Designation: E 1207 – 02

Standard Practice for Sensory Evaluation of Axillary Deodorancy[1]

This standard is issued under the fixed designation E 1207; 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 practice 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 practice includes protocols for the selection and training of judges, selection of subjects, experimental design, and statistical analysis. This practice is limited to assessment of axillary odor by trained judges. 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. Terminology

2.1 *Definitions of Terms Specific to This Standard:*

[1] This practice is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation of Materials and Products and is the direct responsibility of Subcommittee E 18.07 on Personal Care and Household Evaluation.

Current edition approved Nov. 10, 2002. Published January 2003. Originally approved in 1987. Last previous edition approved in 1997 as E 1207 – 87 (1997).

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

2.1.1 *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**).

2.1.2 *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**).

2.1.3 *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**).

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

2.1.5 *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**).

2.1.6 *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.

2.1.7 *judges*—those individuals previously screened and trained, whose purpose during a study is to detect odor differences or evaluate intensity of axillary malodor.

2.1.8 *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.

2.1.9 *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**).

2.1.10 *sequential analysis*—a statistical technique which may be used to screen potential judges for sensory acuity to a specific stimulus. The judge is repeatedly tested until he or she passes or fails the test at a specified level of significance (**1011**).

2.1.11 *subjects*—those individuals recruited to participate in a study as sample carriers.

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

3. Summary of Practice

3.1 The protocols described provide for the designation of panels of individuals suitably selected and trained to perform the functions of judges 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.

4. Significance and Use

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

4.2 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.

4.3 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.

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

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

5. Subject Selection and Restrictions

5.1 *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 7.10. 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.

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

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

5.1.2.1 Based on the control odor scores (see 7.3), 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 10-point scale or 4.0 on a 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 10-point scale or 1.5 on a 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.

5.1.2.2 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).

5.1.2.3 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.

5.1.3 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.

5.1.4 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.

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

5.2.1 *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**).

5.2.1.1 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.

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

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

5.2.1.4 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.

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

5.2.2 *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:

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

5.2.2.2 Pre-laundered wearing apparel (see 7.6) 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.

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

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

5.2.2.5 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.

5.2.2.6 Subjects should not wash the axillae during the test week at home. 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.

6. Judge Selection and Training

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

6.2 Judges employed for assessing body odor intensity should be screened for the following attributes:

6.2.1 Interest and availability;

6.2.2 Qualitative and quantitative olfactory discrimination ability;

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

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

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

6.4 *Judge Training*— In addition to the following points, the recommended procedures are given in Appendix X3 for the training of *in vivo* deodorancy judges.

6.4.1 Judges 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.

6.4.2 *Judge Training for Category Scales*:

6.4.2.1 After being introduced to the rating scale procedure, either 0 to 5 or 0 to 10 category scales, judges should assign ratings to the stimuli in an open discussion to obtain a consensus rating for each stimulus.

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

6.4.3 *Judge Training for Magnitude Estimation*:

6.4.3.1 Given that the subjects have passed the screening criterion for magnitude estimation (see X2.6.6), a short drill in the use of the scale for stimuli such as line length, circle sizes, weights, etc., according to the judge instructions given in 7.9.2.2 will provide a clear understanding of the rating task.

6.5 *Judge Performance Monitoring*—Trained judges should be tested periodically to confirm their ability to discriminate various odors and to rank odor levels. 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 judges are capable of rating these products consistently. The procedure for monitoring judge performance should be carried out at least once a year. More

frequent monitoring may be required if there is some reason to suspect a judge's olfactory acuity. (See X3.3).

7. Test Design

7.1 *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.

7.2 *Conditioning Phase*—Each subject should adhere to the restrictions given in 5.2.1. 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.

7.3 *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 (5.1.2.2). 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 5.1.

7.4 *Post-Treatment Evaluation Interval*—The post-treatment evaluation interval may range from 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.

7.5 *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.

7.6 *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.

7.6.1 The following protocol is suggested for pretest laundry of wearing apparel:

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

7.6.1.2 Wash cycle time should be 10 min.

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

7.6.1.4 Water hardness is not critical for this protocol.

7.6.1.5 Unfragranced detergent[3] should be used in accordance with the manufacturer's recommended dosage.

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

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

7.7 *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 7.10.

7.8 *Test Product Application*:

7.8.1 For deodorant sticks, 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.

7.8.2 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.

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

7.9 *Odor Assessment Rating Scales*:

7.9.1 *Category Scaling of Axillary Odor*:

7.9.1.1 *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 judges 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.

[3] 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.

7.9.1.2 *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].

7.9.1.3 Considerations which arise in the application of category scales to deodorancy testing include the following. Judges 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 judges 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.

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

NOTE 1—It is generally recognized that judges find it difficult to psychologically accommodate more than 10 or 11 points in a scale. With scales consisting of a greater number of points, judges 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 judge 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.

7.9.2 *Magnitude Estimation Scaling of Axillary Odor*:

7.9.2.1 *Introduction*—This section describes the application of magnitude estimation as one subjective rating method for the measurement of axillary malodor. The essence of this rating method is that odor judges assign numerical ratings to axillae in such a way that the ratio of any two ratings reflects the perceived ratio of malodor intensity of the two corresponding axillae. Among the properties which make magnitude estimation either with or without fixed modulus appealing are: *a*) its extensive scientific validation (**14**); *b*) justification for expressing deodorant efficacy in terms of the traditional and intuitive "percent odor reduction," and *c*) conformance with the models underlying parametric statistical methods (*t* tests, analysis of variance, regression, etc.).

7.9.2.2 *Background for the Magnitude Estimation Rating Task*—In the magnitude estimation procedures generally described in the literature (**15**) the rating assigned to the first stimulus (axilla) is arbitrary (see Note 2).

NOTE 2—The experimenter may, for example, allow each judge to assign individual arbitrary starting values using the following instructions: Rate the malodor of the first axilla. Give it a high number if you think it is strong. Give it a low number if you think it is weak. Don't worry about which number you choose, the first number that comes to mind is the best one.

7.9.2.3 The essential requirements of the magnitude estimation method are achieved by giving the judges the following instructions for the second stimulus and all subsequent stimuli: *a*) "Rate the malodor of this second axilla. If you think it is twice as strong as the first, give it a number twice as high. If you think it is half as strong, give it a number half as high. If you think it is ten times as strong, give it a number ten times as high," and *b*) "Rate the malodor of all subsequent axillae similarly, always assigning a number whose ratio to your previous rating reflects the ratio of the strength of the current malodor to that of the previous malodor."

7.9.2.4 Using an arbitrary initial rating, the actual magnitudes of the resulting data values are arbitrary. The only non-arbitrary and, therefore, meaningful information is that conveyed by the ratios of the ratings of any given judge. The statistical analysis described for magnitude estimation in 7.11-7.16.1 uses only those ratios to calculate percent differences and tests for statistical significance among the axillary regimen. If desired, the arbitrary character of the rating magnitudes may be eliminated by calibrating each judge's ratings to a common modulus using standard dilutions of isovaleric acid which are sniffed and rated at specific times during each study. Calibration to a single modulus by this method or some other is required if the researcher elects to enhance study precision by adjusting for the effects of initial malodor level as described in 7.17.1.1, but is not necessary for any other purpose described in this document. Additional discussion of judge training in magnitude estimation has been reported elsewhere (**16**).

7.9.2.5 *Statistical Analysis of Magnitude Estimation Data with Zero Ratings*—The methods for statistical analysis are an application of conventional linear models (see 7.11) after each rating has been transformed by a logarithm (see 7.16.2) with special treatment for zero ratings since the logarithm of zero is undefined. The special treatment for zero ratings is described by the following steps: *a*) Determine separately for each judge the minimum non-zero rating, so that a potentially different minimum non-zero rating is obtained for each judge. *b*) Calculate the "zero substitution values" for each judge as 0.6 times the minimum non-zero ratings, and *c*) Replace all zero ratings with the" zero substitution value" associated with the judge who produced that rating.

7.9.2.6 The procedure given in 7.9.2.5 is suggested by heuristic procedures for analysis of similar data arising in analytical chemistry. In chemical assays a zero determination is not interpreted as indicating that none of the material being assayed is present but merely that the amount present is somewhere between zero and the limit of detection for the assay. For log transformed data, statistical analysis proceeds by replacing each zero measurement by a value which lies

between zero and the limit of detection for the assay. Similarly, in axillary malodor measurement, a zero rating does not indicate that no malodorous material is present but merely that the amount present lies between zero and the threshold of detection for that judge. Statistical analysis proceeds by replacing all zero ratings with non-zero values obtained by steps *a*, *b*, and *c* of 7.9.2.5.

7.9.2.7 The substitution value given in step *c* of 7.9.2.5 was chosen based upon a computer simulation of the censored log-normal distribution over a wide range of conditions likely to occur in deodorant clinical studies, including up to 20 % zero ratings. It was shown to yield virtually unbiased results with a mean squared error generally less than 8 % higher compared to results obtained by analysis of the uncensored data (**17**).

7.10 *Experimental Design Considerations*:

7.10.1 *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").

7.10.1.1 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.

7.10.2 *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).

7.10.3 *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. 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.

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			

7.10.4 *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.

7.10.5 *Order of Evaluation*—The order in which the judges 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.

7.10.6 *Choice of Sample Size*:

7.10.6.1 *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 (**18**), (**19**), and (**20**) for technical discussions of sample size determinations.

7.10.6.2 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).

7.10.6.3 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