are whether judgment about two products of the same category (Section A) comes before or after judgment of a single product (Section B).

Each respondent will be randomly assigned to an experimental condition. A condition is a combination of the three factors described above (labeling, food product, and task). No respondent will see the same labeling scheme or the same food product twice in Sections A and B. For example, if a respondent is assigned to a Section A that asks about a Smart Choices Program symbol on a pair of cereal packages, then she will not be asked about the Smart Choices Program symbol or a cereal product again in Section B. All front panels will be full-color and identify the food (e.g., raisin bran) but not any brand name. Both the front panel and the NF label of a package will be available for all product judgment tasks. The screen will show the front panel with an instruction to allow interested respondents to view the NF label. Since the study focuses on cognitive response to FOP labeling, NF information will be kept constant between labeling conditions for a given food product.

The study will vary the order of Sections A and B to explore if and when respondents use the NF label in making different product decisions. We suspect that respondents are more likely to consult the NF label to rate the levels of nutrients in a product than to make a holistic judgment about two competing products. Therefore, it is possible that respondents will exhibit different NF label viewing behaviors, i.e., whether they will click the button to see the NF label, for the two different tasks (judgment of a single product in Section B and judgment of two products of the same category in Section A). If all respondents perform Section A first and Section B next, or vice versa, then the observed NF viewing behaviors would be confounded by the order of the two tasks. Therefore, we will alternate the order of the two sections to avoid the order effect.

Section A will be used to examine (1) any differences between labeling schemes and (2) any differences between presence and absence of any of the symbols in how they may help respondents identify the healthier product from a pair of products of the same category. The pairs of products included in each category are: shredded wheat and raisin bran, baked crackers and corn chips, and turkey breast dinner and Pepporni pizza. These products have been selected because they are reasonable substitutes for each other within a given category. To avoid order effect, the study will randomize between respondents the relative position (left or right) of the two products on the screen.

Section C will be used to examine whether respondents interpret products of

different categories to be similar nutrition-wise if they bear the same symbol. In this section, a product in the cereal or snack category will be compared to a yogurt and a product in the entrée category will be compared to a salad. Yogurt and salad have been selected because they are reasonable substitutes for cereal and snack and entrée, respectively.

2c. Study 1 (eye-tracking study)

Members on the contractor’s Participant Database will be invited by email to complete a screener for an opportunity to participate in the study. Eligible members will be given the opportunity to sign up for an appointment to go to the contractor’s facility for the study. In the study, the contractor will ask respondents to complete product choice and product evaluation tasks, using a sample of mock product packages that resemble those found in the market with brandnames removed (see Appendix H for eye-tracking questionnaire). The study will ask respondents to perform product judgment and choice tasks and record what specific pieces of FOP labeling information they use and which ones they pay more attention to while performing the tasks. Afterwards, the contractor will collect respondents’ background information such as demographics and food consumption and purchase experiences. We expect each eye-tracking to last about one hour. Since the study is exploratory in nature and its sample size is small, we will not include any randomization of tasks or labels.

2d. Study 2 (experimental study)

Members of the contractor’s online consumer panel will be invited by email to a Synovate Website to complete the study online in one session (Table 4). Each respondent will perform product judgments twice, each time for a different pairs of product in a food category (which will include three products). We estimate that it will take respondents about 15 minutes to complete the study.

Table 4. Structure of Study 2 - experimental study

Section	Topic
A	Which product in a three-product set would be bought; which product in the set would be bought as a healthier product for the family and reason(s) (open-ended), judgment of the chosen “healthier” product in its (1) overall healthiness, (2) relationship with weight gain and risks of chronic illnesses, (3) contribution toward the overall diet, (4) implication on intake, (5) levels of calories, fat, saturated fat, cholesterol, sodium, fiber, sugar, calcium).
B	Repeat Section A for the remaining two products that was not chosen as the “healthier” product in Section A.
C	Perceived usefulness of a labeling scheme (open-ended), other desirable information for product judgments (open-ended), and perception of the scheme.
D	Nutrition consciousness measure

Each respondent will begin the study by first choosing a product to purchase in general, then choosing a healthy product to purchase for her family, both from a set of three products in the same category, and explain the reason for the second choice (see Appendix I for Study 2 questionnaire). The pair may be associated with a hard choice or an easy choice, depending on random assignment. She will then rate the chosen product in terms of its healthiness, relationship with weight gain and risks of chronic illnesses, and other attributes, including nutrient levels. The same procedure will repeat with a pair of products that consists of the product that was not chosen in the previous task and the third product in the category. All three products a respondent sees will have the same labeling scheme. The respondent will then answer questions about when they felt the labeling scheme was helpful and for what purpose (choice or product judgment). All front panels will be full-color and identify the food (e.g., cheerios) and a mock-up brand name. Both the front panel and the NF label of a package will be available for the respondent, except when she is asked to choose between a pair of products. The screen will show the front panel with a button to allow interested respondents to view the NF label. Since the study focuses on cognitive response to FOP labeling, NF information will be kept constant between labeling conditions for a given food product. The basic analytic approach is analysis of variance implemented as a GLM to incorporate both experimental and covariate variables in either categorical or numerical forms. Individual dependent variables; choice accuracy, choice time, reason quality, product perceptions, nutrient level judgments, and nutrient labeling system ratings will be assessed individually. The pattern of effects across the individual dependent variables due to nutrition labeling schemes will be described in detail.

This data collection proposes to collect experimental data in two separate studies, mainly due to resource and time limitations. We have made sure that the approach used for core questions are harmonized between the studies. Core questions are those in Section A of the studies. These questions ask about product choices and related judgments and are the dependent measures needed to test the primary hypotheses regarding FOP effects on product choice, perceived nutrient levels, perceived dietary and health consequences of regular consumption, information truncation (i.e., likelihood of using the NF label), and speed in accomplishing a task (see Part A, Section 2). For these core questions, we have put them at identical position in the respective questionnaire, used identical question wording, and used identical response options and response scales.

Despite our effort in harmonizing the two studies in their treatment of the core questions, we are obligated to include an important caveat about the comparability of the two studies. The degree of comparability in observations about and hypothesis test results of the core questions is uncertain. The two studies use two different contractors and each of them has a different approach toward Web page design (such as color of the background, contrast between background and text,

and layout and location of the question and response option). Therefore, the two studies will look different on a computer screen and respondents will likely have different experience with each of the studies. Existing research suggests that Web design and survey experience differences can cause survey responses to differ for the same survey content. Therefore, it is important that anyone who attempts to compare the two studies understand the caveat and use caution in comparisons.

Other than the core questions mentioned above, the two studies include their own sets of questions that have been designed to obtain similar as well as different information about respondent characteristics and their reactions to a given FOP scheme. We believe all those sets of questions are important, useful, and can provide deeper understanding of symbol responses. Yet, neither of the studies can facilitate the inclusion of these questions due to concern about respondent burden. Some questions in the two studies (e.g., perception of the trustworthiness of an FOP scheme) have been included for the same or similar purpose but given slightly different wording and response options. In those cases, we have made sure that both questions are carefully worded and their responses formatted according to acceptable standards. Overall, we believe our approach toward the different questions will maximize the amount of information and utility we obtain from the two studies, given their constraints, and could also provide an opportunity to examine whether differently worded questions elicit similar responses subject to the caveat mentioned in the previous paragraphs. The latter would be beneficial for our future research of similar topics.

2.3 Degree of accuracy needed for the purpose described in the justification

As discussed in Section A2, the study plans to test hypotheses related to between-symbols, between-food-categories and between-foods differences in product choices, perceived levels of nutrients, product and labeling perceptions, likelihood of using the NF label, and time spent on experimental tasks. We will impose no a priori direction of differences, if any (i.e., we assume all tests are two-tailed).

All perceptions will be measured on a Likert scale (e.g., 1=none/a little, 5=a lot) and are assumed continuous and normally distributed. Product choice and likelihood of using the NF label will be measured as a proportion by a dichotomous variable (0-1). Time will be measured in medians because of the presence of outliers observed in our previous online experiments.

Study 1 (experimental study) - Based on the measures and assumptions mentioned above, the planned sample size is 480 observations for each of the five symbols, and 800 observations for each of the three food categories, respectively. The sizes have been selected so the study can detect a small effect size¹ of mean differences in main effects (i.e., effects of symbol and food category on response)

¹ Cohen, J. A Power Primer. *Psychological Bulletin* 112(1): 155-9, 1992.

and a difference of 10 percentage points between two proportions, with $\alpha = 0.05$ and $\beta = 0.2$. We expect that the sample sizes will also offer a similar power to detect interactive effects of a medium size and for the median test.^{2,3}

Study 2 (experimental study) – Based on the same measures and assumptions, the planned sample size is 440 observations for each of the 11 labeling conditions (9 labeling schemes and two controls), and 1,200 observations for each of the four food categories, respectively. We expect the same degree of power as that for Study 1.

2d. Use of specialized sampling procedures

No specialized sampling procedures are required.

2e. Use of periodic data collection cycles to reduce burden

This is a one-time data collection.

3. Methods to Maximize Response Rates and Deal with Non-response

Our experience with online experimental studies suggests that about 15% of those who are sent invitations will complete a study. The agency will implement several procedures to maximize participation. We will conduct cognitive interviews and pretests to help improve understandability of the questionnaire, to reduce participant burden, and to enhance interview administration (see Appendix M for cognitive interview invitation and Appendix J for pretest questions; the cognitive interview will use the instruments shown in Appendices G and I). We will keep the study questionnaire at a reasonable length to minimize breakoffs.

In addition, the contractors will (1) identify FDA as the sponsor of the study and state the purpose of the study in their invitation and reminder to encourage participation (see Appendices K and L for reminders); (2) provide an email address and a toll-free number for prospective participants to inquire about the authenticity of the interview and other questions; and (3) monitor all interviews and sample assignment and solve any problems daily throughout the course of the collection of information.

4. Tests of Procedures or Methods to be Undertaken

The agency will have SSI and Synovate pretest each of the two experimental instruments with 200 individuals after OMB approval of the collection of

² Montgomery, A.A., T.J. Peters, and P. Little. Design, analysis and presentation of factorial randomized controlled trials. *BMC Medical Research Methodology* 3: p26-5, 2003.

³ Singer, B., Lovie, A., and Lovie, P. Sample size and power. Chapter 7 in A. Lovie Ed. *New Developments in Statistics for Psychology and the Social Sciences*. New York: The British Psychological Society, 129-142, 1986.

information. The pretests will serve to address any unforeseen problems in administration of the interview.

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

Study 1: Survey Sampling International (SSI) and Eyetracking, Inc., will collect the data on behalf of FDA. Phil Guibileo is the project manager for SSI, telephone (203) 567-7328. Cassie Davis is the project manager for Eyetracking, Inc., telephone (619) 265-1840. Analysis and dissemination of the data will be led by Chung-Tung Jordan Lin, PhD, CFSAN, telephone (301) 436-1831.

Study 2: Synovate, Inc. will collect the data on behalf of FDA. Valerie Fuller is the project manager for Synovate, telephone (703) 663-7243. Analysis and dissemination of the data will be led by Alan S. Levy, PhD, CFSAN, telephone (301) 436-1762.