Designation: E1207 – 14

Standard Guide for Sensory Evaluation of Axillary Deodorancy[1]

This standard is issued under the fixed designation E1207; 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.

1. Scope

1.1 This guide provides procedures which may be used in the design and analysis of studies to quantitatively assess the intensity of human axillary odor for the purpose of substantiating deodorant efficacy of personal care products.

1.2 This guide includes protocols for the selection and training of assessors, selection of subjects, experimental design, and statistical analyses. This practice is limited to assessment of axillary odor by trained assessors. Self-evaluation protocols are valid for selected sensory tasks but may be less sensitive.

1.3 With respect to the source of axillary odor, three groups of secretory glands are present in the axillae which participate to a greater or lesser extent in its production—eccrine, apocrine, and sebaceous. Axillary odor has been primarily ascribed to the apocrine gland secretion (**1**).[2] Body odor intensity has been correlated with the volume of the secretory portion of the apocrine gland (**2**) and the density of the glands.

1.3.1 Apocrine glands are found primarily in the axillary vault in conjunction with axillary hairs (**3**). Pure apocrine sweat is sterile and odorless and axillary odor results from degradation of apocrine sweat by resident skin bacteria (**4**). High bacterial populations are found in moist regions of the body, especially in the axillae, providing the appropriate environment for growth (**5**).

1.3.2 Eccrine glands keep the axillae moist through thermally and emotionally induced secretions (**6**).

1.3.3 The sebaceous glands excrete higher molecular weight lipid materials which absorb and retain the volatile materials resulting from bacterial action (**7**). The aerobic diphtheroids are able to produce the typical acrid axillary odor and the micrococcaceae produce an isovaleric acid-like odor when incubated with apocrine sweat (**8**). Therefore, the most undesirable component of axillary odor is caused by degradation of apocrine sweat by particular bacteria normally found in the axillary vault.

1.4 Personal care products are sold and used primarily for their ability to reduce the perception of body odor not only by the individual using the product but also by individuals within the scope of contact. Deodorant protection may be achieved by these products through various modes of action. Antiperspirants achieve their primary efficacy by means of the action of inorganic salts on the eccrine gland production of sweat. Antimicrobial agents achieve deodorancy by inhibiting the growth and activity of the microflora in the axillary vault thus reducing the microbial decomposition of sweat and the consequent production of body odor. Absorbents function either by "binding" available moisture or malodorous substances. Fragrances are effective by altering the perception of malodor and increasing the degree of "pleasantness." Other modes of control become important from time to time, representing changes in the state-of-the-art in product development.

1.5 The studies discussed herein are interpreted through the use of statistical tests of hypotheses. These hypotheses are usually of the form:

The Deodorant Efficacy of Treatment A
= The Deodorant Efficacy of Treatment B

1.5.1 It should be noted that failure to reject this hypothesis at a specified level of significance does not prove the hypothesis, but merely that the weight of evidence provided by the experiment is not sufficient to reject the hypothesis. This could occur because either: *a*) The hypothesis is close to truth and great experimental power would be required to reject it, or *b*) The experiment by design was low in power and, therefore, incapable of rejecting the hypothesis; even when it is far from true. This can occur due to design structure or low sample size. These facts must be taken into consideration when interpreting study results.

2. Referenced Documents

2.1 *ASTM Standards:*[3]

E253 Terminology Relating to Sensory Evaluation of Materials and Products

E1697 Test Method for Unipolar Magnitude Estimation of Sensory Attributes

[1] This guide is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation and is the direct responsibility of Subcommittee E18.07 on Personal Care and Household Evaluation.

Current edition approved March 1, 2014. Published March 2014. Originally approved in 1987. Last previous edition approved in 2009 as E1207 – 09. DOI: 10.1520/E1207-14.

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

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

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

3.1.1 For definitions of terms relating to sensory evaluation, see Terminology E253.

3.1.2 *5-alpha-androst-16-en-3-one* (delta^{16}(5-alpha) androsten-3-one) $C_{19}H_{28}O$—CAS No. 18339-17-7—component of axillary odor which has a "urinous" character and results from the action of certain skin bacteria on apocrine secretion **(9)**.

3.1.3 *5-alpha-androst-16-en-3-alpha-ol* (delta^{16} (5-alpha) androsten-3-alpha-ol) $C_{19}H_{30}O$—CAS No. 14152-27-3—component of axillary odor which has a "musky" character and results from the action of certain skin bacteria on apocrine secretion **(9)**.

3.1.4 *apocrine gland*—a highly coiled tubular system found primarily in axillary epidermis. These glands continuously produce and store apocrine sweat for later excretion onto the skin surface via hair follicles. The excretion is activated by androgenic sympathetic stimuli such as pain or fear **(1)**.

3.1.5 *deodorant efficacy*—the effectiveness or treatment, or both, of a product in reducing axillary malodor.

3.1.6 *eccrine gland*—a simple unbranched tube with a terminal coil. These glands are found in the epidermis over the entire body surface. The glands are controlled by the autonomic nervous system and serve as an evaporative cooling mechanism. Although heat is the primary stimulus, localized eccrine sweating can also occur as a result of emotional stress and other physiological stimuli **(3)**.

3.1.7 *IVA, isovaleric acid* (3-methylbutanoic acid) $C_5H_{10}O_2$; $(CH_3)_2CHCH_2COOH$. CAS No. 503-74-2—component of axillary odor which has a "sweaty, acid" character and results from the action of certain skin bacteria on apocrine secretion.

3.1.8 *right-left imbalance*—a condition of some subjects who have one axilla with notably more intense odor than the other axilla as determined from the control odor evaluation.

3.1.9 *sebaceous gland*—a gland closely related to the hair follicle which produces sebum which combines with apocrine secretion at the base of the follicle. Sebaceous glands are under androgen control **(6)**.

3.1.10 *sequential analysis*—a statistical technique which may be used to screen potential assessors for sensory acuity to a specific stimulus. The assessor is repeatedly tested until he or she passes or fails the test at a specified level of significance **(10, 11)**.

3.1.11 *trigeminal response*—a sensation caused by stimulation of the trigeminal nerve. The sensation is that of a physical feeling, such as burning and tingling.

4. Summary of Guide

4.1 The protocols described provide for the designation of panels of individuals suitably selected and trained to perform the functions of assessors and subjects for the purpose of assessing deodorant efficacy. Details of specific procedures are given in Appendix X1 – Appendix X3. Deodorant products should be tested in a manner which maximizes test sensitivity while still reflecting normal consumer-use conditions. Examples are provided to assist the investigator in the design and performance of test protocols.

5. Significance and Use

5.1 The procedures recommended in this practice can be used to clinically assess axillary deodorant efficacy of personal care products.

5.2 This practice is applicable to the product categories which include deodorant and toilet soap bars, liquid bath soaps and gels, deodorant sticks, antiperspirants, creams and lotions, body talcs, and aerosol and pump delivery deodorants, antiperspirants, and body colognes.

5.3 Procedures of the type described herein may be used to aid in the communication of efficacy within and between manufacturers and to the consumer through the various public communications media. Guidelines are suggested due to the need to determine the relative or absolute performance of experimental materials or of commercial products.

5.4 These procedures may be used by persons who have familiarized themselves with these procedures and have had previous experience with sensory evaluation.

5.5 This practice provides suggested procedures and is not meant to exclude alternate procedures which may be effectively used to provide the same clinical result.

6. Subject Selection and Restrictions

6.1 *Criteria for Selection*—The population should be defined and subjects selected from this population in a random, and unbiased manner according to the experimental design considerations defined in 8.11. If a test is being performed with the product directed at a subset of the consuming population, the subjects should be selected from a population representative of the subset.

6.1.1 The subjects should have a recognizable body odor level when evaluated under the procedures given in this practice.

6.1.2 In situations where it is desirable to enhance test sensitivity, the following criteria may be adopted:

6.1.2.1 Based on the control odor scores (see 8.3), subjects who have low or extremely high odor should not be selected for the test. Subjects may be considered as having a "high" odor relative to a normal population if they develop an odor score in excess of 7.0 on a 0- to 10-point scale or 3.5 on a 0- to 5-point scale. Likewise, subjects may be considered as having a "low" odor relative to a normal population if they develop an odor score below 3.0 on a 0- to 10-point scale or 1.5 on a 0- to 5-point scale. A selection process which excludes "low" odor subjects or "extremely high" odor subjects, or both, must be specified for each test and depends upon the number of subjects required for the test and the relative odor scores of these subjects.

6.1.2.2 There should be no more than a small right-left odor imbalance between axillae of each subject. On the basis of a category, or interval scale, the consensus of the task group was that the control odor score differential should not be greater

than 20 % of the overall scale (that is, 2.0 points on a 10-point scale or 1.0 points on a 5-point scale).

6.1.2.3 Appendix X1 contains additional information on the acceptance/rejection history of experimental subject populations. A selection process which excludes approximately 20 % of the lowest odor intensity individuals of a normal population is generally recognized as appropriate.

6.1.3 Chronic medications such as antibiotics, steroids, etc., which may affect the test, should be restricted during all test phases as deemed appropriate by the sponsor.

6.1.4 In addition to the above restrictions it should be recognized that other factors which contribute to protocol operating efficiency should be emphasized, including interest, cooperation, commitment, and punctuality of the subjects.

6.2 *Subject Restrictions*—In order to achieve appropriate experimental control, the following restrictions should be imposed upon all subjects during the conditioning and test phases.

6.2.1 *Conditioning Phase*—This period is often referred to as the "washout" period and is that portion of the protocol preceding the actual test phase. The duration of the conditioning phase should be a minimum of 7 days. The conditioning phase for antiperspirants shall be 17 days as defined by the FDA monograph on antiperspirants (**11**).

6.2.1.1 Subjects should use no antiperspirants, deodorants, antibiotic creams, antibacterial ointments, or any other cosmetic products on the axillae. No antibacterial products, including deodorant and medicated shampoos should be used. Care should be taken not to expose the axillae to any medicated product or product containing alcohol.

6.2.1.2 Subjects should use only the control cleansing agent(s) provided by the sponsor as instructed for personal hygiene.

6.2.1.3 Swimming should be stopped at least 7 days prior to the test phase and during the entire test phase.

6.2.1.4 Subjects who normally shave their axillae should shave using the control cleansing agent no less than 24 h prior to the control evaluation and abstain from shaving for the duration of the test.

6.2.1.5 Spicy foods, including garlic and onions should be restricted 24 h before the control evaluation and during the test phase.

6.2.1.6 It is acceptable to use smokers as subjects, but they are required to refrain from smoking for 2 h before all evaluations.

6.2.2 *Test Phase*—In addition to the conditions detailed for the subjects during the conditioning phase, the following restrictions are required of the subjects during the test phase:

6.2.2.1 Subjects should use no perfumed substances on the body such as perfume, after shave, lotions, bath oils, and hairspray.

6.2.2.2 Pre-laundered wearing apparel (see 8.6) may be worn by each subject at the option of the test sponsor. Shirts should be collected and laundered in accordance with a uniform laboratory procedure.

6.2.2.3 If specified by the test sponsor, laundry additives such as bleach, fabric softeners, etc., may be used on subjects' outer clothing.

6.2.2.4 Subjects should minimize physical exertion such as tennis and jogging.

6.2.2.5 Subjects should refrain from the use of breath mints, toothpaste, mouth rinses and sprays, chewing gum, and from drinking coffee or tea at least 1 h prior to each evaluation. Smoking should be restricted 2 h prior to each evaluation and alcoholic beverages 8 h before an evaluation.

6.2.2.6 Subjects should not wash the axillae at home for the duration of the test. Axillae should only be washed at the test site in accordance with a supervised wash procedure. Care should be taken not to get the axillae wet during bathing or showering at home.

7. Assessor Selection and Training

7.1 *General*—The selection process should include the principles embodied in Ref (**12**). The assessor's task is to detect differences and rate the intensity of perceived axillary odor.

7.2 Assessors employed for assessing body odor intensity should be screened for the following attributes:

7.2.1 Interest and availability;

7.2.2 Qualitative and quantitative olfactory discrimination ability;

7.2.3 Ability to carry out basic sensory tasks, and competency with the scale used, and

7.2.4 Specific anosmias. While it is desirable to identify any olfactory deficit which an assessor may have, there is experience which indicates that specific anosmias may not detract from accurate odor judgments. (See X2.6.3)

7.3 Recommended procedures are presented in Appendix X2 for the screening and selection of *in vivo* deodorancy assessors.

7.4 *Assessor Training*—In addition to the following points, the recommended procedures are given in Appendix X3 for the training of *in vivo* deodorancy assessors.

7.4.1 Assessors should be exposed to the complete range of quantitative and qualitative malodor stimuli which they will later be asked to rate. This establishes the context in which ratings are to be assigned.

7.4.2 *Assessor Training for Category Scales:*

7.4.2.1 After being introduced to the rating scale procedure, assessors should assign ratings to the stimuli in an open discussion to obtain a consensus rating for each stimulus.

7.4.2.2 Assessors should be drilled until the ratings they independently assign match those obtained by consensus as closely as possible. Assessors whose ratings disagree with the consensus rating much more often than those of most other assessors should be eliminated. The criteria for rejection of individual assessors must be developed in each laboratory. For example, the responses for each assessor can be graphed to determine if they fall within a specified range across time.

7.5 *Assessor Performance Monitoring*—Trained assessors should be tested periodically to confirm their ability to discriminate (rankings, paired comparisons, ratings can be used as appropriate). In order to evaluate rating performance, it is also important to evaluate within- and between-assessor consistency. On a more routine basis, treatments used for the purpose of scale anchors or reference standards can be included in the

regular testing regimen as "unknowns" to determine if assessors are capable of rating these products consistently. The procedure for monitoring assessor performance should be carried out at least once a year. More frequent monitoring may be required if there is some reason to suspect an assessor's olfactory acuity. (See X3.3).

8. Test Design

8.1 *Subject Enrollment*—A sufficient number of subjects should be enrolled for the conditioning phase so that the required number of subjects complete the study. The number enrolled will depend upon the history of the laboratory and the specific selection criteria for the test. In general, it is suggested that at least 20 % more subjects be recruited than will be needed. Each subject should be informed of the responsibilities and obligations of the subjects, provided with a copy of the restrictions and advised of any regulations and consent applicable under the proposed good clinical practices and any applicable regulations covering the obligations of sponsors/investigators.

8.2 *Conditioning Phase*—Each subject should adhere to the restrictions given in 6.2.1. Each subject should be provided with the appropriate control cleansing products for personal hygiene at home during this phase which are to substitute for products normally used, such as liquid soap, bar soap, and shampoo, or all three. These products should contain no antimicrobial ingredients and a minimum level of perfume or no perfume.

8.3 *Control Odor Scores*—This evaluation is conducted to determine baseline axillary odor scores for each subject following a supervised control wash using the control cleansing product. The purpose is to uniformly condition the subjects' axillae prior to the control evaluation. Subjects may then be screened from the test if they have unacceptably low or high odor or have an accentuated right-left imbalance (6.1.2.2). The time interval between the control wash and the control evaluation should be the same as the longest time interval between test product application and axillary odor evaluation. The soap used for the control wash should be the same as the one used by the subject during the conditioning phase. The specified number of subjects will be selected on their control odor scores in accordance with the selection criteria detailed in 6.1.

8.4 *Post-Treatment Evaluation Interval*—The post-treatment evaluation interval may range from immediately after treatment to 30 min to 48 h, or more. The specific interval will be based upon the expected end-product use and the anticipated claim substantiation documentation required. Frequently used post-treatment evaluation intervals are 5, 8, 12, and 24 h.

8.5 *Duration of Test Period (Treatment Cycle Duration)*—During the test phase of the study the subjects are treated with one or more designated test products and evaluated for odor level. Individual product test periods range from 1 to 21 days depending upon the test objective, the test sensitivity desired, the product formulation, and the expected end-product use conditions. Generally, 3 to 5 sequential test days will provide sufficient data to document performance claims.

8.6 *Wearing Apparel*—For studies in which wearing apparel is to be controlled, shirts of uniform fiber content, either cotton or a cotton-polyester blend, but not nylon, should be used. Apparel style may be either T-shirts or dress shirts. All wearing apparel should be laundered immediately prior to use using an unfragranced detergent base. Each subject should be issued a fresh shirt after each product application to be worn at least through the first evaluation point. If successive evaluations are made between applications, the test sponsor should determine if the same shirt is to be worn, a fresh one to be issued, or if the subjects are to be allowed to assume normal clothing habits.

8.7 *Product Assignment*—Test products should be randomly assigned to right and left axilla such that each product is applied to an equal number of right and left axillae. Specific experimental designs are given in 8.11.

8.8 *Test Product Application:*

8.8.1 For deodorant sticks, gels, creams and lotions, body talcs, aerosol and pump delivery deodorants and body colognes, the axillae should be cleansed prior to treatment using a control cleansing agent. It should be determined that such treatment does not impart a residual odor or produce a false treatment effect. Deodorant and toilet soaps and liquid bath soaps and gels provide for normal axillary cleansing during the application process.

8.8.2 All axillary treatments during the test phase should be monitored by a test supervisor. The level of supervision depends upon the experience and number of subjects involved and the product tested.

8.8.3 Specific recommendations for each product category application condition are given in Appendix X4.

8.9 *Test Product Evaluation:*

8.9.1 This is an example of one specific method of evaluation. Odor assessors are positioned in isolated evaluation stations in the odor evaluation room. Subjects (equal to the number of assessors) enter the room and randomly report to the assessors' stations so that each assessor has a subject to evaluate. The subjects stand in front of the designated assessor with their arms held at their sides for 1 min. At the completion of the 1-min interval, a signal is given and the assessors evaluate the subjects in front of them, right arm first followed by the left arm (procedure of right then left is held constant for all subsequent evaluations). During evaluation, subjects raise their right arms and then place their right hands behind their heads. Each assessor takes a sniffing cup (cone-shaped 5-oz paper cup with the pointed end cut off) and places the larger opening of the cup in the center of the right axilla and then sniffs the circumscribed area through the opening at the back end of the cone. Each assessor records the score into the record form while the subjects lower their arms. This procedure is repeated for the left arm. The subjects advance to the next designated assessor and the sniffing process is repeated. Once all the subjects in the first group have been evaluated by each assessor, this group of subjects is released from the evaluation area and the next group of subjects is brought into the room.

8.9.2 Assessors are given breaks after approximately every 20 evaluations, both arms of 10 subjects. Each judge uses a new sniffing cup for each evaluation.

8.9.3 Environmental conditions should be cool room temperatures (68°F) with sufficient air flow but no drafts.

8.10 *Odor Assessment Rating Scale*—Category scaling is very commonly used to rate axillary odor intensity but any scale used in sensory evaluation to rate intensity, including magnitude estimation (see Test Method E1697), is appropriate.

8.10.1 *Category Scaling of Axillary Odor:*

8.10.1.1 *Introduction*—This section describes the use of category scaling as one subjective rating method for axillary malodor measurement. Category scales are the oldest and most frequently used scaling methods for subjective evaluations. The use of category scales for the measurement of axillary malodor was reported in 1967 (**13**). The deodorancy assessors for the studies by Whitehouse and Carter used a 0 to 10 point scale, with "0" meaning no odor, and "10" meaning extremely strong odor. This section discusses background, applications and statistical considerations in using category scales for axillary odor evaluations.

8.10.1.2 *Background*—Category scales applied to deodorancy testing consist of a series of consecutive numbers, each of whose values represent a "level of odor." Two common category scales applied in deodorancy testing are [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] and [0, 1, 2, 3, 4, 5].

8.10.1.3 Considerations which arise in the application of category scales to deodorancy testing include the following. Assessors may tend to use only the low end or the high end of the scale, and not use the entire scale, thus skewing the distribution. There is often an inherent tendency on the part of some assessors not to use the endpoints of the scale. The distribution of category scales is discrete in nature, where often the distribution assumed by the statistical analyses applied is continuous. The psychological difference between two consecutive categories may vary, depending upon their location in the scale.

8.10.1.4 *Application*—Steps may be taken to diminish some of the difficulties encountered in the use of category scales. Training assessors to use the entire scale can reduce problems of skewness and tend to make assessors more consistent with each other in their evaluations. Having assessors compare scores during training sessions will also improve consistency. As assessors gain experience with a particular scale, they tend to mentally anchor the scores to particular odor levels. Another means of improving consistency is to train assessors using calibrated samples of odor as reference points for each category. To reduce problems of discontinuity, it is advisable to use several assessors (at least three) and take the averaged scores as the estimate of odor for a particular axilla.

NOTE 1—It is generally recognized that assessors find it difficult to psychologically accommodate more than 10 or 11 points in a scale. With scales consisting of a greater number of points, assessors may stay in one portion of the scale without using all points available, thereby reducing consistency and adding confusion to the evaluation process. However, scales consisting of a larger number of points reduce discontinuity in the data. Thus, a scale of approximately 10 intervals offers a good compromise between these two considerations. The problem of having consecutive scores represent consistent psychological differences across the entire scale may not be overcome by assessor training. However, in practical terms, these slight distortions are not viewed to be a serious detriment to applying statistical analysis to category scales in deodorancy testing. Category scales provide a heuristic approach to the evaluation of deodorancy odor which has stood the test of time, and are widely held to be an appropriate response variable to which statistical analysis can be applied.

8.11 *Experimental Design Considerations*—Include unidentified controls within the test design. This will help to check assessor performance and may shed light on anomalies within the test.

8.11.1 *Introduction to Relevant Experimental Designs*—Let T_1, T_2, ... T_t symbolize t deodorant treatments. These may include: commercial products, experimental substances, placebo formulations, or a null treatment (an "untreated side").

8.11.1.1 The three experimental designs commonly used in deodorant clinical tests are the Single Pair (1PR) Design, the Each versus Control (EVC) Design and the Round Robin (RRB) Design. Examples of the treatment assignment for each are shown in Table 1.

8.11.2 *Single Pair (1PR) Design*—This design is applicable when only two treatments are compared. Each subject receives either T_1 on the left axilla with T_2 on the right axilla or T_2 on the left with T_1 on the right. The assignment of treatments to axillae is randomized in such a way that each treatment appears an equal number of times on each axillae (or as near to an equal number of times as possible).

8.11.3 *Each Versus Control (EVC) Design*—This design is applicable when three or more treatments are to be compared, and one of the treatments, symbolized by T_t, can be singled out as the control treatment. Carefully consider the choice of the control sample. It may be a different treatment, unfragranced base, treatment with water, or no treatment. The remaining treatments, T_1, T_2, ... T_t–1, are termed test treatments. Each subject receives the control treatment on one axilla and one of the t–1 test treatments on the other axilla. Each test treatment is randomly assigned to an approximately equal number of subjects. The assignment of treatments to the left and right axillae is random, but balanced so that each treatment appears the same number of times on the left as it appears on the right or as near to the same number of times as possible. A group of subjects all of whom receive the same pair of treatments (ignoring left/right assignment) is termed a cell. The EVC design has t–1 cells.

8.11.4 *Round Robin (RRB) Design*—The RRB design is applicable when three or more treatments are to be compared but none of them can be singled out as a control treatment. There are $t(t–1)/2$ possible pairings of t treatments (for example, the three treatments, T_1, T_2, and T_3, generate the $3(3–1)/2 = 3$ pairs T_1T_2, T_1T_3, and T_2T_3). In the RRB design each of the $t(t–1)/2$ possible pairs is randomly assigned to an

TABLE 1 Examples of Treatment Assignment for Three Deodorant Clinical Study Designs

Single Pair			Each vs. Control			Round Robin		
Subject	Left	Right	Subject	Left	Right	Subject	Left	Right
1	T_1	T_2	1	T_1	T_3	1	T_3	T_2
2	T_2	T_1	2	T_3	T_1	2	T_3	T_1
3	T_2	T_1	3	T_3	T_2	3	T_1	T_2
4	T_2	T_1	4	T_1	T_3	4	T_2	T_3
5	T_1	T_2	5	T_3	T_2	5	T_1	T_3
6	T_1	T_2	6	T_2	T_3	6	T_2	T_1
			7	T_3	T_1			
			8	T_2	T_3			

approximately equal number of subjects. As in the other designs, the assignment of treatments to the left and right axillae is random but balanced, so that each treatment appears on the right the same number of times as on the left or as near to the same number of times as possible. Clearly, there are $t(t-1)/2$ cells in a RRB design.

8.11.5 *Order of Evaluation*—The order in which the assessors evaluate the subjects' axillae, either left first then right or right first then left, is held constant throughout any study; thus, the effect of presentation order cannot be estimated independently of left/right effects. Only the sum of the two effects may be estimated.

8.11.6 *Choice of Sample Size:*

8.11.6.1 *Background*—The choice of sample size is an important one, directly affecting the power and the cost of a study. The greater the sample, the more power achieved, and the greater the cost. Below are given some general guidelines for choice of sample size in deodorancy studies. See Refs **(14-16)** for technical discussions of sample size determinations.

8.11.6.2 In general, deodorancy studies will involve 30 to 60 subjects per treatment pair, depending upon the analysis used and the power required. Depending upon the application, one might require as few as 20 panelists for rough approximations, or as many as 100 or more panelists for studies involving many products and requiring high power. If the experimenter, based on past experience, knows that the particular products being tested generally show large differences in efficacy, then a smaller sample may be more cost effective. On the other hand, if he suspects that the products are quite close in deodorant efficacy, then he will want to increase the sample size to enhance the power of the study so that he will be more likely to detect the differences between the products, if in fact meaningful differences exist (see 1.5). A pilot study may be used to determine sample size needs.

8.11.6.3 If the experimenter is testing more than two products, and knows the approximate sample size (for the power required) were he testing only two of these products, using the single pair (1PR) design, the following gives the correct sample size to use for both the Each versus Control (EVC) and the Round Robin (RRB) design:

(a) Each Versus Control Design—To achieve the same precision (standard deviation) in comparing each of several test treatments with a single control that would be obtained by comparing only one of those treatments with the control in a single pair design, requires that the experimenter use a sample size equal to the number of test treatments (excluding the control) multiplied by the number of panelists he would use for the single pair study. If the experimenter would like to compare each test product with another (as opposed to testing the test product with the single control) with the same precision as that obtained in a single pair study, then he must use two times the number of test treatments (excluding the control) times the number of panelists he would use in the single pair study.

(b) Round Robin Design—To obtain the same precision between all pairs of products in a round robin design that would be obtained by testing two of those products in a single pair design requires that the experimenter use a sample approximately equal to "$(t-1)$" times the number of panelists used in the single pair design, where "t" is the total number of products being compared (see Appendix X5).

8.11.6.4 Determining sample size can be difficult, especially in cases where no prior information about the products being tested is available. In this case, it is probably better to overestimate rather than underestimate the sample size, thereby achieving the power required (see Appendix X5).

9. Biasing Effect of Fragrances

9.1 Odor assessors are trained to assign ratings to the intensity of axillary malodor ignoring any fragrance or base odor of the axillary treatment (see X3.2.2). In studies where all axillary treatments have the same fragrance, any effects these fragrances may have upon the ratings of axillary malodor intensity will be the same for all axillary treatments and, therefore, will not bias estimates of the differences in deodorant efficacy of the treatments. In studies where there are noticeable differences in the fragrances of the axillary treatments, the structure of the studies described herein does not preclude the possibility that estimates of the differences in deodorant efficacy of the products will be biased by the fragrance differences, that is, the assessors can't be fully blinded when the axillary treatments have noticeably different fragrances.

9.1.1 Some of the possible biasing effects are given in 9.1.1.1 – 9.1.1.3.

9.1.1.1 *Recognition Effect*—The effect of recognizing the identity of the fragrances as those of commercially available products.

9.1.1.2 *Affective Effect*—The effect of differences in the pleasantness of the fragrances.

9.1.1.3 *Expectation Effect*—The effect of learning part way through the study that some fragrances are usually associated with lower (or higher) malodor so that, by the later subjects, the assessors begin to expect lower (or higher) malodor ratings when those fragrances are recognized.

9.1.2 The potentially biasing effects of axillary treatment fragrances are not precluded by the design of these studies; however, there is no known alternative test method for assessing axillary deodorant efficacy.

APPENDIXES

(Nonmandatory Information)

X1. SUBJECT ACCEPTANCE/REJECTION HISTORY

X1.1 *General*—In an attempt to demonstrate the historical acceptance of subjects onto deodorancy tests, 25 prior studies were reviewed from 1 source and 4 studies from an additional source.

X1.1.1 The data presented are for control odor evaluations that are carried out prior to acceptance onto the study. Subjects have been through several days of abstinence from deodorants, antiperspirants, and deodorant soaps. In addition, a 24-h period has occurred since the axilla have been washed with a non-deodorant soap.

X1.1.2 The scoring scales used to rank the axillary odor were as follows:

X1.1.2.1 *25 Studies:*

0 = No axillary malodor

10 = Very strong and disagreeable malodor

(1 point units are used to rank the scale from 0 to 10)

4 Studies:

0 = No axillary malodor

5 = Very strong and disagreeable malodor

(½ point units are used to rank the scale from 0 to 5)

X1.2 The following tables show the distribution of the accepted/rejected subjects. The basic criteria for acceptance was the highest average scores for those subjects presenting themselves for the control odor evaluation.

TABLE X1.1 Distribution of Accepted/Rejected Subjects

Number of Tests	25
Number of Odor Assessors	4
Total Number of Subjects Screened	1066
Total Number of Subjects Accepted	845
Scoring Scale	0–10

Range of Average Control Odor Scores	Right Axilla, % Total	Left Axilla, % Total
0	0.0	0.0
0.1–1.0	0.2	0.1
1.1–2.0	1.1	1.1
2.1–3.0	6.6	7.5
3.1–4.0	14.4	15.5
4.1–5.0	27.4	24.2
5.1–6.0	30.5	29.2
6.1–7.0	17.2	16.9
7.1–8.0	2.6	4.5
8.1–9.0	0.5	0.3
9.1–10.0	0.0	0.0

X1.3 Subjects may be considered as having a “high” odor relative to a normal population if they develop an odor score in excess of 7.0 on a 0- to 10-point scale or 3.5 on a 0- to 5-point scale. Likewise, subjects may be considered as having a “low” odor relative to a normal population if they develop an odor score below 3.0 on a 0- to 10-point scale or 1.5 on a 0- to 5-point scale. In general, the right-left odor imbalance between axillae of each subject should be no more than 30 % of the overall scale (that is 3.0 units on a 0- to 10-point scale or 1.5 units on a 0- to 5-point scale).

TABLE X1.2 Distribution of Accepted Subjects

Number of Tests	4
Number of Odor Assessors	3
Total Number of Subjects Screened	310
Total Number of Subjects Accepted	310
Scoring Scale	0–5

Range of Average Control Odor Scores	Right Axilla, % Total	Left Axilla, % Total
0–0.4	0.0	0.2
0.5–0.9	0.3	0.2
1.0–1.4	6.0	5.3
1.5–1.9	10.0	9.0
2.0–2.4	21.6	19.0
2.5–2.9	17.5	14.0
3.0–3.4	14.5	14.0
3.5–3.9	12.6	14.0
4.0–4.4	7.5	13.0
4.5–4.9	7.0	5.3
5.0	3.0	6.0

X2. RECOMMENDED PROCEDURE FOR SCREENING AND SELECTING IN VIVO (LIFETIME) DEODORANCY ASSESSORS

X2.1 *Purpose*—The purpose of this series of tests is to screen people who are interested in becoming deodorancy assessors. The screening is for olfactory acuity, specific anosmia to androstenone and androstenol, and interest and availability for testing. To accomplish this purpose, the screening should be divided into two phases, conducted in two different sessions.

X2.2 *Panelist Recruitment*—An adequate number of panelists should be recruited for Phase I testing based on the assumption that half of the people will pass Phase I testing and move on to Phase II testing, from which assessor trainees will be selected. These people should be interested in becoming deodorancy assessors and should be available during the times deodorancy evaluations are conducted.

X2.3 *Pre-Screening Questionnaire*—A brief questionnaire should be administered to the panelists to determine sex, age, and other information needed to confirm their willingness and availability to participate in deodorancy evaluations. Questions updating the assessor's general health status and smoking habits should be included.

X2.4 *Test Location and Scheduling*—Testing should be conducted in individual booths with adequate ventilation to prevent the influence of extraneous odors and sample carryover effects on test performance. If sample sets are to be reused, the panelists should wear odor-free plastic gloves to prevent contamination of the sample container. The tests should be scheduled to minimize panelist fatigue and permit the ventilation of odors from the test area.

X2.5 *Phase I*:

X2.5.1 *Purpose*—Using sequential analysis, potential assessors will be screened for olfactory acuity using deodorant products and isovaleric acid (IVA).

X2.5.2 *Acuity*—A variety of tests may be used to screen the sensitivity of panelists to fragrance materials used in deodorant products, deodorant product perfumes and various levels of IVA mixed with deodorant product perfume solutions. Suggested tests include paired difference, triangle, duo-trio and ranking. Procedures for these and other tests may be found in Ref (**17**) or any basic text on sensory evaluation. An adequate number of tests should be conducted to ensure that the olfactory acuity of the panelists is accurately assessed. Appropriate samples should be determined for the selected tests. The probability levels for each sample set of each test should be established by pretesting and should be no greater than 0.05.

X2.5.3 *Ranking Test*—The following three ranking tests are suggested:

X2.5.3.1 The first test is a series of five standard Isovaleric Acid (IVA) concentrations prepared in distilled water:

Sample Number[A]	IVA Concentration (mL/L)
736	0.014
951	0.058
458	0.22
602	0.89
059	3.57

[A] The assignment of a three-digit code to each test material will follow a computer randomized sequence for all test samples in order to keep the odor assessors from identifying a number series as a specific sample.

X2.5.3.2 The second test is the above series of IVA concentrations with the addition of 0.1 mL of degassed aerosol deodorant (commercially available).

X2.5.3.3 The third test is the standard IVA series with the addition of 0.1 mL of a 5 % solution of a fragranced bar soap (commercially available).

X2.5.3.4 The three tests are administered in sequence to the odor assessor. The assessor is presented the five samples and is requested to rank the samples in order from highest to lowest intensity. Samples should be presented at room temperature in a randomized and balanced order using procedure appropriate for the selected test. Each sample set should be presented twice to obtain 10 to 20 evaluations from each panelist.

X2.5.3.5 Assessor performance in the ranking tests can be evaluated according to pre-established criteria. (Transposition of two adjacent odor concentrations may represent a 1-point odor scale difference and not be significant. However, random ranking of more than two concentrations would indicate difficulty with the overall performance.) Potential assessors are accepted or rejected on the basis of their performance as determined by the percent correct responses or sequential analysis (**18**).

X2.5.4 A screening test which may also be suitable for determining smell function and measuring assessor acuity has been prepared (**19**).

X2.6 *Phase II*:

X2.6.1 *Purpose*—The purpose of this series of tests is to further screen the olfactory acuity of the panelists who passed Phase I testing and to monitor the panelists for specific anosmias relevant to deodorancy testing such as androstenone, androstenol, isovaleric acid, selected members of the methyl ionone family and synthetic and natural musks.

X2.6.2 *Test Sample Selection*—Test samples should be selected according to the guidelines of X2.5.2. The probability level of the selected tests should be the same as those used in Phase I.

X2.6.3 *Determination of Relevant Specific Anosmias*—Compounds commonly tested are androstenone, androstenol and isovaleric acid, although other compounds may also be tested. Procedures for testing specific anosmia to androstenone have been developed (**20**). These procedures may also be adapted for testing other specific anosmias. (The effect of specific anosmia on the ability of assessors to evaluate axillary odor has not been established.)

X2.6.4 *Interview*—Each assessor should be individually interviewed by a test monitor to ascertain motivation, availability and health of each panelist.

X2.6.5 *Data Analysis and Selection of Assessor Trainees*—The results of all testing and interviews for each panelist may be compiled and the panelists should be ranked according to their performance on the tests. The required number of assessor trainees should be selected by choosing those who performed best on the tests and who are the most interested in and available for deodorancy evaluations. Although the relationship between *in vitro* and *in vivo* deodorancy tests has not been established, the performance of the assessors on these screening tests should not fall below a predetermined minimum. It is recommended that an adequate pool of assessors be retained for deodorancy testing.

X2.6.6 Assessors who are to use magnitude estimation should be screened for competency with the scale by having them rate line lengths, some of which are so small that they are forced to use values between 0 and 1 and some which are so close in length that they are forced to use values between two integers (for example, 3.25). Studies have shown (**21**) that magnitude estimation ratings assigned to visual line lengths will virtually be proportional to their actual lengths. Assessors whose ratings depart markedly from this relationship should be eliminated, especially if the reason for the departure is discomfort with or inability to use decimal fractions.

X2.7 The following protocol is suggested for pretest laundry of wearing apparel:

X2.7.1 When using a U.S. style top-loading machine select the high fill level to provide approximately 21.25 gal of water.

X2.7.2 Wash cycle time should be 10 min.

X2.7.3 Select a warm water wash cycle (100 ± 5°F) followed by a cold water rinse cycle.

X2.7.4 Water hardness is not critical for this protocol.

X2.7.5 Unfragranced detergent[4] should be used in accordance with the manufacturer's recommended dosage.

X2.7.6 Wearing apparel should be subjected to one full wash cycle with unfragranced detergent and one full cycle without detergent.

X2.7.7 Dry apparel for 30 min on the permanent press cycle of an automatic clothes dryer, or as required to complete drying.

[4] Standard Detergent 124, No. 8350, available from the American Association of Textile Chemists and Colorists, P.O. Box 12214. Research Triangle Park, NC 27709, has been found satisfactory for this purpose.

X3. IN VIVO DEODORANCY ASSESSOR TRAINING

X3.1 *Purpose*—The purpose of the experimental procedures discussed here is to recommend suggested procedures for training a group of people to be deodorancy assessors. These potential assessors may be selected by means of a series of screening tests and interviews to determine the individuals with the most sensory acuity and most interest in the test as recommended in Appendix X2.

X3.2 Assessor training is accomplished in three phases: orientation, two simulated deodorancy studies, and monitoring of assessor trainee performance during regular deodorancy studies.

X3.2.1 *Orientation*—A brief orientation session may be held for the assessor trainees. The following objectives may be covered.

X3.2.1.1 Introduce the assessor trainees to each other and to test personnel involved with conducting deodorancy testing, explaining the purpose of deodorancy testing in the company;

X3.2.1.2 Orient and train the assessors to the selected rating scale;

X3.2.1.3 Discuss typical testing procedure;

X3.2.1.4 Describe the deodorancy assessor responsibilities, and

X3.2.1.5 Give a tour of the facilities used to conduct deodorancy testing. Any questions that the assessor trainees have may be answered at this time.

X3.2.2 *Simulated Deodorancy Study*—One or more simulated studies may be arranged to give the assessor trainees the opportunity to practice making axillary evaluations. Products for testing should have known differences and may include antiperspirants, deodorants, talcs, deodorant soap bars, and refreshment soap bars. Antiperspirants, deodorants, and talcs are not rinsed from the skin and thus may be easier for the assessor trainee to evaluate. The study may be similar to an actual deodorancy test to make the transition from assessor trainee to expert assessor smoother for those assessors that complete training. Expert assessors may be placed in an evaluation room with several assessor trainees so that the expert is there to answer questions and provide feedback to the trainees. If there is some disagreement or if the assessor trainees have any questions, they may confer with the expert assessors, and if necessary, re-evaluate the subject.

X3.2.2.1 During the training evaluations the assessor trainees may be instructed to evaluate only the intensity of the malodor. The assessors may be asked to think only in terms of the intensity of the odor, not whether or not they liked the odor. They may be instructed to "sniff through" any extraneous odor such as perfume, hairspray, powder, food, etc. They may make a note on the scoresheet of any extraneous odors they smelled but the numeric score is to indicate the strength or intensity of the malodor.

X3.2.2.2 During the training evaluations, assessor trainees may be free to drop from the training at any time. Trainees whose performance is not consistent with the expert assessors and with the group of trainees or does not uphold the assessor responsibilities, such as being on time and remaining quiet during evaluations, may be dropped from further training. Trainees who perform consistently with the expert assessors and with the other trainees may be retained for further practice in regular deodorancy studies.

X3.2.3 The major problems arising in axillary malodor rating using magnitude estimation as well as other scales are that assessors give zero ratings to large percentages of the axillae and that equal ratings to the left and right axillae are given to an excessively large percentage of the subjects. Both of these problems result from insufficient assessor training on how to evaluate axillae. Careful observation shows that neither of these observations has a high actual frequency and that the assessor is not trying hard enough to detect and rate the lower odor levels or discriminate between similar odor intensities. Assessors should be instructed that axillae in which no malodor is initially detected should be re-examined with greater care to assure that absolutely no malodor can be detected. When using magnitude estimation scaling (see Test Method E1697) assessors should be drilled in the use of very small rating values (for example, 0.25, 0.1 etc.) where necessary to describe the lower malodor levels. Similar comments apply to the difference in malodor between the two axillae of any given subject. Although there will usually be a small group of subjects with no detectable difference in odor between the two axillae, assessors may need to be given the opportunity to reevaluate both axillae in a structured fashion or drilled in discrimination between samples of "just noticeable" odor differences. When using the magnitude estimation scaling the assessors should also be drilled in using numbers which differ by only a few percent (for example 3.0 versus 3.1) where necessary to express differences which are small.

X3.2.4 *Monitoring of Assessor Trainee Performance During Regular Deodorancy Studies*—Assessor trainees may be included on regularly scheduled deodorancy studies by exposing them to the full range of olfactory stimuli and to the odor ratings assigned by the members of the pre-existing panel. Trainees may record their scores and the trainees' scores may be statistically analyzed separately from the expert assessors and not used in the reporting of test results. The analysis of the data from their evaluations are used to measure assessor trainee performance. They may be included into the pre-existing panel when their ratings do not disagree with the average panel ratings significantly more often than those of the pre-existing panel.

X3.3 *Assessor Performance*—Assessors' performance on each test can be monitored by conducting an analysis of variance on the collected data to determine if there is a assessor variance.

X4. PRODUCT APPLICATION RECOMMENDATIONS

X4.1 *General Application Recommendations*—It is important that all products be treated in a similar manner. Products should be presented without identifying logos or characteristics as much as practical and should be of uniform size and shape. Product containers should be labeled with subject name or identification number as well as axillae to be treated. The appropriate dosage should be established for each product type based upon intended or projected consumer use-up rates. Water for all treatment should be controlled. It is suggested that a temperature of 105 ± 5°F (40 ± 2°C) and a hardness of 50 to 200 ppm be employed. It is suggested that disposable washcloths/towels be employed in procedures involving application, washing or drying of the axillary vault. It is suggested that whenever possible all the procedural steps be standardized, such as the length of time or amount used to moisten products and towels.

X4.2 *Deodorant and Toilet Bar Soap*:

X4.2.1 Bars should be treated as suggested under X4.1. Bars should be kept in individual, covered, dry soap dishes, or similar containers. The bar soap application procedure presented in X4.2.2 – X4.2.6 is a guideline only.

X4.2.2 The axillae should be thoroughly wetted with a disposable washcloth.

X4.2.3 Soap application may be either by means of direct application of the moistened bar to the axillary vault and lathering for a controlled time (for example 20 to 30 s or by lathering a pre-moistened disposable washcloth for a controlled time, such as for 10 s followed by an axillary scrub for a controlled time, such as 20 s).

X4.2.4 Following product application, the lather is allowed to remain on the axilla which is open to the air for a controlled time (for 15 s to increase the contact time of the skin to the product).

X4.2.5 All of the soap residue is rinsed from the skin using one or more pre-moistened disposable washcloth(s). Irritation may occur if soap residue is allowed to remain. The wash supervisor should check subjects for any remaining product, particularly on the back of their arms.

X4.2.6 After all soap is removed, subjects should blot their axillae dry with a fresh disposable towel.

X4.3 *Liquid Bath Soaps and Gels*:

X4.3.1 Product should be treated as suggested in X4.1.

X4.3.2 The procedure in X4.2 is appropriate except that the product should be applied by means of a pre-moistened, disposable washcloth.

X4.3.3 Pre-measured amounts of the product are used to monitor product application.

X4.4 *Deodorant and Antiperspirant Sticks*:

X4.4.1 Prior to product application, the axilla should be prewashed with the control cleaning agent under appropriate supervision.

X4.4.2 The product should be treated as suggested under X4.1. It is important that the product application be performed by an experienced supervisor throughout the study. To increase test sensitivity, it is recommended that only one supervisor apply the product throughout the study.

X4.4.3 Solid stick products should have approximately ¼ in. cut off the top of the product and the resulting surface rubbed smooth.

X4.4.4 Pre- and post-application weights of the product must be taken to monitor product application.

X4.4.5 Typical application quantity is 0.5 g/axilla.

X4.5 *Deodorant and Antiperspirant Roll-ons, Creams and Lotions*:

X4.5.1 Prior to product application, the axillae should be prewashed with the control cleaning agent, rinsed, and dried under appropriate supervision.

X4.5.2 Product should be treated as suggested in X4.1.

X4.5.3 These products may be applied either by means of the finished package or a syringe-metered amount applied directly to the axilla and distributed over the axillary vault using a glass rod or similar device.

X4.5.4 Pre- and post-application weights of the product must be taken to monitor product application.

X4.5.5 Typical application quantities for roll-on products range from 0.3 to 0.5 g/axilla, cream products from 0.2 to 0.3 g/axilla, and lotion products typically 0.3 g/axilla.

X4.6 *Body Talcs*:

X4.6.1 Prior to product application, the axilla should be prewashed with the control cleansing agent, rinsed and dried under appropriate supervision.

X4.6.2 The product should be treated as suggested in X4.1.

X4.6.3 These products are normally applied by means of a nonporous pad.[5]

X4.6.4 Pre- and post-application weights of the product must be taken to monitor product application.

X4.6.5 Typical application quantities for talc products average 0.3 to 0.5 g/axilla.

X4.7 *Pump and Aerosol Deodorants, Antiperspirants, and Body Colognes*:

X4.7.1 Prior to product application, the axilla should be prewashed with the control cleansing agent, rinsed, and dried under appropriate supervision.

X4.7.2 Product should be treated as suggested in X4.1.

X4.7.3 These products are normally applied in a ventilated hood area to reduce the exposure of subject and supervisor to aerosol build-up. In order to ensure uniform product dosage throughout the study no more than one-half of the contents of any test canister should be consumed during any test. Canisters should be primed before the first use.

X4.7.4 Pre- and post-application weights of the product must be taken to monitor product application.

X4.7.5 Typical application quantity for aerosol products is a timed 2-s spray at a distance of 6 in. (152 mm) or approximately 0.8 to 1.0 g/axilla.

X4.7.6 A typical application quantity for pump products consists of three plunges at a distance of 6 in. (152 mm).

[5] The Webril nonporous pad, distributed by Kleen Test Products, 4425 West Woolworth Ave., Milwaukee, WI 53218, or its equivalent, has been found satisfactory for this purpose.

X5. RELATIVE EFFICIENCY OF DESIGNS

X5.1 A characteristic of experimental designs which is useful as a guide for choosing a design is their efficiency (**14**). A design *X* is said to be twice as efficient as design *Y* if (given an equally precise estimate of experimental error) it requires only half the sample size as design *Y* to yield equally precise estimates of the treatment differences. More generally, design *X* is said to be *E* times as efficient as design *Y* if (given an equally precise estimate of experimental error) it requires only 1/ *E* times the sample size as design *Y* to yield equally precise estimates of the treatment differences.

X5.2 The relative efficiencies of any two of the above described designs may be obtained as the ratio of their corresponding sample sizes shown in Table X5.1. The sample sizes in Table X5.1 are those required by a variety of RRB and EVC designs in order for them to yield treatment difference estimates with precision approximately equal to that of a 1PR design with 10 subjects, assuming equally precise estimates of experimental error in each design. This assumption will be true in most practical deodorant clinical studies since the degrees of freedom for experimental error will typically be large enough that small differences in degrees of freedom will have little effect. (The number of subjects in the 1PR design used here as a benchmark is arbitrary. Ten was chosen for ease of exposition.)

TABLE X5.1 Sample 5 Requirements

Each versus Control Design						
Round Robin Design			Test TRT versus Test TRT		Test TRT versus Control TRT	
Number of Subjects TRT per TRT		Number of Subjects per study	Number of Subjects per TRT	Number of Subjects per study	Number of Subjects per TRT	Number of Subjects per study
2(1PR)	10[A]	10[A]	. . .	. . .	. . .	. . .
3	14	21	20	40	10	20
4	15	30	20	60	10	30
5	16	40	20	80	10	40
6	15	45	20	100	10	50
7	18	63	20	120	10	60
8	21	84	20	140	10	70
9	16	72	20	160	10	80
10	18	90	20	180	10	90

[A]The 1PR study is listed here under RRB Studies with two treatments.

X5.3 As shown in Table X5.1, the sample size requirements for the EVC designs differ depending upon whether interest focuses on a test-treatment-to-test-treatment (test-test) comparison or on a test-treatment-to-control-treatment (test-control) comparison. Test-control comparisons are higher precision since they are within subject comparisons; therefore, their sample size requirements are lower. Test-test comparisons are lower precision since they are between subject comparisons; therefore, their sample size requirements are greater. In practice, the number of assessors employed for a study normally ranges from 3 to 4. Additional assessors may be used as alternates in the event that one of the primary assessors fails to be discriminating during an evaluation session. It is also customary to employ a minimum panel of 30 qualified subjects.

X5.4 Conclusions which may be drawn from examination of Table X5.1:

X5.4.1 A question which frequently occurs when designing a study to compare two treatments is whether to test them in a 1PR design or in an EVC design, comparing each to a control treatment. As shown in Table X5.1, the sample size required by the EVC design is four times that required by the 1PR design in order for those designs to yield treatment difference estimates with equal precision.

X5.4.2 There is a two-fold difference in sample size for an EVC design depending upon whether interest focuses upon test-test or test-control comparisons.

X5.4.3 For laboratories with the capability to analyze RRB designs, the RRB design yields the required level of precision for all comparisons with a total study sample size approximately equal to half that of the low precision test-test comparisons of the EVC design.

X6. STATISTICAL ANALYSIS OF DEODORANT CLINICAL STUDIES

X6.1 The following paragraphs describe the parametric statistical analyses of deodorant clinical studies. These analyses require for their validity that the properties discussed in X6.1.1 – X6.1.3 hold. A non-parametric statistical analysis, which does not require all these properties for validity, for single pair (1PR) designs is given in X7.1.4.3.

X6.1.1 *Independence of the Observations*—The observations are independent when the probability distribution for each observation does not change as a function of the realized values of the other observations. Independence is generally assured by having a single randomization which is applied to each of the following steps of the measurement process: Treatments assignment to subjects, application of substances to axillae, and axillary evaluation.

X6.1.2 *Variance Stability*—The variance is termed unstable when it changes from observation to observation. The most frequent kind of variance instability occurs when variance increases (or decreases) systematically with progressively larger values of the mean.

X6.1.3 *Normality*—The distribution of observations about their mean should approximately follow the normal probability law. The primary way in which this property is violated is when the data are sharply skewed, rather than symmetric.

X7. STATISTICAL ANALYSIS OF SINGLE PAIR DESIGN

X7.1 The subjects in a single pair design may be classified into one of the two "sequence" groups, depending upon the assignment of treatments to the left and right axillae:

Sequence 1 = Treatment 1 on the right, Treatment 2 on the left

Sequence 2 = Treatment 2 on the right, Treatment 1 on the left

X7.1.1 Let Yi_{jk} be the mean odor rating over all assessors for treatment i of subject j in sequence k,

where:

i = 1,2
k = 1,2
j = 1,2 ... n_k
n_k = Total number of subjects in sequence group k.

X7.1.2 Compute the within subject difference:

$$d_{jk} = Y_{1jk} - Y_{2jk}$$

So that d_{jk} is the difference in mean odor rating (treatment 1 minus treatment 2) for subject j in sequence group k.

X7.1.3 An estimate of the difference in odor efficacy between treatments 1 and 2 is computed as:

$$\bar{d}.. = \left(\bar{d}._1 + \bar{d}._2\right)/2$$

where:

$\bar{d}._k$ = the mean of the d_{jk} for the n_k subjects in sequence group k.

NOTE X7.1—In the event that there is a difference in odor intensity due to an inherent left-right difference, unrelated to the treatments, and that n_1 is not equal to n_2, the above estimate is superior to the simple mean of all d_{jk} because it is unbiased by the left-right difference.

X7.1.4 A (1-alpha) confidence interval for the true difference in deodorant efficacy of the two products is obtained as:

$$\bar{d}.. \pm t_{(1-\text{alpha}/2,\, n_1+n_2-2)} \times \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{1}{4}\right)\frac{S^2{}_1(n_1-1)+S^2{}_2(n_2-1)}{n_1+n_2-2}}$$

where:

$$S_k = \sqrt{\sum_{j=1}^{n_k}\left(d_{jk} - \bar{d}._k\right)^2/(n_k-1)}$$

S_k = the standard deviation of the d_{jk} for subjects in sequence group k, and $t_{(1\text{-alpha}/2,n_1+n_2-2)}$ = the standard, tabulated value of the cumulative t-distribution at the (1-alpha/2)th percentile for degrees of freedom equal to $n_1 + n_2 - 2$.

X7.1.4.1 A test of the hypothesis of no difference in deodorant efficacy between treatments at the (1-alpha) confidence level is equivalent to determining if zero lies in the (1-alpha) confidence interval. (All confidence intervals presented here are used in this way to test hypotheses.) If it does not, then zero is not a plausible value for the true difference in treatment efficacy and, therefore, we conclude there to be a significant difference at the (1-alpha) confidence level. Otherwise, the difference is not significant at the (1-alpha) confidence level.

X7.1.4.2 Calculate an estimate of the difference in mean odor rating between the axillae (right minus left) as follows:

$S = \left(\bar{d}._1 - \bar{d}._2\right)/2$, with (1 − alpha) confidence interval:

$$S \pm t_{(1-\text{alpha}/2,\, n_1+n_2-2)} \times \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{1}{4}\right)\frac{S_1{}^2(n_1-1)+S_2{}^2(n_2-1)}{n_1+n_2-2}}$$

X7.1.4.3 When the error distribution of the d_{jk} departs markedly from normality (X6.1.3), a non-parametric analysis will be more appropriate. For single pair (1PR) designs a simple test which can be used to assess the significance of efficacy differences between treatments and which adjusts for a possible difference in malodor level between the left and right axillae is the Wilcoxon rank sum test (also known as the Mann-Whitney test) **(14)** applied as if to test the hypothesis that the median of the d_{j1} equals negative the median of the d_{j2} for log-transformed data. (This may at first appear incorrect; however, the median of the d_{j1} estimates:

A = (right axilla − left axilla) + (Treatment 1 − Treatment 2),

and the median of the negative of the d_{j2} estimates:

B = (right axilla − left axilla) + (Treatment 2 − Treatment 1),

and note that if Treatment 2 = Treatment 1, then A = B, no matter what the size of (right axilla − left axilla).)

X7.1.5 Compute the Wilcoxon rank sum test using the following steps:

X7.1.5.1 Pool all of the d_{j1} with $-d_{j2}$;

X7.1.5.2 Assign ranks to these pooled values so that the lowest is assigned the rank of 1 and the highest the rank of $n_1 + n_2$;

X7.1.5.3 Generally many ties will occur in the data. These tied observations are given ranks that would be equal to their average rank had they been distinct values coming at the same location in the sorted data;

X7.1.5.4 Obtain W_k which is the sum of the ranks associated with the observations which come from sequence group k. (k is arbitrary. Either group can be used.)

X7.1.5.5 Calculate the test statistic as follows (see Ref. **(14)** for discussion of sample size restrictions):

$$Z = \frac{W_k - \frac{1}{2}n_k(N+1)}{\left[n_1 n_2(N+1)/12\right] - C}$$

where:

N = $n_1 + n_2$
C = correction for ties (usually negligible in these studies)

$$= \frac{n_1 n_2 \sum_{i=1}^{e}(f_i - f_o)}{12\,N\,(N-1)}$$

where:

e = number of distinct values in the data set, and the f_i are the multiplicities of each distinct value.

For example, in the data set 2, 2, 4, 9, 9, 9 there are $e = 3$ distinct values; 2, 4, and 9 with multiplicities $f_i = 2, 1, 3$.

X7.1.5.6 The hypothesis of equal deodorant efficacy is rejected when $|Z|$ exceeds the critical value associated with the desired confidence level for the standard, tabled normal distribution, (for example, 1.28 for 80 % confidence, 1.64 for 90 % confidence, 1.96 for 95 % confidence, 2.58 for 99 % confidence).

X8. REPORTING OF AXILLARY DEODORANCY RESULTS

X8.1 Different clinical testing laboratories have favored different forms in which to present study results. The simplest, called "absolute body odor differential" is to simply report $d..$ and its (1-alpha) confidence limits ($d..{}_1$, $d..{}_u$).

X8.1.1 Calculate absolute body odor differential as follows:

$$(d..{}_1,\ d..{}_u) = d.. \pm t_{(1\ -\ \text{alpha}/2\ ,\ n_1 +\ \ n_2\ -\ 2)} \times \text{SE}\,(d..)$$

where:

$$\text{SE}\ (d..) = \sqrt{(1/n_1 + 1/n_2)\left(\frac{1}{4}\right)\frac{S^2{}_1(n_1 - 1) + S^2{}_2(n_2 - 1)}{n_1 + n_2 - 2}}$$

= the standard error of $d..$

X8.2 A second approach, perhaps the one used most frequently, called "percent body odor reduction," is to express both $d..$ and its confidence limits as a percent of the observed mean of that member of the treatment pair considered to be the control treatment.

X8.2.1 Let C be the mean for the control treatment, then calculate percent body odor reduction as follows:

$$\text{Percent Body Odor Reduction} = d../C,$$

with (1-alpha) confidence limits + ($d..{}_1/C$, $d..{}_u/C$).

NOTE X8.1—It should be noted that some researchers object to the above method when category scales are used since those scales are thought to lack the ratio properties necessary for rigorous validity of ratio calculations.

X8.3 A third approach is to present mean treatment differences normalized by their standard error. This has the advantage of carrying information on how significant the difference was in the study. For example, in studies of typical sample size, a value of 1.7 implies a significant difference with 90 % confidence, 2.0 implies 95 % confidence, and 3.0 implies 99 % confidence.

X8.3.1 Calculate standardized treatment differences as follows:

$$\text{Standardized Treatment Difference} = d../\text{SE}\,(d..)$$

X9. STATISTICAL ANALYSIS OF EACH VERSUS CONTROL DESIGN

X9.1 Each subject in an EVC design receives one of the $t-1$ test treatments on one side and the control treatment on the other. Therefore, the subjects may be cross classified by two factors:

X9.1.1 *Test-Treatment*—Which of the $t-1$ test treatments he received, or

X9.1.2 *Sequence*—Whether the control treatment is on the right or the left.

Sequence 1 = control treatment on right, and

Sequence 2 = control treatment on left.

X9.1.2.1 The ANOVA model for the EVC (each versus control) design analyzed with the GLM procedure should include the following effects: Treatment, Sequence, and the Treatment-by-Sequence interaction.

X10. STATISTICAL ANALYSIS OF ROUND ROBIN DESIGN

X10.1 The ANOVA model for the RRB design analyzed with the GLM procedure should include the following effects: Subject, Treatment, Sequence, and the Treatment-by-Sequence interaction.

X11. RESPONSE SURFACE AND FACTORIAL TREATMENT STRUCTURE

X11.1 There is no restriction on the nature of the treatments included in the RRB design or included among the test-treatments in the EVC design; therefore, response surface and factorial studies (**22**) may be carried out. Extensive computer analysis by a knowledgeable analyst is required.

X11.2 *Log Transformed Data*—For some rating methods (for example, magnitude estimation) one may prefer to take logarithms of all individual assessor ratings prior to the statistical computations. This is often done to achieve one or more of the following goals:

X11.2.1 Transform treatment and sequence effects to a multiplicative rather than additive form (for example, "Treatment *i* reduces odor by P %" rather than "Treatment *i* reduces odor by P scale units".) Multiplicative effects are sensible only when the rating scale used by the assessors has the ratio property as has been claimed for magnitude estimation (**23**).

X11.2.2 *Stabilize the Variance*—Often the inherent variance of observations increases as the mean increases. Such data violate the underlying assumptions for all of the statistical methods described above. A logarithmic transformation will often eliminate that trend, thereby permitting analysis by the methods in Appendix X9 – Appendix X11.

X11.2.3 *Normalize the Data*—Sometimes the distribution of observations about their mean is not normal but is skewed toward higher values. This is also a violation of the underlying assumptions of the statistical methods described above. Logarithmic transformation will often convert such distributions to approximate normality, thus, permitting analysis by the above methods.

X11.3 When statistical computations are carried out on the log transformed data, the most convenient form in which to report the estimated treatment effects and their confidence limits is in terms of percent odor reduction (of the geometric mean odor ratings).

X11.4 Let $\bar{d}$ be any of the above described estimates of the difference in deodorant efficacy between two treatments (treatment i minus treatment i') and let d_u and d_1 be the upper and lower confidence limits for $\bar{d}$. If the data have been log transformed prior to computation, then the (geometric) mean reduction in odor rating from treatment i' to treatment i expressed as a percent of the treatment i' geometric mean is as follows:

$$\mathrm{PR} = \left(1 - \exp\left(\bar{d}\right)\right) \times 100\,\%,$$

with confidence limits:

$$\left[\left(1\exp\left(\bar{d}_u\right)\right) \times 100\,\%, \left(1 - \exp\left(\bar{d}_1\right)\right) \times 100\,\%\right]$$

X11.5 For any of the above described estimates of mean treatment differences, some algebraic manipulation will show that the above expressions are equivalent to percent reduction in terms of geometric means. One disadvantage of taking logarithms of all ratings is that there are generally ratings of zero in all data and the logarithm of zero is undefined. A logically consistent way in which to treat these ratings is described in Test Method E1697.

X11.6 Other statistical analyses exist which can be correctly applied to the designs of 8.11.

NOTE X11.1—It should be recognized that, in addition to the designs suggested in 8.11, other designs appropriate for addressing questions regarding deodorancy efficacy exist.

X12. FACTORS FOR ENHANCEMENT OF STUDY PRECISION

X12.1 The following protocol features should be carefully evaluated for their ability to improve test sensitivity. However, the use of these enhancement features may make the results more test specific. The determination of the study protocol, including any and all test enhancements, should be based upon the objective of the study and the magnitude of the expected difference in test product efficacy (see 1.5). In most cases no scientific data can be cited which confirms that these enhancements will directly affect the test results; however, they reflect the consensus of the task group.

X12.1.1 Subjects may be preselected for axillary malodor (see 6.1). Estimates of percent odor reduction, as measured by any of the procedures described herein, are generally larger among subjects with higher initial odor levels than among subjects with lower initial odor levels. Statistical control of this factor is recommended to assure that a substance will not be estimated to have low or high efficacy due solely to low or high initial odor levels in the panel in which it is tested.

X12.1.2 The right-left imbalance of the panel may be reduced (see 6.1).

X12.1.3 The climate and time of year may be controlled. Erratic results can be obtained during periods of severe weather fluctuation, therefore, a period of stable climatic conditions may be the time of year which would give the least variable results.

X12.1.4 A pre- and postapplication holding period may aid in normalizing test product contact and in stabilizing the subjects.

X12.1.5 Subjects may be asked to refrain from the use of highly fragranced fabric conditioners and treatment products, such as laundry aids and rinses, during the testing period.

X12.1.6 Subject selection criteria may assess the subject's general cognition of personal hygiene.

X12.1.7 As much evidence as possible may be obtained that a subject's general attitude and interest is high and that all restrictions will be closely observed.

X12.1.8 Shaving of the axillae may be considered obligatory for the protocol.

X12.1.9 All abnormally rigorous physical exertion may be minimized during the test period, specifically excluding participation in any sporting activity.

X12.1.10 The test period duration may be extended in order to reduce the influence of certain test variances. The relevant statistical considerations should be considered.

REFERENCES

(1) Shelley, W.B., "The Role of Apocrine Sweat in the Production of Axillary Odor," Journal of the Society of Cosmetic Chemists, Vol 7, 1956, pp. 171–175.

(2) Shehadeh, N., and Kligman, A.M., "Variations in Axillary Odor," Journal of the Society of Cosmetic Chemists, Vol 14, 1963, pp. 605–607.

(3) Pillsbury, D.M., Shelley, W.B., and Kligman, A.M., Dermatology, W.B. Saunders Company Publishers, Philadelphia, PA, 1956.

(4) Shelley, W.B., Hurley, H.J., and Nicholas, A.C., "Axillary Odor: Experimental Study of the Role of Bacteria, Apocrine Sweat, and Deodorants," Archives of Dermatology and Syphilogy, Vol 68, 11953, pp. 430–446.

(5) Aly, R., and Maibach, H.I., "Aerobic Microbial Flora of Intertrigenous Skin,," Applications of Environmental Microbiology, Vol 33, 1977, pp. 97–100.

(6) Kuno, Y., Human Perspiration, Charles C Thomas Publisher, Springfield, IL,, 1956, pp. 157–170.

(7) Labows, J.N., Preti, G., Hoelzle, E., Leyden, J., and Kligman, A., "Analysis of Human Axillary Volatiles: Components of Exogenous Origin," Journal of Chromatography, Vol 163, 1979, p. 294.

(8) Leyden, J.J., McGinley, K.J., Hoelzle, E., Labows, J. N., and Kligman, A.M., "The Microbiology of the Human Aaxilla and its Relationship to Axillary Odor," Journal of Investigative Dermatology, Vol 77, 1981, pp. 413–416.

(9) Labows, J.N., McGinley, K.J., and Kligman, A.M., "Perspectives on Axillary Odor," Journal of the Society of Cosmetic Chemists, Vol 34, 1982, p. 193.

(10) Amerine, M.A., Pangborn, R.M., and Roessler, E.B., Principles of Sensory Evaluation of Food, Academic Press Inc., New York, NY, 1965.

(11) "The Tentative Final Order for OTC Topical Antimicrobial Products," Federal Register, Vol 43, Jan. 6, 1978, p. 1210.

(12) ASTM Committee E-18, Guidelines for the Selection and Training of Sensory Panel Members, ASTM STP 758, 1981 , 35 pp.

(13) Whitehouse, H.S., and Carter, R.O., "Evaluation of Deodorant Toilet Bars," Proceedings of the Scientific Section of the Toilet Goods Association., No. 48, 1967, p. 31.

(14) Snedecor, G.W., and Cochran, W.G., Statistical Methods, 6th ed. The Iowa State University Press, Iowa.

(15) Cochran, W.G., Sampling Techniques, John Wiley and Sons, Inc., New York, NY, 1953 .

(16) Mace, A.E., Sample Size Determination, Reinhold Publishing Corp., New York, NY, 1964.

(17) ASTM Committee E-18, Manual on Sensory Testing Methods, ASTM STP 434, 1968, 77 pp.

(18) Amerine, M.A., and Roessler, E.B., Wines, Their Sensory Evaluation, Freeman, San Francisco, CA, 1976.

(19) Doty, R.L., Shaman, P., Applebaum, S.L., Giberson, R., Siksorski, L., and Rosenberg, L., "Smell Identification Ability: Changes with Age," Science, Vol 226, 1984 , pp. 1441–1442.

(20) Amoore, J., "Directions for Preparing Aqueous Solutions of Primary Odorants to Diagnose Eight Types of Specific Anosmia," Chemical Senses and Flavor, Vol 4, No. 2, 1979, pp. 153–161.

(21) Baird, J.C., and Noma, F., Fundamentals of Scaling and

Psychophysics, John Wiley & Sons, Inc., New York, NY, 1978.
(22) Box, G.E.P., Hunter, W.G., and Hunter, J.S., Statistics for Experimenters, John Wiley & Sons, Inc., New York, NY, 1978.
(23) Stevens, S.S., Psychophysics: Introduction to its Perceptual, Neural and Social Prospects, John Wiley & Sons, Inc., New York, NY, 1975.
(24) Marks, L.E., Sensory Processes: The New Psychophysics, Academic Press, New York, NY, 1974.
(25) Pearce, J.H., Korth, B., and Warren, C.B., "Evaluation of Three Scaling Methods for Hedonics," Journal of Sensory Studies, Vol 1, pp. 27–46.
(26) Levine, M., "Two Computationally Simply Procedures for the Analysis of Deodorant Clinical Efficacy Studies Using a Mixed Linear Model: A Simulation Study," Shulton Research Report, American Cyanamid, Clifton, NJ, 1984.
(27) Cramer, S.G., and Swanson, M.R., "Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods," Journal of the American Statistics Association, Vol 68, 1973, pp. 66–74.
(28) Scheffe, H. "An Analysis of Variance for Paired Comparisons," Journal of the American Statistics Association, Vol 47, 1952, pp. 381–400.