

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

For the study, 11,750 (750 for the pretests, 11,000 for the main study) participants will be recruited from a panel of one million households. Each panel member will complete a prescreening questionnaire (screener), and we will recruit participants who indicated that they have heartburn or acid reflux disease, regardless of whether or not they take prescription medicine for the condition (see Appendix E for the screener and Appendix F for the recruitment and reminder emails). The target population is the adult noninstitutionalized population in the US with access to the Internet and who self-report recent experience with heartburn or acid reflux disease. The contractor will use quota sampling with the goal of yielding 500 respondents in each of 22 test conditions for a total of 11,000 completed interviews. Initially, survey invitations will be sent to panel members in proportion to the 2007 National Health Inventory Survey's (NHIS) distribution of persons with heartburn. Upon entry into the survey, respondents will be screened for heartburn (see screener). Only those panelists who have experienced heartburn during the last three months will continue with the survey. The contractor will perform a slow release of sample by sending email invitations containing the survey link to 10% of the total sample on the initial day of the field period. Additional invitations will be released in waves using varying proportions to account for different qualification and yield rates among the demographic groups of panel members. The goal will be to have a final sample with similar demographic proportions of heartburn sufferers as seen in the 2007 NHIS data. The 2007 NHIS allows us to calculate the demographic proportions of heartburn sufferers for gender, age, race, and Hispanic origin. After

qualifying for the survey, each respondent will be randomly assigned to an experimental condition. Assignment to condition only after qualifying for the survey ensures equal and unbiased allocation of the respondents to experimental condition. Because the sample is not nationally representative, we cannot estimate population parameters (that is, we will not make statements such as “X% of heartburn sufferers in the US think X”). The interpretation of results will include a discussion of the generalizability of the findings given this sampling method.

2. Procedures for the Collection of Information

Design Overview

This study will be conducted in two concurrent parts; one examining variations on the benefit information presented in DTC print advertisements and the other examining variations on the risk information presented in DTC print advertisements. The factors studied will be the type of information (i.e., the addition of quantitative and qualitative information in a box format) and the level of efficacy or risk. We will vary the level of efficacy and risk such that the largest effect is noticeably different from the placebo, whereas the smallest effect is minimally different from the placebo. We plan to use pretests to determine the number of levels and the content of the levels (e.g., the differences used) to be included in the main study. We will also pretest whether participants should have access to the ad while completing the questionnaire (see Appendix D for pretest questionnaires). The following design includes the maximum number of levels we would include. These factors will be combined in a factorial design as follows:

Table 5. Benefit Design (4 x 5 + 2):

Information Type	Efficacy Level				
	Smallest Effect	Smaller Effect	Mid-size Effect	Larger Effect	Largest Effect
(1) Absolute Frequency	19% vs. 18%	39% vs. 18%	59% vs. 18%	79% vs. 18%	99% vs. 18%
(2) Absolute Frequency + Qualitative Label	More 19% vs. 18%	More 39% vs. 18%	More 59% vs. 18%	More 79% vs. 18%	More 99% vs. 18%
(3) Absolute Difference + Qualitative Label	More (1 percentage point)	More (21 percentage points)	More (41 percentage points)	More (61 percentage points)	More (81 percentage points)
(4) Absolute Frequency + Absolute Difference + Qualitative Label	More (1 percentage point) 19% vs. 18%	More (21 percentage points) 39% vs. 18%	More (41 percentage points) 59% vs. 18%	More (61 percentage points) 79% vs. 18%	More (81 percentage points) 99% vs. 18%

Note. Qualitative label example: “more people taking drug X had heartburn relief.”

(5) No information	(6) Qualitative Label only (More)
--------------------	-----------------------------------

Table 6. Risk Design (4 x 5 + 2):

Information Type	Risk Level				
	Smallest Effect	Smaller Effect	Mid-size Effect	Larger Effect	Largest Effect
(1) Absolute Frequency	3% vs. 2%	23% vs. 2%	43% vs. 2%	63% vs. 2%	83% vs. 2%
(2) Absolute Frequency + Qualitative Label	More 3% vs. 2%	More 23% vs. 2%	More 43% vs. 2%	More 63% vs. 2%	More 83% vs. 2%
(3) Absolute Difference + Qualitative Label	More (1 percentage point)	More (21 percentage points)	More (41 percentage points)	More (61 percentage points)	More (81 percentage points)
(4) Absolute Frequency + Absolute Difference + Qualitative Label	More (1 percentage point) 3% vs. 2%	More (21 percentage points) 23% vs. 2%	More (41 percentage points) 43% vs. 2%	More (61 percentage points) 63% vs. 2%	More (81 percentage points) 83% vs. 2%

Note. Qualitative label example: “more people taking drug X had side effect Y.”

(5) No information	(6) Qualitative Label only (More)
--------------------	-----------------------------------

In the benefit design, we will use the mid-size effect for the risk information in all conditions and vary the information type to match the benefit information type (e.g., participants who see absolute frequency benefit information will also see absolute frequency risk information). Similarly, in the risk design, we will use the mid-size effect for the benefit information in all conditions and vary the information type to match the risk information type.

The test product will be for the treatment of gastroesophageal reflux disease (GERD) and modeled on an actual drug used to treat this condition. Participants will be consumers who have heartburn or acid reflux disease. They will be randomly assigned to read one ad version. After reading the ad, participants will answer a series of questions about the drug. We will test how the information type affects perceived efficacy, perceived risk, behavioral intention, and accurate understanding of the benefit and risk information. The questionnaires for the risk and benefit designs will have identical questions; however, the order will differ. In the risk design, questions about risk will appear before questions about benefits; in the benefit design questions about benefits will appear before questions about risks.

Procedure

All parts of this study will be administered over the internet. A total of 11,750 interviews will be completed. Participants will be randomly assigned to view one version of a DTC prescription drug print ad which consists of a display page and the accompanying brief summary page. Following their perusal of this document, they will answer questions about their recall and understanding of the benefit and risk information,

their perceptions of the benefits and risks of the drug, and their intent to ask a doctor about the medication.

Demographic and numeracy information will be collected. In addition, participants will answer questions about their familiarity with their medical condition. The entire procedure is expected to last approximately 20 minutes. This will be a one-time (rather than annual) information collection.

Participants

Data will be collected using an Internet protocol. Approximately 11,750 consumers who have heartburn or acid reflux disease will be recruited for the study. Because the task presumes basic reading abilities, all selected participants must speak and read English fluently. Participants must be 18 years or older.

Hypotheses

Information Type Hypotheses:

1. There will be greater variance in accuracy in the no information condition than in the absolute frequency, absolute frequency + qualitative label, absolute difference + qualitative label, and absolute frequency + absolute difference + qualitative label conditions (this tests whether adding quantitative information increases accuracy).
2. There will be greater variance in accuracy in the qualitative label only condition than in the absolute frequency + qualitative label, absolute difference + qualitative label, and absolute frequency + absolute difference + qualitative label conditions (this tests whether adding quantitative information to a qualitative label increases accuracy).

3. Because the presence of quantitative and/or qualitative information may affect perceptions, we will test whether perceived efficacy and perceived risk differ between (1) the no information condition and all other information type conditions and (2) the qualitative label only condition and all other conditions.
4. We will test whether perceived efficacy and perceived risk differ between the absolute frequency condition and the absolute frequency + qualitative label and absolute frequency + absolute difference + qualitative label conditions (this tests whether adding qualitative information to absolute frequency changes perceptions).

Efficacy/Risk Level Hypotheses:

5. Perceived efficacy will increase as efficacy level increases. Perceived risk will increase as risk level increases.
6. Perceived risk will decrease as efficacy level increases. Perceived efficacy will decrease as risk level increases.
7. Behavioral intention will increase as efficacy level increases. Behavioral intention will decrease as risk level increases.
8. Accuracy will not differ across efficacy/risk levels.
9. Accuracy will be better in all efficacy/risk levels when compared with either control group (i.e., no information or qualitative label only).
10. In the absence of quantitative information, people may assume a high level of efficacy; therefore, we predict that perceived efficacy and behavioral intention will be higher in the two control groups (i.e., no information and qualitative label only) compared to all efficacy level conditions. In the absence of risk

information, people may assume a low level of risk; therefore, we predict that perceived risk will be lower and behavioral intention will be higher in the two control groups (i.e., no information and qualitative label only) compared to all risk level conditions.

11. Risk comprehension and benefit comprehension will not differ across efficacy/risk level conditions.

Efficacy/Risk Level * Information Type Hypotheses:

12. The qualitative label will help people process the numerical information; therefore, the efficacy/risk level effects predicted above for perceived efficacy, perceived risk, behavioral intention, and accuracy will be more pronounced in the absolute frequency + qualitative label condition than in the absolute frequency condition.
13. Similarly, the numeric information will help people understand the qualitative label; therefore, the effects above for perceived efficacy, perceived risk, behavioral intention, and accuracy will be more pronounced in the absolute frequency + qualitative label condition, the absolute difference + qualitative label condition, and the absolute frequency + absolute difference + qualitative label condition, compared with the qualitative label only condition.

Numeracy Hypotheses:

14. High numeracy participants will have better accuracy than low numeracy participants in all efficacy/risk level conditions and in all information type conditions except the no information and qualitative label conditions.

15. Numeracy may moderate the effects predicted above. For instance, the effects predicted in hypotheses 5-9 may be more prominent in high numeracy, compared to low numeracy participants (because high numeracy participants may understand the quantitative information better than low numeracy participants). On the other hand, the effects predicted in hypothesis 4 may be more prominent in low numeracy, compared to high numeracy participants (because low numeracy participants may be more affected by the addition of qualitative labels to quantitative information).
16. Because the addition of quantitative information may be overwhelming for low numeracy participants, we will test whether risk comprehension and benefit comprehension differ between the information type conditions and the control conditions, by numeracy.

All other comparisons are exploratory.

Analysis Plan

The following analysis plan pertains to both the benefit design and the risk design. For hypotheses regarding efficacy or risk level, we will conduct tests within each efficacy (risk) level as well as across efficacy (risk) levels (main effects). We will use Levene's Test of homogeneity of variances to test hypotheses 1 and 2. For all other hypotheses, we will conduct ANOVAs or linear regressions with continuous dependent variables (i.e., perceived efficacy, perceived risk, behavioral intentions, benefit comprehension, risk comprehension) and chi-square tests and logistic regressions with categorical dependent variables (i.e., accuracy). For instance, we will examine linear

regression models (predicting each dependent variable from efficacy or risk level) to test hypotheses 5-7 and 11. We will conduct these analyses both with and without covariates (e.g., demographic and health characteristics) included in the model. In addition, we will test whether effects are moderated by numeracy (see hypotheses 14-16). If a main effect is significant, we will conduct pairwise-comparisons to determine which conditions are significantly different from one another. We will also conduct planned comparisons in line with our hypotheses (see above).

Power

The following assumptions were made in deriving the sample size for the study: 1) 0.90 power, 2) 0.05 alpha or 0.0125 alpha (Bonferroni-adjusted for four comparisons) and 3) an effect size between small and medium. The table below shows the sample size required to detect differences with effect sizes ranging from conventionally “small” ($f = 0.10$) to “medium” ($f = 0.25$) for the comparison between the no information group and each of the information type conditions.

Table 7. Power Analysis Calculation.

A priori power analysis to determine sample size needed in F tests (ANOVA: fixed effects, main effects, and interactions) to achieve power of 0.90 (Faul et al., 2007). ¹							
		Effect size f^*			Effect size f^*		
Input		0.10	0.15	0.25	0.10	0.15	0.25
	α error probability	0.05	0.05	0.05	.0125	.0125	.0125
	Power ($1 - \beta$ error probability)	0.90	0.90	0.90	0.90	0.90	0.90
	Numerator df	1	1	1	1	1	1
	Number of groups	2	2	2	2	2	2
Output							
	Critical F	3.85	3.86	3.89	6.25	6.27	6.34

¹ Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

	Denominator df	1,050	466	168	1,429	635	229
	Sample size per cell	527	235	86	716	319	116

*An effect size of 0.10 is traditionally considered small, whereas an effect size of 0.25 is considered medium (Cohen, 1988).² Here we have shown three different effect sizes centering around small to medium effects.

We will have 250 participants per cell, with a total of 11,000 participants in the 44 cells represented in the tables (two 4 x 5 + 2 designs). With this sample size, we will be able to detect small to medium effects with an unadjusted *p*-value of .05 and medium effects with a Bonferroni-adjusted *p*-value of .0125.

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

Response rates can vary greatly depending on many factors including the sample composition, panel type, invitation content, time of day and incentive offering. In addition, outside factors including email filters, recipient Internet service provider (ISP) downtime and general conditions on the Internet can impact response rates. We will calculate response rate as ratio of the number of surveys completed to the number of panelists contacted by invitation. To help ensure that the participation rate is as high as possible, FDA and the contractor will:

- Design an experimental protocol that minimizes burden (short in length, clearly written, and with appealing graphics);
- Administer the experiment over the Internet, allowing respondents to answer questions at a time and location of their choosing;
- Sending out two email reminders after the initial invitation.
- Provide respondents with a helpdesk link that they can access at any time for assistance.

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

Additionally, the Panel leverages the social media concept and has developed ‘panel communities’ in order to maximize member engagement and overcome challenge of declining survey response rates and multi-panel membership.

4. Test Procedures

The contractor will ask nine participants to go through the procedure to assess blatant glitches in questionnaire wording, programming, and execution of the study. We will also conduct pretests with 750 consumers before running the main studies to ensure that stimuli and questionnaire wording is clear. Finally, we will run the main studies as described elsewhere in this document.

5. Individuals Involved in Statistical Consultation and Information Collection

The contractor, Synovate, will collect the information on behalf of FDA as a task order under Contract HHSF223200510007I. Valerie DiPaula, Ph.D., is the Project Director for this project, 703-663-7243. Data analysis will be conducted primarily by the Research Team, Division of Drug Marketing, Advertising, and Communications (DDMAC), Office of Medical Policy, CDER, FDA, and coordinated by Helen W. Sullivan, Ph.D., M.P.H., 301-796-4188, and Amie C. O’Donoghue, Ph.D., 301-796-0574.