

Superimposed Text in Direct-to-Consumer Promotion of Prescription Drugs
OMB Control No. 0910-NEW

SUPPORTING STATEMENT Part B: Statistical Methods (used for collection of information employing statistical methods)

1. Respondent Universe and Sampling Methods

The eligible population for this study will be U.S., noninstitutionalized adults ages 18 or older who speak English and have not participated in a focus group or interview in the past 3 months. To provide geographic diversity across the study sample, participants will be recruited from three U.S. locations using two recruitment firms: Schlesinger Associates in Los Angeles, CA, and L&E Research in Cincinnati, OH, and Tampa, FL. Schlesinger Associates maintains a database consisting of over 200,000 residents in the Los Angeles, CA area. L&E Research maintains a database of approximately 89,800 residents in the Cincinnati, OH, area and 22,000 in the Tampa, FL, area. In addition to our aim for regional variation, we selected these three cities with the aim of recruiting a sample that is diverse in gender, race/ethnicity, education, and age characteristics. It will not be nationally representative of the entire population. We will recruit a total of 1,512 individuals for the pretest and main study combined. **Table 3** shows the overall pretest and main study sample design and sample sizes.

Table 3. Sample Design and Sample Sizes

Category	Number of Participants
Pretest	240
Main study	1,272

2. Procedures for the Collection of Information

Part A of the supporting statement described the rationale for conducting the study.

General Research Questions

1. Does the size of the superimposed text, the contrast behind the superimposed text, and/or the device type influence the noticeability, recall, and perceived importance of the super information?
2. Does the size of the superimposed text, the contrast behind the superimposed text, and/or the device type influence the recall of and attitudes toward the promoted drug?
3. Are there any interaction effects among any combination of independent variables?

Design

To test these research questions, we will conduct one randomized controlled study. We will examine reactions to supers in a fictitious DTC prescription drug promotional video on two

types of viewing devices with a general population sample. The study design will be a $3 \times 2 \times 2$ factorial design, where participants are randomly assigned to one of 12 experimental study arms differentiated by:

- Super text size (small, medium, large);
- Device type (television, tablet);
- Super text contrast (high, low).

Table 1. Design and cell sizes for main study.

Device Type:		TV			Tablet			Total
Super Size:		Small	Medium	Large	Small	Medium	Large	
Contrast:	High	106	106	106	106	106	106	636
	Low	106	106	106	106	106	106	636
Total		212	212	212	212	212	212	1,272

Note: The sample will be split across three cities (Los Angeles, CA; Cincinnati, OH; and Tampa, FL).

We will also conduct a pretest to (1) test consumer perceptions of superimposed-text size with the aim of choosing perceptibly different levels of size (small, medium, large) for use in the main study; and (2) testing our planned procedures for implementation of the intervention (TV and tablet) and in-person data collection. The pretest will have a three-part design, which each participant will complete in a single in-person session. Part 1 of the pretest will have a 5 (super size) $\times 2$ (device) between-subjects design, yielding ten experimental arms (see Table 2). Part 2 of the pretest will have a within-subjects design, where participants will view a sequence of still images from each of the 5 ads (small, medium-small, etc.) and answer 3 repeated-measures questions related to perceptions of super size per image. For Part 3, participants will view 5 more still images and rank them in order of perceived super size, from smallest to largest (ties will be allowed). The images will not differ by content type in the ranking task, only by size of the superimposed text.

Table 2. Pretest Part 1 between-subjects design and sample size

Device Type	Super Size					Total
	Small	Medium-Small	Medium	Medium-Large	Large	
TV	24	24	24	24	24	120
Tablet	24	24	24	24	24	120
Total	48	48	48	48	48	240

For both the pretest and main study, we will work with two market research firms to recruit adult participants and conduct in-person data collection in three U.S. cities: Los Angeles, CA, Cincinnati, OH, and Tampa, FL. In addition to our aim for regional variation, we selected these three cities with the aim of recruiting a sample that is diverse in gender, race/ethnicity, education, and age characteristics.

Participants from the general population will be invited to a market research facility to watch one video for a fictional prescription drug that treats asthma. In-person administration of study procedures will enable us to control the television and tablet watching experience in terms of size, distance, and other variables. Participants will watch the video once and then answer questions addressing recall of risks and benefits, perceptions of risks and benefits, and questions regarding the salience of information in text. The questionnaire is available upon request. Participation is estimated to take approximately 25 minutes.

To examine differences between experimental conditions, we will conduct inferential statistical tests such as analysis of variance (ANOVA). Pretesting will take place before the main study to select super sizes for the main study and to evaluate the procedures and measures that will be used. We will exclude individuals who work in healthcare or marketing settings because their knowledge and experiences may not reflect those of the average consumer.

Specific Hypotheses

Perceived Visual Salience of Supers

Hypothesis 1.a The perceived visual salience of supers will increase as super size increases.

Hypothesis 1.b The perceived visual salience of supers will be greater for participants exposed to high contrast supers than those exposed to low contrast supers.

Memory of Supers

Hypothesis 2.a Recall and recognition of information in supers will increase as super size increases.

Hypothesis 2.b Recall and recognition of information in supers will be greater for participants exposed to high contrast supers than those exposed to low contrast supers.

Perceived Importance of Information in Supers

Hypothesis 3.a The perceived importance of information in supers will increase as super size increases.

Hypothesis 3.b The perceived importance of information in supers will be greater for participants exposed to high contrast supers than those exposed to low contrast supers.

In addition to these proposed analyses, we will test the effects of device type on study outcomes; analyses by device type are exploratory.

Analysis Plan

Descriptive Analysis

During descriptive analysis, we will calculate frequency distributions and check the apparent validity of the data (i.e., range checks, frequency of missing responses, or response distribution). For continuous/ordinal variables, statistical output will include means, medians, standard deviations, ranges, and counts. For categorical variables, output will include counts and percentages.

In addition to frequency distributions, we will conduct three other types of analyses during this step. First, we will calculate reliability of composite variables and multi-item scales to determine if the individual items hang together as composite measures. Specifically, we will calculate Cronbach's alpha for each composite variable. If alpha for a composite measure or scale does not meet our pre-established threshold of 0.75, we will discuss whether to use single-item measures rather than the composite or to consider such composites as indices (because of a theoretical reason to consider an aggregate measure regardless of item correspondence) in hypothesis testing.

Second, we will conduct a content analysis of responses to the open-ended risk, benefit, and other information recall questions (main study only). We will develop a codebook to guide classification of responses based on their match with risk and benefit claims made in the ad. To ensure consistent and reliable coding of open-ended data, we will develop and implement an inter-rater reliability protocol before proceeding to code the full content.

Finally, we will conduct a non-response analysis to determine if individuals who do not respond to the study's invitation differ from those who complete the study. We will compare responding individuals to invited but non-responsive individuals on key demographics—such as age, sex, race, and education—to see if significant differences exist. Specifically, we will conduct t tests comparing the proportions of respondents and non-respondents using a standard significance threshold of $p=0.05$.

Hypothesis Testing

We will test hypothesized relationships implied by our central research questions by conducting one of several statistical tests as outlined below. In most cases, we plan to conduct an overall test of the relationship between the independent and dependent variables and then conduct hypothesis-specific planned comparisons to assess whether the data support predicted differences among experimental groups. If a significant interaction is observed, we will conduct follow-up analyses to describe the interaction. Foremost, we will test for the predicted pattern of means across manipulated experimental conditions (i.e., super size in the pretest; super size, contrast, and device type in the main study). When applicable, we will develop planned contrast equations for this purpose corresponding to each of our research hypotheses. Alternatively, we will use post-hoc pairwise comparisons to test for differences between experimental groups.

For hypotheses examining continuous or scale outcomes (e.g., perceived super size), we will conduct ANOVAs to detect significant relationships. For dichotomous outcomes, we will use logistic regressions, followed-up by chi-squared tests for equal proportions when more specific pairwise differences between groups are of interest. In the pretest, we will explicitly test for simple effects by super size (small, medium-small, medium, medium-large, large) and device type (TV, tablet). In main study, we will test for effects by super size (small, medium, large), device type (TV, tablet), contrast (low, high, and their interactions. Statistical output will include, as appropriate, F or chi-square statistics, test degrees of freedom, p values, mean or proportional differences, and standardized effect sizes (e.g., Cohen's d) for the main effects of each independent variable as well as any interaction effects. We will conduct planned comparisons based on hypothesized relationships or post-hoc comparisons to identify significant differences between specific experimental groups.

Power

The following assumptions were made in deriving the sample size for the main studies: (1) 0.90 power, (2) 0.05 alpha level for main effects and interactions, and (3) a small effect size. We use Cohen's¹ conventional thresholds for interpreting the magnitude of effect sizes. Effects with f values in the order of 0.40 and greater are large, from 0.25 to but not including 0.40 are medium, and from 0.10 to but not including 0.25 are small. Corresponding effects measured with d statistics equal to or greater than 0.80 are large, from 0.50 to but not including 0.80 are medium, and from 0.20 to but not including 0.50 are small. Given 12 experimental groups in a 3x2x2 factorial design, our proposed main-study sample size is 1,272 (106 participants per experimental arm). Our F tests for main effects and interactions will be able to detect small effects ($f \geq 0.10$). Using the same power and alpha levels, this corresponds to between-group tests with enough sensitivity to detect small differences in group means when following-up significant main effects of a single factor ($d \geq 0.22$) and 3-way interactions ($d \geq 0.44$). With our sample size, we will also be able to detect pair-wise differences in proportions between experimental groups on dichotomous outcome variables as low as 22 percentage points, and differences between levels of a single factor as low as 9 percentage points.

We will conduct the pretest with a smaller sample size than the main study ($N = 240$). The pretest will also have a different, three-part design. Each part has different statistical assumptions, so we conducted separate power analyses for them:

- **Part 1 (between-subjects design, analyzed 5-level one-way ANOVA).** Assuming a conventional alpha level of 0.05 and power set to 0.80, the established pretest sample size of 240 participants would be sufficient to detect a medium or larger main effect of super size (i.e., $f \geq 0.23$). Pairwise comparisons between any two of the five size groups ($n = 48$ per group) using independent-samples t tests without Bonferroni adjustment would be sensitive to detect medium or larger effects ($d \geq 0.58$).
- **Part 2 (within-subjects).** Given an alpha level of 0.05, power set to 0.80, and a total sample size of 240,² the within-subjects experiment proposed for Part 2 of the pretest

¹ Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Erlbaum.

² For purposes of estimating sensitivity, we also assumed moderate correlations among repeated measures (i.e., $r=0.5$) and a nonsphericity correction of 1.

would be sensitive to detect small effects ($f \geq 0.07$). Dependent-samples t tests without Bonferroni adjustment would be sensitive enough to detect small pairwise differences between text-size ratings ($d \geq 0.18$); applying a Bonferroni correction for 10 comparisons (alpha=0.005) would reduce the sensitivity of follow-up tests, but would still be able to detect moderately-small or larger effects ($d \geq 0.24$).

- **Part 3.** Analysis of the ranking data will be descriptive only (e.g., tabulating the proportion of participants who correctly ranked all five screenshots). Because we do not intend to use the data from Part 3 for statistical inference, we did not conduct a separate power analysis for this section.

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

Both the pretests and main study will be administered in-person. To provide geographic diversity across the study sample, participants will be recruited from three U.S. locations using two recruitment firms. The recruitment firms maintain databases comprising individuals in the three cities: Los Angeles, CA, Cincinnati, OH, and Tampa, FL. To help ensure that the participation rate is as high as possible, FDA and the contractor will:

- Design a protocol that minimizes burden (short in length, clearly written, and with appealing graphics);
- Screen people for eligibility prior to inviting them to participate and offer eligible participants several dates and times to schedule an appointment to complete the study;
- Use incentive rates that meet industry standards for each study location. In addition to offsetting respondent burden, using market-rate incentives tends to increase response rates, reduce sampling bias, and reduce nonresponse bias.

Participants in the pretest and main studies will be convenience samples, rather than probability-based samples of U.S. adults. The sample will not be representative of the U.S. population to accommodate in-person administration of data collection protocols, but we will aim to recruit a mix of participants in terms of race/ethnicity, age, education, gender and other characteristics. Rather, the strength of the experimental design used in this study lies in its internal validity, on which meaningful estimates of differences across manipulated conditions can be produced and generalized. This is a counterpoint to observational survey methodologies where estimating population parameters is the primary focus of statistical analysis. The recruitment procedures in this study are not intended to fit the criteria for survey sampling, where each unit in the sampling frame has an equal probability of being selected to participate. In an observational survey study, response rates are often used as a proxy measure for survey quality, with lower response rates indicating poorer quality. Nonresponse bias analysis is also commonly used to determine the potential for nonresponse sampling error in survey estimates. However, concerns about sampling error do not generally apply to experimental designs, where the parameters of interest are under the control of the researcher—rather than being pre-established characteristics of the participants—and each participant has an equal probability of being assigned to any of the experimental conditions.

Generally, there are several approaches to conducting a nonresponse bias analysis, such as comparing response rates by subgroups, comparing respondents and nonrespondents on frame variables, and conducting a nonresponse follow-up study³. For the proposed project, we will examine nonresponse for its descriptive value by performing two steps: comparing response rates on subgroups and comparing responders and nonresponders on frame variables.

To the extent that information from the recruitment databases is available about potential participants who are invited to participate, we will first identify subgroups of interest, such as age and gender. At the end of the data collection, we will calculate response rates by subgroup. At the end of data collection, we will review frame data to determine if any variables are associated with the key survey estimates, such as age. We will then compare the frame information for the full sample compared with respondents only. Differences between the full sample and the respondents are an indicator of potential bias.

4. Test of Procedures or Methods to be Undertaken

Nine cognitive interviews will have been conducted to assess questionnaire flow and wording. After this round of cognitive testing, we plan to conduct pretests on a larger scale to ensure the main study will run smoothly. We propose to test 240 individuals in the pretest.

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

The contractor, RTI, will collect and analyze the data on behalf of FDA as a task order under Contract HHSF223201510002B. Jessica DeFrank, Ph.D., 919-485-2661, is the Project Director for this project. Data analysis will be overseen by the Research Team, Office of Prescription Drug Promotion (OPDP), Office of Medical Policy, CDER, FDA, and coordinated by Amie C. O'Donoghue, Ph.D., 301-796-0574, and Kevin R. Betts, Ph.D., 240-402-5090.

³ Office of Management and Budget, *Standards and Guidelines for Statistical Surveys*, September, 2006. www.whitehouse.gov/sites/default/files/omb/inforeg/statpc. Last accessed April 18, 2013.