J. Soc. Cosmetic Chemists, **15**, 579–592 (1964)

STATISTICAL EVALUATION OF QUANTITATIVE ANTIPERSPIRANT DATA (I)

By W. M. Wooding, B.Ch.E., H. E. Jass, Ph.D., and I. Ugelow, M.S.*

ABSTRACT

An improved method for the quantitative measurement of perspiration on human subjects was developed. It consisted of the use of small perforated containers of dehydrated silica gel, designed to be preweighed, then taped to subjects' skins so that sweat output could be collected at specific areas. The investigation described in this report was concerned with the evaluation of the ability of these devices to distinguish between typical antiperspirant treatments and other experimental variables. In the present case, the antiperspirants were applied in the form of aqueous solutions of active materials. A major object of the study was to test the application of a conventional statistical analysis (analysis of variance) to data obtained with the silica gel device. The results indicated that the measurement was sufficiently precise to allow comparisons of the efficacy of some antiperspirant materials with a blank and with each other and that a balanced experimental design plus an analysis of variance was satisfactory, statistically sound and very useful in interpreting the resulting data.

Introduction and Nature of Problem

Methods of measurement of human perspiration have generally suffered from a lack of precision and flexibility. Earlier methods depended upon visualization of sweat droplets by color reagents, and quantitation often consisted of a comparative grading of observed results. Thus, starch-iodine mixtures which turned blue at the sweating site were used by Minor (1) and modified variously by many other workers (2–5). Indicators other than starch-iodine, such as bromphenol blue (6) and phenolphthalein (7), have also been utilized by some investigators in a similar manner.

The use of electrical conductivity methods (8, 9) have yielded quantita-

* Revlon Research Center, Inc., Bronx 10473, N. Y.

tive results which appear to have reasonably good precision and accuracy. These methods, however, have the disadvantages of being cumbersome, not portable, indirect, not conducive to multiple simultaneous determinations, difficult to use for long continuous periods and requiring frequent and tedious standardizations.

Direct weighing methods have also been used in the past. These have comprised the absorption and weighing of sweat on tared absorbent paper (10) or cotton balls (11). Although such procedures would appear to offer the advantages of direct measurement and ease of use, the materials require extremely careful handling, including immediate and rapid weighing to avoid loss of the absorbed sweat. Further, the precision and probably the accuracy of such procedures appear quite highly dependent upon the quantity of sweat collected. Finally, statistical design and evaluation procedures, to improve rigor and sensitivity in the interpretation of results, were apparently not used.

A new method for the direct, quantitative measurement of human perspiration is described below. Based upon experimental work, some of which is presented here, it is believed that this method is more accurate than previously-reported procedures, and it has been found to be precise enough to distinguish readily among various pretreatments of the sweating site. It is easy to use and minimizes losses due to experimental errors in handling the apparatus.

The new method comprises the use of silica gel as an absorbent, contained in weighed easily-handled containers. The use of an orthogonal (balanced) experimental design, calculated to yield data suitable for the application of conventional parametric statistical analyses, is considered to be an inherent and necessary part of the new procedure.

The particular experimental work and subsequent analysis described in this paper had the following objects:

1. To test the new method, in order to determine whether it was capable of distinguishing among typical antiperspirant agents and a blank.
2. To test the use of a conventional factorial design of the experiment and an analysis of variance of the data from the new method.

This experiment was planned as the initial one of a series and consisted of the first formal work following development of the silica gel method. Subsequent experiments and their analyses were more sophisticated and included quantitative studies of perspiration produced at various anatomical sites, determination of the relative efficiency of a number of pure and formulated antiperspirants, studies of antiperspirant product formulation parameters, investigations of the duration of antiperspirant effects and more advanced experimental designs and statistical analyses. It is expected that some of these results will be published later.

Sweat Collection Containers

The silica gel absorbent is contained in desiccator pans of the type used to maintain a dry atmosphere in analytical balance cases.* They consist of press-fitted aluminum pans and covers about $2^3/_4$ in. in diameter, perforated on both sides. An outer metal canister is used for storing each unit to prevent loss or gain of moisture by the silica gel when it is not in use.

When the absorbent is saturated with moisture, it is regenerated by heating in an open dish on a hot plate at about 350°C until the original blue color of an indicator contained in the gel has returned, then transferring it, hot, back into the inner can and outer canister, and finally cooling and storing it until use in a desiccator.

In use, the entire assembly is preweighed in the outer canister on an analytical balance to the nearest milligram. It had been found that there was no appreciable weight change after storing containers in their unopened outer canisters in the atmosphere for periods of a half hour. Since this exceeded the total weighing time by a considerable margin, no significant errors in the data due to this operation were anticipated.

After preweighing, the containers are removed from their outer canisters and placed on a predetermined skin site which may be pretreated with a test material if the procedure is part of an experiment. Each container applied to the skin is masked completely with surgical tape to prevent moisture pickup from the air or from the surrounding skin. Following a wearing period, each container is removed from the skin, replaced in its canister and reweighed. The difference in weights is then calculated and recorded.

Experimental Design

As previously indicated, the principal object of this experiment was to test the effectiveness of the new method, used in conjunction with a statistical design and analysis, for distinguishing among various experimental conditions (in this case, typical antiperspirant agents, subjects and exposure times) such as would be likely to be encountered in general antiperspirant work. The following four "treatments" were selected with this in mind:

(T_1) Blank (20% aqueous sodium chloride)
(T_2) 20% aqueous aluminum chlorhydrate
(T_3) 20% aqueous solution of an experimental antiperspirant agent
(T_4) 20% aqueous sodium aluminum chlorhydroxy lactate

Eight male subjects were selected for the experiment. On the basis of

* Davison Silica-Gel Air Dryer, Davison Chemical Co., Div. W. R. Grace and Co., Baltimore, Md.

preliminary work, all of those chosen were believed to be reasonably normal as regards perspiration output. It was not planned to expose the subjects to exercise or excessive heat, or otherwise to attempt to induce abnormal sweating rates, because, in this case, it was desired that the test be made at levels reflecting normal use conditions. The sample size of eight subjects was, of course, quite small in view of the potentially great variation possible in a population, and some of the subsequent experiments used larger and more diverse groups. To demonstrate the utility and practicability of the new method and the analysis, however, it was only necessary in this case to be able to show differences, without unusual regard for the sampling problem; therefore it was felt that the selection of subjects used was quite suitable in this instance.

Each of the above four treatments was repeated on each of the eight subjects on five successive periods a day apart, and measurements were made at these intervals, beginning three hours after the first application of the treatments. The time factor was introduced to determine, if possible, whether there were detectable differences in effect with the repetitive applications, and whether there were detectable interactions between treatments and application history. This constituted a second factor in the design of the experiment, identified as follows:

(D_1) Measurements made three hours after first set of treatments
(D_2) Measurements made three hours after second set of treatments, one day later
(D_3) $\rightarrow$ (D_5) Same, for third, fourth and fifth treatments

Each new treatment was associated with a separate measurement, so that the need of the statistical analysis for independent observations was fulfilled.

Details of the experimental design, using the above two factors, were as follows:

Each subject received all four treatments (T_1 through T_4) on his back. To eliminate any possible bias among the treatments which might have been caused by variations in normal perspiration output among the different treatment sites, each treatment was assigned to a position on each subject which was predetermined by the use of two "Latin Squares," each using a group of four subjects arranged in a "block" of four "plots" each. Each plot was the site of one treatment, and each group of four subjects constituted a complete 4×4 Latin Square. Each treatment thus appeared a total of eight times on each day, twice in each of four possible positions. This arrangement is made clear in Fig. 1.

In carrying out the actual experimental work, each back was thoroughly cleaned, then dried gently with a towel. Each treatment was applied by swabbing with a cotton ball, ensuring that each swabbed area was a little

larger than the contact area required by the silica gel cup, and counting "swabs" so that about the same amount of material was applied in each case. The swabbed areas were outlined with indelible pencil and allowed to air-dry for about 15 minutes. The preweighed silica gel cans were then attached to each treated area as described above. After three hours, each container was removed, replaced in its canister and weighed. On the following day, each treatment was reapplied and fresh preweighed containers of silica gel attached, then again removed and reweighed after three hours. This procedure was repeated until five successive sets of data had been gathered over a period of four days. For a time just prior to and during

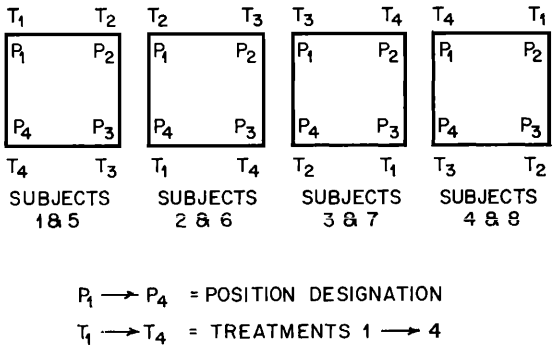

Figure 1.—Position-treatment-subject arrangement (Latin square).

the actual three hours of application, each subject was exposed to controlled temperature and humidity conditions. In so far as possible, other factors with a possible effect upon the results (time of day, time since last meal, physical exertion, water intake, etc.) were held constant.

After weighing and recording the data, calculating the weight gains and tabulating the results, the difference weights in grams were coded by multiplying each by 100. This was done to facilitate subsequent computation.

Preliminary work had given no evidence that there were any substantial position effects, and the use of the above Latin Square design for applying treatments was only a precautionary procedure. For this reason, it was planned to average the four data for each treatment in each of the two Latin Squares, thus yielding two means of four data for each treatment at each period of observation. Any position effects were thus included in the apparent effect of treatments, but in a balanced way, so that, provided real differences among treatments were detected at all, they could be confidently assumed to be due to treatments and not to position differences. It is possible in an analysis of this kind to assess separately the row and column effects (subjects and positions) in the Latin Squares, thus separating them from the treatment effects (14, chapt. 5), in this way increasing the

sensitivity of the experiment to treatment differences. In this case, however, it was not felt that this was necessary, and since some treatment differences were later demonstrated (see below), this procedure was found to have been justifiable.

Based upon the concept of analyzing the sets of means of four data each as described above, the format of the variance analysis to be used was that shown in Table I (13, p. 131–132; 15). For those not familiar with the symbolism used, an explanation is given in the Appendix to this paper.

TABLE I—FORMAT OF ANALYSIS OF VARIANCE

Source of Variation	DF	Expected Mean Square
Among treatments (T)	3	$\sigma_\epsilon^2 + (2)\ \sigma^2_{td} + (2)\ (5)\ \sigma_t^2$
Effect of reapplication (days) (D)	4	$\sigma_\epsilon^2 + (2)\ \sigma^2_{td} + (2)\ (4)\ \sigma_d^2$
Treatments × days (TD)	12	$\sigma_\epsilon^2 + (2)\ \sigma^2_{td}$
Replication (residual)	20	σ_ϵ^2
Total	39	

With respect to the arrangement of Table I, some comments are pertinent:

1. The basic requirements theoretically justifying the use of a variance analysis are:

(a) Normality of the error distribution about the observations
(b) Variance for error to be derived from a single population (homoscedasticity)
(c) Independence of the separate observations

Of these, the first was assumed because of the nature of the measurements (weights). The homogeneity of the variances was checked as described later. The requirement of independence was observed as described above, since each original datum represented a fresh weighing and a new treatment application. Thus, it was believed that the use of variance analysis for these data was theoretically sound. As a matter of fact, moderate departures from normality (the first requirement, which was not checked directly) usually have no serious influence upon the validity of conclusions drawn from an analysis of variance (13, 14, 17 chapt. 10).

2. This experiment and the subsequent analysis were considered to be a "Model II," and the expected mean squares shown in Table I were derived under this assumption. In such an analysis, the several levels of each variable tested are considered to be samples from a large population of possible levels, rather than as representing complete populations. The decision in this matter is sometimes self-evident, but frequently, as in the present case, it depends upon the nature of the questions the experimenter is asking. In the present instance, the object, as stated previously in

slightly different language, was to test typical antiperspirant agents in order to evaluate the ability of the test method to find differences; the four treatments chosen were thus considered to represent a sample taken from a large possible number, many others of which could have served the same purpose. In the same way, the observation periods were similar samples. On the other hand, had interest lain in the particular four treatments and periods used and in no other, the analysis would then have been considered to be a "Model I" and the expressions for the expected mean squares would have been different. These expressions dictate the method of significance testing used after the variance analysis is complete, and it is therefore important to predetermine them.

Data Obtained

The experiment was run as described, and the weight differences obtained in each case were coded and recorded. The four data for each treatment, each day and each Latin Square were then averaged and tabulated. These means of four were the raw materials for the analysis of variance used, and they are presented in Table II.

Table II—Means of Complete Latin Squares of Four Subjects Each

Treatment and Time	Square I (Subjects 1–4)	Square II (Subjects 5–8)
T_1D_1 (3 hr)	25.63	21.60
T_1D_2 (27 ")	21.08	15.30
T_1D_3 (51 ")	25.53	20.03
T_1D_4 (75 ")	22.80	15.63
T_1D_5 (99 ")	20.55	25.13
T_2D_1	18.48	16.35
T_2D_2	13.38	11.65
T_2D_3	18.10	14.15
T_2D_4	12.88	10.40
T_2D_5	10.80	13.83
T_3D_1	24.35	13.28
T_3D_2	16.90	10.48
T_3D_3	15.43	14.00
T_3D_4	12.90	11.20
T_3D_5	13.95	15.45
T_4D_1	30.38	23.23
T_4D_2	20.55	14.78
T_4D_3	24.18	19.78
T_4D_4	32.38	16.75
T_4D_5	28.28	21.58

Analysis of Data

A conventional analysis of variance was carried out on the data of Table II. The computational work included a Bartlett test for homoscedasticity (18, pp. 38–39), and significance tests as detailed below (the Bartlett test

gave no evidence of heterogeneity). The residual sum of squares was computed directly from the replicates represented by the two groups of subjects. The analysis of variance table is presented below as Table III.

The significance tests (F ratios) were done as follows, guided by the expressions for the expected mean squares shown in Table I:

(1) The TD interaction was tested against the residual and was found nonsignificant.
(2) The D (times) effect was tested against the interaction mean square and found significant (99% level). The reason for this choice of error term here and for treatments is explained below.
(3) The T (treatments) effect was also tested against the interaction mean square and found significant (99.9% level).

The use of the residual mean square as an error term to test the significance of the interaction effect is conventional, and in this case there is little risk in concluding that the interaction, if it exists, is small. Having drawn this conclusion, it now became proper to test the main effects against either the mean square for interaction or that for the residual (13, pp. 131–132). In most cases, the use of the residual would be preferable, since it is usually based upon so many more degrees of freedom than that of the interaction that its precision is much higher. In the present case, however, there is little differentiation on this basis, since the difference in precision between an estimate of error based upon 20 degrees of freedom and one based upon 12 is quite small.* In such cases it is preferred in this laboratory to utilize the interaction mean square for error, on the basis that it is more realistic because of the terms composing it, even when the reality of the interaction variance has not been demonstrated. The two possible alternatives thus become (a) using the residual as error, or (b) pooling the residual and the interaction terms. The latter, although commonly done, results in an improper estimate of error if the interaction term is real and should be used with caution (13, p. 132; 20). In the present case, for example, such a pooled error term would have the components ($\sigma_\epsilon^2 + 0.75\ \sigma^2_{td}$), compared to that for the interaction ($\sigma_\epsilon^2 + 2\sigma^2_{td}$). The pooled term would therefore inflate the significance test if $\sigma^2_{td} \neq 0$.

TABLE III—ANALYSIS OF VARIANCE

Source of Variation	*DF*	*SS*	*MS*	*F*	Error Term	Significance Found
Among treatments (T)	3	638.33	212.78	32.63	MS_{td}	99.9%
Exposure times (D)	4	173.02	43.26	6.63	MS_{td}	99.0%
Treatments × times (TD)	12	78.25	6.52	0.35	MS_{rep}	none shown
Replication (residual)	20	377.33	18.87			
Total	39	1266.93				

* For example, the tabular F value at $P = 0.05$ for 3/12 DF is 3.5, while that for 3/20 DF is 3.1.

The results of the analysis of variance are shown in Table III:

Having shown significance for the two main effects, it was then necessary to make individual comparisons to determine which differences among the levels of each variable were responsible for this finding. For this purpose, a Tukey test was necessary (this is a modified t test for contrasting any pair out of k means) (13, 19). The error variance estimate needed was derived from the mean square term used for error in the analysis of variance; *viz.*, 6.52. The standard deviation for error was thus the square root of 6.52, or 2.55, based upon 12 degrees of freedom. The computation of the Tukey least significant difference is similar to the computation of a least significant difference using t, but with the substitution of a new value, q, for the product (t) $(\sqrt{2})$. When k (i.e., the number of means from which any pair is to be compared) is greater than 2, q becomes larger than (t) $(\sqrt{2})$. In the present case:

q_{95} for 4 means, with standard deviation estimated at 12 DF, = 4.20
q_{95} for 5 means, with standard deviation estimated at 12 DF, = 4.51

The least significant difference is

$$(LSD)_{95} = (S_{\bar{x}i})\ (q_{95})/\sqrt{n}$$

$(LSD)_{95}$ = Tukey least significant difference, 95% confidence level
$S_{\bar{x}i}$ = estimated standard deviation
n = number of items in each mean
q_{95} = multiplier described above.

Applying the formula,

$(LSD)_{95}$ for any pair of 4 means $= (2.55)(4.20)/\sqrt{10} = 3.38$*
$(LSD)_{95}$ for any pair of 5 means $= (2.55)(4.51),\sqrt{8} = 4.06$*

The four means for treatments and the five for exposure times were computed from the data of Table II and gave the following values:

Treatments (means of 10)	Exposure Times (means of 10)
$\bar{T}_1 = 21.33$ (blank)	$\bar{D}_1 = 21.66$ (first treatment)
$\bar{T}_2 = 14.00$ (al. chlorhydrate)	$\bar{D}_2 = 15.52$ (second ")
$\bar{T}_3 = 14.79$ (experimental)	$\bar{D}_3 = 18.90$ (third ")
$\bar{T}_4 = 23.19$ (sod. al. chlor. lac.)	$\bar{D}_4 = 16.87$ (fourth ")
	$\bar{D}_5 = 18.70$ (fifth ")

Comparing the above least significant differences to these data, it is apparent (with a probability 0.05 of being incorrect), that $\bar{T}_1 > \bar{T}_2$, $\bar{T}_1 > \bar{T}_3$, $\bar{T}_4 > \bar{T}_2$ and $\bar{T}_4 > \bar{T}_3$. No other treatment contrasts can be shown to be significant. Also, $\bar{D}_1 > \bar{D}_2$ and $\bar{D}_1 > \bar{D}_4$ and no other exposure time contrasts can be shown significant.

This concluded the statistical examination of the data.

* Note that the *LSD* values, as well as the means following, are expressed in coded test units (grams of perspiration $\times 100$).

Interpretation of Results

Having performed the statistical analysis, it was then necessary to "translate" the numerical conclusions into English, as detailed below.

Based upon the data of this experiment, accepting the 0.05 probability level as evidence that a real difference exists, and assuming that the subjects tested were reasonably typical in their physiology, the following conclusions may be drawn:

1. The silica gel weighing method is capable of distinguishing among treatments having varying antiperspirant efficacy.
2. The antiperspirant effect of an active agent, used as described, apparently requires a period of time to develop.
3. The results suggest that treatment of a previously untreated skin site may have less antiperspirant effect than the same treatment applied a second time, i.e., that the skin "remembers" the initial application.
4. Beyond the initial period of a few hours, repeated applications of an antiperspirant compound do not appear to increase its effectiveness. On the other hand, there is no evidence of a decrease over the period of time used in the experiment.
5. The use of analysis of variance with silica gel test results appears to give reasonably sensitive detection of differences among application variables.

These observations led to the conclusion that further work utilizing the silica gel method in conjunction with adequate experimental design and statistical analysis was justified.

Discussion

It is quite evident from the statistical evaluation that the method described in this paper yielded reliable and precise results in the study of antiperspirant effects. Ease of use and the ability of the silica gel to absorb large quantities of moisture are two further advantages over previous procedures. By minor changes in the design of the collection containers, they may be, and eventually were, used to study axillary sites, thus permitting quantitative comparisons of perspiration produced by axillae to that produced on the back or other areas.

Several of the references cited at the end of this paper, as well as many others not listed, describe methods which yield data in quantitative terms. In view of this fact, it is surprising that almost no evidence of the use of statistical procedures has been found. Such procedures are in very common use elsewhere, for example, in chemical, medical and biological product development work. In the absence of a statistical treatment of experimental data resulting in conclusions associated with confidence levels, the experimenter subjects himself to a severe handicap, because he has no

measure of the dependability of his conclusions. Statements or numerical computations showing relationships among relative effectiveness of several antiperspirants, for example, do not necessarily imply real differences in antiperspirant action, since no account of the magnitude of the associated experimental errors is taken in a simple comparison of averages or of individual results.

Some of the published work utilized methods for inducing perspiration (such as treadmills) presumably in order to increase the quantities of sweat produced and thus increase the ease of measurement. Such experiments are of course of great interest, but it is felt that these results have not been proved to reflect conditions which may exist when the perspiration rate is more moderate; this might be especially true when preparations or ingredients intended for every day use are being tested. It is possible that conditions of heavy, induced perspiration may be associated with a different response to a given antiperspirant, or may show a different relationship among two or more test substances, than might be found under normal use conditions.

It was interesting to note that no significant difference was found, in the work described in this paper, between the amount of perspiration produced by skin treated with sodium aluminum chlorhydroxy lactate and that produced by skin treated with a blank consisting of aqueous sodium chloride, especially in view of the effectiveness of the aluminum chlorhydrate. In our present state of relative ignorance of the mechanism of the operation of antiperspirants, however, no worthwhile explanation seems available. Of course, this investigation was directed toward the evaluation of a method and the use of statistics therein rather than to such questions. Some further work is being done in the investigation of fundamental mechanisms, and a rationale for the chemistry of an effective antiperspirant may eventually be evolved.

It is hoped that this presentation will be of assistance to others in carrying out experimentation in the field of antiperspirants and of the physiological (and possibly psychological) mechanisms of human sweating.

Acknowledgment: This work was done under the guidance and encouragement of the late Raymond Stetzer, President of the Revlon Research Center, Inc.

References

(1) V. Minor, *Deutsch Z. Nervenheilk.*, **101,** 302 (1927); *Zbl. Haut Geschlechtskrankh.*, **44,** 727 (1933).
(2) W. C. Randall, *J. Clin. Invest.*, **25,** 761 (1946).
(3) S. A. Muller and R. R. Kierland, *J. Invest. Dermatol.*, **32,** 126 (1959).
(4) C. Papa and A. M. Kligman, *Ibid.*, **36,** 167 (1961).
(5) Y. Takahashi and M. Wada, *Ibid.*, **38,** 197 (1962).
(6) R. Brun, *J. Soc. Cosmetic Chemists*, **10,** 70 (1959).
(7) E. G. Helton, E. W. Daley, and J. E. Ervin, *Proc. Sci. Sect. Toilet Goods Assoc., No. 26,* 27 (1956).

(8) W. J. O'Malley and J. E. Christian, *J. Am. Pharm. Assoc.*, **49,** 398 (1960).
(9) E. W. Rosenberg, H. Blank, and S. Resnick, *J. Am. Med. Assoc.*, **179,** 809 (1962).
(10) W. G. Fredell and R. R. Read, *Proc. Sci. Sect. Toilet Goods Assoc., No. 15,* 23 (1951); *Ibid., No. 25,* 32 (1956).
(11) M. J. Rodman, *J. Am. Pharm. Assoc.*, **42,** 550 (1953).
(12) W. M. Wooding, *Tappi,* **39,** No. 6, 417 (1956).
(13) O. L. Davies, *Statistical Methods in Research and Production*, 3rd Ed., Hafner Publishing Co., New York, 1961.
(14) O. L. Davies, *Design and Analysis of Industrial Experiments*, Hafner Publishing Co., New York, 1956.
(15) D. S. Villars, *Statistical Design and Analysis of Experiments for Development Research*, W. C. Brown Co., Dubuque, Iowa, 1951.
(16) W. M. Wooding and H. E. Jass, Unpublished paper, First Annual Clinic and Conference on Statistics and Quality Control in the Consumer Product Industries, Louisville Sect. A. S. Q. C., Louisville, Ky., June, 1963.
(17) H. Scheffe, *The Analysis of Variance*, John Wiley and Sons, Inc., New York, 1961.
(18) K. Brownlee, *Industrial Experimentation*, 4th Ed., Chemical Publishing Co., New York, 1953.
(19) E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. I, Cambridge University Press, Cambridge, England, 1954.
(20) L. P. V. Johnson and E. S. Keeping, *Appl. Statistics*, **1,** 202 (1952).

Appendix

Explanation of Statistical Terms and Procedures

The following is a brief summary of some statistical terms and procedures, for the convenience of those to whom they may be unfamiliar. It is obviously impractical to be thorough. For further information, references (13) and (14) will be found particularly lucid and well-adapted to the practical investigator's point of view.

Full Factorial Design: An experimental arrangement comprises a full factorial design if all levels (separate values) of each experimental variable (such as treatments in the present case) are run with each level of each of the other variables (such as days in this experiment).

Balance: An experimental design is considered to be balanced if each combination of levels of the variables occurs the same number of times as each of the others.

Analysis of Variance: This is a statistical procedure for the analysis of data resulting from experiments. It separates sources of variation in the data due to the controlled factors, their interactions with each other, and experimental error due to one or more sources. Quantities representing variation due to the controlled variables are then compared with a measure of variation due to error, to determine whether the apparent effect of the controlled factors or variables may reasonably be concluded to be greater than that accounted for by error alone. A variable believed to be real is said to be "significant," and statements of significance are accompanied by a probability statement referring to the chance that the observed effect is due to chance alone (i.e., to experimental error), and that the conclusion of reality is untrue.

Latin Square: A two-dimensional arrangement of quantities in a table having the same number of rows as columns, and having certain special characteristics, principally "orthogonality," or balance. For example, in the present experiment, the assignment of the four treatments to four positions on the subjects' backs was controlled by two duplicate Latin Squares, viz:

	Positions			
Subject	1	2	3	4
1	T_1	T_2	T_3	T_4
2	T_2	T_3	T_4	T_1
3	T_3	T_4	T_1	T_2
4	T_4	T_1	T_2	T_3

The orthogonality of the above 4 × 4 Latin Square lies in the fact that each treatment appears once in every position (column) and once on every subject (rows).

Mean Square: A measure of variation composed of one or more "pure variances" which in turn are numerical expressions representing variation due to one or more causes. For example, σ_ϵ^2 represents the variance due to all experimental error effects; i.e., all sources of variation in the experimental data not accounted for by the controlled variables or factors. Similarly, σ_t^2 represents the variance due to differences among the several treatments. If no real difference exists, a variance will, of course, be theoretically equal to zero. However, in practice, true variances are unknown and must be estimated; such estimates are also subject to uncertainty.

Replication: Repetition of a measurement or a set of measurements in order to provide an estimate of experimental error.

Interaction: Two factors are said to show a significant interaction if the effect on the measurement due to the differences between two or more levels of the first is different at some levels of the second. For example, in an experiment to measure the effect on the yield in a chemical reaction when temperature is varied and two different catalysts are used, an interaction would be present if the change in yield with a given temperature change were greater with one catalyst than with the other.

(Received February 13, 1964)