



Designation: E1958 – 25

Standard Guide for Sensory Claim Substantiation¹

This standard is issued under the fixed designation E1958; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

Formats or standards for testing related to sensory claim substantiation cannot be considered without a frame of reference of where that format or standard would fit within the legal framework that surrounds the topic. Product sensory claims tests are performed for three basic reasons: (1) *Comparison of Products*—Determines how one product compares to another, usually a competitor or earlier version of itself. (2) *Substantiation of Claims*—Enables marketing personnel to use positive references through advertising or packaging, or both, in the presentation of the product to the consumer. (3) *Test Performance*—Ascertains and establishes the tested product performance within the scope of its intended use.

The risk associated with each claim is assessed when considering claims substantiation. Compelling and aggressive claims are sure to be scrutinized closely by competitive firms, and if inconsistencies are found through competitive test data, the claims could be challenged in one or more of the following venues: (1) National Advertising Division (NAD) of the Advertising Self-Regulatory Council (ASRC), (2) one or more media, such as print, broadcast, or electronic media, (3) Consumer Advocacy Organizations, and (4) Civil or Federal courts. No single test design or standard test will prevent challenges. The criteria used by each of the potential forums are not identical and are constantly evolving. With the introduction of new technologies coupled with changing consumer demands, testing processes and protocols that were sufficient five or ten years ago may not hold up under today's criteria and scrutiny. Testing requirements of the future can only be a matter for speculation. The one constant is that, as advocates of their clients' positions, attorneys will defend their clients' testing processes and protocol while questioning with great detail every aspect of their competitor's protocol in the attempt to sway the arbiter to agree that their clients are in the right. Legal counsel should be part of any team developing claim substantiation.

This guide demonstrates what a group of professionals who are skilled in the science of testing consider appropriate from a scientific and technical standpoint, and represents an effective method for both defendant and challenger to determine the viability of a sensory claim. The key word is "appropriate." If a particular aspect of a test, or method, is not appropriate for a specific application, it should not be used. Care should be taken to clearly define the reasons and data supporting a deviation from the standard, as any departure invites scrutiny. Since departures are inevitable, the word "should" is used in this guide to indicate when other techniques may have applications in certain unusual circumstances. Whenever a test protocol has been completed, it should be critiqued for weaknesses, including whether experts in the relevant field would consider the research objectively designed, conducted, and analyzed, using procedures that give accurate and reliable results. If weaknesses are found, corrective action should be taken, since the competition may point out any weakness or discrepancy and challenge the study.

While the scientific and technical community identifies the appropriateness of a research method used to support a sensory claim, the legal community evaluates substantiation for legal claims using "reasonableness" as the criterion. With the importance of having a legal "reasonable basis" for a claim, the question remains, "What is reasonable?" Unfortunately, there is no specific answer to that legal question, as it will depend on the type of claim, product application and use, applicable regulations where the product is sold, and other factors. These considerations, market pressures (such as timing), and testing budgets can influence and impact the protocols to support a specific claim. This guide provides principles and considerations that need to be addressed for good sensory and consumer testing practices.

1. Scope

1.1 This guide covers reasonable practices for designing and implementing sensory tests that validate claims pertaining only to the sensory or perceptual attributes, or both, of a product.

This guide was developed for use in the United States and must be adapted to the laws and regulations for advertisement claim substantiation for any other country. A claim is a statement about a product that highlights its advantages, sensory or perceptual attributes, or product changes or differences compared to other products in order to enhance its marketability. Attribute, performance, and hedonic claims, both comparative and non-comparative, are covered. This guide includes broad principles covering selecting and recruiting representative consumer samples, selecting and preparing products, constructing product rating forms, test execution, and statistical handling of data. The objective of this guide is to disseminate good sensory and consumer testing practices. Validation of claims should be made more defensible if the essence of this guide is followed.

Table of Contents

Introduction	
Scope	1
Referenced Documents	2
Terminology	3
Basis of Claim Classification	4
Consumer Based Affective Testing	5
Sampling	5.1
Sampling Techniques	5.2
Selection of Products	5.3
Sampling of Products When Both Products Are Currently on the Market	5.4
Handling of Products When Both Products Are Currently on the Market	5.5
Sampling of Products Not Yet on the Market	5.6
Sample Preparation/Test Protocol	5.7
Test Design—Consumer Testing	6
Data Collection Strategies	6.6
Interviewing Techniques	6.7
Type of Questions	6.8
Questionnaire Design	6.9
Instruction to Respondents	6.10
Instructions to Interviewers	6.11
General/Overall Questions	6.12
Positioning of the Key Product Rating Questions	6.13
Total Test Context and Presentation Matters	6.14
Specific Attribute Questions	6.15
Classification or Demographic Questions	6.16
Preference Questions	6.17
Test Location	7
Test Execution by Way of Test Agencies—Food and Non-food Testing	8
Documents to Retain in Sensory Claims Substantiation Research	9
Laboratory Testing Methods	10
Types of Tests	10.2
Advantages and Limitations of the Use of Trained Descriptive Panels in Claims Support Research	10.3
Test Design—Laboratory Testing	11
Product Procurement	11.6
Experimental Design	11.7
Data Collection	11.8
Data Analysis	11.9
Questionnaire Construction	12

Test Facility	13
Statistical Analysis	14
Paired-Preference Studies	14.1
Superiority Claims	14.2
Equivalence Claims	14.3
Unsurpassed Claims	14.4
Paired Comparison/Difference Studies	14.5
Analysis of Data from Scales	14.6
Keywords	15
Commonly Asked Questions About ASTM and Claim Substantiation	Appendix X1

1.2 This international standard was developed in accordance with internationally recognized principles on standardization established in the *Decision on Principles for the Development of International Standards, Guides and Recommendations* issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.

2. Referenced Documents

2.1 ASTM Standards:²

- E253 Terminology Relating to Sensory Evaluation of Materials and Products
- E1885 Test Method for Sensory Analysis—Triangle Test
- E2164 Test Method for Directional Difference Test
- E3314 Guide for Protection of Respondents and Informed Consent for Sensory Evaluation Studies

3. Terminology

3.1 Definitions:

3.1.1 Terms used in this guide are in accordance with Terminology E253. Additional terms are listed below.

3.2 Definitions of Terms Specific to This Standard:

3.2.1 α (alpha) risk, n —the probability of concluding that a sensory claim is supported when, in reality, the sensory claim is false.

3.2.1.1 Discussion— α risk also is known as Type I Error or significance level.

3.2.2 β (beta) risk, n —the probability of concluding that a sensory claim is not supported when, in reality, the sensory claim is true.

3.2.2.1 Discussion— β risk also is known as Type II Error.

3.2.3 central location test (CLT), n —a test where stimuli are evaluated by consumers at a common designated location(s); the stimuli preparation and presentation are usually controlled.

3.2.3.1 Discussion—Examples of designated locations include market research facilities, academic laboratories, grocery stores, or hotel conference rooms.

3.2.4 comparative claim, n —designed to compare similarities and differences between two or more products. The basis for comparison can be within the same brand, between two brands, or between a brand and other products in the category.

3.2.5 context effect, n —effect upon the perception of a stimulus arising from its interrelationship with other stimuli, such as in a presentation set or in the experiment.

¹ This guide is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation and is the direct responsibility of Subcommittee E18.05 on Method Applications - Specific Uses or Assessors.

Current edition approved Aug. 1, 2025. Published September 2025. Originally approved in 1998. Last previous edition approved in 2022 as E1958 – 22. DOI: 10.1520/E1958-25.

² For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at www.astm.org/contact. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.2.6 *directional difference test, n*—a paired comparison or 2-AFC (Alternative Forced Choice) method in which assessors select the stimulus from a pair of stimuli that is perceived to be higher or lower in intensity of a specified sensory attribute.

3.2.7 *documenting, v*—the process of capturing information to create a record (see: **record**).

3.2.7.1 *Discussion*—Documenting produces records necessary for substantiating a sensory claim.

3.2.8 *equivalence claim, n*—in an equivalence claim, two products are claimed to be equivalent in one or more particular feature.

3.2.9 *home use test (HUT), n*—a test where stimuli are evaluated by consumers at home or in the environment typical of the actual use situation; the stimuli preparation and presentation are usually self-administered.

3.2.9.1 *Discussion*—Home use test and in-home use test are used interchangeably.

3.2.10 *monadic or single product tests, n*—product tests where only one product is experienced and rated.

3.2.11 *power (statistical), n*—the probability that a sensory claim will be supported, when, in reality, the sensory claim is true.

3.2.11.1 *Discussion*—Power = $1 - \beta$, where β is the Type II error rate.

3.2.12 *record, n*—any combination of text, images, data, video, audio, or other information in physical, electronic, or other format that is created, maintained, and retrieved in a controlled manner (based on (1)).³

3.2.12.1 *Discussion*—In most cases the person who created the record and the date should both be included. The time should be included, if appropriate.

3.2.12.2 *Discussion*—A record is evidence of something done in the past. In limited circumstances, a record can be modified following good documentation practices where original and changed aspects of the record are unambiguous. For example, on a paper record new text could be added over a single line crossing through the original text. Next to the change, the initials of the person who made the change are unambiguously identified, and the date of the change.

3.2.13 *self-administered questionnaire, n*—questionnaires independently completed by the respondent are referred to as self-administered.

3.2.14 *superiority claim, n*—a superiority claim asserts a higher level of performance or liking or preference relative to another brand; a superiority claim can be opposed to a competitive brand (for example, “cleans better than brand Z”) or opposed to an earlier formula of the brand (for example, “now more cleaning power than before”); a superiority claim is supported if a statistically significant proportion of the respondents prefer the advertiser’s product.

3.2.15 *unsurpassed claim, n*—an unsurpassed claim indicates that the product(s) selected for comparison is not better/

higher (or greater than) in some way to the target product(s) for which the analysis is executed.

4. Basis of Claim Classification

4.1 A fundamental step in advertising claim substantiation is creating an explicit statement of the claim prior to actual testing. The statement is then forwarded to all parties concerned in the substantiation process. Concerned parties could include marketing, marketing research, legal, consumer testing, sensory evaluation, research suppliers, etc. The statement is essential as it can encourage collaboration in terms of corporate resources, confirms the selection of appropriate test methods, and has the potential to maximize the chance of making reliable business decisions about the proposed claim, pending the results of substantiation research. Collaboration among all involved parties prior to executing substantiation research is critical in achieving the best results. All involved parties should meet and agree (perhaps several times) prior to implementing the substantiation research.

4.2 Familiarity with the general classification of advertising claims is important in developing clear statements of claims at an early stage and for developing a rational plan for testing. This familiarity also facilitates the process of selecting appropriate testing methods, among the many types of methods available to the consumer/sensory science professional. Each method answers specific questions and may support one type of claim but not another. Therefore, the consumer/sensory science function provides an important source of information and experience in claim substantiation and will provide much of the definition of testing methodology. There are multiple ways to support claims depending on the characteristics of the claim. Two approaches are consumer based and trained panel based evaluations.

4.3 Advertising claims can be divided into two fundamental classifications: comparative and non-comparative. The distinction between the two classifications is whether a comparison is made relative to an existing product (advertiser’s or competitor’s) or to itself.

4.4 *Comparative Claims*—Comparative claims are designed to compare similarities and differences between two or more products. The basis for comparison can be within the same brand, between two brands, or between a brand and other products in the category.

4.4.1 Comparative claims generally take one of three forms: superiority, equivalence, or unsurpassed. Superiority, equivalence, or unsurpassed claims are further sub-classified into two central areas of application: hedonic and attribute/perception. Hedonics broadly concern measuring the degree of liking and preference; either liking overall or liking that is limited to one or more specific attributes. Attribute/perception claims apply to intensity when measuring one or more specific product attributes.

4.4.2 *Equivalence Claims*—Equivalence claims are claims that assert equivalent levels of performance or liking when comparing a particular product to another product. In general, equivalence claims are made relative to a market/category leader. Two products are not claimed to be exactly equal in one

³ The boldface numbers in parentheses refer to the list of references at the end of this standard.

or more particular features; rather, the products are claimed to be sufficiently similar as to be equivalent from a practical perspective.

4.4.2.1 *Hedonic:*

(1) “Tastes as good as brand X.”

4.4.2.2 *Attribute/Perception:*

(1) “Our product reduces odors as much as brand X.”

(2) “Our product lasts as long as brand X.”

(3) “Our cake is as moist as the leading brand.”

4.4.2.3 *Overall Equivalence:*

(1) “We’re just the same, except for the price.”

(2) “You’ll never know the difference between us and brand X.”

4.4.3 *Unsurpassed Claims*—In unsurpassed claims, the claim stated indicates that the product(s) selected for comparison is not better/higher (or greater than) in some way to the target product(s) for which the analysis is executed. Examples of unsurpassed claims include the following types:

4.4.3.1 *Hedonic:*

(1) “No other product is better than our product.”

(2) “No other product is more liked for butter flavor.”

4.4.3.2 *Attribute/Perception:*

(1) “No other cake is more moist than ours.”

(2) “No other product has more butter flavor than ours.”

(3) “No other product reduces odors more than our product.”

(4) “No other product lasts longer than our product.”

(5) “No other product is thicker than our product.”

(6) “No other product cleans faster than our product.”

4.4.4 *Superiority Claims*—Superiority claims assert a higher level of performance or liking relative to another brand. Superiority claims can be opposed to competitive brands (for example, “cleans better than brand Z”) or opposed to an earlier formula of the brand (for example, “now more cleaning power than before”). Examples of superiority claims include:

4.4.4.1 *Hedonic:*

(1) “Our product tastes better than brand X.”

(2) “Our product tastes better than any other.”

(3) “Our product is preferred over any other brand.”

4.4.4.2 *Attribute/Perception:*

(1) “Our cake is more moist than any other.”

(2) “Reduces odors more than brand X.”

(3) “Lasts longer than any other product.”

(4) “Thicker than brand X.”

(5) “Cleans faster than any other product.”

4.4.4.3 In superiority claims, combinations of hedonic claims and attribute/perception claims can sometimes be found, when superiority claims are established based on overall liking and for specific attributes (for example, “Our hosiery is preferred over Brand X for overall liking and it offers more support and comfort.”).

4.5 *Statistical Considerations for Comparative Claim Types:*

4.5.1 From a statistical perspective, it may require a smaller consumer sample to support claims of unsurpassed than a claim of superiority or equivalence.

4.5.2 If hypotheses related to multiple attributes are tested independently using the Type I error rate (level α) with the

intention of making the claims that are supported, then it elevates the risk of making one or more false claims above the nominal Type I error rate. To control such risk, an appropriate statistical procedure must be used to adjust for multiplicity. Typically, each hypothesis needs to be tested using an adjusted alpha level that is less than the Type I error rate. This testing procedure must be determined and documented prior to data collection. Consult a statistician if required.

4.5.2.1 If hypotheses are related to multiple attributes and no claim will be made unless all of the test results support the claim, then a multiplicity adjustment is not required. This testing procedure must be determined and documented prior to data collection. A longer questionnaire is often needed to test multiple claims. For considerations related to test design, see Section 6 (especially 6.13).

4.5.2.2 If one hypothesis is related to an attribute evaluated by a trained sensory panel and another hypothesis is related to an attribute evaluated by a consumer panel and no claim will be made unless all of the test results support the claim, then a multiplicity adjustment is not required. This testing procedure must be determined and documented prior to data collection.

4.5.3 An advertiser can use a test result to evaluate hypotheses sequentially without adjustment for having conducted multiple tests if the hypotheses are tested using the same approach and differ only in their critical values. Specifically, if one claim is supported, then a subsequent claim can be evaluated without multiplicity adjustment if all possible test results that support the subsequent claim also support the first claim. If using the statistical methods in Section 14, then the following statements hold:

4.5.3.1 Every superiority claim is also an unsurpassed claim. An advertiser can evaluate an unsurpassed claim as a primary hypothesis and a superiority claim as a secondary hypothesis. No claim can be made if the unsurpassed claim is not supported. If the unsurpassed claim is supported, then the unsurpassed claim can be made, regardless of whether the superiority claim is supported. If the superiority claim is supported, then the superiority claim can be made. The unsurpassed test (14.4) and superiority test (14.2) use the same approach and differ only in their critical values. As every test result that supports a superiority claim also supports an unsurpassed claim, it is valid to evaluate both tests using the Type I error rate (level α) without adjusting for multiplicity.

4.5.3.2 Every equivalence claim supported by the test in 14.3 is also an unsurpassed claim that will also be supported by the test in 14.4. These tests use the same critical value for the lower bound. The equivalence test from 14.3 uses an additional critical value for the upper bound. An advertiser can evaluate an unsurpassed claim as a primary hypothesis and an equivalence claim as a secondary hypothesis. No claim can be made if the unsurpassed claim is not supported. If the unsurpassed claim is supported, then the unsurpassed claim can be made, regardless of whether the equivalence claim is supported. If the equivalence claim is supported, then the unsurpassed claim can be made. As every test result that supports an equivalence claim also supports an unsurpassed claim, it is valid to evaluate both the unsurpassed and equivalence claims using the Type I error rate (level α) without adjusting for multiplicity.

4.5.3.3 If using statistical methods other than those recommended in Section 14, then these statements do not necessarily hold, so additional justification is required.

4.5.4 Methods for hypothesis testing can be found in the sections related to the analysis plan (5.2.10) and to statistical methods (Section 14).

4.5.5 An advertiser may create a data analysis plan (5.2.10) in which data are analyzed prior to the end of the study. Most studies evaluate whether a particular sensory claim is substantiated. Some studies have multiple parts, where each part can evaluate one or more sensory claims. In this section (4.5.5), the best practices for interim analysis of a study apply equally to each part of a study.

4.5.5.1 An interim data analysis can be used to stop a study early or to modify a study. The purpose of stopping a study early or modifying a study is to avoid wasting resources.

4.5.5.2 Interim data analysis should be pre-specified and documented in the data analysis plan (5.2.10). Data should only be used in ways specified. Data should not be used in ways not pre-specified.

4.5.5.3 A study can be stopped for futility, for example, where the probability of obtaining the desired result is deemed too low to continue the study.

(1) When a futility analysis is incorporated into the data analysis plan (5.2.10), the advertiser should describe how the interim data analysis will guide the decision to stop the study. An analysis plan could establish certain thresholds to be considered when deciding whether to continue or stop the study (for example, the advertiser might specify getting a majority of the preferences, observing less than 20 % no-preference responses, observing no unexpected adverse reactions, etc.). Binding thresholds provide strict rules indicating interim analysis outcomes that will lead to stopping or continuing the study. Nonbinding thresholds are generally preferred for making decisions more flexibly while incorporating new information (for example, an unexpected deviation in interim analysis base size, results from a relevant new study, a change in the advertiser's risk tolerance, budget, claim strategy, etc.).

(2) If a study is stopped early for futility, then responses collected for the study that was stopped cannot be used to support any claim.

4.5.5.4 A study can also be stopped for success, for example, if the claim is demonstrated based on the test result of the interim data analysis.

(1) Stopping for success requires a technical rationale. It must be demonstrated that the range of relevant respondent characteristics (5.2) is sufficient for substantiating the sensory claim based on the interim data analysis. The statistical approach used in the analysis plan must be justified. It is strongly recommended that a statistician be consulted for designing the data analysis plan (5.2.10) and analyzing the data.

4.5.5.5 A study can be redesigned based on the test result from an interim data analysis.

(1) Changing the study design based on the interim data analysis, such as a change to the number of consumers in the study, requires a strong technical rationale. It must be demon-

strated the range of relevant respondent characteristics (5.2) after the design change is sufficient for substantiating the sensory claim. The statistical approach used in the analysis plan must be justified. It is strongly recommended a statistician be consulted for designing the data analysis plan (5.2.10) and analyzing the data.

4.5.5.6 A study or part of a study can be continued after an interim data analysis.

(1) If the data analysis plan allows only the possibility of continuing or stopping for futility, then the analysis of the complete data can proceed conventionally since in this case the interim data analysis does not affect the Type I error rate for the complete data analysis.

(2) If the data analysis plan allows the possibility of stopping for success or redesigning the study based on the interim data analysis, then Type I error rate of the complete data analysis is affected even if the original test design is preserved. The tables in this guide do not apply when the data analysis plan allows the possibility of stopping for success or redesigning the study based on the interim data analysis. If these possibilities will be allowed, it is strongly recommended a statistician be consulted to develop a valid data analysis plan.

4.5.5.7 In some cases, the specified number of consumers is not exactly realized in spite of best efforts. If the reason the number of consumers is slightly over or under the target number of consumers is unrelated to interim data analysis, then no statistical rationale is needed to adjust for not getting the planned number of consumers; however, the analysis will need to take the observed number of consumers into account when conducting interim and/or final statistical analyses of test data.

4.5.5.8 *Example: An Interim Data Analysis for Futility.* Pre-product launch consumer testing showed 42 % prefer the advertiser's product, 40 % prefer the competitor's product, and 18 % have no preference. After launching the product, the advertiser plans to make an unsurpassed claim, or, if possible, a superiority claim (4.5.3.1). Data from the pre-product launch study are already known and cannot be used at all to substantiate this claim. The data collection plan specifies a study with 500 consumers in five regions with 100 results per region. Based on Table 5, an unsurpassed claim is substantiated if the choice count for the advertiser's product is 244 of the 500 consumers (48.8 %). Data will be collected in two regions (200 consumers), then from the remaining three regions (300 consumers). The plan includes an interim data analysis after the first stage of data collection based on responses from 200 consumers. The rationale for this interim data analysis is to provide a decision point that allows the advertiser to make one of two choices:

(i) Cancel the second stage of data collection and abandon the claim to save resources, or

(ii) Proceed to the second stage of data collection after which the claim is evaluated based on the results from all 500 consumers, as specified in the data collection plan.

(1) *Decision Rule for Example*—The advertiser creates a data analysis plan stating the study can be stopped for futility for a variety of reasons after the interim data analysis where the stopping thresholds are nonbinding (4.5.5.3.1). The advertiser establishes the following threshold: to stop the study for futility

if the advertiser's choice count in the interim analysis is 95 or fewer out of 200 (after no-preference choices are handled; 14.1.2). The threshold is nonbinding and set to avoid allocating resources to complete a study with an unsubstantiated claim.

(2) *Example Stopping Scenario 1*—The interim data analysis reveals 74 consumers chose the advertiser's product, 88 chose the competitor's products, and 38 have no preference. Since the advertiser's choice count (including half of the no-preference choices) is 93 out of 200, the study is stopped for futility based on the nonbinding threshold.

(a) None of the data from the 200 consumers analyzed in the interim data analysis can ever be used to support any claim. For example, if, in one city, 42 consumers chose the advertiser's product, 36 chose the competitor's product, and 22 expressed no preference, the advertiser cannot reuse these data, which are already known, in a new study to support a regional claim for the advertiser's product.

(3) *Example Stopping Scenario 2*—The interim data analysis reveals 90 consumers chose the advertiser's product, 98 chose the competitor's products, and 12 express no preference. The advertiser has a total of 96 out of 200 favorable answers (including half of the no preference choices). Although the advertiser's choice count (96) exceeds the nonbinding threshold for stopping the study in the data analysis plan (95), the advertiser stops the study because the advertiser has become more risk adverse and sensitive to unanticipated costs associated with the study.

(4) *Example Stopping Scenario 3*—The interim data analysis reveals 97 consumers chose the advertiser's product, 89 chose the competitor's products, and 14 express no preference. The advertiser has a total of 104 out of 200 favorable answers (including half of the no preference choices). Although the outcome is very favorable to the advertiser, the study is stopped due to a change in marketing strategy.

(5) *Example Continuation Scenario*—The interim data analysis reveals 88 consumers chose the advertiser's product, 76 chose the competitor's products, and 36 have no preference. Since the advertiser's choice count of 106 exceeds the non-binding threshold of 95, the study continues with the second stage of $n = 300$ consumers across the remaining three regions.

(a) The study concludes with results from only 495 consumers, not 500 consumers. The deviation in sample size was due to recruitment issues arising from severe weather in the second stage of the study. At the completion of the study, the advertiser's choice count is 245 (49.5 %). Based on Table 5, the advertiser requires a claim count of 242 (48.9 %) to make an unsurpassed claim with 495 consumers. This claim can be made because the consumer count for the advertiser's product is sufficiently high, best efforts were made to get 500 consumers, and getting fewer consumers was not influenced by the interim data analysis.

4.6 *Non-comparative/Communications Claims*—The objective of the non-comparative/communications claim is to convey something specific about the product, usually a product benefit or difference, and in general, does not seek to provide comparative claims relative to other products. For example, the statement "provides long-lasting flavor" or "smells strong for one month" tells us something about the product, but not in a

comparative sense relative to an existing product. These types of claims are common in new product types, but also are used to bring attention to specific product benefits. Examples of non-comparative/communications claims include the following types.

4.6.1 *Hedonic:*

(1) "Tastes great."

(2) "Makes your laundry outdoor-fresh."

(3) "Leaves a long-lasting freshness you will like."

4.6.2 *Attribute/Performance:*

(1) "Removes odors for 60 days."

(2) "Leaves glass streak-free."

(3) "Leaves no residue on surfaces."

(4) "Works fast."

NOTE 1—In the above attribute examples, some of these could be approached either as a non-comparative claim, since no other product is mentioned, or as a comparative claim versus an appropriate standard (streak-free glass, residue-free surface, odor-free room).

4.7 *Selecting an Appropriate Ad Claims Test*—Product claims made in print or on radio, TV, or the Internet require valid data that supports the intended claim. As with most sensory testing, it is necessary to first identify the project and test objectives for the study. The claim statement should indicate whether the claim is based on consumer or laboratory sensory methods or, in fact, some instrumental or chemical test. Sensory claims for preference or liking ("preferred over the leading brand" or "better than the competition") require consumer tests with the preference or liking questions to support the claim. Claims about product attribute(s) or performance can be based on data from consumers, who are asked about the specific attribute, or from laboratory sensory tests designed to measure the specific attribute(s). In some cases, both types of testing (consumer and laboratory) can be used together to support the same claim. The ad claims team needs to determine the type of claim, the claim statement, the target population, and the aspect(s) of the product that is the focus of the claim. Only then can the test to support the claim generate data with the right focus and weight to support the claim.

5. Consumer Testing: Selection of Study Respondents and Product

5.1 *Study Respondent Sampling:*

5.1.1 Representative sampling of relevant study respondents is a key element of a defensible claim study. The type of claim and the target audience of the advertising are two determining factors for the appropriate respondent sample, within practical considerations of the category and claim. Respondents should be typical users and not be restricted to any specific subset, such as heavy or light users, unless it is a qualified claim about that specific subset. It is also important to distinguish between purchasers or users of the product to ensure sampled respondents align with the claim.

5.1.1.1 Screening based on recent category usage is recommended to identify target respondents. A simple product statement should be used to describe the product to be tested with just enough information so that it is clear what the product is. The product statement should not contain puffery, benefits, aesthetic description, or ingredients, unless these are unusual enough to require an explanation. Sometimes respondents need

to be screened to a more narrowly defined criterion. For example, a claim about strawberry yogurt should recruit respondents who like and eat strawberry yogurt, not just any yogurt with fruit. Conversely, respondents typically do not need to be restricted to use of only that product within the category, they may also use alternative products. For the same example, a claim about strawberry yogurt does not mandate that they be exclusive to eating strawberry yogurt. In some cases, the screening criteria needs to be broadened, for example a criterion for inclusion may be that all the participants have the same need, even though that need may be met by products that are designed very differently or for other purposes (for example, widely different products may be used for make-up removal).

5.1.1.2 If a claim is aimed at drawing future category users in addition to present users, future intent to use should be included as a screening criterion. If recent category usage is not applicable (more than the past three to six months), such as with seasonal products or products with long purchase-repeat cycles, or when a product is innovative and part of a newly created category, identifying study respondents based on positive future category usage intent (for example, concept or prototype acceptors) may be necessary. More detail than the simple product statement should only be used when it is necessary to clarify what the product is because of its potential to bias respondents in advance of the claims substantiation test. When the product is new to the market without any comparable product, a formal concept may be used to screen respondents for the claims substantiation testing.

5.1.2 Hedonic or affective (such as liking, acceptance, preference, or purchase intent) claims should always apply to the target user population. Sampling from any population other than the users to whom the claim is focused, such as purchasers, may require a qualified claim to limit its generality. The test protocol should state clearly whether a claim is being made for one or more of the following:

Population	Examples
Product recommenders	"doctors prefer..."
Purchasers	Pet food by pet owners "Choosy moms choose Jif" ^A is a claim specific to the purchaser, the user is the child
The ultimate consumer of a product	"now more refreshing"

^A Jif is a registered trademark of the J.M. Smucker Company.

5.1.3 Perception claims are based on product attribute(s) or performance, or both, that are detectable in the target user population. Claims that state or imply detectability (for example, "more chocolate flavor"), include study participants from the target user population. Note, trained and validated panel data can be used, if data related to the claim from the target population already exists, and it can be demonstrated that the trained panel's measurements can be used to predict the target population (see 10.3.2 and 10.3.3).

5.1.4 For category usage claims, respondents may be recruited by screening for brand usage, but care should be taken during screening to ensure respondents cannot determine which brands are targeted for testing. This can be accomplished by providing a large list of brands with the brand or brands of interest embedded in the questionnaire. Brand usage and

frequency of use data also can be collected to help match the incidence of these variables in the respondent sample and target population. Target users can be defined by their responses to several questions, including:

5.1.4.1 "What one brand of this product type do you use most often?"

5.1.4.2 "What brands have you used in the last (insert time period appropriate for the category)?"

5.1.4.3 "What brand have you used most often in the last 'x' months?"

5.1.4.4 If the frequency of use is one of the criteria for recruitment, then the respondent also may be asked how often they use the product or how many times they have purchased the product within a specific time frame. More questionnaire discussion can be found in 6.9.

5.1.5 Potential allergies or sensitivities to product ingredients should be assessed based on the product being tested and nature of exposure. Participants in product studies need to be screened by self-report for potential allergens, sensitizers, or lifestyle factors, or combinations thereof, that may exclude them from the test (for example, tree nuts, wheat, or special diets for religious or ethical reasons for food and beverage products, nickel, and some perfume components for non-food products).

5.2 Study Respondent Sampling Techniques:

5.2.1 The type of claim should be kept in mind when determining the number of respondents. For example, an equivalence claim may require more respondents than a superiority claim (see Section 14). Furthermore, the number of respondents required to represent the range of relevant respondent characteristics may be larger than the minimum that is required to achieve desired statistical power.

5.2.2 The characteristics of the respondent sample should be consistent with those of the target population (that is, about whom the claim is being made). The characteristics may include the population in terms of age, gender, product usage, culture, and language. For example, a constrained demographic sample should be employed when it is consistent with the stated claim and normal product usage (for example, a particular gender or age).

5.2.3 Using quotas on relevant respondent characteristics is important to achieve a match between sampled respondents and the intended target population. Relevant respondent information should be collected to demonstrate the appropriateness of the sample of the target population. Respondents chosen should reflect the target population. For example, representation of age and gender should reflect the target population.

5.2.4 Recruiting of the respondent sample must be stated in the test protocol and should be as objective as possible. Develop and document a rationale for both the number of respondents and for the inclusion and exclusion criteria. If the number of respondents is less or more than determined by statistical criteria alone, this should be included in the rationale. Consumer product testing respondents are usually obtained from stratified (not convenience) sampling procedures. Quotas are often used because they tend to produce a sample of respondents that is fit for the purpose of testing with the target user population of interest. When there are multiple testing

locations (see 5.2.7.2), then these quotas are applied within each of the locations. Probability sampling for consumer product testing is rarely practical because it would consider all consumers in the population of interest as potential participants in the test. In reality, many of these consumers cannot be located or are unwilling to participate.

5.2.4.1 Screening criteria should not be revealed to potential respondents during recruitment, and the standard security screening questions (for example, whether family members work in advertising or marketing or other related fields, including that of the test product) should be included.

5.2.5 *Sources for Selecting Respondents*—Potential respondents may be obtained from companies that provide testing services or sell marketing information. These databases are built using a variety of techniques, without targeting a particular manufacturer’s consumers. In some cases, a company may maintain its own branded database of product users. If using these databases, it is important to consider potential biases, for example, the repeated use of the same individuals, or the influence of the known sponsor of the research.

5.2.6 Respondents should be screened specifically for the claims substantiation test to ensure they currently meet the inclusion and exclusion criteria in the study design. To minimize over-testing with the same respondents, it is common to screen for past study participation, excluding those who have tested a specific product in the past three months, or tested in the product category in the past six months. Other approaches to minimize sampling bias may include conducting testing across a range of days of the week and times of the day, and varying the locations where potential respondents are recruited.

5.2.7 *Geographic Area Recruiting*—Some claims will require geographic considerations as a function of the target population and the nature of the claim. The study design and rationale may need to include variables that reflect the geography of the claim. Variation in product use may be due to geographic variability in demographics. Product performance may be impacted by environmental factors such as temperature, humidity, water hardness, pollution, or hours of sunshine. For example, use of certain sauces and spices may vary geographically, laundry detergent performance and preferences may be impacted by water hardness, hair product performance may be impacted by humidity, and suntan lotion use may vary by geographic region.

5.2.7.1 Some preference claims and usage claims may have a potential for geographic and demographic dependencies. For example, such a dependency exists if consumer preference or usage varies by region or by socioeconomic factors, such as urban versus suburban versus rural. The evidence for or against such dependencies could come from patterns in product sales, or usage, or both. These considerations need to be included in the respondent sampling rationale when they apply.

5.2.7.2 When geographic region is assumed to be a factor relevant to a claim, the geography of respondents should be consistent with the scope of the claim. A rationale should be developed when testing country-wide or regionally, ensuring a fair representative sample of the defined target population in multiple markets dispersed across the geography stated or implied in the claim. Various approaches for selection of test

locations are possible; they need to be included in the rationale. Some examples of how test sites could be selected follow.

(1) *Recruiting Example Based on Geography: National Claim*—An advertiser intends to make a national claim for a product in the United States. A consumer test is conducted via Central Location Testing (6.6.1) in four major regions (West, East, Central, and South) with two cities per region, with a sufficiently large number of consumers included from each of the cities. In general, it is best to recruit a large enough sample of consumers in each city to have the range of relevant characteristics represented in the consumer sample and to achieve the desired statistical power (see Section 14). If the consumer sampling requirement is met by having 40 consumers per city, this would yield a total sample of 320 consumers (40 consumers $\times$ 8 cities). If the consumer sampling requirement is met by having 60 consumers per city, this would yield a total sample of 480 consumers (60 consumers $\times$ 8 cities). Consumer sampling depends on both the range of relevant characteristics needed to be included in the consumer sample to achieve representativeness and the risk of challenge.

(2) *Recruiting Example Based on Geography: National Claim*—An advertiser intends to make a national claim for a product in the United States. A consumer home use test (HUT; 6.6.2) is conducted using zip codes to ensure that a sufficiently large number of consumers are included in multiple markets across geographies. In general, it is best to recruit a large enough sample of consumers to have the range of relevant characteristics represented in the consumer sample and to achieve adequate statistical power (see Section 14). To achieve the targeted number of consumers, it will likely be necessary to recruit some additional consumers. Consumer sampling depends on both the range of relevant characteristics needed to be used in the recruiting/screening of the consumer sample to achieve representativeness and the risk of challenge. Relevant characteristics refer to the criteria used in selecting respondents to serve in a sensory/consumer product test. For example, these criteria can pertain to factors such as demographic characteristics, product usage (“users of automatic dish detergents,” “users of bottled barbecue sauce”), purchasing habits or potential to purchase a product or product category, flavor likers, attitudes, or behaviors relevant to the need for a product (for example, frequent gym-goers desiring gym apparel that reduces odors; parents with young children who need healthy snacks in portable packaging).

(3) *Recruiting Example Based on Geography: Regional Claim*—An advertiser intends to make a regional claim in the western United States for a product that is distributed only in the western United States. Three cities in western U.S. states where the advertiser and competitor products have comparable market share are selected. The consumers sampled are users who are representative of consumers with respect to relevant factors (that is, demographic, socio-economic and product usage). A test is conducted via CLT with a sufficiently large number of consumers from each of these three cities. For example, a comparative taste claim pertaining to a mayonnaise product that is distributed only west of the Mississippi River would require mayonnaise consumers to be sampled from cities in states west of the Mississippi River.

(4) *Recruiting Example Based on Brand Development Index in Limited Geographies*—An advertiser intends to make a regional claim for a hand cream product that is designed to protect and soothe hands exposed to cold temperatures. Sales are concentrated in regions in northern U.S. states, mostly in northern Minnesota, upstate New York and northern Maine. Since there are three regions where the Category Development Index (CDI) is highest, a test is conducted via consumer CLT in three cities in these regions, one city per region. The cities are selected so that taken together each brand's Brand Development Index (BDI) is approximately proportional to that brand's BDI for the regions in which the claim will be made. The test includes a sufficiently large number of consumers from each city to represent target users of this category of products.

5.2.8 *Recruiting Example Based on Market Share*—If market share is a criterion, be cautious when selecting markets and ensure that the test adequately represents the people residing in the geographic territory on which the claim is based. In categories with strong geographic differences in market share, the total market share should be approximated by representing high, low, and average share markets in the study. The number of respondents per region may vary, reflecting their contribution in terms of population density, but not heaviness of usage.

5.2.9 *Recruiting Examples Based on Product Usage*—Respondent selection may be based on product usage or product category usage rather than geographical considerations. Examples follow.

5.2.9.1 Sensory testing is needed to support a claim that a product delivers relief from itchy eyes, the primary need for selecting a representative sample of respondents is to have users of the category independent of geography. A large sample of users of products in this category can serve as a representative consumer sample in this case.

5.2.9.2 A product claim pertains to hair coloring products. Only users of product in the hair coloring category need to be included in the respondent sample.

5.2.10 *Analysis Plan*—Claims substantiation is confirmatory testing, not exploratory testing. Once a target population is defined in the study protocol and represented adequately by sampling, results from the total respondent sample are used for substantiating a claim. Results among some subgroup may not correspond to overall results because the number of respondents in subgroups are smaller, and therefore may not be as statistically reliable. Moreover, since there is a risk of false positives and false negatives in testing any hypothesis, analysis of multiple subgroups will increase the overall error rate. Therefore, the analysis plan reflects the intended claim(s) and any subgroup testing should be determined before testing begins (for example, ensure an adequate number of respondents for any analysis and minimize the number of hypotheses to test). Any subgroup analyses that were not planned and properly designed may not be a sound analytical practice for claims substantiation (see Section 14). If the number of respondents who complete the test is lower than what was planned, researchers need to provide a rationale for the

shortage or a rationale for why the achieved number of respondents is sufficient to provide adequate data to support the claim.

5.2.11 Be prepared to defend the documented rationale for respondent sample criteria if challenged.

5.3 *Selection of Products:*

5.3.1 If a test is being conducted to support a competitive claim that is not brand-specific (for example, versus "other leading brands"), then the competitive brands should be the two brands with the highest national market share. If the market is highly fractionated, such that the top two national brands control less than 50 % of the market, then more competitors must be included in the test. Either the three leading national brands or any brand that is among the top two in the four major geographic regions of the country must be tested. Unless the product is tested against brands representing at least 85 % of the national market, it is recommended that claims should be made against specific brands in lieu of general superlative claims. Eighty-five percent (85 %) of the market is defined as all products within said category, including the brand making the claim.

5.3.2 Competitive brands should be in the same market segment as the brand for which the claim is being made. If a brand straddles market segments, then products most similar in a reasonable competitive context should be used.

5.3.3 When competing products are sold in more than one form, the products being tested must be of the same form or in the form most relevant to the claim. If a powdered drink mix is being compared with a competitor's product that also comes in a powdered drink mix and as a reconstituted liquid, both brands would have to be tested in their reconstituted from powdered forms. The specific directions for preparation given on each product must be followed. If there is substantial crossover use of different forms, a claim involving different forms may be desired. The forms tested must be stated explicitly as part of the claim, for example, "instant tastes as good as ready-made."

5.4 *Sampling of Products When Both Products are Currently on the Market:*

5.4.1 For central location consumer tests, commercial products to be used for competitive claims testing should be purchased at the end of the distribution chain to ensure the product is representative of the product the consumer would purchase. Some products are made at different or multiple manufacturing sites. In those instances, the product should be purchased from a distribution center that services the particular test areas.

5.4.2 For other test methods in which the test product is manufactured at one location, samples can be purchased from any high volume store. Products should be sourced at the same time from the same store(s) in each local testing area. Products should reflect the choice available to local consumers. Care should be taken to include a variety of production sites and dates that typically are found on the retail shelf.

5.4.3 In cases where competitive products are not sold in the same stores (for example, fast food restaurants and private label products) test products should be sourced as close in time as possible from locations that reflect choices available to local consumers. It is important that the geographic identity of

samples match that of local test participants. This way, if national products manufactured in more than one site have been formulated differently to appeal to regional differences in sensory preferences, appropriate products will be tested against relevant regional competitors. It is critical that all information regarding product sourcing be documented.

5.4.4 Competitive products should be purchased in the standard size package with the highest unit volume or in similar size, or both, to the test product. Trial size and club-store oversized product packages should not be used unless the package meets the specific target of the claim.

5.4.5 Every effort should be made to obtain competitive products of representative freshness found in the marketplace. All products in the test should be of typical age. A freshly-made product should not be compared against a product nearing its expiration date.

5.5 *Handling of Products When Both Products are Currently on the Market:*

5.5.1 After procurement but prior to testing, handling, length of storage, and storage conditions of all products must be identical and consistent with normal consumer practice.

5.5.2 Competitive samples must not show any signs of mishandling or abuse. If products become non-homogeneous during handling, in that they cannot be returned to their original state (precipitates may be returned to solution, but fractured pieces cannot be made whole), then test samples should be remedied for such defects. For example, the last serving or two from a box of cereal that may have a disproportionate share of fines should be discarded or screened.

5.5.3 To minimize the likelihood of product recognition by respondents, manufacturers sometimes try to “blind” the competitive product. Manipulations beyond labeling the original package should be approached with extreme caution. Repackaging of product would need to be supported by instrumental and sensory tests demonstrating no impact on the product. Any alteration of the product itself to minimize recognition could potentially impact acceptability and should be applied with the utmost discretion. It may be feasible to remove a product from its identifying package, but altering the structure of a product, such as grinding cereals to mask their shape, may change a product beyond the point where the competitive assessment is credible. When a product is instantly recognizable by its appearance, shape, or design, then cognitive factors due to brand recognition or previous experience with the product may contribute to the ratings obtained in the study.

5.6 *Sampling of Product Not Yet on the Market:*

5.6.1 If the manufacturer’s product is not yet on the market at the time of testing, the product should represent commercial production, and either be typical retail age of competitive products or expected age of the product when the cycle of the manufacturer’s distribution is observed. The competitive product should be selected to represent average retail age at the time of testing. If a suitable product is not available in the test city, the product should be sourced from a nearby location.

5.6.2 To ensure that the claimed benefit of the new product results from the product itself and not from special handling during limited scale production, it is desirable, but may not always be practical, for the new product to have been made at

the production facility. A new product, therefore, should be made at its intended manufacturing site, preferably on the same equipment and under normal operating conditions that will be used to manufacture the product. If pilot plant material must be used for claim support, then supplemental testing, for example, discrimination test for similarity, must be conducted to demonstrate that the claim benefits extend to material made at the production facility.

5.7 *Sample Preparation/Test Protocol:*

5.7.1 To minimize bias, it is essential that all samples for testing are prepared and served in a manner that will have limited impact on the perception of the products and in a manner that treats all of the products fairly.

5.7.2 For claims substantiation tests in particular, samples should be prepared and served under reasonably realistic conditions, that is, in a manner consistent with normal consumer practice. Samples should not be prepared in any fashion that would mask or alter various product characteristics.

5.7.3 All samples should be tested blind and with unbiased codes, such as three-digit codes. The respondents should have no leading or biasing information about the products that they are testing or about the overall objective of the study.

5.7.4 A decision must be made regarding the manner in which the samples will be presented to the respondents. For example, the samples can be served as pairs or one at a time (monadic presentation). Differences among samples are more likely to be detected when two or more samples are presented together; however, monadic presentation generally is considered more representative of the consumer experience.

5.7.5 The order of sample presentation must also be considered prior to testing and this must be designated according to a statistical design. Various psychological factors can influence judgment, for example, the impact for which the following order effects must be accounted:

5.7.5.1 *Context/Contrast Effect*—As defined in Terminology **E253**.

5.7.5.2 *Positional Bias*—Respondents may be more sensitive to differences in specific samples in a series, such as the first or last sample.

5.7.5.3 *Pattern Effect*—Any pattern in order will be detected quickly.

5.7.5.4 *Ceiling Effects*—This typically occurs when the majority of the scores occur towards the top of a rating scale. When the products are well-liked, there is not a sufficient amount of scale available to the respondents to differentiate the products. Variation in rating scores is compressed, making mean-based statistical tests misleading. Therefore, analysis should be performed using a more robust statistical model that does not have distributional requirements and is less prone to outlier influence such as multinomial logistic regression.

5.7.6 It is essential to balance the order of presentation to distribute these effects across all products.

5.7.7 The test and questionnaire should be designed to be free of all forms of bias. Bias during testing may come from the samples, the test protocol, including the questionnaire, or the test environment, or a combination thereof. Other sections of this guide discuss these issues.

6. Test Design—Consumer Testing

6.1 Monadic designs are those in which a single product is rated by respondents at a time.

6.1.1 Sequential monadic designs require each respondent to evaluate products one at a time and in consecutive order.

6.1.2 Protomonadic tests consist of providing one product, obtaining ratings of that product on a variety of attributes, removing the first product, and replacing with a second product. No monadic ratings are obtained on the second product; instead, a paired-comparison test is conducted.

6.2 Comparative test designs are those in which two or more products are presented to the same respondents to compare the products to each other.

6.3 Comparative claims imply, but are not limited to, comparative designs, where each respondent evaluates two or more products. For comparative claims, paired comparisons are used most frequently. Simultaneous presentation provides the most direct comparison of the products. In some situations, sequential presentation may be needed that introduces execution and sensitivity issues, so there should be a rationale for choosing a sequential (monadic) presentation.

6.3.1 In cases where there are multiple products to be compared, the respondents may be able to evaluate all of the products (complete balanced block design) or a subset of products (an incomplete block design) or only a single product (monadic design). When the products are evaluated in subsets, overlapping product blocks may be constructed using techniques such as balanced incomplete blocks and partially balanced incomplete blocks. These incomplete block designs may require specialized analysis procedures to construct the correct averages, as outlined in Cochran and Cox (2) and other statistical references.

6.4 Since monadic testing is not the most direct method for making comparisons, it is not always the most desirable approach. Nevertheless, sometimes it may be the only practical method to support comparative claims. For example, some products may require long periods of repeated usage to provide a consumer benefit, which can undermine the ability to make direct comparisons. In this case, product performance can be assessed by giving each product to a different group of consumers and conducting statistical analysis on the ratings. In monadic designs, respondents, as well as products, contribute to the total variation, rendering it less sensitive and larger differences or larger sample sizes are required for significance. It is critical that the groups be matched adequately.

6.5 Non-comparative claims may be supportable by either monadic or sequential-monadic test designs. While a monadic rating may provide a measure free from influences inherent in multi-product, sequential-monadic designs, either approach is sufficient to meet the “reasonable basis” required to make a claim.

6.5.1 Qualitative research, such as focus groups, is not acceptable for claims support since one cannot project their findings to a larger population of consumers.

6.5.2 Both central location (CLT) and home use (HUT) test methods can be acceptable, depending on the specifics of the category and usage. CLTs include all locations other than

respondents’ homes. These locations may include sensory facilities, mall facilities, field sites, supplier’s premises, community centers, or others. Each type of location has some benefits and limitations that must be taken into consideration when projecting results.

6.6 Data Collection Strategies:

6.6.1 *Central Location Testing (CLT)*—This method of testing provides maximum control over product preparation and usage. Central location testing assures that the participant actually evaluated the product in question and provides his or her own opinion immediately following evaluation, rather than relying on past usage or recollection. Blind testing often precludes the need to repackage product. In addition, CLTs can provide direct product comparisons, isolate specific attributes, such as color or crunchiness, vanilla flavor, and so forth, and accommodate complex evaluation protocols. They are appropriate for superiority, equivalence, or unsurpassed claims.

6.6.1.1 Key limitations are that central location tests usually involve a single product exposure with small amounts of product under conditions that may not closely duplicate typical usage. Questions about whether such exposure can exaggerate trivial differences or whether CLTs provide a basis for forming a preference have been raised. Other limitations that can be controlled are potential for respondents to overhear one another, and testing at times of day that are inappropriate for the product, for example, breakfast cereal in the evening. Where these issues outweigh the limitations inherent to central location testing, home use testing can be considered.

6.6.1.2 Respondents can be intercepted from a public area if they meet the screening criteria or they may be pre-recruited and scheduled for testing (useful when testing is targeted to a specific time of day or where incidence is low). Tests that require special equipment have limited shelf life, or shortened project schedules may not be feasible in mall or intercept type facilities, and are better handled with pre-recruiting.

6.6.2 *Home Use Testing*—The term “home use test” (HUT) refers to tests that allow respondents to use the products in a more natural environment, rather than the controlled environment of a CLT. Since there is still experimenter intervention (product placement and questionnaires) the HUT is not a truly normal use environment; but, it comes closer to how consumers actually use and evaluate products. These HUTs allow for product use that is more typical of normal use conditions, as respondents typically use the products where, when, and how they normally would. HUTs are particularly useful when an overall evaluation of the product cannot be realistically conducted in a CLT environment, or when a feature or benefit must be experienced under normal usage conditions.

6.6.2.1 The choice to use a HUT rather than a CLT to substantiate or evaluate a claim should be determined by the nature of the claim, the amount, type, and length of usage of the product. A very narrow claim about a particular flavor of a pre-made product could very well be evaluated in a CLT, while in an overall claim for suitability would usually require more extended use in the home environment, as with an air freshener.

6.6.2.2 Even if the product as a whole may require an HUT, certain visual, tactile, aural, or olfactory properties of the product may be evaluated in a CLT when the objective is to

evaluate salient, non-use characteristics of the products. As an example, respondents could evaluate the look or hand and skin feel, or a combination thereof, of products such as toilet paper and other toiletries, feminine care products, or the aroma of a product. If a claim is being made concerning the context or setting of the actual use, or both, it would still need to be proven on a case-by-case basis that testing a given product outside of the home use environment does not artificially influence consumer behavior or perception.

6.6.2.3 When deciding between a CLT or HUT, one needs to consider the issue of realistic product performance and the ability to generalize the study results to the population that is being targeted by the claim. Certain product categories, such as moisturizing creams, lotions, and acne preparations may require usage over an extended period of time for respondents to evaluate product performance realistically. In such instances, HUT may be the most feasible method for providing realistic performance and evaluations that can be generalized to the population that is being targeted by the claim.

6.6.2.4 A key difference between a CLT and a HUT is the limited experimental control in the HUT. As the HUT provides a more realistic use and evaluation environment, the experimenter has less control over product preparation or use, and must rely on the respondent's ability to recall features of the product use. As in normal use, this recall may be influenced by comments received from family and friends, and a respondent's overall impression of a product may influence his/her recall of particular attributes (for example, halo effect). Often, the usage experience will require sequential product placement and usage in a sequential design. The products may be compared at the end (as in a paired comparison or ranking design) or evaluated after each use (as in sequential monadic design or blocked design). These sequential designs may not be appropriate when the product substantially changes the test environment, so the environment in which a second or later product is applied would not be comparable to the first. Example products could include drain cleaners, mold or mildew removers, or shoe polish. This effect may require that respondents use only a single product (for example, a "monadic" or unblocked ANOVA).

6.6.2.5 Certain test conditions may compensate for some of the issues mentioned in 6.6.2.4 by using simultaneous or split-sample designs. An example could be the case where a respondent cleans one-half of a surface with one product and the other half with a second product. Other cases might include shampoo on different sides of the head or skin cleansers or treatments on different sides of the body or face. Care must be taken so split-sample use is counterbalanced across respondents to avoid potential limitations due to handedness and other biases.

6.7 Interviewing Techniques:

6.7.1 Self-administered:

6.7.1.1 Questionnaires independently completed by the respondent are referred to as self-administered. Responses can be collected on paper questionnaires, from on-site computerized questionnaires, or from questionnaires administered over the Internet. Paper copies have the advantage of keeping the original data in its real state for an indefinite period of time.

Paper copies of questionnaires can be re-examined as needed if any questions about the data arise. Automated data collection and Internet questionnaires have the advantage of being a direct record of consumer rating, uninfluenced by any possible human bias. The biggest risk in data collection is in the home use environment due to the lack of control over who answers the questionnaire, and therefore, whom the information actually represents, whether collected from paper or automated questionnaires.

6.7.1.2 Self-administered questionnaires can be used in both CLTs and in HUTs. Trained panelists exclusively use self-administered questionnaires.

6.7.1.3 A self-administered questionnaire must be understandable by the respondents with minimal to no verbal instructions by a test administrator. The questionnaire is simple and structured in a logical and unbiased manner. When the questionnaires do not meet these criteria, one-on-one interviewing may be required.

6.7.2 One-on-one Interviewing Techniques:

6.7.2.1 One-on-one interviews involve eliciting answers or opinions, or both, from a single respondent through an interviewer, either face-to-face or by means of telephone.

6.7.2.2 Interviewer training and instruction with practice ensure consistent and flawless execution by all interviewers at all test sites. Instructions include spelling out all actions and their contingencies so that no decisions need to be made by the field agency or the interviewer. Interviewers are thoroughly briefed and practiced before beginning data collection. It is strongly recommended that instructions be tested.

6.7.2.3 Interviewers record respondent responses to questions after they are exposed to a stimulus. The stimulus could be asking a question or testing a product.

6.7.2.4 Interviewer bias can be a major concern and a potential disadvantage with this technique. Double blind testing, where neither the respondent nor the interviewer knows the identity of the sponsor or the products, is imperative. Interviewer bias can be further minimized by using unique code numbers for test products to better mask their identity and make trends more difficult for interviewers to discern.

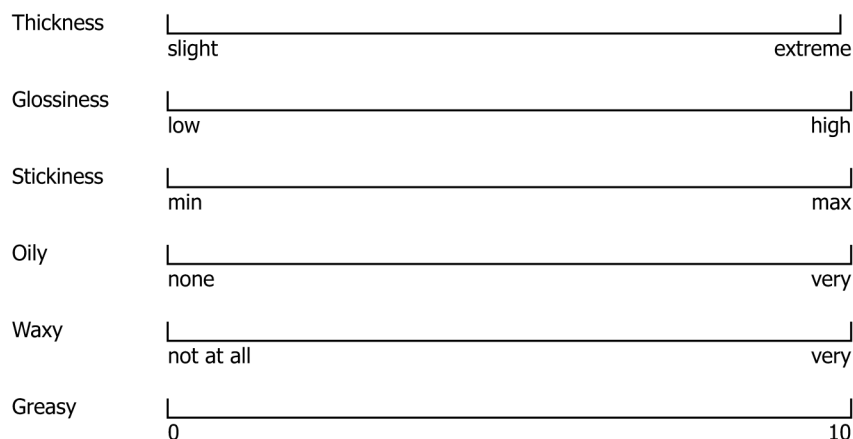
6.7.2.5 If the questionnaire has several questions, the one-on-one interviewing format is preferred over self-administered questionnaires, since interviewing will prevent respondents from reading ahead or going back, which may influence their answers to other questions.

6.7.2.6 When a claim substantiation study questionnaire involves skipping questions based on the answers to previous questions, referred to as skip patterns, the one-on-one format is recommended over the self-administered format, unless computerized interviewing software is used to ensure correct skips.

6.7.3 Telephone:

6.7.3.1 Use of the telephone for claim substantiation support usually will be limited to studies where respondents are not immediately reacting to a stimulus, as they would in a taste, visual, or tactile evaluation, but rather voicing their opinion of a product's performance during actual use or over an extended period of time.

6.7.3.2 Responses can be collected over the telephone from a self-administered questionnaire completed during product



NOTE 1—These scales are for example only. Disparate scales, that is, some using numbers and some words, are not recommended. Consistent scale style is the norm.

FIG. 1 Examples of Intensity Scales

usage, or interviewers can ask questions based on respondents' recall of their product experience.

6.8 Type of Questions:

6.8.1 *Rankings*—When respondents can compare three or more products, a simple way to establish superiority, equivalence, or unsurpassed within that set of products is through the use of ranking. In this case, a respondent is presented with a set of products simultaneously and asked to rank the products for preference or other attributes (see section on data analysis). In cases where the advertiser is comparing more products than can be accurately ranked, blocking designs, such as balanced and partially balanced incomplete block designs may be used, according to Meilgaard et al. (3). Consult a statistician for assistance in constructing these designs.

6.8.2 *Preference*—The choice among two or more alternative products is the most direct way to establish superiority, equivalence, or unsurpassed, given adequate sample size.

6.8.3 *Acceptance*—The nine-point hedonic scale traditionally is used for sensory acceptance measurements because it is reliable, valid, and of practical value. In addition to measuring degree of liking of a single product or multiple products evaluated sequentially, it measures degree of differences in acceptance and direction of liking. A large enough difference in mean ratings on an acceptance scale might lead to the researcher making an inference about preference. The hedonic acceptance scale can be used with a wide variety of products and with minimal respondent instruction. Absolute levels of liking can change over time and between groups, but scalar differences between products are reproducible with different groups of respondents. Resulting data lends itself to powerful parametric statistics. Other structured, semi-structured, and numerical scales can be used effectively for acceptance testing. When using other scales, care should be taken that the distributions are relatively normal so parametric statistics can be used. If not, nonparametric statistics should be applied.

6.8.4 *Attribute/Diagnostic*—There are four types of attribute/diagnostic questions in general use: (1) hedonic, (2) preference, (3) just right, and (4) intensity.

6.8.4.1 Hedonic scales measure the degree of liking of the level of an individual attribute in a product (for example, measuring the degree of liking of the level of fragrance of a product).

6.8.4.2 Attribute preference scales present questions about individual product attributes, such as the fit of a pair of jeans and the preference between the fit of two products.

6.8.4.3 Just right scales measure the appropriateness of the individual attribute level, for example, too sweet, just right or not sweet enough.

6.8.4.4 Intensity scales measure the strength of an individual attribute, for example, no sweetness to extremely sweet, and questions measuring which product has more or less of a specific attribute(s). In consumer hedonic testing, the researcher must have information that demonstrates that consumers truly understand the meaning of the sensory attribute. For example, consumers may confuse “sourness” and “bitterness” or interpret “creaminess” to mean creamy flavor, creamy texture, or both. See Figs. 1 and 2 for examples of intensity scales.

6.8.4.5 It would be inappropriate to use “just right” scales to support an intensity claim for a specific product attribute. Intensity claims must be validated by using intensity scales, where “0” is the anchor for none of the attribute and a higher number such as “9,” “11,” “15,” or “100” is the anchor for an extreme amount of the attribute. For example, the claim “more butter flavor than Brand X,” shall only be supported by a significant difference in butter flavor using an appropriate scale for the intensity of butter flavor.

6.9 Questionnaire Design:

6.9.1 Format:

6.9.1.1 Once the type of response (for example, acceptance, preference, diagnostics, and specific attributes) and attribute terms have been selected, attention should be given to the questionnaire format.

6.9.1.2 Although there is not one perfect questionnaire format, the format of the questionnaire is determined by several considerations.

"Our sunscreen leaves less sticky residue on your skin."

1) Paired comparison method

In front of you are two samples. Please rub the first sample on your left arm, followed by rubbing the second sample on your right arm. You may reuse on a different portion of your skin as often as you wish.

Which samples leaves less sticky residue on your skin? [] 212 [] 791

Attribute rating

Please use sample # _____ and rate it on the scale below.

Amount of sticky residue
None at all large amount
0 10

Clean your hands with alcohol and cotton ball.

Now Please use sample # _____ and rate it on the scale below.

Amount of sticky residue
None at all large amount
0 10

FIG. 2 Example of Paired Comparison With Intensity Scales

6.9.1.3 In general, a well-designed questionnaire has the following characteristics:

- (1) Includes key components (questions) relevant to the claim,
- (2) Excludes questions not needed to support the claim (this will preclude any potential biasing effect of any one question on any other question),
- (3) Provides sufficient explanations and clarity to the consumer on its use,
- (4) Looks organized and professional,
- (5) Is easy to decode, and
- (6) Is appropriate to its interviewing method (self- or interviewer-administered).

6.9.1.4 It is recommended that the final questionnaire be tested prior to its use in the claims test. If consumers do not understand a required task or do not comprehend a given attribute, the questionnaire can be modified prior to the quantitative test. Optimally, a small group of consumers (10 to 20) should be used for this purpose; however, company employees not related to the project and untrained in sensory testing can be asked to participate in the assessment of the questionnaire, but not to participate in the study.

6.9.2 *Components*—Generally, there are four major components in a consumer questionnaire:

- 6.9.2.1 Instructions for respondents or interviewers, or both (if using interviewer-administered questionnaires),
- 6.9.2.2 General/overall questions,
- 6.9.2.3 Specific attribute questions, and
- 6.9.2.4 Classification or demographic questions.

6.10 *Instructions to Respondents*—If the questionnaire is self-administered and no orientation or verbal instructions are given to respondents, the written instructions should be

complete, concise, and clear. If the questionnaire is interviewer-administered or an orientation is given, or both, the written instructions only need to be a summary of the evaluation process and directions. Because many consumers do not take enough time to read and understand directions carefully, an orientation, together with brief written instructions, is the recommended procedure. In general, written instructions should be located at the beginning of the questionnaire and include the following items:

6.10.1 The type of product and number of products to be evaluated,

6.10.2 The task manipulation or procedure to be followed by consumers (for example, bite, chew, rub, compress, wipe, smell, apply),

6.10.3 Special directions in handling or using or removing product, or a combination thereof, if required,

6.10.4 An indication of the overall flow or components of the questionnaire,

6.10.5 Examples of the rating technique or questionnaire usage, and

6.10.6 Instructions as to what consumers should do after completion of a sample evaluation and the whole test.

6.11 *Instructions to Interviewers*—These instructions must be clear enough to ensure consistent and flawless execution by all interviewers in all test sites. Adequate instructions spell out all actions and their contingencies so that no decisions need to be made by the field agency or the interviewer. It is strongly recommended that instructions be tested, and that interviewers are thoroughly briefed and practiced before beginning data collection. Interviewers should read from a prepared script to ensure consistency across interviewers, test sites, and sessions.

6.12 *General/Overall Questions*—Under this category, there are the questions that address general or overall impression. Usually, these questions are the most important questions in the test and need to come first. In general, ask only the key questions related to the claim to minimize potential bias in asking additional questions irrelevant to the claim. Examples of general/overall questions include:

- 6.12.1 Overall acceptance or liking,
- 6.12.2 Acceptance/liking of broad sensory dimensions, and
- 6.12.3 Overall preference.

6.13 *Positioning of Key Product Rating Question*—Product tests almost always have an overall question, such as overall liking, acceptance, ranking, or preference. Placement in the questionnaire for this overall measure is very important in a claim test.

6.13.1 In tests where only overall acceptance/liking or preference is asked, these questions come first by default. Asking multiple overall questions runs the risk of obtaining conflicting results; however, in a more complex questionnaire, for example, with attributes, the position of these questions has to be decided.

6.13.2 In general, questions asked first are assumed to be more free of influences or biases that may affect questions appearing later. The extent to which ratings truly represent product performance is critical if a claim is challenged. When claims are challenged, methodologies are scrutinized, question order and flow are reviewed, and a judgment is made about the extent to which the overall liking/acceptance/ranking/preference rating is free from other-item influences or biases. Questions appearing first will stand up to such scrutiny. In a claims test, more confidence will be placed in data obtained from first-asked questions.

6.13.3 *Recommendation Regarding Where to Position Questions:*

6.13.3.1 *Monadic or Single Product Tests*—Product tests where only one product is experienced and rated.

(1) One question presented at a time by paper, computer, or interviewer. The key question pertaining to the claim should be positioned first. It will be free of influences of other questions and most defensible under scrutiny.

(2) *Multiple Questions (Self-administered)*—When the questionnaire allows all the items to be read or reviewed, the key question should be placed in the most logically appropriate position. It should appear first if what is needed is the consumer overall and immediate hedonic reaction without consideration of attributes.

6.13.3.2 The key claims question could also be presented at the end of the set, an example would be if all attributes need to be judged or the product used in a specific fashion prior to making a decision. Examples are a personal care product such as shampoo, or a household product such as dish detergent. Other questions can influence individual items since the respondent can read and review the self-administered questionnaire at will.

6.13.3.3 *Multi-product Tests*—When more than one sample is to be evaluated by a respondent in a monadic sequential presentation, after the first product is evaluated, subsequent ratings will be affected by earlier products seen and the

attributes that have been rated. Products must be sequenced (balanced for order of presentation or randomized presentation) to minimize effects of sensory adaptation, fatigue, and contextual effects. The effects of the attributes can only be overcome by having the liking or acceptance question at the end of the questionnaire so that the influence of the attribute ratings affects all products equally. In any multi-product test, placement of the key question must be consistent from product to product.

6.13.3.4 *Two-sample Comparative Tests*—These tests, where preference or ranking data are obtained, are special cases of multi-product tests. Comparative questions that are to serve as the key data to support a claim should appear first. Therefore, these measures will be free of the influence of other attribute questions that may be asked, and thus will be able to withstand scrutiny.

6.14 *Total Text Context and Presentation Matters*—When designing claims research, the number of products, evaluation methods, and questionnaire development should be considered. Some formats allow only one item to be presented at a time as in interviewer- or computer-administered questionnaires. Other formats allow all questions to be reviewed or considered as in a self-administered paper questionnaire.

6.14.1 Single product studies yield evaluations free of influences from other products. In multiple product tests, the first product experienced and the first question answered is the only rating free of influence and potential bias. Presentation and sampling of all the products in a pretest warm-up session can mitigate some of the position, order, and carryover effects in a multi-product test. The position of a key rating question among many is more important when a single question is presented at a time in a preplanned order. In self-administered paper questionnaires, item order matters less, since all questions are available for review at any time and potentially can influence all other items.

6.15 *Specific Attribute Questions*—If claims are to be based on the attributes, direct questions can be asked. It is important that they be asked alone or positioned first in the questionnaire to avoid potential bias. Attribute questions are of three types and include the following:

- 6.15.1 Attribute hedonic/liking questions,
- 6.15.2 Attribute intensity or attribute diagnostic questions, and
- 6.15.3 Attribute preference.

6.15.4 The attribute hedonic/liking questions collect liking information on specific attributes, for example, liking of the herb combination, sweetness level, absorbency, comfort, or hair shine.

6.15.5 The attribute diagnostic questions collect information on the perceived intensity/level of that attribute, for example, intensity/level of fruitiness, saltiness, and oiliness/warmness. Attribute diagnostic questions are asked using either an absolute intensity scale, for example, none to extreme or a just-about-right scale, for example, too low/just about right/too high. The latter is not very useful for claims support, and deviations from 100 % “just right” are likely to be highlighted by challengers. If the claim has to do with a specific amount of an attribute, then an intensity scale should be used.

6.15.6 Attribute preferences can be determined by questions, such as, “which do you prefer for (state attribute of interest)?”

6.15.7 These attribute questions are used either alone or in combination. When more than one is asked, for example, liking and intensity, the same attribute term should be used. The selection of these terms is critical. However, asking about an attribute in more than one way increases the risk of results that could be viewed as inconsistent, for example, a difference in preference without a difference in liking.

6.15.8 The format used for the attribute questions should allow consumers to properly understand and respond to these questions. To achieve this goal, some considerations include the following:

6.15.8.1 The same type of scale should be used throughout the questionnaire, for example, a nine-point hedonic scale for all attribute liking questions.

6.15.8.2 The same anchors and positioning of the anchors in the hedonic scales should be used.

6.15.8.3 The anchors for the diagnostic questions should be placed in the same positions for all questions.

6.15.8.4 If both attribute liking and diagnostic questions are used, the format and position of both questions should be kept constant all through the questionnaire, for example, both questions for the same attribute positioned side-by-side throughout the questionnaire, or attribute liking question followed by the attribute intensity question throughout the questionnaire.

6.15.9 *Selection of Scale*—The two types of measurement data that can be obtained for attributes are rating and ranking. The selection of a scale is made based on the advantages and disadvantages of each, the ease of its use by consumers, and the type of data to be collected.

6.16 *Classification or Demographic Questions*—These questions are critical to demonstrating congruence between the target population and the target sample. Standard questions include age, gender, ethnicity, income range, frequency/heaviness of use, use of related product formats, for example, homemade versus ready to eat, and brand used most often. Within the questionnaire, questions involving specific brands or product formats must come after product evaluation or there is risk that responses to these questions can impact respondents’ behaviors. For example, after a respondent commits to a favorite brand, they may look for and choose that product in a preference test.

6.17 *Preference Questions*—A procedure for asking preference questions is not easily chosen. It generally is accepted that the most effective way to ask the preference question is to ask the respondents which of the products tested they preferred without any reference to the degree of preference that the respondent may have had. The question of offering a no-preference choice is subject to various opinions. This area has been discussed for years and likely will continue to be the subject of discussion in the future. Currently, some television networks, and some courts have taken the position that respondents should be given the opportunity to respond directly to an asked “no preference” alternative in the questionnaire. While this approach is generally accepted, it is not

without its shortcomings. It is possible that respondents offered a “no preference” choice will choose that option as a way to avoid making a choice but it is also possible that respondents equally prefer both products.

6.17.1 A “no preference” option should be included, because respondents may not always have a product preference. If there are a high number of no preferences for the product category or attribute, making a preference claim is risky; a statistical risk assessment should be conducted.

6.17.2 It is important that users of this guide remember that the above recommended method is one of many approaches currently suggested and opponents may question the validity of a claim based on the above procedure, because they may have conducted testing using a different approach. Also, it is possible that within a given section of industry, there may be a consensus on a particular test format and that preference would be given to that test design over others.

7. Test Location

7.1 When central location consumer tests are conducted in mall facilities, particularly for intercept recruitment, or at the premises of the research supplier or interviewing service (for pre-recruited respondents), a third party location, such as a hotel, may be used. The venue should not have signs or other cues that indicate the sponsor of the test. Testing conducted at the manufacturer’s facilities is never acceptable for claims substantiation.

7.2 When the geographic region is suspected to be a factor relevant to a claim, national or regional claims tests should be conducted across a number of geographically dispersed locations. Local claims should sample more than a single site (see 5.2).

7.3 Test facilities must be staffed by an experienced and professional interviewing organization. To avoid bias and achieve double blind testing, the people who prepare the test products should not conduct interviews for any part of that study, unless products are blinded well enough that brand identities cannot be determined (for example, completely repackaged products as opposed to overwrapped). Field supervisors must not identify the test sponsors to any staff involved with the test, and preparers must not discuss the identity of the test products with the interviewers.

7.4 Preparation activities must not impact the interviewing process. The preparation areas must operate quietly to avoid distracting the respondents and interviewers. Ventilation should be adequate to prevent odors from the preparation area to be detectable in the interviewing area, for example, if a personal care item has a fragrance, or a food item is accidentally burnt. In addition, ventilation systems should provide adequate turnover of air between samples, as well as between respondents to minimize inappropriate carryover. The preparation area must not be visible to respondents. With the exception of tobacco testing, smoking should be forbidden in the interviewing area.

7.5 The testing area should have separate interviewing stations that are sufficiently isolated to avoid voice or visual influence of ongoing interviews on each other.

7.6 Testing often requires refrigeration capacity or cooking facilities and other accoutrements found in most households. Lighting must be adequate to allow the full visual impact of the test products, unless the test calls for intentional masking of appearance.

7.7 Adequate electrical outlets will be needed to test the product. Water supply is necessary for most laundry, cleaning, and food or beverage preparation, skin testing, or personal care product usage.

7.8 The ability to provide good traffic flow is often overlooked. Rooms with a separate entrance and exit may help.

7.9 Each test has different facility requirements and the agency needs to know the specific requirements for the proposed test.

8. Test Execution by way of Test Agencies—Food and Non-food Testing

8.1 Each test is unique in its requirements and execution. Thorough preparation for a study includes clear strategies for collecting data and benchmarks for planning, preparing, and completing the study. Supplying the product to the agency on the agreed upon date makes the study run smoothly. A well-run study maximizes the potential for accurate and functional study results. Meeting personally with an agency representative to discuss processes, study requirements, and reporting criteria makes it possible for the agency to serve the needs of the client effectively.

8.2 *Protocol Documentation*—Protocol documentation is supplied to all contracted agencies as early as possible prior to testing. Two weeks is recommended when possible. It is a detail of the study content, including procedural requirements such as screening requirements, storage of test product, supplies, and any other specific study constraints. Fundamental to all Protocol Documentation is communicating all procedural specifics and in as much detail as possible.

8.3 *Time Constraints*—Time constraints include the planned or expected test date, the length of time required for each respondent to complete the test, the number of days required for the test, and the time of day. Time constraints help the agency determine the test location (or locations), the personnel required to execute the study, the amount of time to complete the study, and to negotiate the date the client should expect to receive the study results.

8.4 *Test Design*—Test design provides specific documentation to enable the agency to complete the test effectively. Proprietary details may or may not be necessary, but a confidentiality document is highly recommended. The agency needs to know the number of respondents required, the number of products to be tested, expanded directions, including temperature of product, application, or function (how the product(s) is to be employed, for example, applied to skin, eaten, or dispensed), and methods desired in documenting study results. Test choices may include paired comparisons, sequential monadic designs, or one of many other multiple product designs. Randomization of the products in the study to minimize position bias is required.

8.5 *Respondent Recruiting/Screening*—The agency will need instructions regarding the respondent's demographic information that may include, amount other things, age range, income, gender, ethnicity, category and brand usage, usage incidence, family size, and regional habitat. Food allergy status should also be documented when appropriate. Other instructions may be necessary for targeted claims. Test timing is factored in to determine respondent availability to meet and complete the test requirements.

8.5.1 The agency is responsible for confirming that the respondents understand and accept their responsibilities before they participate in the study. Informed consent and confidentiality agreements are signed and retained as part of the study.

8.5.2 Criteria for qualifying and terminating (non-qualifying) respondents is the responsibility of the manufacturer or client of the agency. The agency is responsible for maintaining records of qualifying and non-qualifying (terminated) respondents that include clear reasons respondents were terminated.

8.6 *Personnel Requirements*—The agency is responsible for having sufficient personnel at the test site to administer the test. Comprehensive instructions detailing the various roles individuals may have to perform to execute the study successfully are the client's responsibility. On-site product preparation, special handling, serving, storage, and other variables could require additional agency personnel.

8.6.1 Most claims supports require a double blind format if preparation is a part of the product presentation. It is desirable that the preparer and the interviewer be separate individuals to minimize product knowledge.

8.6.2 Complicated questionnaires may require additional personnel to conduct and additional staff to supervise the interviewers. Some sensitive products such as products specific to male or female consumers may require additional training of the interviewers or extra supervision, or both. The manufacturer, client, or agency should role-play the execution or the questionnaire several times to determine a reasonable estimate of the time required per respondent. This enables the testing agency to assign sufficient personnel and to assure the test is conducted appropriately and the respondent does not feel rushed. Self-administered interviews require role play as well, since determining a reasonable amount of the time to complete the questionnaire has to be considered in order for the agency to plan facility space and personnel.

8.7 *Product Requirements*—Consumer tests commonly require product shipment before the test date. The agency needs to know when product is expected to be shipped, expected arrival date, storage requirements (ambient, air-conditioned, refrigerated, or frozen), the length of time the product needs to be under the prescribed storage conditions, and how the product must be handled once the agency receives the product, plus any special instructions. Assurances of product safety, such as microbial and allergen statements, should be supplied, if appropriate.

8.7.1 Advance planning by the agency is required if there are special instructions in product handling. For example, if a product must be shipped frozen, thawed, and then prepared for a study, the agency needs to schedule its personnel in order to

follow their clients' directions. Also, if a product requires assembly that requires a specific skill set, the client must include the time expected for the assembly, provide assembly instructions to the agency, and any special skills required for assembly. Products requiring preparation may also require specialized equipment. The amount of equipment, size, cleaning instructions, temperature, lighting, noise level, and other critical factors are important to communicate to the agency as well. The instructions should clearly indicate how the product is to be presented or displayed, or both, and how the product will be served, as well as portion sizes and other controls that may be necessary.

8.7.2 After the respondent has finished with the product, the agency needs complete disposal instructions to protect the clients' proprietary information. Specify whether the product is secured and must be returned, whether the product can be reused, or how it must be discarded.

8.8 *Facility Requirements*—Facility requirements reflect product handling requirements. Product preparation, length of time each panelist needs, type of interview, and various other factors determine facility requirements. Consideration for facilities for nonfood items includes fragrance testing booths and accessibility to home type appliances or rooms (for example, washing machines, stovetops, and toilets). The client is responsible for selecting an agency that meet all the requirements of the test and communicating those requirements adequately to the agency.

8.9 *Interviewer Scripts*—Interviewer scripts maintain consistency in data collection. Deviation from the script could impact study results; therefore, the script must be followed verbatim—no additions, subtractions, or tonal variation.

8.9.1 Some agencies offer scriptwriting services and can work with the client to develop suitable scripts, as well as train agency personnel to execute the interview as stipulated. Whether the agency personnel execute the interview or the client brings in an interviewer, rehearsing the interview questions can protect the integrity of the study.

8.10 *Questionnaires*—Questionnaires can be read by an interviewer or self-administered by the respondent either in paper form or by using a computer. Each method may require a different amount of time to complete, so the client must clarify with the agency which method is to be used and a reasonable estimate of time each respondent will need to complete the questionnaire should be determined.

8.10.1 Written instructions on each questionnaire must be consistent so each respondent receives the same stimulus. Slight nuances in instructions can influence the respondents' perspective of the study or of the product, creating another (albeit unintended) variable that can impact study results.

8.10.2 All special instructions must be on each questionnaire, such as the type of writing utensil (for example, No. 2 pencil if questionnaires will be electronically scanned). Techniques such as applying creams, lotions, or cosmetics, wait times if delayed responses are recorded, and any other special instruction should be incorporated into the questionnaire.

8.10.3 Should a computer be employed for the study questionnaire, then data management, formatting, and the

method(s) of transferring data must be incorporated into the study documentation provided to the agency.

8.11 *Data Recording and Verification*—Data recording and verification instructions should be unambiguous to ensure the required data is gathered and retained. The client must communicate to the agency whether the answers to study questions and products tested must be linked to each respondent so the agency can incorporate identification methods. If voluntary statements are solicited, the method of recording those statements must be predetermined. Safeguards must be in place to ensure the agency is not creating data or padding data by interviewing people who did not participate in the study.

8.11.1 Third party observers can be used to verify links between products, questionnaires, and respondents. For claim support, it is wise to incorporate third party observers, even if the observers are only required to confirm all questions have been answered on the questionnaire or just to observe the study. Validating a minimum of 10 % of the cases by phone is standard practice in the industry. If there are anomalies in these 10 %, then up to 100 % of the data should be validated.

8.12 *Data Submission*—Data submission guidelines are critical for the client to provide the agency. Examples of guidelines needed are: who receives the data, when the data must be submitted, whether interim reports are required, and how to format the data for submission (for example, web-based, written hard copy, email, and so forth). The database used to collect data must be compatible with the data analysis system, and for claim support, the original questionnaires need to be returned to the client as final verification.

9. Records in Sensory Claims Substantiation Research

9.1 *Reasons for Retaining Documents*—Documenting (3.2.7) is essential for robust scientific work. Documenting should be done contemporaneously, that is, at the time the work is done. Records (3.2.12) should be specific and comprehensive enough to allow an independent person to replicate the study. Records should allow an internal or external reviewer (for example, legal reviewers or other stakeholders, including government agencies) to evaluate the merits of the research design and relevant activities performed. If records pertain to a sensory claim, then records must include the claim, study plans, execution, and a test result providing clear evidence substantiating the claim.

9.2 *How to Retain Records*—Consult your legal counsel or company's record retention policy for additional guidance, for example, how long to retain records, where to retain the records, and who will retain the records. Some companies use version control or memos-to-file to capture revisions to records. Records may need to be retained by the company making the claim, the research supplier, and/or another organization.

9.3 *Which Records to Retain*—Records may include, but are not limited to, the information described in subsections below. Records to retain depend on the type of testing and/or product category.

9.3.1 *Statement of Desired Claim(s)*—Records should include a complete statement of the proposed claim(s) as well as any modifications of the desired claim(s) made after test completion.

9.3.1.1 The desired claim(s) should be specified in the analysis plan (5.2.10) before testing begins.

9.3.1.2 If, after testing, a modified claim will be made, then the record stating the original claim must still be maintained (see 3.2.12.2). In this case, the modified claim should be documented, usually in a new record. It is also acceptable to modify the original record following good documentation practice, including showing both the original and modified claim.

(1) *Example 1*—After substantiating an unsurpassed preference claim, the advertiser decides to restrict marketing communication to the equivalency of preference (see 14.4). Following best practices, the original dated statement of the desired claim is retained and a new addendum record documenting the revised claim statement is signed and dated, then submitted to be retained with other records associated with this claim.

(2) *Example 2*—After substantiating a preference claim, the advertising wording that was documented (a superiority claim stating “our product tastes superior to X”) is modified to use an alternate wording (a superiority claim stating “our product tastes better than X”). In this particular case, the wordings (“superior to” and “better than”) are interchangeable. Since the original record should not be changed, a new record is created to give the revised wording. The new record is signed, dated, and retained with the other records for this claim.

(3) *Example 3*—An advertiser seeks to make an unqualified superiority claim (“consumers prefer our product to X”). In testing, the test result indicates 65 % prefer the advertiser’s product, 10 % prefer the competitor’s product, and 25 % of consumers have no preference. Since the proportion of no-preference responses is too high to make an unqualified claim (14.2.2), the advertiser creates a new record with a modified claim statement for a qualified claim (“among consumers with a preference, more consumers prefer our product to X”). The new record is signed, dated, and retained with the other records for the claim study.

9.3.2 *Test Objective and Analysis Plan*—Records should include a statement of the specific test objective and the analysis plan (5.2.10) guiding all potential data analyses.

9.3.3 *Products*—Records should include a complete description of each product and how it was procured and handled throughout the study. Products used in claim substantiation research may be in-market commercial products or prepared noncommercialized products (5.3 – 5.6).

9.3.3.1 Records of each product purchased in-market should include brand, name, product form, size, and package (type). Photographs of the product(s) and product label(s), including ingredient statements, directions for preparation/use instructions, and product codes should be taken and stored. Include any product lot numbers, best before dates, expiration dates, or production date where available. Include the agency that purchased the products and the location(s) where the

products were sourced. Records should include information on how the products were blinded.

9.3.3.2 Records of each noncommercialized product should include relevant information regarding product size, ingredients, and where and when it was made. If the claim focuses on noticeable differences between the advertiser’s own in-market product versus a new product, then consider adding a general statement to reflect the change (for example, where applicable, state whether there were any differences in formulation, processing, packaging, and/or label changes). Where applicable, also provide information on how each product was blinded.

9.3.3.3 For test materials included in the research (for example, utensils used by assessors to evaluate product samples), records should indicate where each test material was sourced, shipped, stored, and handled throughout the entire study.

9.3.3.4 If applicable, records should indicate how products were prepared and served, including the quantity or amount of each sample, and a description of any utensil(s) provided or serving vessel used. The blinding codes used should be indicated. If the remaining or unused product samples are retrieved from the assessors after evaluations, this information should also be documented, including the state of the returned product(s). For example, if an opaque sleeve was affixed to a can or bottle for blinded evaluation, then photographs of the returned containers provide evidence that the assessors did not remove the opaque sleeves to learn what products were tested.

(1) Photograph a representative example of what was evaluated by the respondents. This might include a sample, a package, or a carrier. For example, jam may be presented on bread, or a defined amount of hand lotion may be presented to a respondent in a disposable weigh boat.

(2) *Assessors*: Records should include information about the assessors and their qualifications for participation in this test.

(3) A record may include one or more of the following elements: number of assessors, screening and selection criteria, quota groups, recruiting method, geographical locations (for example, city, region), agency that did the recruiting, descriptions of the database(s) from which participants were drawn, and considerations for assessor protection and informed consent (see Guide E3314).

9.3.4 *Test Design and Execution*—Records should include a complete description of method selection, test plan, design, and execution. When documenting the method or methods used in the study, include any guidelines or standards followed. Document why the testing procedures are appropriate and fair and what steps were taken to avoid biasing the test result.

9.3.4.1 A record may include one or more of the following elements:

(1) *Type of Test*:

(a) Assessor type (for example, consumer or trained panel).

(b) Location (for example, CLT or HUT).

(c) Test method and sample presentation format (for example, paired comparison with simultaneous presentation of samples).

(2) References to standards or literature guiding the research design and experimental procedures and protocols, including any relevant ASTM or published sensory methodology.

(3) Test design, including experimental design, evaluation order, presentation protocol, or randomization used to balance the sample presentation order (if applicable).

(4) Test environment (for example, picture of the CLT sample preparation environment and evaluation area, including for example, a description of lighting, temperature, pressure gradients, and air circulation).

(5) Lab certification(s), if applicable.

(6) Questionnaire(s)/ballot(s).

(7) Fielding/study placement instructions, including product preparation and serving instructions, briefing records.

(8) Script used by interviewer or person orienting the assessors to their task.

(9) Staff training protocols.

(10) Screener, re-screener.

(11) Informed consent form (template and signed forms).

(12) Description of palate cleanser(s) or other ancillary items.

9.3.5 Data Collection and Integrity—Records should provide information on how and where the data were collected. The following points should be documented or demonstrated:

(1) The questionnaire/ballot was not altered after data collection commenced.

(2) Any deletions or modifications to data were tracked and explained (for example, if data were collected by staff to confirm the questionnaire was functioning correctly, then a record should indicate the date, time, reason, and person responsible for deleting their own data).

Otherwise, a written confirmation that the data are complete and unmanipulated can be created.

(1) Time(s) and date(s) of the evaluation session(s).

(2) Description of data collection method.

(a) *Paper Testing*: returned consumer questionnaire.

(b) *Interviewer Testing*: interviewer qualifications and note taking procedures.

(c) *Computerized Testing*: software used (including version number), and a copy of the electronic ballot.

(3) Average amount of time for a respondent evaluation, if available.

9.3.6 Data Analysis—Records should provide a complete description of how the data were analyzed as well as the results.

(1) Data validation method (for example, if paper ballots are entered in a database, a procedure for double entering the data to verify accuracy could be used).

(2) Data analysis method, which should be documented in the analysis plan (5.2.10).

(3) Raw data.

(4) Data processing, if applicable.

(5) Results.

(6) Data analysis software used, including version number.

9.3.7 Final Report—Records should include the final report. If possible, include information about the intended audience for the report, including the recipient(s) of the report.

9.4 Archiving, Retrieving, and Distributing Records:

9.4.1 Archiving Records—Records (3.2.12) must be archived in a manner that is controlled and organized where the integrity of the records is preserved (9.2).

9.4.2 Retrieving Records—Records associated with a claim must be retrievable in a timely manner upon request. If a sensory claim is challenged, it must be possible to identify and gain access to all records associated with the sensory claim.

9.4.3 Distributing Records—The integrity of records must be maintained in distribution. Reasonable justification must be provided to explain any withholding of particular records, such as to comply with legal requirements or protect secret or business-sensitive information. If particular information in a record will be redacted prior to its distribution, then the original record must remain intact and the information redacted only in a copy of the record. It must be visually obvious where information is redacted. Reasonable justification must be provided to explain the rationale for redaction.

9.4.3.1 Example. In an advertising challenge, a competitor requests records from the advertiser of a sensory claim. Most records associated with the claim are distributed in a timely manner by the advertiser's legal representative to both the competitor and to the regulator. In this example, some reasons why records might not be disclosed follow.

(1) As is standard practice, potentially identifying information about assessors who participate in sensory tests is withheld. Although the agency that executed the study has archived records identifying consumers who participated in the consumer sensory test, the advertiser neither retrieves nor distributes these records. Instead, the advertiser provides records relevant for demonstrating the appropriateness of the consumer sample to both the competitor and regulator.

(2) The purpose of the advertising challenge is to determine the validity of the claim, not to disclose the advertiser's business secrets or to provide the competitor with highly business-sensitive information (for example, product formulation, proprietary process) that allows the competitor to improve its processes and competitiveness. For these reasons, the advertiser withholds such information from both the competitor and the regulator. Instead, the advertiser provides justification indicating the type of information withheld and why it is withheld.

(3) Some records contain relevant evidence for substantiating the sensory claim that, if disclosed to the competitor would provide valuable business information (for example, business know-how, consumer insights). To allow the regulator to review this information, the advertiser elects to distribute such records selectively to the regulator, but not to the competitor. Instead, the advertiser provides justification indicating the type of information and why this information should be distributed selectively.

(4) A record contains many pieces of information, including one piece of information that discloses secret business information. To provide reasonable transparency without disclosing such information, the advertiser elects to provide a copy of the record where the secret business information has

been redacted. The record is accompanied by a justification for redaction and the appearance of the missing information after redaction (for example, a black rectangle).

10. Laboratory Testing Methods

10.1 Laboratory sensory methods that include discrimination and descriptive test methods are intended to determine if a difference exists in the sensory properties of products, and in the case of descriptive methods, to describe and quantify those differences. These methods provide objective data regarding what humans can perceive without regard for personal preference, and are not appropriate for claims of preference or acceptability.

10.1.1 The laboratory sensory methods' application to claim support is intended to be used to communicate:

10.1.1.1 Product attributes,

10.1.1.2 Overall claims of increase, decrease, or equivalence in a specific attribute(s), and

10.1.1.3 Claims for magnitude of difference between products.

10.1.2 The appropriate application of these methods to claim substantiation requires careful consideration of these factors:

10.1.2.1 It is mandatory that panelists are trained and experienced in the use the selected test method.

10.1.2.2 Panelists must be familiar with the meaning of product attribute descriptors used in the test.

10.1.3 Lack of experience with the test method or misunderstanding about the meaning of attribute descriptors can contribute to inappropriate conclusions being made from the data. ASTM Manual 13 (4) and ASTM Manual 26 (5) contain information on the appropriate application and interpretation of laboratory panel data.

10.2 Types of Tests:

10.2.1 *Overall Difference/Discrimination Tests*—These tests determine if a perceptible sensory difference exists between samples. This difference can occur due to any number of reasons including ingredient differences, processing changes, packaging changes, and so forth. Common overall discrimination tests include the following:

10.2.1.1 *Triangle Test*—Three blind-coded samples are presented either simultaneously or successively to panelists. Two of the samples are the same, representing a single sample, while the third represents a different sample. The panelist is required to identify the different sample.

10.2.1.2 *Duo-Trio Test*—The basic set of samples is the same as in the triangle test, but one of the identical samples is labeled as the “reference.” The panelist is asked to identify the coded sample that is either the same or different from the reference.

10.2.2 *Attribute Difference Tests*—In these test methods, the attribute of interest is defined prior to testing, and the panelists are trained to be able to identify the attribute in question and select or rate the relative intensity of that attribute. It is not necessary to evaluate every occurring attribute, only the attributes being addressed in the claim.

10.2.2.1 *Directional Difference Test*—As defined in E253 Standard Terminology Relating to Sensory Evaluation of Materials and Products.

10.2.2.2 *Attribute Difference Rating Test*—This test also determines if one or more specific attributes differ between two samples. The intensities of the attributes are measured on rating scales showing several degrees of intensity. One or more specific attributes of the product that relate to the claim are rated. Samples are presented, and the panelists' task is to evaluate and assign each test sample an intensity to reflect the amount of the designated attribute(s).

10.2.3 *Descriptive Analysis Test*—A descriptive test is a complete, detailed, and objective characterization of a product's sensory attributes, measuring some or all of the sensory parameters found in a product or material (visual, auditory, olfactory, kinesthetic, and so forth) using screened, qualified panelists who have been specifically trained for this purpose. This method provides information on perceived sensory attributes and the intensities or strength of each sensory attribute, thus identifying specific differences between products in quantitative terms. See ASTM Manual 13 (4) for details on descriptive methodology.

10.3 Advantages and Limitations of the Use of Trained Descriptive Panels in Claims Support Research:

10.3.1 Laboratory panels are useful in objectively determining if and how differences in sensory characteristics are perceived when the claim is attribute or performance focused, not preference or acceptance based. Attributes must be objectively measurable (more butter flavor) as opposed to subjective (better butter flavor). For example, an overall difference test can demonstrate that there is no change to the sensory characteristics of a beverage due to a packaging change, thus allowing the claim “tastes the same as bottled.” A trained descriptive panel can show that a specific formulation delivers a claimed benefit, for example, measuring the duration of a fragrance compared to another formulation, allowing the claim “now longer lasting.”

10.3.2 Laboratory panels are sensitive tools for detection of both large and small product differences. This sensitivity and precision also is its limitation. Laboratory panelists may find product characteristics and detect differences that typical, untrained consumers cannot. Claims are designed for the consumer so that the consumer should expect to experience the product in the same ways as stated in the claim. If the consumer cannot perceive it, then the claim should not imply that the perceived difference is one they would notice.

10.3.3 If the claim in question is intended to be interpreted as representing consumer experience, then such a claim is tenable only if the relationship between the trained panel's response to products and consumers' evaluation is known. The more descriptive and consumer data converge, the more convincing the claim. In short, converging descriptive data and consumer data make a claim significantly less vulnerable to criticism compared to claims based on descriptive panel data or consumer data that stand alone.

10.3.4 In correlating descriptive and consumer panel data, care should be taken to ensure that there is reasonable translation of terms. For example, trained panel data separates

basic taste sweetness from aromatic sweetness, whereas consumers would integrate both together. Correlations may not be possible in cases where consumers do not have the necessary skills to measure or evaluate the attribute(s) in question. For example, trained panel data may support a claim of “more saffron flavor,” but most consumers would not be able to measure this claim. Correlations between the trained panel and the consumer may not be necessary in cases where the claim is used to bring public attention to an attribute that might be new or unique. For example, “We’ve got buzz in every bite.”

10.3.5 Note that trained descriptive panelists are different from “experts” who are drawn from personnel who have extensive experience with the product or product category. Experts may or may not be able to express the perception of differences or descriptions regarding products in terms that can be referenced by standards or treated statistically, and are not appropriate for use with ad claim substantiation.

11. Test Design—Laboratory Testing

11.1 The primary goal of laboratory panels, including descriptive, discrimination, and attribute testing, is to provide an objective sensory evaluation of a product. For claims substantiation, evaluation usually focuses on just one or two product attributes rather than a full product description. These tests can be used to support claims about specific product attributes, such as “Ours’ is thicker,” “It’s less sweet,” “It has more cheese flavor.”

11.2 The test design and questionnaires for laboratory tests should ensure that descriptive/difference data are gathered in an objective and systematic fashion. The test objective and hypothesis should be defined clearly prior to the start of testing. All test procedures should be focused on this objective, such that the test design answers the specific claim that is desired in a short and concise format.

11.3 Panelists used for claim substantiation should be very familiar and experienced with the test method. Descriptive panelists should have extensive training in the descriptive methodology, considerable experience evaluating products, and should be trained specifically on the product under study. References should be used during descriptive training relevant to the product and attribute(s) being evaluated. There should be some documentation of the experience level and type of training received.

11.4 Consider the source of panelists for claims substantiation. If a panel is used that routinely tests the product, there may be some potential for bias. If the panel is familiar with the product, the panel may inadvertently describe it differently; for example, score a particular margarine higher in dairy flavor because the panel is more familiar with that flavor, than if the panel had never seen this product before. If such bias is anticipated, a panel internal to the company is not recommended to substantiate a claim.

11.5 The test design should be reviewed with other members of the technical team and the legal department to ensure accountability for all potential pitfalls.

11.6 Product Procurement:

11.6.1 A laboratory panel must test representative product samples. A representative sample is best accomplished by testing replicate samples of each brand that have been obtained at several representative locations and from several different distribution venues. Sample procurement and handling should occur following a strict protocol. All such information should be documented carefully.

11.6.2 Samples should be selected and handled in the same rigorous manner described in Section 5.

11.7 *Experimental Design*—The exact statistical design will need to be determined on a case by case basis; however, the following describes some of the more important issues that must be considered when a statistical plan is being designed.

11.7.1 Design of discrimination testing is dependent on the selected test method. The number of panelists is key to ensuring adequate power of the test, and should be fully discussed with statisticians and legal prior to conducting the test. The number of samples to be evaluated is dictated by the method; for example, three samples are evaluated in a triangle test. ASTM Manual 26 (5) provides general guidelines for discrimination test designs. Triangle testing is covered in detail in Test Method E1885 and directional difference testing is covered in Test Method E2164.

11.7.2 Design of descriptive panels is covered in ASTM Manual 13 (4). Replications are an essential part of descriptive panel testing and the number of subjects and replications should be determined prior to the test. Three primary types of variability must be accounted for in the design for claims substantiation. These include the following:

11.7.2.1 *Measurement Error*—Repeatability within the individual panelist. This error can be accounted for by having each panelist test a particular sample more than once.

11.7.2.2 *Experimental Error*—Variability between the panelists. This error can be accounted for by using more than one panelist to test each sample.

11.7.2.3 *Product Variability*—Batch-to-batch variation. This error can be accounted for by testing multiple and representative batches of a product.

11.7.3 The number of samples a descriptive panelist evaluates in a session is important, as too many samples could create sensory fatigue. These issues are not likely to be of much consequence in a claims test, due to the fact that the number of samples and the number of attributes being evaluated are usually quite limited.

11.8 Data Collection:

11.8.1 For discrimination tests, the type of data collected is determined by the test method.

11.8.2 For claims substantiation, descriptive panelists should individually evaluate each sample. A group consensus format should not be used with descriptive analysis, as it will be subject to questions regarding the potential of group bias.

11.8.3 It is essential to be explicit about the technique the panelists should use to evaluate the samples. During the data collection phase, the panel leader should ensure that the test protocol is strictly followed.

11.9 Data Analysis (see Section 14):

11.9.1 Any analyses of data should be reviewed with a statistician.

11.9.2 Data should be analyzed according to the statistical design. A typical analysis for descriptive data would be an initial calculation of means and statistical deviations. Next, analysis of variance is performed to determine significant effects. Finally, a multiple comparison technique, such as Tukey's HSD, is used to determine which samples differed significantly.

11.9.3 The analysis of a duo-trio is based on the probability that, if there is no detectable difference, the different sample will be selected by chance one-half of the time. Analysis of a triangle test is based on the probability that, if there is no detectable difference, the different sample will be selected by chance one-third of the time. Data are analyzed using the binomial or chi-square test.

11.9.4 Analysis of a paired comparison is based on the probability that, if there is no detectable difference, the different sample will be selected by chance one-half of the time. Data are analyzed using a binomial test.

12. Questionnaire Construction

12.1 Questionnaires used in discrimination tests are specified by the test method. The focus of the questionnaire is on the selection of a sample based on its difference from other samples, either for its overall difference or for its higher (or lower) intensity in a specific attribute.

12.2 The main objective of descriptive panel tests is to provide an accurate description of a product in terms of its perceived attributes and their intensities. Questionnaires for trained panels should ensure that data are collected with the goal of obtaining all the necessary information in a concise and easily understood format.

12.3 There are several ways to construct a descriptive panel questionnaire; but more importantly, the questionnaire should be brief, including only the attributes necessary to establish or support the claim(s). A recommended procedure is to use questions from a more comprehensive, established evaluation ballot, and then elect specific attributes needed for the claims testing, thus eliminating or reducing any special training time.

12.4 If specific training is necessary, the training should be accomplished with relevant products or materials, or both, that reference the specific product under study. Select panelists with experience in evaluating similar products or attributes, or both, necessary to authenticate and defend the claim. Pilot testing should be conducted to detect any questionnaire or methodological deficiencies and to confirm applicability and accuracy.

13. Test Facility

13.1 Environment:

13.1.1 When selecting a test facility for a claims test, take into consideration such environmental aspects as color, lighting, and air control, specifically temperature and humidity. The evaluation area should also be free of distractions from other panelists, laboratory personnel, or general noise.

13.1.2 It is optimal to use natural colors when selecting the furnishings in the testing area. Make the walls of the evaluation area off-white to prevent unwanted effects of color on the sample product.

13.1.3 Most testing does not require special lighting. Shadow-free illumination at intensity typical of an office area is suitable for most studies. The exception is when visual attributes are being evaluated, requiring the specification and documentation of lighting conditions.

13.1.4 Ideally, the sensory testing area should be maintained at approximately 72 °F, with a relative humidity between 45 and 55 %. Ventilation should be such that extraneous odors are eliminated, particularly if fragrances or aromas/flavors are being evaluated.

13.2 *Facility Design*—The facility design and overall space requirements depend on the number and nature of tests conducted and on the type of products. Different designs and layouts are illustrated in ASTM STP 913 (6).

14. Statistical Analyses

14.1 *Paired-preference Studies*—This section describes how to analyze data from paired preference studies which are designed to establish that there is sufficient evidence to substantiate a sensory claim. More than statistical criteria are involved in developing a sampling plan for product tests designed to support advertisement claims. It is widely recognized that attempting to collect a simple random sample is impractical and that cluster samples, for example, multiple city CLTs, with quotas are accepted alternatives. Proper representative sampling may require sample sizes greater than the sample sizes that are discussed in this section. Subsections 5.1 and 5.2 detail approaches to sampling to ensure they adequately approximate the population to which the claim is intended to apply. Section 14 focuses on the analysis of the paired preference results, addressing the three forms of the claim: superiority, equivalence, and unsurpassed, under the assumption that the data sample can be treated as arising from a simple random sample.

14.1.1 *Binomial Test*—Paired preference data are usually analyzed using a binomial test. The sample size is n . The number of people who give a response for the advertiser's product is n_1 . The number of people who give a response for the competitor's product is n_2 . The binomial test is conducted at the 5 % significance level (95 % confidence level). For each claim type and test, how to calculate critical values in Microsoft Excel⁴ and in R (7) will be shown.

14.1.2 *Handling No-preference Results*—If an odd number of people expresses no preference, the extra no-preference response should be awarded to the competitor. Specifically, let n_1 be the number of people preferring the advertiser's product and n_0 be the number of people who expressed no preference.

Then set the advertiser's choice count to $x = n_1 + \frac{n_0}{2}$ when n_0 even and to $x = n_1 + \frac{(n_0 - 1)}{2}$ when n_0 is odd. No-difference responses can be allocated identically. Although the exact

⁴ "Microsoft" and "Microsoft Excel" are registered trademarks of Microsoft Corporation in the United States and/or other countries.

distribution is no longer binomial after splitting no-preference responses, this allocation of no-preference responses in a superiority test yields a Type 1 error that is never greater than the nominal value (8).

14.1.3 Consideration of a Seemingly Contradictory Claim—It is noted that test results (especially for tests with very large sample sizes) could be used to substantiate claims that seem to be contradictory. For example, an advertiser could observe a test result that substantiates a superiority claim with a result of 52.5 % preference for the advertiser and a sample size of approximately 1100 or more. However, a challenger could use the same data to substantiate an equivalence claim because they could show that the result obtained falls significantly within the equivalence bounds of 55 %/45 %. A challenger could also claim that their product is unsurpassed by the advertiser because they could show that 47.5 % (the competitor's choice proportion) is significantly greater than 45 %, the lower bound for establishing an unsurpassed claim. Given the different action standards chosen to define superiority, equivalence and unsurpassed claims in the guide, the advertiser should be prepared to defend their claim against these potential challenges should their test result put them in the situation where seemingly contradictory claims are supported simultaneously.

14.2 Superiority Claims—A superiority claim is supported if a statistically significant proportion of the respondents prefer the advertiser's product at the effective sample size.

14.2.1 Unqualified Superiority Claims—This type of claim is based on responses from all consumers. If the percentage of people expressing no preference is 20 % or more, an unqualified preference claim should not be made (a strong technical rationale is needed to exceed this guideline). The advertiser can still make a qualified superiority claim.

14.2.1.1 A binomial test can be used to analyze the data. Let n be the total number of people participating in the study. Split no-preference responses according to 14.1.2 to obtain the advertiser's choice count x .

14.2.1.2 Determine the critical value for significance based on the 95 % quantile of the binomial distribution with probability $P = 0.5$ at sample size n . A critical value for a sample size n that is not found in Table 1 can be obtained in Microsoft Excel⁴ or in R (see below).

Excel: =BINOM.INV(n , 0.5, 0.95) + 1

R: qbinom(.05, n , 0.5, lower.tail=FALSE) + 1

NOTE 2—The Excel formulae provided assume that the U.S. English regional setting is used, where the period is the decimal separator and the comma separates function arguments. The formulae provided might require modifications on some computers (depending on regional settings).

14.2.1.3 If x is greater than or equal to the critical value, then the unqualified superiority claim is supported at the 5 % (one-tailed) level of significance. Table 1 shows critical values for a superiority claim (up to $n = 539$).

14.2.2 Qualified Superiority Claims—If an unqualified claim cannot be made based on the analysis of the total data (for example, if percentage of people expressing no preference is 20 % or more or the unqualified claim is not supported), then the advertiser can still make a qualified superiority claim,

provided that the analysis, excluding the no-preference responses, shows a significant result, and the advertisement includes a suitable reference to the fact that the claim is based on "those who expressed a preference."

14.2.2.1 A binomial test can be used to analyze the data. Discard no-preference responses and let the sample size be the number of people participating in the study who expressed a preference ($n_1 + n_2$). Let the advertiser's choice count be $x = n_1$.

14.2.2.2 Determine the critical value for significance based on the 95 % quantile of binomial distribution with probability $P = 0.5$ at sample size $n_1 + n_2$. A critical value for a sample size $n_1 + n_2$ that is not found in Table 1 can be obtained in Microsoft Excel⁴ or in R (see below).

Excel: =BINOM.INV($n_1 + n_2$, 0.5, 0.95) + 1

R: qbinom(.05, $n_1 + n_2$, 0.5, lower.tail=FALSE) + 1

14.2.2.3 If x is greater than or equal to the critical value, then the qualified superiority claim is supported at the 5 % (one-tailed) level of significance. Table 1 shows critical values for a superiority claim (up to $n = 539$).

14.2.3 Other analytic approaches can be used but require justification. The approach described in 14.1 and 14.2 is recommended for all cases. If another analytic approach will be used, then it should be documented and justified by the advertiser prior to conducting the sensory test for the claim. In some product use testing (for example, cleaning or absorbent products), consumers may experience many, but not all of the product use situations, and will be unable to detect a difference between the products they are comparing or the products may appear so similar that they are unable to form a preference. (See example in 14.2.4.4.)

14.2.4 Examples:

14.2.4.1 Example 1: Unqualified Superiority Claim—Example 1: Unqualified Superiority Claim—In a preference test among 205 consumers (n), 100 consumers choose the advertiser's product (n_1), 91 choose the competitor's product and there are 14 no preferences (n_0). Splitting the no preferences according to 14.1.2 gives the advertiser's product a count of $x = 107$. With a sample size of 205 (n), we determine from Table 1 that we require 115 preference judgments in favor of the advertiser to declare significance at the 95 % confidence level. The advertiser has not met this requirement and cannot claim superiority based on this test.

14.2.4.2 Example 2: Unqualified Superiority Claim—In a preference test among 221 consumers (n), 120 consumers choose the advertiser's product (n_1), 90 choose the competitor's product and there are 11 no preferences (n_0). Splitting the no preferences according to 14.1.2 gives the advertiser's product a count of $x = 125$ and the competitor's product a count of $n - x = 96$. (Note that the allocation of ties according to 14.1.2 assigns one more no-preference response to the competitor than to the advertiser.) With a sample size of 221 (n), we determine from Table 1 that we require 124 preference judgments in favor of the advertiser to declare significance at the 95 % confidence level. The advertiser has met this requirement and can claim superiority based on this test.

14.2.4.3 Example 3: Qualified Superiority Claim—In a preference test among 281 consumers (n), 118 consumers choose the advertiser's product (n_1), 87 choose the competitor's

TABLE 1 Critical Values for Superiority Claim^A

NOTE 1—The sample size is found by adding the row and column headers. The critical value is obtained at the intersection of the row and column. To declare the advertiser's product superior to a competitor's product at the 95 % confidence level, the choice count for the advertiser's product must equal or exceed the critical value indicated at the sample size.

<i>n</i>	0	1	2	3	4	5	6	7	8	9
10	9	9	10	10	11	12	12	13	13	14
20	15	15	16	16	17	18	18	19	19	20
30	20	21	22	22	23	23	24	24	25	26
40	26	27	27	28	28	29	30	30	31	31
50	32	32	33	33	34	35	35	36	36	37
60	37	38	38	39	40	40	41	41	42	42
70	43	43	44	45	45	46	46	47	47	48
80	48	49	49	50	51	51	52	52	53	53
90	54	54	55	55	56	57	57	58	58	59
100	59	60	60	61	61	62	62	63	64	64
110	65	65	66	66	67	67	68	68	69	69
120	70	71	71	72	72	73	73	74	74	75
130	75	76	76	77	78	78	79	79	80	80
140	81	81	82	82	83	83	84	84	85	86
150	86	87	87	88	88	89	89	90	90	91
160	91	92	92	93	94	94	95	95	96	96
170	97	97	98	98	99	99	100	100	101	101
180	102	103	103	104	104	105	105	106	106	107
190	107	108	108	109	109	110	111	111	112	112
200	113	113	114	114	115	115	116	116	117	117
210	118	118	119	119	120	121	121	122	122	123
220	123	124	124	125	125	126	126	127	127	128
230	128	129	130	130	131	131	132	132	133	133
240	134	134	135	135	136	136	137	137	138	138
250	139	140	140	141	141	142	142	143	143	144
260	144	145	145	146	146	147	147	148	148	149
270	150	150	151	151	152	152	153	153	154	154
280	155	155	156	156	157	157	158	158	159	159
290	160	161	161	162	162	163	163	164	164	165
300	165	166	166	167	167	168	168	169	169	170
310	170	171	172	172	173	173	174	174	175	175
320	176	176	177	177	178	178	179	179	180	180
330	181	181	182	183	183	184	184	185	185	186
340	186	187	187	188	188	189	189	190	190	191
350	191	192	192	193	193	194	195	195	196	196
360	197	197	198	198	199	199	200	200	201	201
370	202	202	203	203	204	204	205	205	206	207
380	207	208	208	209	209	210	210	211	211	212
390	212	213	213	214	214	215	215	216	216	217
400	217	218	218	219	220	220	221	221	222	222
410	223	223	224	224	225	225	226	226	227	227
420	228	228	229	229	230	230	231	231	232	233
430	233	234	234	235	235	236	236	237	237	238
440	238	239	239	240	240	241	241	242	242	243
450	243	244	244	245	246	246	247	247	248	248
460	249	249	250	250	251	251	252	252	253	253
470	254	254	255	255	256	256	257	257	258	258
480	259	260	260	261	261	262	262	263	263	264
490	264	265	265	266	266	267	267	268	268	269
500	269	270	270	271	271	272	272	273	274	274
510	275	275	276	276	277	277	278	278	279	279
520	280	280	281	281	282	282	283	283	284	284
530	285	285	286	286	287	288	288	289	289	290

^A Table adapted from "Tools and Applications of Sensory and Consumer Science," Table 1, p. 164 (9), with permission from The Institute for Perception, Richmond, Virginia, 2018.

product (n_2), and there are 76 no preferences (n_0). An unqualified claim cannot be made because more than 20 % of consumers provided a no-preference response. With an effective sample size of 205 ($n_1 + n_2$), we determine from **Table 1** that we require 115 preference judgments in favor of the advertiser to declare significance at the 95 % level. The advertiser has met this requirement and can make a qualified superiority claim based on this test, in which the claim references "those who expressed a preference".

14.2.4.4 Example 4: Use of Different Analytic Approach—An advertiser of an absorbent product plans to make a superiority claim in which the percentage of people expressing no preference is expected to exceed 20 %. Although the advertiser will use a robust study design, only some of the consumers will experience leakage during testing; whether they experience leakage would likely influence their evaluation of the product performance. Regardless of the reasons that might lead to a no-preference response, the advertiser decides

to use an analytic approach that is different from the approach that is recommended in Section 14. Prior to commencing consumer data collection, the advertiser documents and provides strong justification for the analytic approach that will be used (as described in 14.2.3).

14.2.5 Performance Characteristics of Preference Tests for Superiority Testing—The ability to detect departures from 50 %:50 % preference improves as the number of respondents increase. The number of respondents is under the control of the advertiser, and it is the advertiser who risks missing the opportunity to make a superiority claim when too few respondents participate in the test. As such, this guide does not specify a minimum number of respondents for a preference test to support a superiority claim. To help the advertiser select the number of respondents, Table 2 contains the minimum values of the observed preference proportions required to support a superiority claim for various numbers of respondents. Also presented in Table 2 are two ways to assess the sensitivity of the test for various numbers of respondents. The third column in Table 2 contains the probabilities of detecting a 55 %:45 % preference split for the various numbers of respondents considered. The final column of the table contains the preference percent that has an 80 % probability of being detected for the various numbers of respondents considered. Both of the last columns demonstrate the advantage that larger sample sizes present to the advertiser. These pieces of information can be used together with an assessment of the testing resources available to the advertiser to decide how large of a test needs to be run (see 14.2.5.2).

14.2.5.1 Interpretation of Non-significant Result—In the methods used for superiority testing, failure to conclude that a significant difference in preference exists between two products does not prove that two products are equivalently preferred. The failure to achieve statistical significance may result from using an insufficient number of respondents, thus yielding an insensitive test. It is also possible that failure to find a statistically significant preference may be due to sampling error in which sampling procedures resulted in a group of respondents who exhibited no preference. Furthermore, observing a

preference percent slightly less than 50 % does not prove that equivalence does not exist.

14.2.5.2 Sample Size—For superiority claims, the advertiser assumes the risk of an insensitive test. Larger numbers of respondents are preferred because they both protect the competitor and provide an advantage to the advertiser.

14.3 Equivalence Claims—Since one cannot prove the null hypothesis, an equivalence claim must be specified in terms of an acceptable range of possible difference within which the products can be considered to be essentially equivalent. For paired preference and (more commonly) difference testing in which the instruction is to choose the product which is most preferred or has the most of some specified attribute (for example, sweetness or smoothness), one may choose a 45 %:55 % split for the population as a limit on the meaning of equivalence. If either product equals or exceeds 55 % of the choices, then the products are not considered to be equivalent in the meaning of this section. Table 3 provides the required lower choice count to claim equivalence at the 95 % confidence level for sample sizes of 300 to 1999.

14.3.1 An equivalence claim is always based on responses from all consumers. There is no limitation on the percent acceptable no-preference responses when making a claim of equivalence. The advertiser cannot drop no-preference responses to make a qualified claim of equivalence.

14.3.1.1 A binomial test can be used to analyze the data. Let n be the total number of people participating in the study. Split no-preference responses according to 14.1.2 to obtain the advertiser's choice count x and the competitor's choice count $n - x$.

14.3.1.2 Determine the critical value for significance based on the binomial distribution with probability $P = 0.45$ at sample size n . The critical value is the 95 % quantile of this distribution. A critical value for a sample size n that is not found in Table 3 can be obtained in Microsoft Excel⁴ or in R (see below).

Excel: =BINOM.INV(n , 0.45, 0.95) + 1

R: qbinom(.05, n , 0.45, lower.tail=FALSE) + 1

14.3.1.3 If the lesser of x and $n - x$ is greater than or equal to the critical value, then the equivalence claim is supported at the 5 % (one-tailed) level of significance. (This is the same as saying that the equivalence claim is supported at the 5 % (one-tailed) level of significance if both x and $n - x$ are greater than or equal to the critical value.) Table 3 is constructed so that the counts control Type 1 error at 5 % or less. (Sample sizes in Table 3 are larger than sample sizes in Tables 1 and 5 because testing an equivalence hypothesis requires a larger sample size than testing a superiority or an unsurpassed hypothesis for reasons related to statistical power.)

14.3.2 Other Analytic Approaches Can Be Used but Require Justification—The approach described in 14.1 and 14.3 is recommended for all cases. If another analytic approach will be used, then it should be documented and justified by the advertiser prior to conducting the sensory test for the equivalence claim.

14.3.3 Example: Equivalence Claim—A paired comparison test with 1000 consumers is conducted in which consumers choose the sweetest sample and 470 chose one product and 530

TABLE 2 Performance Characteristics of Preference Tests for Superiority Testing

n	P_C^A	Power ^B	80 % Detect ^C
100	59.0	24.1	62.6
200	56.5	36.2	59.2
300	55.0	52.4	57.2
400	54.2	63.8	56.2
500	53.8	72.1	55.6
600	53.5	78.2	55.1
700	53.3	82.9	54.8
800	53.0	87.9	54.4
900	52.9	90.4	54.2
1000	52.7	93.2	54.0

^A P_C —Minimum observed percent preference required to claim superiority at the 95 % confidence level.

^B Power—Probability of claiming superiority when the actual preference for the advertiser's product is 55 %.

^C 80 % Detect—Actual preference for the advertiser's product that has an 80 % probability of being detected.

TABLE 3 Critical Values for Equivalence Claim

NOTE 1—The sample size is found by adding the row and column headers. The critical value is obtained at the intersection of the row and column. To declare the advertiser's product equivalent relative to a competitor's product at a given sample size with 95 % confidence, the lesser of the two choice counts (the choice count for the advertiser's product and the choice count for the competitor's product) must equal or exceed the critical value indicated at the sample size. (At some sample sizes, such as $n = 301$, no test result is consistent with an equivalence claim; in these cases, a dash is shown in place of the critical value.)

<i>n</i>	0	1	2	3	4	5	6	7	8	9
300	150	-	151	-	152	-	153	-	154	154
310	155	155	156	156	157	157	158	158	159	159
320	160	160	161	161	162	162	162	163	163	164
330	164	165	165	166	166	167	167	168	168	169
340	169	170	170	171	171	171	172	172	173	173
350	174	174	175	175	176	176	177	177	178	178
360	179	179	179	180	180	181	181	182	182	183
370	183	184	184	185	185	186	186	187	187	188
380	188	188	189	189	190	190	191	191	192	192
390	193	193	194	194	195	195	196	196	196	197
400	197	198	198	199	199	200	200	201	201	202
410	202	203	203	204	204	204	205	205	206	206
420	207	207	208	208	209	209	210	210	211	211
430	211	212	212	213	213	214	214	215	215	216
440	216	217	217	218	218	219	219	219	220	220
450	221	221	222	222	223	223	224	224	225	225
460	226	226	227	227	227	228	228	229	229	230
470	230	231	231	232	232	233	233	234	234	234
480	235	235	236	236	237	237	238	238	239	239
490	240	240	241	241	242	242	242	243	243	244
500	244	245	245	246	246	247	247	248	248	249
510	249	249	250	250	251	251	252	252	253	253
520	254	254	255	255	256	256	256	257	257	258
530	258	259	259	260	260	261	261	262	262	263
540	263	264	264	264	265	265	266	266	267	267
550	268	268	269	269	270	270	271	271	271	272
560	272	273	273	274	274	275	275	276	276	277
570	277	278	278	278	279	279	280	280	281	281
580	282	282	283	283	284	284	285	285	285	286
590	286	287	287	288	288	289	289	290	290	291
600	291	292	292	292	293	293	294	294	295	295
610	296	296	297	297	298	298	299	299	299	300
620	300	301	301	302	302	303	303	304	304	305
630	305	306	306	306	307	307	308	308	309	309
640	310	310	311	311	312	312	313	313	313	314
650	314	315	315	316	316	317	317	318	318	319
660	319	320	320	320	321	321	322	322	323	323
670	324	324	325	325	326	326	326	327	327	328
680	328	329	329	330	330	331	331	332	332	333
690	333	333	334	334	335	335	336	336	337	337
700	338	338	339	339	340	340	340	341	341	342
710	342	343	343	344	344	345	345	346	346	347
720	347	347	348	348	349	349	350	350	351	351
730	352	352	353	353	353	354	354	355	355	356
740	356	357	357	358	358	359	359	360	360	360
750	361	361	362	362	363	363	364	364	365	365
760	366	366	367	367	367	368	368	369	369	370
770	370	371	371	372	372	373	373	373	374	374
780	375	375	376	376	377	377	378	378	379	379
790	380	380	380	381	381	382	382	383	383	384
800	384	385	385	386	386	386	387	387	388	388
810	389	389	390	390	391	391	392	392	393	393
820	393	394	394	395	395	396	396	397	397	398
830	398	399	399	399	400	400	401	401	402	402
840	403	403	404	404	405	405	406	406	406	407
850	407	408	408	409	409	410	410	411	411	412
860	412	412	413	413	414	414	415	415	416	416
870	417	417	418	418	419	419	419	420	420	421
880	421	422	422	423	423	424	424	425	425	425
890	426	426	427	427	428	428	429	429	430	430
900	431	431	432	432	432	433	433	434	434	435
910	435	436	436	437	437	438	438	438	439	439
920	440	440	441	441	442	442	443	443	444	444
930	444	445	445	446	446	447	447	448	448	449
940	449	450	450	451	451	451	452	452	453	453
950	454	454	455	455	456	456	457	457	457	458
960	458	459	459	460	460	461	461	462	462	463
970	463	463	464	464	465	465	466	466	467	467
980	468	468	469	469	469	470	470	471	471	472

TABLE 3 *Continued*

<i>n</i>	0	1	2	3	4	5	6	7	8	9
990	472	473	473	474	474	475	475	476	476	476
1000	477	477	478	478	479	479	480	480	481	481
1010	482	482	482	483	483	484	484	485	485	486
1020	486	487	487	488	488	488	489	489	490	490
1030	491	491	492	492	493	493	494	494	494	495
1040	495	496	496	497	497	498	498	499	499	500
1050	500	501	501	501	502	502	503	503	504	504
1060	505	505	506	506	507	507	507	508	508	509
1070	509	510	510	511	511	512	512	513	513	513
1080	514	514	515	515	516	516	517	517	518	518
1090	519	519	519	520	520	521	521	522	522	523
1100	523	524	524	525	525	525	526	526	527	527
1110	528	528	529	529	530	530	531	531	531	532
1120	532	533	533	534	534	535	535	536	536	537
1130	537	537	538	538	539	539	540	540	541	541
1140	542	542	543	544	544	544	544	545	545	546
1150	546	547	547	548	548	549	549	550	550	550
1160	551	551	552	552	553	553	554	554	555	555
1170	556	556	556	557	557	558	558	559	559	560
1180	560	561	561	562	562	562	563	563	564	564
1190	565	565	566	566	567	567	568	568	568	569
1200	569	570	570	571	571	572	572	573	573	574
1210	574	574	575	575	576	576	577	577	578	578
1220	579	579	580	580	580	581	581	582	582	583
1230	583	584	584	585	585	586	586	586	587	587
1240	588	588	589	589	590	590	591	591	592	592
1250	592	593	593	594	594	595	595	596	596	597
1260	597	598	598	598	599	599	600	600	601	601
1270	602	602	603	603	604	604	604	605	605	606
1280	606	607	607	608	608	609	609	610	610	610
1290	611	611	612	612	613	613	614	614	615	615
1300	616	616	616	617	617	618	618	619	619	620
1310	620	621	621	622	622	622	623	623	624	624
1320	625	625	626	626	627	627	628	628	628	629
1330	629	630	630	631	631	632	632	633	633	634
1340	634	634	635	635	636	636	637	637	638	638
1350	639	639	640	640	640	641	641	642	642	643
1360	643	644	644	645	645	646	646	646	647	647
1370	648	648	649	649	650	650	651	651	652	652
1380	652	653	653	654	654	655	655	656	656	657
1390	657	657	658	658	659	659	660	660	661	661
1400	662	662	663	663	663	664	664	665	665	666
1410	666	667	667	668	668	669	669	669	670	670
1420	671	671	672	672	673	673	674	674	675	675
1430	675	676	676	677	677	678	678	679	679	680
1440	680	681	681	681	682	682	683	683	684	684
1450	685	685	686	686	687	687	687	688	688	689
1460	689	690	690	691	691	692	692	693	693	693
1470	694	694	695	695	696	696	697	697	698	698
1480	699	699	699	700	700	701	701	702	702	703
1490	703	704	704	704	705	705	706	706	707	707
1500	708	708	709	709	710	710	710	711	711	712
1510	712	713	713	714	714	715	715	716	716	716
1520	717	717	718	718	719	719	720	720	721	721
1530	722	722	722	723	723	724	724	725	725	726
1540	726	727	727	728	728	728	729	729	730	730
1550	731	731	732	732	733	733	734	734	734	735
1560	735	736	736	737	737	738	738	739	739	739
1570	740	740	741	741	742	742	743	743	744	744
1580	745	745	745	746	746	747	747	748	748	749
1590	749	750	750	751	751	751	752	752	753	753
1600	754	754	755	755	756	756	757	757	757	758
1610	758	759	759	760	760	761	761	762	762	763
1620	763	763	764	764	765	765	766	766	767	767
1630	768	768	768	769	769	770	770	771	771	772
1640	772	773	773	774	774	774	775	775	776	776
1650	777	777	778	778	779	779	780	780	780	781
1660	781	782	782	783	783	784	784	785	785	786
1670	786	786	787	787	788	788	789	789	790	790
1680	791	791	791	792	792	793	793	794	794	795
1690	795	796	796	797	797	797	798	798	799	799
1700	800	800	801	801	802	802	803	803	803	804
1710	804	805	805	806	806	807	807	808	808	809
1720	809	809	810	810	811	811	812	812	813	813

TABLE 3 *Continued*

<i>n</i>	0	1	2	3	4	5	6	7	8	9
1730	814	814	814	815	815	816	816	817	817	818
1740	818	819	819	820	820	820	821	821	822	822
1750	823	823	824	824	825	825	826	826	826	827
1760	827	828	828	829	829	830	830	831	831	831
1770	832	832	833	833	834	834	835	835	836	836
1780	837	837	837	838	838	839	839	840	840	841
1790	841	842	842	843	843	843	844	844	845	845
1800	846	846	847	847	848	848	849	849	849	850
1810	850	851	851	852	852	853	853	854	854	854
1820	855	855	856	856	857	857	858	858	859	859
1830	860	860	860	861	861	862	862	863	863	864
1840	864	865	865	866	866	866	867	867	868	868
1850	869	869	870	870	871	871	871	872	872	873
1860	873	874	874	875	875	876	876	877	877	877
1870	878	878	879	879	880	880	881	881	882	882
1880	883	883	883	884	884	885	885	886	886	887
1890	887	888	888	888	889	889	890	890	891	891
1900	892	892	893	893	894	894	894	895	895	896
1910	896	897	897	898	898	899	899	900	900	900
1920	901	901	902	902	903	903	904	904	905	905
1930	905	906	906	907	907	908	908	909	909	910
1940	910	911	911	911	912	912	913	913	914	914
1950	915	915	916	916	916	917	917	918	918	919
1960	919	920	920	921	921	922	922	922	923	923
1970	924	924	925	925	926	926	927	927	928	928
1980	928	929	929	930	930	931	931	932	932	933
1990	933	933	934	934	935	935	936	936	937	937

TABLE 4 Performance Characteristics of Difference Tests for Equivalence Testing

<i>n</i>	<i>Pc Range</i> ^A	<i>Power</i> ^B	<i>LL95</i> ^C
300	[50.0, 50.0]	4.6	45.1
400	[49.3, 50.8]	27.4	45.8
500	[48.8, 51.2]	43.9	46.2
600	[48.5, 51.5]	56.2	46.6
700	[48.3, 51.7]	65.5	46.8
800	[48.0, 52.0]	75.7	47.0
900	[47.9, 52.1]	80.6	47.2
1000	[47.7, 52.3]	86.3	47.4
1100	[47.5, 52.5]	90.3	47.5
1200	[47.4, 52.6]	93.1	47.6
1300	[47.4, 52.6]	94.4	47.7
1400	[47.3, 52.7]	96.0	47.8
1500	[47.2, 52.8]	97.2	47.8
1600	[47.1, 52.9]	98.0	47.9
1700	[47.1, 52.9]	98.6	48.0
1800	[47.0, 53.0]	99.0	48.0
1900	[46.9, 53.1]	99.3	48.1

^A *Pc Range*—Range of preference percentage of advertiser's product required to claim equivalency at the 95 % confidence level.

^B *Power*—Probability of claiming to be equivalent when the advertiser and competitor are truly equal (50 %) in the attribute being evaluated.

^C *LL95*—Lower limit of a one-tailed 95 % confidence interval that represents how low the actual percent for the advertiser's product may be when a 50 % proportion is observed in the study.

chose the other product. Since the required lower choice count must be at least 477 and the experiment resulted in a value of 470, we cannot declare support for the equivalence hypothesis. However, if 481 consumers had chosen one of the products, we would have concluded in favor of the equivalence hypothesis.

14.3.4 Performance Characteristics of Difference Tests for Equivalence Testing—Advertisers who are planning an equivalence test should review the performance characteristics given in **Table 4**. This table contains the percentages required to support an equivalence claim for various numbers of respondents that protect the competitor with 95 % certainty when the

true proportion is either less than 45 % or greater than 55 %. The table also gives the probability that a test result based on various numbers of respondents will support the equivalence claim when the true probability for the advertiser's product is exactly $P = 50\%$. The final column of **Table 4** shows how low the lesser of the advertiser proportion and competitor proportion may be, with 95 % certainty, when a 50 % result is observed in a test. It is important to note that statistical power is very low, even at $n = 400$; reasonable statistical power is only reached at much higher sample sizes. The information in **Table 4** illustrates the advantage of larger sample sizes for the

TABLE 5 Critical Values for Unsurpassed Claim

NOTE 1—The sample size is found by adding the row and column headers. The critical value is obtained at the intersection of the row and column. To declare the advertiser's product unsurpassed to a competitor's product at a given sample size at the 95 % confidence level, the choice count for the advertiser's product must equal or exceed the critical value indicated at the sample size.

<i>n</i>	0	1	2	3	4	5	6	7	8	9
10	8	9	9	10	10	11	11	12	13	13
20	14	14	15	15	16	16	17	17	18	18
30	19	20	20	21	21	22	22	23	23	24
40	24	25	25	26	26	27	27	28	28	29
50	29	30	30	31	31	32	32	33	33	34
60	34	35	35	36	36	37	37	38	38	39
70	39	40	40	41	41	42	42	43	43	44
80	44	45	45	46	46	47	47	48	48	49
90	49	50	50	51	51	52	52	53	53	54
100	54	55	55	56	56	57	57	58	58	59
110	59	60	60	61	61	62	62	63	63	63
120	64	64	65	65	66	66	67	67	68	68
130	69	69	70	70	71	71	72	72	73	73
140	74	74	75	75	76	76	77	77	78	78
150	79	79	80	80	80	81	81	82	82	83
160	83	84	84	85	85	86	86	87	87	88
170	88	89	89	90	90	91	91	92	92	93
180	93	93	94	94	95	95	96	96	97	97
190	98	98	99	99	100	100	101	101	102	102
200	103	103	104	104	105	105	105	106	106	107
210	107	108	108	109	109	110	110	111	111	112
220	112	113	113	114	114	115	115	115	116	116
230	117	117	118	118	119	119	120	120	121	121
240	122	122	123	123	124	124	125	125	126	126
250	126	127	127	128	128	129	129	130	130	131
260	131	132	132	133	133	134	134	135	135	135
270	136	136	137	137	138	138	139	139	140	140
280	141	141	142	142	143	143	144	144	145	145
290	145	146	146	147	147	148	148	149	149	150
300	150	151	151	152	152	153	153	154	154	154
310	155	155	156	156	157	157	158	158	159	159
320	160	160	161	161	162	162	162	163	163	164
330	164	165	165	166	166	167	167	168	168	169
340	169	170	170	171	171	172	172	172	173	173
350	174	174	175	175	176	176	177	177	178	178
360	179	179	179	180	180	181	181	182	182	183
370	183	184	184	185	185	186	186	187	187	188
380	188	188	189	189	190	190	191	191	192	192
390	193	193	194	194	195	195	196	196	196	197
400	197	198	198	199	199	200	200	201	201	202
410	202	203	203	204	204	204	205	205	206	206
420	207	207	208	208	209	209	210	210	211	211
430	211	212	212	213	213	214	214	215	215	216
440	216	217	217	218	218	219	219	219	220	220
450	221	221	222	222	223	223	224	224	225	225
460	226	226	227	227	227	228	228	229	229	230
470	230	231	231	232	232	233	233	234	234	234
480	235	235	236	236	237	237	238	238	239	239
490	240	240	241	241	242	242	242	243	243	244
500	244	245	245	246	246	247	247	248	248	249
510	249	249	250	250	251	251	252	252	253	253
520	254	254	255	255	256	256	256	257	257	258
530	258	259	259	260	260	261	261	262	262	263
540	263	264	264	264	265	265	266	266	267	267
550	268	268	269	269	270	270	271	271	271	272
560	272	273	273	274	274	275	275	276	276	277
570	277	278	278	278	279	279	280	280	281	281
580	282	282	283	283	284	284	285	285	285	286
590	286	287	287	288	288	289	289	290	290	291
600	291	292	292	292	293	293	294	294	295	295
610	296	296	297	297	298	298	299	299	299	300
620	300	301	301	302	302	303	303	304	304	305
630	305	306	306	306	307	307	308	308	309	309
640	310	310	311	311	312	312	313	313	313	314
650	314	315	315	316	316	317	317	318	318	319
660	319	320	320	320	321	321	322	322	323	323
670	324	324	325	325	326	326	326	327	327	328
680	328	329	329	330	330	331	331	332	332	333
690	333	333	334	334	335	335	336	336	337	337
700	338	338	339	339	340	340	340	341	341	342
710	342	343	343	344	344	345	345	346	346	347

TABLE 5 *Continued*

<i>n</i>	0	1	2	3	4	5	6	7	8	9
720	347	347	348	348	349	349	350	350	351	351
730	352	352	353	353	353	354	354	355	355	356
740	356	357	357	358	358	359	359	360	360	360
750	361	361	362	362	363	363	364	364	365	365
760	366	366	367	367	367	368	368	369	369	370
770	370	371	371	372	372	373	373	373	374	374
780	375	375	376	376	377	377	378	378	379	379
790	380	380	380	381	381	382	382	383	383	384
800	384	385	385	386	386	386	387	387	388	388
810	389	389	390	390	391	391	392	392	393	393
820	393	394	394	395	395	396	396	397	397	398
830	398	399	399	399	400	400	401	401	402	402
840	403	403	404	404	405	405	406	406	406	407
850	407	408	408	409	409	410	410	411	411	412
860	412	412	413	413	414	414	415	415	416	416
870	417	417	418	418	419	419	420	420	421	421
880	421	422	422	423	423	424	424	425	425	425
890	426	426	427	427	428	428	429	429	430	430
900	431	431	432	432	432	433	433	434	434	435
910	435	436	436	437	437	438	438	438	439	439
920	440	440	441	441	442	442	443	443	444	444
930	444	445	445	446	446	447	447	448	448	449
940	449	450	450	451	451	452	452	452	453	453
950	454	454	455	455	456	456	457	457	457	458
960	458	459	459	460	460	461	461	462	462	463
970	463	463	464	464	465	465	466	466	467	467
980	468	468	469	469	469	470	470	471	471	472
990	472	473	473	474	474	475	475	476	476	476

advertiser. It also indicates that the advertiser must pay special attention to statistical power when designing an equivalence test, so as not to waste resources on severely underpowered tests.

14.4 Unsurpassed Claims—Every equivalence claim implies an unsurpassed claim (subject to considerations in 14.1.3). The difference between an unsurpassed claim and an equivalence claim is that an unsurpassed hypothesis uses only one of the two limits that were used to define an equivalence hypothesis. An unsurpassed claim does not imply an equivalence claim; however, it is possible in the case of hedonic variables to conduct an unsurpassed test and limit communication to the equivalence region. For instance, when the test result supports a hedonic unsurpassed claim, the advertiser can instead make a hedonic equivalence claim (for example, “preferred as much as” or “tastes as good as”) because the latter claim type only excludes a region where the advertiser may be superior to what is claimed. However, in the case of sensory claims such as “as sweet as” or “as strong as” limiting the unsurpassed claim to the equivalence region may be misleading because the advertiser would not have tested beyond the upper limit for equivalence where consumers may perceive the product to be sweeter or stronger. For statistical reasons, sample size requirements for an unsurpassed claim are generally much lower than for an equivalence claim, which is why advertisers might prefer this approach for hedonic claims where it can be used. Table 5 provides the minimum choice counts for the advertiser’s product to make an unsurpassed claim at the 95 % confidence level. Notice that in Table 5 when sample sizes approach and exceed 300, experimental results in which the advertiser’s product receives less than 50 % of the choices can support an unsurpassed claim because at this sample size it is possible to establish that significantly more

than 45 % of the population will choose the advertiser’s product. In fact, at a sample size of 800, the advertiser could obtain results such as 48 % (advertiser): 52 % (competitor) and still claim to be unsurpassed by the competitor.

14.4.1 An unsurpassed claim is always based on responses from all consumers. There is no limitation on the percent acceptable no-preference responses when making a claim of unsurpassed. The advertiser cannot drop no-preference responses to make a qualified claim of unsurpassed.

14.4.1.1 A binomial test can be used to analyze the data. Let n be the total number of people participating in the study. Split no-preference responses according to 14.1.2 to obtain the advertiser’s choice count x .

14.4.1.2 Determine the critical value for significance based on the binomial distribution with probability $P = 0.45$ at sample size n . The critical value is the 95 % quantile of this distribution. A critical value for a sample size n that is not found in Table 5 can be obtained in Microsoft Excel⁴ or in R (see below).

Excel: =BINOM.INV(n , 0.45, 0.95)+1

R: qbinom(.05, n , 0.45, lower.tail=FALSE)+1

14.4.1.3 If x is greater than or equal to the critical value, then the unsurpassed claim is supported at the 5 % (one-tailed) level of significance. Table 5 is constructed so that the counts control Type 1 error at 5 % or less.

14.4.2 Other Analytic Approaches Can Be Used but Require Justification—The approach described in 14.1 and 14.4 is recommended for all cases. If another analytic approach will be used, then it should be documented and justified by the advertiser prior to conducting the sensory test for the unsurpassed claim.

14.4.3 Example 1: Unsurpassed Claim—In a preference test among 205 consumers (n), 100 consumers choose the advertiser's product (n_1), 91 choose the competitor's product and there are 14 no preferences (n_0). Splitting the no preferences according to 14.1.2 leads to 107 (x) counts for the advertiser's product. With a sample size of 205 (n), we determine from Table 5 that we require 105 preference judgments in favor of the advertiser to declare significance at the 95 % confidence level. The advertiser meets this requirement and can claim to be unsurpassed based on this test. (The same test results do not support an unqualified superiority claim; see the example given in 14.2.4.1.)

14.4.4 Performance Characteristics of Preference Tests for Unsurpassed Testing—If the observed preference for the advertiser's product is at least 50 %, based on a 300 respondent test, then the competitor can be 95 % certain that the true preference for the advertiser's product is no lower than 45 %. Increasing the respondent base above 300 allows the advertiser to support an unsurpassed claim with observed preferences slightly less than 50 %, while still protecting the competitor (with 95 % certainty) that the true preference for the advertiser's product is not lower than 45 %. Table 6 contains the minimum preference percentages required to support an unsurpassed claim for various numbers of respondents that protect the competitor from the worst case 45 % preference with 95 % certainty. Table 6 also presents the probability that preference tests based on various numbers of respondents will support the unsurpassed claim when the true preference for the advertiser's product is at exactly $P = 50\%$. The final column of Table 6 shows how low the actual preference proportion may be, with 95 % certainty, when a 50 % preference result is observed in a study. The information in Table 6 illustrates the advantage of larger sample sizes for the advertiser.

14.5 Paired Comparison/Difference Studies—The techniques described in 14.2 through 14.4 are also used for analyzing data from a paired comparison or paired difference

study. In a paired comparison study, each respondent is presented with two samples and is asked to select the sample that has more (or less) of the characteristic of interest. In a sense, a paired preference study is just a special case of a general paired comparison study in which the characteristic of interest is preference.

14.5.1 The same criteria used in the paired preference study for determining numbers of respondents and the number of correct answers needed to support either a superiority, an equivalence or an unsurpassed claim are used in a paired comparison study. That is, Table 1, Table 3, and Table 5 can be used to analyze the data from a paired comparison study, substituting the characteristic of interest for "preference," where the term occurs in the tables.

14.6 Analysis of Data From Scales:

14.6.1 Data from acceptance tests, descriptive-panel studies collected using unstructured line scales, magnitude estimation, or category scales with at least five points are commonly analyzed as continuous data using parametric statistical methods such as analysis of variance. Analysis of variance is used to statistically compare the average ratings of the products in the test, one response at a time.

14.6.2 Both acceptance tests and descriptive analysis panels vary widely in the number of samples involved in the study and in how the samples are distributed to the respondents who participate in the study. These issues determine the form of the analysis of variance model that is appropriate for analyzing the data from the study (see Meilgaard et al. (3) or ASTM Manual 26 (5)). For complicated or irregular product-presentation schemes, it may be necessary to consult a statistician to determine the appropriate model to analyze the data.

14.7 Guidance for Other Analyses and Claims Types:

14.7.1 Statistical analyses described in this section are written to be usable by a broad population of practitioners. These data analyses are practical and reasonable, even if they are not always theoretically optimal. For example, no-preference/difference responses can be allocated according to 14.1.2 for superiority and unsurpassed claims if there are fewer than 20 % no-preference/difference responses. The possibility of using more sophisticated models and corrections is provided because it is known that the data distribution increasingly departs from the binomial with allocation of ties according to 14.1.2 as the proportion of no-preference/difference responses increases.

14.7.2 This section provides guidance for analyzing the most common types of superiority, equivalence, and unsurpassed sensory claims. It does not preclude making sensory claims that are not described in this Guide; however, these claims should be well supported with appropriate technical justification.

15. Keywords

15.1 advertisement claim(s); claim substantiation; consumer testing; descriptive testing; sensory laboratory testing; sensory testing

TABLE 6 Performance Characteristics of Preference Tests for Unsurpassed Testing

n	Pc^A	Power ^B	LL95 ^C
100	54.0	24.2	41.4
200	51.5	36.2	44.0
300	50.0	52.3	45.1
400	49.2	63.7	45.8
500	48.8	71.9	46.2
600	48.5	78.1	46.6
700	48.3	82.8	46.8
800	48.0	87.8	47.0
900	47.9	90.3	47.2
1000	47.7	93.6	47.4

^A Pc —Minimum percent preference required to claim to be unsurpassed at the 95 % confidence level.

^B $Power$ —Probability of claiming to be unsurpassed when the actual preference for the advertiser's product is 50 %.

^C $LL95$ —Lower limit of a one-tailed 95 % confidence interval that represents how low the actual percent preference may be when a 50 % preference proportion is observed in the study.

APPENDIX

(Nonmandatory Information)

X1. COMMONLY ASKED QUESTIONS ABOUT ASTM AND CLAIM SUBSTANTIATION

X1.1 *What is ASTM?*

X1.1.1 Since it was first organized in 1898, ASTM International has grown into one of the largest voluntary standards development systems in the world. ASTM International is a nonprofit organization which provides a forum for producers, users, ultimate consumers and those having a general interest (representatives of government and academia) to meet on common ground and write standards for materials, products, systems, and services. The purpose of ASTM according to its charter is “the development of standards on characteristics and performance of materials, products, systems and services, and the promotion of related knowledge.”

X1.1.2 ASTM International believes that technically competent standards result when a full consensus of all concerned parties is achieved and rigorous due process procedures are followed. This philosophy and standards development system ensure technically competent standards have the highest credibility when critically examined and used as the basis for commercial, legal, or regulatory actions. ASTM International standards are developed and used voluntarily. Standards become legally binding only when a government body references them in regulations or when they are cited in a contract. Any item that is produced and marketed as conforming to an ASTM International standard must meet all applicable requirements of that standard.

X1.1.3 From the work of 131 standards-writing committees, ASTM International has published more than 12 000 standards each year. These standards and other related technical information are sold by ASTM International throughout the world. An ASTM International standard is subject to revision at anytime by the responsible technical committee and must be reviewed every five years, and if not revised, either reapproved or withdrawn.

X1.1.4 Committee E18 is a technical committee of ASTM International. The purpose of Committee E18 is to promote knowledge, stimulate research, and develop principles and standards for the sensory evaluation of materials and products. Committee E18 is comprised of nearly 300 industry and academia professionals—food scientists, sensory scientists, psychophysicists, statisticians, psychologists, and other professionals, representing the world’s leading universities and Fortune 500 companies. These professionals are at the forefront of new product development technology, designing and applying the appropriate sensory methods for the evaluation of food, beverage, tobacco, household and personal care products, worldwide.

X1.1.5 This guide was recommended, developed and approved by the collective membership of ASTM Committee E18, individuals who are intimately involved with the design and analysis of studies to assess product performance, and who are responsible for the interpretation and communication of

their research results to the business and professional communities. As a standard, the recommendations put forth in this document are subject to review by the Society at regular intervals, to assure up-to-date and accurate information.

X1.2 *Why ASTM Committee E18 Developed This Guide:*

X1.2.1 In November of 1990, Committee E18 held a discussion on the increased interest in sensory testing to support advertising claims. Although a number of individuals and groups had made recommendations on how to effectively conduct sensory tests for advertisement claims, there were many inconsistencies between groups.

X1.2.2 Because Committee E18 is composed of sensory professionals whose purpose is to write voluntary industry standards for this field, it seemed logical that they should attempt to review, combine, and filter individual and group recommendations into one document. Those contributing to this document represent both large and small corporations, academicians, and consultants in a wide variety of consumer products categories. The categories include but are not limited to food, beverage, cosmetics, health and beauty aids, and other related products.

X1.2.3 The goal is to provide a document that is straightforward, easy to understand, and implement. The members contributing to this guide bring together many years of experience in designing, implementing, and analyzing these types of tests. The intent is to provide a technically sound document that will be equitable for all including the advertiser, the challenger, and ultimately the consumer.

X1.3 *How Are the Members of This Committee Recruited?*

X1.3.1 After the subcommittee had been approved by the Executive Committee of E18, a general call at the main committee meeting and through ASTM publications was made to all members that this committee was now ready to begin work. Anyone, members of E18 or other interested parties, was invited to participate. The only criteria for members to receive “working documents” is that participants be “active” members, fully participating in both the decision and production processes. Members who do not wish to fully participate are welcome at any meeting to participate in the discussion and vote on issues. At each meeting the members are asked to encourage anyone in their respective companies for input or to attend the meetings personally.

X1.4 *Who is the Intended User for This Guide?*

X1.4.1 This guide is written for all those who are involved in evaluating products from a sensory perspective and supporting product claims based upon those evaluations. This encompasses anyone from those who set up product tests, to the end-users of those product test results.

X1.4.2 Within the industries devoted to developing new products or maintaining the competitive edge of existing products, the intended users include sensory evaluation and consumer research professionals, product formulators or developers, marketers, advertisers and copywriters, as well as the consumer advocates and legal professionals who may question or defend such claims.

X1.4.3 Based on the consensus of those in the forefront of current practice, this guide will direct the inexperienced practitioner or peripheral professional through the detailed heart of a complex process.

X1.5 What is the Intended Use of This Guide?

X1.5.1 Claims research usually will be scrutinized by competitors who will critically evaluate all aspects of the methodology and findings. Research must be conducted in a scientifically sound manner, or a claim based upon it will be in jeopardy. Claims research requires expertise in several disciplines, including experimental design, sampling, and statistical data analysis. In addition, methodological expertise also is required because executional factors, and question content can affect the outcome of the research. This guide recommends best practices from a technical perspective based on the expertise and experience of research professionals.

X1.5.2 Ultimately, the advertising media, and in the case of disputes, arbiters, determine the adequacy of research as substantiation for a claim. This guide will not alter these roles. The intent is to assist and strengthen decisions by claimants, competitors, and those who need to evaluate research by identifying technically sound practices, which comprise valid research.

X1.5.3 As a set of guidelines, this guide is not intended to be prescriptive. In many cases, there may be more than one reasonable approach, and the pros and cons of each option must be weighed carefully to determine the best approach. This guide is an aid to judgment, and it is hoped that it will help those with a vested interest in claims substantiation research be knowledgeable about the subject.

X1.6 What Are the Applications of This Guide?

X1.6.1 This guide can help those considering advertising claims by discussing the key factors, which can impact the validity of claims substantiation research. As such, it can help readers decide whether to pursue a claims test and design valid research that will have the best chances of withstanding challenge. Another application is to help critically evaluate existing research. This application can be used in one's own research to decide whether it should be used to substantiate a claim or to evaluate others' research to decide whether a challenge is worth pursuing. Media clearance personnel, attorneys, and arbiters can use this guide to help develop positions on the adequacy of research in question.

X1.7 What Are the Limitations of This Guide?

X1.7.1 Unlike many physical tests for which ASTM standards have been written, the scope of this guide is too diverse for a uniform specification. It provides guidelines for practices,

which comprise scientifically valid claims research. Since no single universal method is specified, claimed conformity with the guidelines cannot substitute for detailed description of the research methodology.

X1.7.2 This guide is not intended to serve as a template or "cookbook" for all situations. Each situation is unique and what is reasonable will be determined by the objectives of the test. There is no panacea; a rationale will be required, and research will always need to be tailored to the situation at hand.

X1.7.3 Discussion of specific methodologies is not intended to limit the types of approaches or methodologies, which could be used in claims substantiation research. Ultimately, any reasonable, methodologically sound approach should be considered for claims support. As in other fields of research, there are a number of issues upon which qualified practitioners do not agree. Where this is the case, the pros and cons of some alternatives are discussed.

X1.8 How Are Statistical Criteria Determined?

X1.8.1 The statistical criteria related to paired-preference claims have been developed through extensive discussions and consensus decisions of the task group participants. For example, a paired-preference test becomes more sensitive as the number of respondents increase. "Sensitivity" in a rigorous statistical sense is based on three criteria: (1) the smallest difference in preference proportions (that is, the advertiser's versus the competitor's) that is deemed to be meaningful, (2) the probability that the test will be significant when the difference between the preference proportions equals the meaningful difference (that is, the "power" of the test), and (3) the level of risk that is deemed acceptable for falsely concluding that a difference in preference exists when, in fact, it does not. Once values for these three criteria are selected, the number of respondents necessary to deliver that level of "sensitivity" can be computed using basic statistical techniques.

X1.8.2 For superiority, equivalence, or unsurpassed claims, it has been decided to protect the competitor against adverse outcomes resulting from insensitive tests. The advertiser has control over the sensitivity of the test, and therefore, is free to increase the number of respondents to values that correspond to his selected levels of acceptable risk without compromising the fair levels chosen for the competitor by the task group.

X1.8.3 Historically the term "parity" appeared in this guide; however, parity describes multiple claim types. The guide now uses the more specific terms equivalence and unsurpassed, which provide greater clarity and simplify the presentation of the appropriate statistical tests.

X1.9 When is Descriptive Analysis the Best Method to Use for Claim Support?

X1.9.1 When desiring to demonstrate the strength of one sensory attribute (for example, color, minty, sweet, shine, sticky) is more, less, or equivalent to that of a competitor.

X1.9.2 When desiring to demonstrate that treatment with the product increases or decreases a specific perceived property (for example, underarm odor, peanut flavor, dry skin).

X1.9.3 Descriptive analysis is not a good method if it is desirable to know about or make a claim about liking, goodness, preference, or any other subjective consumer-type response.

X1.10 How Does Descriptive Analysis Differ From Tests With Regular Consumers?

X1.10.1 Descriptive panels are highly trained and behave more like analytical tools or instruments in that they only describe what attributes are perceived and how strong they are. There is no indication of preference or liking.

X1.11 How Many People Participate in a Panel?

X1.11.1 Descriptive panels can have smaller base sizes than studies using consumers because they use trained panelists. Trained panelists can show differences between or among samples after effective training and validation because attributes are clearly defined and panelists are familiar and practiced with attribute evaluations. These help reduce data variability. For actual numbers of panelists used in typical descriptive studies, the reader is referred to recent articles in refereed publications such as *Journal of Sensory Studies* or *Food Quality and Preference*.

X1.12 How Many Attributes Are Evaluated by the Panel to Make an Advertising Claim?

X1.12.1 Only the sensory attributes (terms, properties, characteristics) about which a claim is to be made should be rated for intensity by a panel.

X1.13 Can a Descriptive Panel in One Geographic Area Test a Product That is Sold or Used in Another?

X1.13.1 Any descriptive analysis panel, that has been properly trained, can test a product or sample from anywhere in the

world. The panel does not represent some segment of the population, but rather, represents the ability of humans to discriminate (detect) and describe properties and their strength.

X1.14 How Much Training is Necessary to Prepare a Descriptive Panel?

X1.14.1 The amount of training depends on a number of factors, such as objectives and product complexity.

X1.14.2 If the panel has been trained and validated previously, the training for complex products with complex attributes will still require adequate training for the product type and research objectives. For further information, see ASTM Manual 13 (4).

X1.15 What Other Means Are Available to Ensure That a Claim is Defensible?

X1.15.1 The members of the Guide E1958 Sensory Claims Substantiation Workgroup acknowledge an early contribution by Ron Smithies, then director of the National Advertising Division (NAD) of the Council of Better Business Bureaus.⁵ He met with individuals who led the emergent workgroup to share some of his expertise on claim support. He outlined a principle for substantiating a sensory claim using evidence-based on three types of data: consumer data, instrumental data, and trained panel data. (Although instrumental data are not discussed in this guide, this type of data can be very helpful in support of a sensory claim.) He advised that a sensory claim that is substantiated by (at least) two of these three types of data makes a very strong case for the claim. Serious consideration should be given to designing studies accordingly.

⁵ As of June 2019, the NAD is a self-regulatory unit in the BBB National Programs, Inc.

REFERENCES

- (1) Code of Federal Regulations, Title 21 Food and Drugs (Vol. 1). Part 11 Electronic records; electronic signatures. National Archives and Records Administration's (NARA) Office of the Federal Register (OFR). January 1, 2024.
- (2) Cochran, W. G., and Cox, G. M., *Experimental Designs, 2nd Edition*, Wiley, New York, NY, 1957.(Reissued 1992 as part of Wiley Classics Library.)
- (3) Meilgaard, M., Civille, G. V., and Carr, B. T., *Sensory Evaluation Techniques, 5th Edition*, CRC Press, Boca Raton, FL, 2016.
- (4) ASTM MNL 13, *Descriptive Analysis Testing for Sensory Evaluation*, Second Edition, ASTM International, West Conshohocken, PA, 2020.
- (5) ASTM MNL 26, *Sensory Testing Methods*, Second Edition, ASTM International, West Conshohocken, PA, 2020.
- (6) *Physical Requirement Guidelines for Sensory Evaluation Laboratories*, ASTM STP 913, ASTM International, 2009.
- (7) R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2021. Available: www.R-project.org
- (8) Ennis, D. M. and Ennis, J. M., "Accounting for no-difference/preference responses or ties in choice experiments," *Food Quality and Preference*, Vol 23, No. 1, 2011, pp. 13–17.
- (9) Ennis, D. M., Rousseau, B., and Ennis, J. M., *Tools and Applications of Sensory and Consumer Science*, The Institute for Perception, Richmond, VA, 2018.

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), or through the ASTM website (www.astm.org/contact). Permission rights to photocopy the standard may also be secured from the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Tel: (978) 646-2600; <http://www.copyright.com/>