

## B. Statistical Methods

### 1. Respondent Universe and Sampling Methods

As previously mentioned, the first phase of the study under OMB #0910-0497 used focus groups to gather information about current device labeling, including how it is currently used and how it is understood by the health care professionals (HCPs). The second phase of the study under OMB #0910-0715 obtained HCPs' feedback on a standard content and format of an abbreviated version of medical device labeling, developed with HCPs' feedback from the first phase of the study. The third phase of the study will compare a proposed standardized content of medical device labeling to the current manufacturer device labeling in terms of ease of use and location of key pieces of information.

The subject population for this study includes HCPs who are users or potential users of at least one of the medical devices being tested. A screening tool (located in the initial contact email document) will be used to determine the eligibility of each HCP in the pool to test the labeling of the devices being used in the study. The goal is to have 12 practitioners at each facility test the labeling for each device; 6 would test the standardized labeling, and 6 would test the manufacturer labeling. It should be noted as this is an applied cognitive or human factors/usability investigation, the power analyses typically used to determine sample size in clinical trials or surveys would not apply, because their assumptions are not valid for this type of study. Typically, in most usability studies, a sample size of five to eight per group is considered sufficient to discover the vast majority of serious usability problems (see, for example, AAMI/ANSI HE-75. (2009) Annex A, p. 441). We have chosen 6 per group (an even number) because we expect to have two or three types of practitioners testing each device, and we would like to have an even number of each type of practitioner in the test. Further, the experimental design (discussed later) is a repeated measures design, which provides additional confidence that the data collected will be sufficient to test the primary hypothesis.

Using data gathered via the screening tool, we would attempt to control for the effects of 1) experience (user or potential user), and 2) practitioner type (e.g., nurse, physician, technician) in the sample by matching participants in the labeling groups (draft standardized labeling versus manufacturer labeling) for each device. Each participant will test between one and three devices. The particular devices each will test will be selected based on the screening information, and the amount of time required to test each device, relative to the designated length of the testing session.

Participants will therefore be a self-selected sample from among those recruited and will be motivated to complete the study. They will also be offered a nominal incentive for participating, which is customary when recruiting participants from a hard-to-recruit population.

## 2. Procedures for the Collection of Information

Participation in this data collection is voluntary and the respondent's information will remain private to the fullest extent allowed by law. Participants will be recruited from the CDRH MedSun program, or drawn from locally available clinicians found at participating hospitals, based on their experience with the devices being used in this study. The MedSun program will enable us to find potential participants (from the three key medical device labeling user groups: physicians, nurses, and therapists) in the DC area. There is no evidence to suggest that performance in cognitive usability testing is a function of geographic location and therefore the performance recorded in the DC location should not differ from practitioners in other localities. Subjects will be tested individually in sessions lasting no more than 90 minutes. Subjects will be audiotaped. Subjects will be able to see each device that they test since it will be placed in front of them, for reference purposes only, along with either the standardized labeling outline (i.e., Table of Contents) for that device or the manufacturer's labeling outline. Each subject will be given a series (approximately 8 to 12) of scenarios posing questions a user might be expected to ask about labeling, and to respond to each scenario by indicating the section of the labeling outline which he/she believes would contain the answer to the question. If the subject's answer is correct, the next scenario is then presented. If the subject's answer is incorrect, the subject is asked seek another section of labeling in which they think the answer might be found. The process continues until the subject makes the correct response or gives up. The dependent measures for each scenario will be the time elapsed between the end of the reading of the scenario and the correct response (timed via stopwatch) and the number of incorrect responses prior to the correct response.

When all scenarios are completed, subjects will be asked to explain their thought process for any scenarios for which they gave at least one incorrect answer, and will also be asked a number of device-specific labeling questions that could not be answered via the scenarios.

If a subject is testing the labeling for more than one device, the above process would be repeated for each device labeling tested.

We do not anticipate having any missing data, assuming each subject arrives for his/her testing session. If a subject misses, or is unable to attend, their testing session and cannot be rescheduled, we would return to the pool to acquire a replacement subject.

The two primary hypotheses of the study are that a) subjects in the standardized labeling group will have faster response times than subjects in the manufacturer's labeling group; and b) subjects in the standardized labeling group will make fewer errors, prior to the correct response, than subjects in the manufacturer's labeling group.

Data will be analyzed separately for each device, using repeated measures analysis of variance, to test the two primary hypotheses, using packages available to the FDA. Additional analyses will be conducted post hoc, if it appears there are differences among practitioner types or level of experience that should be explored or other trends

observed in the data. Investigators will also conduct a non-quantitative error analysis for each scenario, and an analysis using descriptive statistics on the post-testing interview data.

3. Methods to Maximize Response Rates and 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. While outside factors such as email filters, recipient ISP downtime, and general internet conditions can impact potential participants from receiving the screener, these issues are unlikely to alter the response rate for participants once they agree to the study as the study is performed in person. To help ensure the response rate is as high as possible FDA will:

- Recruit participants based on knowledge and experience using the devices under study.
- Provide nominal incentive to increase participation from study participants.
- Conduct the study at the work place. Two-thirds of the participants will complete the study at their place of work, decreases non-response related to travel.
- Use an experimental protocol that minimizes burden (short in length, clearly written, easy to view font styles).
- Provide participants with contact information to send questions about the study to the FDA at any time.

4. Test of Procedures or Methods to be Undertaken

Pretests will be completed with CDRH HCPs to provide input on scenarios and the use of standardized labeling outline based on their device experience, and to more closely determine the appropriate number of scenarios and devices that feasibly be tested in the time allotted for the testing sessions.

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

CDRH statisticians were consulted to confirm the appropriate testing methodology and confirmed the number of individuals needed to achieve statistical significance.

Reference:

ANSI/AAMI HE75. (2009). *Human factors engineering – Design of medical devices*. Arlington, VA: AAMI.

Attachments:

Attachment A—Initial Contact Letter with Screener

Attachment B—Consent Form In-Person

Attachment C—Follow-Up Letter

Attachment D—EIR Standard Content Draft

Attachment E—Sample Scenarios Acuvue

Attachment F—Sample Labeling FDA Standard Content TOC Only Acuvue

Attachment G—Sample Labeling current mfr TOC Only Acuvue

Attachment H—Sample Complete Labeling in Standard Content Format Acuvue