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

1. Respondent Universe and Sampling Methods

Participants will be recruited by professional recruitment facilities in Raleigh, NC (pre-test and main study) and the Washington, DC area (cognitive interviews). Each facility will contact individuals listed in their internal database that have agreed to participate in occasional studies on various topics. These firms will identify promising individuals, screen them for eligibility by phone, and invite eligible individuals to participate in the study on pre-determined days.

Recruitment for the pre-test and main study in Raleigh, NC will be from the general population, with participants meeting in the following age groups: 18-25 years, 40-50 years, 60-74 years, or over 75 years old, with sample sizes for each determined by the breakout described in Table 1 below.

Recruitment for the cognitive interviews in the Washington, DC area will be from the general population, with participants in the following age groups: 18-25 years or 60-74 years.

Participants in all phases (the pre-test and main study in Raleigh, NC and the cognitive interviews in the Washington, DC area) also will meet the following criteria:

- speak English;
- have not participated in a focus group or interviews in the past 3 months;
- are not employed as a healthcare provider or by a pharmaceutical company, the Department of Health and Human Services, or a marketing or advertising firm.

Individuals who have participated in one phase of the study (cognitive interviews, pre-test) will not be eligible to participate in a later phase (pre-test, main study).

2. Procedures for the Collection of Information

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

General Research Questions

1. How do hearing and cognitive declines in older adults affect comprehension of DTC television ads, and the major statement in particular?

2. How do the frequency, speed, and complexity of the major statement influence the comprehension of the major statement and DTC ads as a whole?
3. How do hearing and cognitive declines interact with the frequency, speed, and complexity of the major statement to affect the comprehension of DTC ads?

Design Overview

To test these research questions, we will examine four groups of adults and manipulate three variables as shown in Table 1.

Table 1.

Age	Speed	Voiceover Frequency				Total
		Male (Low Frequency)		Female (High Frequency)		
		Organization of Major Statement		Organization of Major Statement		
		Simple	Complex	Simple	Complex	
Young adults (YA; 18-25)	Low Speed	33	33	33	33	132
	High Speed	33	33	33	33	132
Middle-aged (MA; 40-50)	Low Speed	33	33	33	33	132
	High Speed	33	33	33	33	132
Young-older (YO; 60-74)	Low Speed	33	33	33	33	132
	High Speed	33	33	33	33	132
Old-older (OO; 75+)	Low Speed	33	33	33	33	132
	High Speed	33	33	33	33	132
Total		**264**	**264**	**264**	**264**	**1,056**

Cognitive interviewing and pretesting will take place before the main study to evaluate the procedures, described below, and measures used in the main study. We will recruit adults who fall into one of four age brackets shown in Table 1. We will exclude individuals who work in healthcare or marketing settings because their knowledge and experiences may not reflect those of the average consumer. A priori power analyses revealed that we need 640 participants for the pretest to obtain 80% power to detect a small effect size, and 1,056 participants for the main study to obtain 90% power to detect a small effect size. Data collection will take place in person.

Procedure

Prior to data collection, participants will read and sign the informed consent form (Appendix A).

For the pretest and main study, within each age group, participants will be randomly assigned to one of eight experimental conditions in a 2 (speed) x 2 (frequency) x 2 (complexity) design, as depicted in Table 1. Study participants will first participate in a complete hearing evaluation, including audiometric measurement of individual hearing ability to help determine if hearing declines account for any age group differences in reported comprehension or retention of ad information. In order to achieve high quality hearing evaluations, these assessments will be conducted by audiologists from the UNC Hearing and Communication Center. The audiometric testing will include three parts:

- *Self-report questionnaire***.** The questionnaire will use a single question format (Do you feel you have a hearing loss?) that has been reported to have good sensitivity and specificity in a most populations (Sindhusake et al., 2001), plus some contextual questions.
- *Otoscopy and visual inspection*. Otoscopy will be performed prior to pure tone testing. If conditions likely to affect testing (e.g., excessive cerumen) or indicative of active disease process (e.g., drainage) are observed, the individual will be advised of the condition and disqualified from participation.
- *Pure tone testing*. Pure-tone air conduction audiometric thresholds in decibels hearing level (dB HL) will be obtained for each ear at 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz using calibrated audiometers (MA 40, MA 42, MA 25, Maico; GSI 18, Grason-Stadler) with insert earphones (E•A•Rtone 3A; Etymotic Research) or supra-aural headphones (TDH-39; Telephonics).

The testing specifications were informed by discussion with audiology experts Dr. Blake Wilson and Dr. Sharon Williams, as well as Dr. Nancy McKenna and Dr. Stephanie J. Sjoblad at the UNC Hearing and Communication Center.

After completing the audiometric measurement, participants will move to a different room where they will watch one of the eight manipulations of a fictitious DTC television ad twice, and answer questions in a survey.

Participation overall in the pretest and main test is estimated to take approximately 45 minutes.

Questionnaire measures are designed to assess, for both risk and benefit information, verbatim memory, comprehension, gist memory, and confidence in memory and comprehension judgments. The draft questionnaire is included in Appendix B.

To examine differences between experimental conditions, we will conduct inferential statistical tests such as analysis of variance (ANOVA).

Analysis Plan

Testing most of the hypotheses and research questions listed above will require multiple comparison procedures. For example, in addition to examining the overall 4×2 interaction of age and voiceover speed, we could test interaction contrasts of differences in the effects of voiceover speed between pairs of age groups, and simple main effects of voiceover speed within age groups. Our proposed approach for handling multiple comparisons in the current factorial design reflects statistical principals from Jaccard (1998). To control for experimentwise error rates across multiple tests, we will need to use adjusted significance thresholds. We plan to use the Holm-modified Bonferroni adjustment, which is less conservative than the traditional Bonferroni method and offers sufficient statistical power for maintaining experimentwise error rates at an appropriate alpha level (usually .05). In the Holm-Bonferroni approach, significance thresholds depend on the number of contrasts (or tests) being performed. For example, if we observe a significant interaction of age and voiceover speed, and we want to examine the simple main effects of voiceover speed at each age level, we would calculate 4 contrasts, 1 for each simple main effect within the 4 age groups. Each of those contrasts would have a corresponding p-value. Applying the Holm-Bonferroni method, we would rank order those p-values from smallest to largest and compare them to the following thresholds, respectively, which are also rank ordered from smallest to largest: $p < .013$ (.05/4), $p < .017$ (.05/3), $p < .025$ (.05/2), and $p < .05$ (.05/1). The first p-value threshold equals the traditional alpha level of .05 divided by the total number of contrasts (4). Subsequent p-value thresholds use the same numerator of .05, but change the denominators to reflect the remaining number of contrasts in descending order (i.e., 3 contrasts, 2 contrasts, 1 contrast).

We propose applying the Holm-Bonferroni correction to distinct families of contrasts, rather than across all possible contrasts in the full analysis. For example, we will apply the adjustment to each outcome, rather than across outcomes. We will also treat parameter estimates for each independent variable in a model as a separate family (e.g., age, voiceover speed, and the interaction of age and voiceover speed will constitute separate families). For significant main effects of independent variables, we will treat planned pairwise comparisons as a family. If we observe significant interaction effects, we will treat planned interaction contrasts as a family, and planned simple main effects as a family.

When using multiple comparisons procedures, Jaccard (1998) recommends that investigators specify contrasts that are of theoretical interest. Identifying a subset of the most important contrasts will also allow for larger significance thresholds using the Holm-Bonferroni correction (or other corrections for Type I error rates). Given that the OO (75+) age group is a central focus of the current research, it would be appropriate and theoretically interesting to prioritize contrasts that compare each of the YA, MA and YO age groups to the OO group.

Power

Using the PASS software program (Hintze, 2011)[1], we estimated the sample sizes needed to have sufficient statistical power to detect differences in means for the outcome variables (i.e., recall of major statement) based on the study design: 2 (speed) x 2 (frequency) x 2 (complexity).

As shown in Table 1 above, the main study will include 1,056 participants (33 per cell). For this phase, the design will result in 90% power with an alpha of 0.05 to detect a small effect size (f=0.20), where f is the standard deviation of the standardized population means. When applying the Holm-modified Bonferroni correction as described above, the p value thresholds change to $p < .013$, $p < .017$, $p < .025$, and $p < .05$. Given the sample size of 1,056 participants and assuming 90% power, these corrected values will result in medium (Cohen, 1988) effect sizes of $f= 0.24$ ($p < .013$), $f= 0.23$ ($p < .017$) and $f= 0.22$ ($p < .025$).

The pretest will include 640 participants equally split across the four age groups (with 160 in each age group). For the pretest, the design will result in 80% power with an alpha of 0.10 to detect a small effect size, where f is defined as noted above.

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

This experimental study will rely on the services of professional facilities in Raleigh, NC (pre-test and main study) and the Washington, DC area (cognitive interviews) to recruit participants. These facilities will contact individuals in their internal databases who have agreed to participate in occasional studies on various topics, and screen them for eligibility in the proposed study. Though the use of a panel of individuals who have actively chosen to be included may decrease the potential for non-response by its opt-in nature, we recognize that additional measures to encourage full participation are helpful in increasing participation rates. As with other similar studies, to help ensure that the participation rate is as high as possible, FDA will:

- Design an experimental protocol that minimizes burden (clearly written and with appealing graphics);
- Work with facilities that have a record of success recruiting for similar studies;
- Offer a range of days and times for participation to give eligible participants options that are convenient for them;
- Make telephone reminder calls to identified participants with detailed information and answers to FAQ (e.g., time and date, study procedures, directions);
- Provide on-site study staff at the facility to answer questions or assist participants who may have questions or technical difficulty as they complete the survey.

[1] Hintze, J. (2011). PASS 11 NCSS, LLC. Kaysville, Utah, USA. www.ncss.com

Despite these procedures, we realize that not all identified participants will complete the study, and we recognize that less than full participation could bias study results if the nonrespondents would have answered differently had they completed participation in the study. Hence, to reduce the potential for such bias, we will adjust the sample design weights within cells indexed by gender, age and race/ethnicity. These adjustments will be applied to each respondent within the cells, and the sum of the product of the adjustment times the design-based weights will be equal to the population characteristics of any group they were recruited from. The process forces each respondent to represent a portion of the nonrespondents within the cell. These weight adjustments will be calculated using our Generalized Exponential Model (GEM) software (Folsom & Singh, 2000)[2]. GEM is a raking procedure that is a generalization of the logic-type model, which has been proven to produce less-variable weights than do the traditional weighting class methods (Levy & Lemeshow, 2008)[3].

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.[4] To measure the impact of nonresponse for this study, we propose to measure nonresponse bias by comparing the distribution patterns of responders with known population distributions. These comparisons will be limited to those variables for which we have population information (e.g., sex, age, race/ethnicity).

4. Test of Procedures or Methods to be Undertaken

Nine cognitive interviews have been conducted with individuals in the young (18-25 years) and young-old (60-74 years) age groups to assess questionnaire flow. We propose to interview another nine individuals, this time with participants in the middle-aged (40-50 years) and old-old (75+) age groups, which will allow us to test the audiometric procedures as well as the survey flow (materials included in Appendix C). Moreover, 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 640 individuals in the pretest, or 20 per cell and age group, to enable us to obtain 80% power to detect a small effect size with an alpha of .10 and a small effect size.

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

[2] Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. ASA Proceedings, Survey Research Methods Section. 598–602.
[3] Levy, P., & Lemeshow, S. (2008). *Sampling of populations*. 4th ed. New York: Wiley.
[4] 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 contractor, RTI, will collect and analyze the data on behalf of FDA as a task order under Contract HHSF223201400478G. Brian Southwell, Ph.D., 919-541-8037, 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 Helen W. Sullivan, Ph.D., M.P.H., 301-796-4188.