Designation: E1207 – 14 (Reapproved 2022)

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.

1.6 *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.*

[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 Oct. 1, 2022. Published October 2022. Originally approved in 1987. Last previous edition approved in 2014 as E1207 – 14. DOI: 10.1520/E1207-14R22.

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

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

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

[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.

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

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			

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 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).