B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

In Study 1, invitations will be sent to a random selection of 2,770 participants from a prescreened sample of 40,823 general population participants for a total of 1,800 (200 for the pretest; 1,600 for the main study) completed surveys. In Study 2, invitations will be sent to a random selection of 2,677 participants who have experienced seasonal allergies within the past 12 months out of a prescreened sample of 5,984 individuals with seasonal allergies for a total of 2,202 (462 for the pretest; 1,740 for the main study) completed surveys. Study invitations will be sent to individuals within the existing GfK panel (see Appendix C for the study invitation and reminder emails).

GfK will take the following steps:

- 1. Identify individuals at the rate of no more than one per household who are appropriate for each study, including those with seasonal allergies for the Study 2;
- 2. Randomly assign the panel into replicates and then release as many replicates as they think will be necessary under the most optimistic scenario;
- 3. After a short time in the field (somewhere between a few days and a week), re-evaluate the cooperation rate and then release additional replicates as needed to achieve the required number of completed interviews.

Weighting procedures

Recruitment will begin with an equal probability sample. The data will be weighted to adjust for known unequal selection probabilities, for unequal response rates, and for any remaining deviations between the sample and population distributions. In the final step, we will use poststratification to calibrate the sample distribution to known population distribution to reduce bias due to frame undercoverage. We believe that poststratification should reduce undercoverage bias to some extent for the same reasons that weighting adjustment reduces nonresponse bias. Population counts for use in poststratification will be based on demographic distributions from the most recent data from the Current Population Survey (CPS) and benchmark distributions for Internet access among the U.S. population of adults obtained from the most recent special CPS supplemental survey measuring Internet access (October 2010). Available variables on which to weight include gender, age, race, ethnicity, education, census region (Northeast, Midwest, South, West), household income, home ownership status, metropolitan area (Yes, No), and Internet access (Yes, No).

2. Procedures for the Collection of Information

Design Overview

Study 1. In this study, individuals in a general population sample of 1,600 adults of varying education levels will answer an internet survey designed to explore whether consumers recognize composite scores in DTC ads and their understanding of composite scores. The survey will be conducted with a probability-based consumer panel of U.S. adults.

As part of the survey, participants will view a print ad that contains claims based on composite scores and respond to questions about the ad to assess whether they recognized that composite scores were used. Other outcomes will include ad comprehension, perceived efficacy, and perceived risk as they relate to their understanding of composite scores. We will also examine whether and in what ways participants' perceived efficacy and perceived risk change after they are given a definition and examples of composite scores. Questions will also explore consumers' understanding of how the effectiveness of drugs is measured in general.

This exploratory survey will not be used to test specific hypotheses about the outcome measures. However, we will explore the differences in responses to the ad before and after information about composite scores is provided. We will also examine differences in the comprehension of the composite score concept and in the features of the ad by education level and age because literature suggests that less-educated and older consumers may not understand this type of information as well.¹

Study 2. Unlike Study 1, Study 2 will be a randomized, controlled study. Study 2 will examine 1) different ways to present the information that arises from a composite score and 2) different ways to explain the concept of a composite score (an educational intervention). Outcome measures will include consumers' awareness and comprehension of the composite score concept, perceived drug efficacy, and risk recall. Participants will be randomly assigned to experimental arms in a 3 x 2 design as shown in Table 1.

Educational Intervention	General Indication	List of Symptoms	Composite Definition	TOTAL
Absent	Arm 1	Arm 2	Arm 3	870
	(n=290)	(n=290)	(n=290)	
Present	Arm 4	Arm 5	Arm 6	870
	(n=290)	(n=290)	(n=290)	
TOTAL	580	580	580	1,740

Table 1. Study design for study 2.

¹ Fagerlin, A., & Peters, E. (2011). Quantitative information. In B. Fishoff, N. T. Brewer, & J. S. Downs (Eds.), *Communicating risks and benefits: An evidence-based user guide*. Food and Drug Administration, U.S. Department of Health and Human Services. Retrieved from http://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/ucm268078.htm

This study will manipulate two variables: three types of information presentations and the presence or absence of an educational intervention. In terms of information presentation, there are many aspects of composite scores that could be communicated and one research project cannot test them all. In this study, we have chosen to examine three different information presentations that may or may not help consumers understand the composite score concept. These different information presentations were chosen based on a review of the literature and a review of past DTC submissions.

The three different information presentations are described below.

General Indication. The first information presentation is the indication of the product. In this condition, participants will see the drug indication, but will not see any explicit statement that the drug's benefits are based on a composite score. This is a common way that composite scores are currently communicated. An example of this presentation is: "Drug A treats and helps prevent seasonal nasal allergy symptoms."

List of Symptoms. The next information presentation will include the drug indication and all of the symptoms that are used to make up the composite score. This condition, like the general indication condition, will not include an explicit statement referencing composite scores. This is also a common way that composite scores are currently communicated. An example of this presentation is: "Drug A treats and helps prevent seasonal nasal allergy symptoms: congestion, runny nose, nasal stuffiness, nasal itching, and sneezing."

Composite Definition. The final information presentation will present the indication, describe that the drug's benefits are based on a composite score, and explicitly define a composite score. To our knowledge, this would be a new way to communicate composite scores. An example of this presentation is: "Drug A treats and helps prevent seasonal nasal allergy symptoms.

Drug A's effectiveness is based on a composite score. A composite score is a single measure of how well a drug works based on a combination of symptoms. Drug A may not be as effective in addressing each factor individually."

We will also manipulate whether or not participants see a specific educational intervention. This intervention was developed from prior focus groups (OMB Control No. 0910-0677) where it was found to resonate with participants. In these focus groups, medical examples were confusing, so non-medical examples were explored. This example will feature the decathlon as an educational example of a composite score. For example, "Drug A's effectiveness is based on a composite score. A composite score is like a decathlon. In that event, athletes compete in 10 events, such as the long jump, the shot put, and the 50 yard dash. An athlete may not win all events, but if he or performs well enough in some events, he or she may be the winner based on a combination of scores for each event."

We will test whether the educational intervention, the information presentation, and the interaction of the two affect outcomes such as consumers' awareness and comprehension of the composite score concept; perceived drug efficacy; and risk recall. We will test whether numeracy and literacy moderate any significant relations.

The sample for the second study will include approximately 1,740 participants who have been diagnosed with seasonal allergies. The protocol will take place via the internet. Participants will be randomly assigned to view one print ad for a fictitious prescription drug that treats seasonal allergies and will answer questions about it. The entire process is expected to take no longer than 20 minutes. This will be a one time (rather than annual) collection of information.

Hypotheses

Study 1

1. Participants who have lower levels of education will exhibit poorer comprehension of composite endpoint scores than participants who have higher levels of education.

2. Older participants will exhibit poorer comprehension of composite endpoint scores than younger participants.

Although we do not have directional hypotheses, we also believe that presenting the composite score information will change/affect the following variables: clarity of the advertisement, perceived efficacy, trust in ad information, source credibility, and attitudes toward the drug. We do not expect presenting the composite score information to affect perceived risk.

Study 2

Study 2 manipulates two factors (i.e., type of information and presence/absence of an educational intervention). We can test the levels of each factor (e.g., list of symptoms versus composite score definition) and we can test the interaction between the factors. The primary study hypotheses are found below.

- Participants who see the educational intervention will show greater awareness and comprehension of the concept of composite scores than participants who do not see the intervention. The educational intervention's effect on perceived efficacy and risk are exploratory.
- 2. Within the education-absent conditions, participants who see a composite definition (3rd level of type of information) will show greater awareness and comprehension of the concept of composite scores as compared with the general indication and the list of symptoms conditions. No difference is expected between the general indication and list of symptoms conditions on these outcomes. The role of type of information on perceived efficacy and risk is exploratory.

- 3. Educational intervention and type of information will interact such that participants in the composite definition/education-present condition will show the greatest awareness and comprehension of the composite score concept more often than participants in any other group. We will explore whether there are interaction effects for perceived benefit and risk.
- 4. Effects on benefit and risk recall are exploratory.

Analysis Plan

Prior to main analyses in both studies, we will perform an outlier analysis by flagging inactive (i.e. spend more than 10 minutes on a question) and multi-session (i.e., close survey browser) participants and examining whether their responses on key outcome variables are significantly different than other participants. We will also examine whether there are differences in responses on key outcome variables for participants who pass and who fail the attention filter item and assess the extent of any missing information to determine the data quality. Descriptive statistics will afford a look at the frequency of certain key responses, such as the number of respondents who can accurately identify a composite score. Regression will allow us to assess the effect of various independent variables on key outcomes when we control for several variables simultaneously. We can use ordinary least squares regression for continuous variables and logistic regression where the outcome variables are categorical in nature. Regression also offers a framework within which we can assess potential moderating factors, such as education or age, on the relationship between consumer understanding of composite endpoint scores and perceived efficacy. For questions of mediation, structural equation modeling will allow us to build an overall path analysis model.

We will implement analyses using SUDAAN (version 10) to take into account the complexity of the study design as a result of weighting, clustering, and stratification. We will also

reproduce some of the analysis results using the SPSS computer program; the Complex Samples add-on module will be used to account for the complex survey design and appropriately use the analysis weights.

For Study 2, we will test whether there is a main effect of the educational intervention (present/absent), a main effect of type of information (general indication, list of symptoms, or composite definition), and/or an interaction between educational intervention and type of information on our main dependent variables (e.g., consumers' awareness and comprehension of the composite score concept, perceived drug efficacy, and risk recall) using ANOVAs. We will conduct ANOVAs both with and without covariates (e.g., demographic characteristics) included in the model. In addition, we will test whether effects are moderated by other measured variables (e.g., health literacy). 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 Analysis

Study 1

The proposed study design consists of 1,600 participants. All power calculations are for a two-tailed test. We assumed a design effect equal to 1. If the study involves over or under sampling, the design effect will increase and reduce the power.

Hypothesis 1: Education Level and Comprehension. Participants with lower levels of education will exhibit poorer comprehension of composite endpoint scores than participants with higher levels of education.

The G*Power program² was used to calculate the power of a t-test for independent groups to detect differences in comprehension between less educated and more educated participants. We defined less educated as a participant who did not have any education past high school. We assumed that this study would have a similar percentage of less educated participants as we had in a recent, similar FDA-funded study (20.6%). The power results for Hypothesis 1 are shown in Table 2.

Table 2. Power for Comprehension Comparisons between Less Educated and More Educated Participants

	Sample Size Per Group				
Education	Less Education	More Educatio	Probability of Type I	Cohen's Effect Sizo ³	Dowor
High School or Less		11	Error (a)	0.2	0.90
vs. More than High School	330	1,270	0.05	0.5 0.8	0.99

Hypothesis 2: Age and Comprehension. Older participants will exhibit poorer

comprehension of composite endpoint scores than younger participants.

The G*Power program⁴ was used to calculate the power of a t-test for independent groups to detect differences in comprehension between younger and older participants. We defined an older participant as someone who was at least 55 years old. We assumed that this study would have a similar percentage of older participants as we had had in a recent, similar FDA-funded study (63.3%). The power results for Hypothesis 2 are shown in Table 3.

² Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

³ Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Lawrence Erlbaum Associates: Hillsdale, New Jersey.

⁴ Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

	Sample Gr	Size Per oup	Probability of Type I	Cohen's Effect	
Age	Younger	Older	Error (α)	Size	Power
				0.2	0.97
18-54 Years vs. 55+ Years	587	1,013	0.05	0.5	0.99
				0.8	0.99

Table 3. Power for Comprehension Comparisons between Older and Younger Participants

Study 2

The proposed study design includes 1,740 participants and assigns 290 subjects to each of the 6 arms. Each of the two educational inventions consists of 3 arms resulting in a sample size of 870 per intervention. Each of three categories of information consists of both educational interventions which results in a sample size of 580 per category of information. To provide conservative estimates of power, a Bonferroni correction was used when there were multiple comparisons and one of the proportions was assumed to be 0.5. All power calculations are for a two-tailed test and equal sized groups. We assumed a design effect equal to 1. If the study involves over or under sampling, the design effect will increase and reduce the power.

Binary Outcomes

The G*Power program⁵ was used to calculate the power to detect differences between the six experimental arms. The Bonferroni correction was used to adjust for the required 15 comparisons; therefore, the corrected α =0.003. This correction results in a loss of power and the results for the arm-to-arm comparisons are shown in Table 4.

⁵ Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

	Probability of a		
Percentage Difference	Type I Error	Sample Size	
Between 2 Proportions	(α)	Per Arm	Power
20%	0.003	290	0.98
15%	0.003	290	0.76
10%	0.003	290	0.29

The power to detect differences between the different methods of presenting the information also required a Bonferroni correction; however, there are only 3 comparisons so the corrected α =0.017. The power results for the method of presentation are shown in Table 5.

Table 5. Power to Detect Differences in Presentation Methods					
Percentage Difference		Sample Size			
Between 2	Probability of a	Per Presentation			
Proportions	Type I Error (α)	Method	Power		
20%	0.017	580	0.99		
15%	0.017	580	0.99		
10%	0.017	580	0.85		
5%	0.017	580	0.25		

Table 5. Power to Detect Differences in Presentation Methods

The power to detect differences between the educational intervention groups (i.e. absent and present) did not require a correction for multiple tests and has the highest sample size per group; therefore, these comparisons have the highest power as shown in Table 6.

	Probability of a	Sample Size	
Percentage Difference	Type I Error	Per Educational	
Between 2 Proportions	(α)	Intervention	Power
20%	0.05	870	0.99
15%	0.05	870	0.99
10%	0.05	870	0.99
5%	0.05	870	0.55

Table 6. Power to Detect Differences in Educational Interventions

Continuous Outcomes

The G*Power program⁶ was used to calculate the power to detect differences between the 6 groups described above and assuming a small ANOVA effect size (d=0.1) as suggested by Cohen (1988).⁷ As shown in Table 7, we will have excellent power to detect differences for the main effects of educational intervention and type of information. We will have good to fair power to detect a small interaction effect. We will have very low power (\leq 0.5) to detect very small effect sizes (d=0.05).

 Table 7. Power for Continuous Outcome

Probability of	Total Sample	Cohen's	Power
а Туре I	Size (Equal	Effect	

⁶ Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

⁷ Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Lawrence Erlbaum Associates: Hillsdale, New Jersey.

		Sized		
	Error (α)	Groups)	Size	
Educational Intervention	0.05	1740	0.1	0.99
Type of Information	0.05	1740	0.1	0.97
Educational Intervention * Type of Intervention	0.05	1740	0.1	0.80

Methods to Maximize Response Rates and to Deal with Issues of Nonresponse

The survey in Study 1 will use an existing Internet panel to draw a sample. The panel (described in B.1) comprises individuals who share their opinions via the Internet regularly. 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);
- Administer the survey over the Internet, allowing respondents to answer questions at a time and location of their choosing;
- Email a reminder to the respondents who do not complete the protocol 4 days after the original invitation to participate is sent;
- Provide a toll-free hotline for respondents who may have questions or technical difficulty as they complete the survey.

In the absence of additional information, response rates are often used alone as a proxy measure for survey quality, with lower response rates indicating poorer quality. However, lower response rates are not always associated with greater nonresponse bias (Groves 2006)⁸. Total survey error is a function of many factors, including nonsampling errors that may arise from both

⁸ Groves, R., "Nonresponse Rates and Nonresponse Bias in Households," *Public Opinion Quarterly*, vol. 70(5), pp. 646–675, 2006.

responders and nonresponders. (Biemer and Lyberg 2003)⁹. A nonresponse bias analysis can be used to determine the potential for nonresponse bias in the survey estimates from the main data collection.

There are several approaches to address the potential for nonresponse bias analysis in this study, such as comparing response rates by subgroups, comparing respondents and nonrespondents on frame variables, and conducting a nonresponse follow-up study.¹⁰ For the proposed project, we will perform two steps: comparing response rates on subgroups and comparing responders and nonresponders on frame variables.

We will first identify the subgroups of interest, such as age and gender. At the end of the data collection, we will calculate response rates by subgroup. If the response rates are the same within subgroups, then nonresponse bias should not affect the results related to those group categories. For example, if the response rate for males and females is the same, then there will not be a large nonresponse bias in the survey estimates for gender.

To the extent that information is available about all sample cases on the frame and that information is associated with the key survey estimates, this approach can provide additional information about the potential for nonresponse bias. At the end of data collection, we will review the sampling frame 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. For example, if the median age of the full sample is 45, but the median age of the respondents is 60, there is likely bias in the estimates due to age if age is correlated with any of the survey estimates.

 ⁹ Biemer, P. and L. Lyberg, "Introduction to survey quality," New York: Wiley, 2003.
 ¹⁰ 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.

The experimental Study 2 will also use an existing Internet panel to draw a sample. To help ensure that the participation rate is as high as possible, FDA and the contractor will:

- Use 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;
- Email a reminder to the respondents who do not complete the protocol four days after the original invitation to participate is sent;
- Provide a toll-free hotline for respondents who may have questions or technical difficulty as they complete the experiment.
- 3. Test of Procedures or Methods to be Undertaken

Prior to pretesting, nine participants will respond to the survey in Study 1 while explaining their thoughts and responses, enabling us to assess blatant glitches in questionnaire wording, programming, and execution of the study. Another nine participants will respond to the experimental protocol in Study 2. We will also conduct pretests for each study before running the main study to ensure that the questionnaire wording is clear. Finally, we will run the main study as described elsewhere in this document.

4. <u>Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing</u> Data The contractor, RTI International, will collect the information on behalf of FDA as a task order under Contract No. HHSF223201110333G. Pam Williams, Ph.D., is the Project Director, 919-316-3936. Data analysis will be conducted by both RTI and 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 Helen W. Sullivan, Ph.D., M.P.H., 301-796-4188.