J. Soc. Cosmet. Chem., **26**, 255–275 (May 1975)

A Critical Comparison of Two Procedures for Antiperspirant Evaluation

W. M. WOODING, B.Ch.E.* and PAUL FINKLESTEIN, Ph. D.†

Presented December 2, 1974 New York City

Synopsis—This paper is concerned with EXPERIMENTAL DESIGN and DATA ANALYSIS PROCEDURES associated with GRAVIMETRIC AXILLAR ANTIPERSPIRANT TESTS. The most widely known of these tests requires the estimation of "control ratios" for each subject, based upon data obtained prior to the evaluation of product application results. These ratios are subsequently employed as correction factors to modify similar ratios in sweating responses, which are calculated during the product evaluation phase of the test. It will be shown that an improved procedure, which has been in use by this laboratory for more than 7 years, is substantially more economical with regard to time and effort than the above procedure. The new procedure does not require preliminary testing or the establishment of "control ratios." In addition, it is conceptually more rigorous, allows more definitive conclusions to be drawn, and conforms to established principles of statistical design and analysis.

In summary, it is claimed that the new method is easier, faster, more economical, and can be shown to be unbiased. The use of the two procedures will be illustrated with examples.

Introduction

Although several methods have been used to estimate the effectiveness of antiperspirants in the axillae, the gravimetric procedure is the most widely practiced one. This involves collection and weighing of axillar sweat under controlled conditions. It has the advantages of simplicity and of applicability to fairly large panels of human subjects and gives quantitative results suitable for mathematical analysis.

The earliest literature dealing with gravimetric procedures seems to be the well-known Fredell and Read paper, which was published in 1951 (1). These

* Carter Products Research, Carter-Wallace, Inc., Cranbury, N.J. 08512.
† Health Care Division, Johnson & Johnson, Inc., New Brunswick, N.J. 08901.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

workers appear to have been the first to introduce the concept of using a test material in one axilla of each subject while leaving the other untreated as a control. For a period of time before the initial use of the actual test materials, they made daily sweat collections from both axillae of each subject taking part in the test. They then calculated pretest ratios (i.e., milligrams of sweat from the right axillae–milligrams of sweat from the left axillae) for each subject. These pretest or control ratios were subsequently compared with posttest right/left ratios, the magnitude of the differences being taken as indicative of the degree of efficacy of the antiperspirant material under test. Fredell and Longfellow in 1958 (2) suggested standardization of some of the test conditions, but continued to use right/left ratios in the same way. However, methods of calculation were not clearly defined, and although several pretest determinations were recommended, the number was not made explicit. Also in 1958, Daly (3) described a ratio method of calculation, but his procedure was not clear.

Laboratory protocols used by most workers today are similar to those described by these early workers, but the ratio which now appears to be universal is that of milligrams of sweat from treated axilla per milligrams from control axilla, regardless of the side treated. The pretest ratio in this case, of course, would be that of milligrams yielded by the axilla destined to be treated per milligrams yielded in the opposite axilla. The method of adjusting posttest ratios with the pretest ratios has been described by Majors and Wild (4). The methods used by the older workers employing right/left ratios were not entirely clear.

In the literature, we have been unable to find a description of the evolution of the modern axillar evaluation procedure using ratios, but protocols for tests of antiperspirant with control, as well as the computation and use of treated side/control side ratios, are discussed in the Majors and Wild paper (4).

One of the authors (5) has described the analysis of gravimetric data collected in experiments resembling those discussed by Majors and Wild, but without the use of pretest ratios. In that paper, certain aspects of the population distributions of axillar milligram data were discussed, and a statistical model was proposed. Although the differences between that procedure and the ratio method do not seem to have attraced much attention, we believe it possible to show conclusively that that method is superior to the ratio method on the bases of statistical rigor, ease of correct computation, simplicity, and economy.

The purposes of this paper are as follows:

1. to propose a standard method for comparing an antiperspirant with a control, which does not require the use of pretest ratios, and which will give statistically correct and unbiased estimates of per cent reduction and confidence intervals;

2. to show that commonly used data analysis methods, involving the deter-

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

mination of pretest ratios and the adjustment of posttest ratios thereby, are incorrect and produce incorrect results;

3. to demonstrate a statistically correct method of analyzing ratio data;

4. to show that point estimates of per cent reduction obtained with the ratio model disagree with those of the standard method, even when correct calculations are used; and

5. tc briefly describe experimental designs for comparing more than one antiperspirant with control, using tests requiring 5 days, without the use of pretest ratios.

Experimental Procedure

For purposes of identification we will call our method the Sides Subjects Effects Model (SSEM) and the ratio method the Ratio Model (RM). As far as we can determine, we are the only workers consistently using the SSEM. Minor variations in the clinical procedures employed exist among users of the RM, and our description below is typical of, but not necessarily identical with, those used by a given investigator.

Both methods are concerned with the same problem: the comparison of an antiperspirant material with a control in order to assess efficacy; and obtaining this information as precisely as possible, without interference or distortion due to known "natural" differences between sides or among subjects (Majors and Wild have produced numerical evidence of a commonly hypothesized disparity between quantities of sweat produced in right and left axillae, which may be related to whether a subject is right or left handed (4)).

For simplicity, we will confine all of our detailed comparisons of the methods to "two-sample" tests in which one antiperspirant material is compared with control, although there is no theoretical reason why more than one antiperspirant may not be used, and designs of this kind will be described briefly in a later section.

The two methods are essentially identical in the clincal procedures used, with the exception of the use of pretest control runs in the RM; they may vary in some details, however. The object of both methods is to produce a point estimate of per cent reduction for antiperspirant relative to control, with confidence limits as a measure of its precision.

The SSEM procedure is done as follows. First 36 subjects are usually used. They are required to abstain from the use of any antiperspirant product for at least 4 weeks before the test begins, although they may use deodorants. The subjects are arbitrarily numbered before the test begins; then, using a table of random numbers (or a set of random numbers generated by a computer), 18 are randomly assigned to receive antiperspirant in the left axilla and control in the right, leaving the other 18 assigned to the opposite configuration. "Control" usually implies no treatment rather than a placebo. Applications of product are made daily for 4 days. Methods of application are stan-

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

dardized in detail, but depend upon the physical form of the antiperspirant material used. When both "ambient" and "hot-room" tests are to be run, the ambient test is done on the fourth day, beginning 1 hour after the fourth application of the product. A preweighed "ambient" pad of non-woven fabric is placed in each axilla of each subject using a harness device or tape to hold it in position. The subjects go about their normal business for 3 hours, then return to the clinic, where the pad is removed, reweighed, and the weight inincrease recorded. The subjects return to the clinic on the fifth day, receive a fifth application of the product and 1 hour later enter a hot room maintained at 105°F and 50% RH. They are seated in a random spatial order around the hot room, with a nonwoven fabric pad inserted in each axilla. They are asked to sit quietly with their hands in their laps and both feet on the floor throughout the hot-room period. After 40 min., the pads are removed and discarded (pad A). A second pad, this time preweighed, is then inserted (pad B), left in place for 20 min., then removed, reweighed, and the weight gain recorded. Finally, a third pad (pad C) is inserted and handled in the same manner as pad B. This completes the test, and the subjects are dismissed.

A number of details are not included in the above account, such as exact timing, randomization of the order of treating axillae, etc.

In the RM method, the procedures are similar, but one, two, or more days of pretest control data are obtained before treatments with antiperspirant begin. Confining the discussion to hot-room tests only for simplicity, these pretest runs are made both with pad B and pad C. It should be remembered that there are many variations of the RM test. Some conduct a test very similar to the SSEM just described. Others treat all subjects on one side, wait 2 weeks, then treat all on the other side. These variations, however, are variations in the experimental design and are unrelated to the central question to which we address ourselves. Therefore, to demonstrate differences between the two models on an equal basis, we will assume that both protocols are identical except, in the case of the RM, for the provision for the use of pretest ratios.

Statistical Considerations

General

The correct computational procedures for either model are dictated within rather narrow limits by the nature of the experimental design (including type of randomization) and the procedure of the protocol. Unless these particular computations are done, incorrect values of per cent reduction and/or confidence limits will be obtained. The computational procedures are therefore as important as the clinical ones and must follow valid statistical principles, take careful account of all underlying assumptions, and be compatible with the nature of the randomizations used in the experiments. These requirements

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

should be noted, as the use of incorrect randomization procedures, and computations are very common.

The procedures to be described must be applied separately to data from B and C pads. Because of the probable lack of independence of the errors of the two sets of data, it is not correct to regard them as replicates. Of course, averages may be used, if desired, or a slope analysis done. We prefer to analyze them separately. Actually, it is not necessary to use two pads, as we have found very close agreement between them in hundreds of tests.

SSEM Analysis (Crossover)

There is more than one possible model that could be adopted for the SSEM. The one we use at present, as reflected by the randomization procedures employed in assigning treatments to subjects and the subsequent handling of the subjects, is a crossover design (unlike many crossovers, however, the rows of the design represent sides rather than time periods). This design is illustrated in Figure 1. The columns represent subjects, the two rows sides, and the letters T_1 and T_2 the antiperspirant and control treatments. The statistical population model describing this design and its analysis is

$$y_{ijk} = \mu + \alpha_i + \sigma_j + \tau_k + \epsilon_{ijk}$$

where

y_{ijk} = ln mg of sweat for axilla i, subject j, and treatment k
μ = general mean
α_i = axilla i $\quad i = 1, 2$
σ_j = subject j $\quad j = 1, 2, \ldots n$
τ_k = treatment k $\quad k = 1, 2$
ϵ_{ijk} = error for axilla i, subject j, and treatment k

α fixed; $\quad \Sigma\alpha_i = 0$
σ random; $\quad \sim N(0, \sigma_e^2)$, independent
τ fixed; $\quad \Sigma\tau_k = 0$

As shown in the above model, the usual statistical assumptions underlying the design and analysis are implied: the errors are independent, randomly and normally distributed and homogeneous; treatments and sides represent fixed populations. The data analysis equates certain interactions with error under the assumption that they do not represent "real" effects in the population, as is done in Latin square designs. The analysis of data under this model follows that outlined briefly in a previous paper (5). It is assumed that the data are tabulated to show milligram values identified by subject, side, and treatment. The purpose of the analysis is to obtain estimates of the "treatments effect" and associated error uncontaminated by the influence of either the sides or subjects effects. Details of the analysis will be given later in the form of an example. Its major feature is that it requires transformation of the original milligram values to their logarithms before the statistical computa-

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

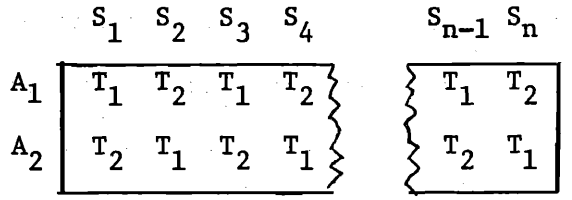

Figure 1. SSEM design represented as a crossover: A_1 represents left axillae; A_2 represents right axillae; S_1, S_2 . . . S_n represent subjects; T_1 represents antiperspirant; T_2 represents control. (Appropriate randomization not shown.)

tions begin in order to provide conformance to the assumptions of normality and homogeneity of variance. (The work reported in the earlier paper (5) showed this to be necessary, and has been confirmed by a large number of subsequent experiments.) A second important feature is the type of randomization procedure used, which is appropriate to the crossover design.

RM Analysis

The SSEM analysis, coupled with correct experimental design and randomization, can be shown to produce statistically unbiased estimates of per cent reduction and experimental error. The RM uses a similar design, but the analysis normally used assumes that sides effects are removed by the adjustment procedure. For the RM to remove side effects fully, however, it would be necessary that the pretest ratios be constants.

It is easy to observe by examination of any set of pretest ratios done repeatedly on the same subjects (see Table I) that the ratios are not constants. The fact that they are more uniform than milligram values of sweat produced is irrelevant, since ratios, not milligrams, are used in the adjustment procedure.

As an indicator of the degree of variability of pretest ratios with time, a correlation coefficient is a suitable statistic, although certain kinds of bias will remain undetected thereby. We carried out such tests on a number of pretest ratios determined 1, 3, and 21 days apart with the same subjects, using a rank correlation procedure to avoid violation of the statistical requirements of normality and homogeneity of variance. We obtained values ranging from less than 0.50 to 0.87 (a value of 1.00 would have indicated perfect correlation between successive measurements on the same subjects).

In addition to the above, we noted that the variance of adjusted mean posttest ratios is a function of the number of pretest measurements made and averaged, which are then used in the adjustment procedure. It is possible, with the use of a sufficient number of pretest measurements, to exercise considerable control of the experimental error of the final mean ratios. In one case, for example, the width of the confidence limits about the mean per cent reduction

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

Table I
Typical Set of Two-Sample Test Data: Pad B Only, Four Applications before Posttest Readings (Milligrams and Ratios)

		PRETEST DATA				POSTTEST DATA (after 4 applications)	
		Day 1		Day 2			
Subject Number	Treatment On (side)	Right	Left	Right	Left	Right	Left
13	R (mg)	606	599	690	704	263	630
	(ratio)	1.012		0.980		0.417	
14	R (mg)	657	606	776	695	445	670
	(ratio)	1.084		1.117		0.664	
15	R (mg)	630	555	646	593	304	380
	(ratio)	1.135		1.089		0.800	
16	R (mg)	356	262	415	310	217	420
	(ratio)	1.359		1.339		0.517	
17	R (mg)	400	409	489	546	336	497
	(ratio)	0.978		0.896		0.676	
18	R (mg)	210	350	394	556	288	557
	(ratio)	0.600		0.709		0.517	
19	L (mg)	789	332	1060	809	850	365
	(ratio)	0.421		0.763		0.429	
20	L (mg)	710	589	750	568	400	257
	(ratio)	0.830		0.757		0.643	
21	L (mg)	725	607	825	684	460	261
	(ratio)	0.837		0.829		0.567	
22	L (mg)	809	663	312	243	430	200
	(ratio)	0.820		0.779		0.465	
23	L (mg)	587	612	745	860	788	325
	(ratio)	1.043		1.154		0.412	
24	L (mg)	618	461	547	523	555	283
	(ratio)	0.746		0.956		0.510	

varied, as the number of pretest measurements used to compute a geometric mean pretest ratio was increased, as follows:

Number of Pretest Measurements	Width of 95% CL about Per Cent Reduction Computed From Adjusted Posttest Ratios
1	26.4 (% reduction units)
2	22.6 (% reduction units)
4	17.2 (% reduction units)

This reduction of the confidence limits about the PR is sufficient so that, even with geometric means of only two pretest ratios, it is often possible to equal or exceed the precision obtained with the SSEM when a single posttest determination is used with the latter. This was the case in examples 2, 3, 4, and 5 of Table VII (as will be shown later). Of course, the precision of the SSEM

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

can likewise be controlled by additional replication (for example, by averaging data obtained on successive days or by the use of additional test subjects*).

Aside from the above, the RM consistently yields different point estimates of per cent reduction than those obtained with the SSEM, using the same posttest data. Since the latter can be shown to be correct due to the balance of the experimental design, the question arises as to the source of the disagreement. Either both estimates are correct, or the use of the RM introduces a bias into the estimates. Since the confidence intervals obtained by either method (with correct calculations) generally include both point estimates, this question cannot be answered with certainty without further investigation. It seems clear, however, that the safe procedure is to use the SSEM, in view of its known validity (in addition to its practical advantage of requiring less time and effort).

The above remarks apply to the comparison when statistically correct methods are used in analyzing RM data. However, there appear to be as many methods of data analysis as there are practitioners, and all of those we have examined are incorrect and produce incorrect results. There are three common errors, which are (1) lack of recognition of the nonnormal character of the milligram-weight ratios and per cent reduction values; (2) lack of correct randomization prior to and during the clinical work; and (3) the use of a design implying a model not reflected by the analysis. The first of these errors can be remedied by an appropriate transformation of the ratios. Since it has been shown that the milligram weights used to form the ratios are log normally distributed (5), it follows that the adjusted ratios are also log normally distributed. Thus the proper transformation of the ratios is logarithmic. The problem of finding an appropriate transformation in order to validate the assumptions underlying the ordinary statistical procedures is a very common one and has been treated extensively in the literature of applied statistics (6–12).

The second error can be corrected by the use of the appropriate randomization procedures and the third by the performance of a suitable analysis.

Summary

Since the removal of the sides effect from error is a property of the analysis used with the SSEM and since the balance in the design guarantees an unbiased estimate of the per cent reduction, questions inherent in the use of the

*If this is attempted, however, care must be taken to do the analysis correctly, as measurements on the same subjects on successive days are likely to be correlated.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

RM are easily avoided by using the SSEM. We will now present examples to illustrate the calculations for the SSEM in detail and to compare these with RM calculations.

Examples

General

Using data from several sources, we have analyzed 5 sets of two-sample data in three different ways. None of these methods will efficiently detect interactions among sides, treatments, and subjects, although a new design, discussed briefly later on and now being studied in this laboratory, has this capability. The most interesting of these interactions would be those involving treatments, of course, but it is not known at present whether they exist and, if so, whether their magnitudes are important.

We will illustrate the following procedures in this section.

A. SSEM procedure: This is the recommended analysis and gives unbiased estimates of per cent reduction. In this analysis, the effects of sides, subjects, and treatments are removed from the error estimates used to compute confidence intervals. The method does not require pretest ratios.

B. RM procedure (incorrect numerical analysis): This illustration will use one of the several incorrect procedures commonly employed to obtain per cent reduction estimates and confidence intervals, using adjusted ratios. The procedure is included to illustrate the distortion of results brought about by its violation of the statistical assumptions of normality and homogeneity of variance, as well as its failure to take residual sides effects into account.

C. RM procedure (correct analysis): This illustration will show the analysis of adjusted ratios when the above assumptions and a possible residual treatments times sides effect are accounted for. Its disagreement with the correct results yielded by the SSEM analysis (see (A) above) will be shown.

We shall describe the computations for each of the above procedures in some detail, using a single small set of data to allow the reader to compare the arithmetic involved in each case. We shall then summarize the results we obtained in applying the same procedures to each of the four other sets of similar two-sample data.

Table I gives a set of milligram data from 12 subjects, using the design of Fig. 1, with both pretest and posttest weights, for an antiperspirant known to be effective. For simplicity, we present data from only 1 pad and 1 hot-room collection posttest, the latter done after 4 antiperspirant applications. Note that, as was explained earlier, SSEM tests done for this laboratory normally use only 1 day of hot-room testing, done after 5 daily antiperspirant applications, and our test is completed in 5 days.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

The data of the example represent a portion of a larger study, which included more than 1 active product. As was indicated earlier, the recommended minimum number of subjects for the SSEM procedure is about 36.

A. *SSEM Calculations*

1. For the SSEM calculations, the milligram posttest data are first transformed to their natural logarithms, which are shown in Table II.

2. An analysis of variance (ANOVA), mixed model (treatments and sides fixed), is done on the transformed data of Table II, using the crossover model. The ANOVA is shown in Table III, with back-transformed mean milligrams for treatment and control.

Table II
Natural Logs[a] of Posttest Data of Table I

Subject Number	Treated On	Right	Left
13	R	5.572	6.446
14	R	6.098	6.507
15	R	5.717	5.940
16	R	5.380	6.040
17	R	5.817	6.209
18	R	5.663	6.323
19	L	6.745	5.900
20	L	5.991	5.549
21	L	6.131	5.565
22	L	6.064	5.298
23	L	6.669	5.784
24	L	6.319	5.645

[a]Rounded to nearest three decimals for convenience; the actual analysis by computer used ten. It is generally sufficient to use two or three places if hand calculations are done.

Table III
ANOVA of Data of Table II

SOURCE OF VARIATION	DF	SS	MS	F	P
Treatments (Product versus control)	1	2.2795	2.2795	107.477	<0.0001
Sides (Left versus right)	1	0.0385	0.0385	1.816	<0.20
Subjects	11	1.0611	0.0965		
Error	10	0.2121	0.0212		
TOTAL	23	3.5912			

	Means of 12 Data	
	Based upon transformed data	Back-transformed to mg (antilogs)
Treatment	5.6656	288.8
Control	6.2820	534.9
Left axillae	5.9338	377.6
Right axillae	6.0139	409.1

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

3. The mean per cent reduction is computed as follows:

$$PR = \frac{T_2 - \overline{T}_1}{T_2} 100 = \frac{534.9 - 288.8}{534.9} 100 = 46.01$$

or

$$\overline{PR} = \left(1 - \frac{\overline{T}_1}{T_2}\right) 100 = \left(1 - \frac{288.8}{534.9}\right) 100 = 46.01$$

where T_2 represents control mean in milligrams, and $\overline{T}_1$ represents the treatment mean in milligrams.

Note that the ANOVA F tests indicate a strong antiperspirant effect and a small sides effect, if any, in this case.

4. Confidence limits about the population mean per cent reduction* are estimated as follows:

(a) Using the error mean square (emr) from the original ANOVA (done on logs of the data), compute the standard error of the difference between the treatment and control means

$$SED = \bar{S}_d = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

where S_1^2 represents the variance for antiperspirant data; S_2^2 represents the variance for control data; n_1 represents the number of data in the antiperspirant mean; and n_2 represents the number of data in the control mean (under the assumption of homoscedasticity, which was satisfied by the log transformation $S_1^2 = S_2^1$). Because of the experimental design, equal numbers of axillae were used for antiperspirant and control; therefore $n_1 = n_2$. The above expression therefore reduces to

$$\bar{S}_d = \sqrt{\frac{2S^2}{n}}$$

where S^2 is the emr in the ANOVA.
Therefore

$$\bar{S}_d = \sqrt{\frac{2(0.0212)}{12}} = 0.0594$$

*95% confidence limits used in this way are indicators of the precision of the test but *a posteriori*, do not imply a probability that the population mean is included in the interval. This interpretation can only be used *a priori*, when one can say that (if $(1\text{-}\alpha) = 0.95$) the population mean will be included in 19 out of 20 identical experiments. A *posteriori*, the mean either is or is not included (12), and a probability estimate is inappropriate.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

(b) Compute the difference between the antiperspirant and control means (in terms of logs)

$$\overline{T}_1 = 5.6656: \qquad \overline{T}_2 = 6.2820$$
$$\overline{D} = 5.6656 - 6.2820 = -0.6164$$

note that the difference used is antiperspirant minus control, not the reverse.

(c) Compute the 95% confidence interval about the population difference

$$CL_{0.95} = \overline{D} \pm \bar{S}_d\, t_{0.05}$$

where $\overline{D}$ represents the above difference; $\bar{S}_d$ represents the standard error of the difference; and $t_{0.05}$ represents Student's t at $\alpha = 0.05$.
In the example

$$\overline{D} = -0.6164$$
$$\bar{S}_d = 0.0594$$
$$t_{.05} = 2.228^*$$

so that

$$CL_{0.95} = -0.6164 \pm (0.0594)(2.228)$$
$$= -0.6164 \pm (0.1323)$$

or

$$-0.4841 \text{ to } -0.7487$$

(d) A difference between the logarithms of two numbers is the ratio of the two numbers when antilogs are taken. Thus

$$\text{antilog } \overline{D} = \frac{\overline{T}_1}{\overline{T}_2}$$

where $\overline{T}_1$ and $\overline{T}_2$ are in milligrams. If we take antilogs of the two confidence limits, we obtain a maximum and minimum ratio of antiperspirant to control milligrams, representing 95% confidence limits about the population mean ratio:

$$\text{antilog } -0.4841 = 0.6163$$
$$\text{antilog } -0.7487 = 0.4730$$
$$PR_1 = (1\text{-}0.6163)\ 100 = 38.37$$
$$\overline{PR} \quad \text{(from step 3)} = 46.01$$
$$PR_2 = (1\text{-}0.4730)\ 100 = 52.70$$

*At 10 df (the error df in the ANOVA).

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

B. *RM Calculations (Without Transformations)*

The method to be illustrated manipulates individual adjusted ratios and is commonly used. Another, not illustrated here, uses individual per cent reduction values computed from single adjusted ratios as the basic statistics. Both methods are incorrect due to the fact that they use statistics (arithmetic means, estimates of standard deviations, and standard errors), which yield biased results because of the nonnormal character of the ratios and per cent reduction values. An additional source of bias in the error estimates is lack of provision for the influence of residual sides effects, since it is unlikely, as discussed above, that the adjustment procedure removes them completely. Procedure B is illustrated, therefore, solely to show the departure from the correct values given by method A. For simplicity, in our examples of the RM we use mean pretest ratios derived from two successive days of pad B runs, and a single hot-room test using pad B only is illustrated.

This procedure is commonly done as follows:

1. Individual pretest and posttest ratios are computed from the milligram data. For the data of the example, these were given in Table I.

2. Arithmetic means of the pretest ratios are calculated, and the adjusted ratios are computed by dividing the posttest ratios by these mean pretest ratios. For example, for subject 19 in Table I (chosen because of the fairly large disagreement between the two pretest ratios)

$$\begin{aligned}
&\text{Pretest ratio, day 1} = R_{p1} = 0.421\\
&\text{Pretest ratio, day 2} = R_{p2} = 0.763\\
&\text{Mean pretest ratio} = R_p = 0.592\\
&\text{Posttest ratio} = R_t = 0.429\\
&\text{Adjusted ratio} = \overline{R'} = R_t/R_p = 0.725
\end{aligned}$$

In a similar way, $\overline{R'}$ values for each of the 12 subjects of Table I are computed. These are shown in Table IV (last column).

The overall mean posttest ratio $\overline{R'}$ is obtained by averaging the adjusted ratios R′ in Table IV ($\overline{R'} = 0.6173$).

The variance, standard deviation, and standard error of the mean are computed from the adjusted ratios $\overline{R'}$

$$\begin{aligned}
&\text{Estimated variance} = S^2 = 0.0235\\
&\text{Estimated standard deviation} = S_{x_i} = 0.1533\\
&\text{Estimated standard error} = S_{\bar{x}} = 0.0442
\end{aligned}$$

95% confidence limits about $\overline{R'}$ are computed using $S_{\bar{x}}$ and Student's t at $p = 0.05$ and 11 degrees of freedom

$$\begin{aligned}
CL_{95} &= \overline{R'} \pm S_{\bar{x}}\, t_{0.05}\\
&= 0.6173 \pm (0.0442)\,(2.201)\\
&= 0.7146 \text{ to } 0.5200
\end{aligned}$$

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

Table IV
Summary of Ratios for Method B

Subject Number	R_{p_1}	R_{p_2}	$\overline{R_p}$	R_t	R′
13	1.012	0.980	0.996	0.417	0.419
14	1.084	1.117	1.101	0.664	0.603
15	1.135	1.089	1.112	0.800	0.719
16	1.359	1.339	1.349	0.517	0.383
17	0.978	0.896	0.937	0.676	0.721
18	0.600	0.709	0.655	0.517	0.790
19	0.421	0.763	0.592	0.429	0.725
20	0.830	0.757	0.794	0.643	0.810
21	0.837	0.829	0.833	0.567	0.681
22	0.820	0.779	0.800	0.465	0.582
23	1.043	1.154	1.099	0.412	0.375
24	0.746	0.956	0.851	0.510	0.599

The upper and lower confidence limits about the mean per cent reduction (PR_1, PR_2) are computed from these, and the mean per cent reduction $\overline{PR}$ is computed from $\overline{R'}$

$$\text{UCL for } \overline{R'} = 0.7146$$
$$\text{mean ratio, } \overline{R'} = 0.6173$$
$$\text{LCL for } \overline{R'} = 0.5200$$
$$PR_1 = (1\text{-}0.5200)100 = 48.00$$
$$\overline{PR} = (1\text{-}0.6173)100 = 38.27$$
$$PR_2 = (1\text{-}0.7146)100 = 28.54$$

C. *RM Calculations With Transformation*

This method takes into account the fact that correct estimates of the mean ratios and other statistics cannot be obtained by method B because neither the ratios nor the milligram values from which they are derived are normally distributed. In addition, despite the adjustment procedure, some sides effects may remain in the data, and therefore the error estimate must be obtained from the data after accounting for these. Since method C also uses adjusted ratios, however, it cannot compensate for any distortion in the per cent reduction estimates, which may be introduced by the adjustment procedure. Therefore, although the method is more nearly correct than B, it still may not give correct results. Note also that, quite aside from these considerations, the error estimate yielded by method C has a different composition than that of method A, although it is not necessarily incorrect.

The normality problem is handled by transforming the ratios to their logarithms before manipulating them, then back-transforming them. Since the

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

Table V
Summary of Ratios for Method C

Subject Number	Treatment on	R_{p_1}	R_{p_2}	$\bar{R}_G$	R_t	R′	R'_L
13	R	1.012	0.980	0.996	0.417	0.419	−0.870
14	R	1.084	1.117	1.100	0.664	0.604	−0.504
15	R	1.135	1.089	1.112	0.800	0.719	−0.330
16	R	1.359	1.339	1.349	0.517	0.383	−0.960
17	R	0.978	0.896	0.936	0.676	0.722	−0.326
18	R	0.600	0.709	0.652	0.517	0.793	−0.232
19	L	0.421	0.763	0.567	0.429	0.757	−0.278
20	L	0.830	0.757	0.793	0.643	0.811	−0.209
21	L	0.837	0.829	0.833	0.567	0.681	−0.384
22	L	0.820	0.779	0.799	0.465	0.582	−0.541
23	L	1.043	1.154	1.097	0.412	0.376	−0.978
24	L	0.746	0.956	0.844	0.510	0.604	−0.504

original milligram data are log normally distributed, it can be shown that their ratios, as well as the adjusted ratios (which are ratios of ratios) are also log normal. Interestingly, if the milligram data had been normal, the resulting ratios would not have been log normal, and a much more complex transformation would then have been required.

Referring to Table I, and again using the data from subject 19, we get the following:

1. For each subject, obtain the geometric mean of the pretest ratios

$$\begin{aligned}
&\text{Day 1 pretest ratio} && = 0.421; && \ln 0.421 && = -0.865\\
&\text{Day 2 pretest ratio} && = 0.763; && \ln 0.763 && = -0.270\\
&\text{Arithmetic mean of logs} && = (-0.865 - 0.270)/2 && && = -0.568\\
&\text{Antilog of mean} && = \text{geometric mean} && = && 0.567
\end{aligned}$$

(The close agreement between the mean of the logs and its antilog is a result of the particular value of the mean.)

2. Adjust the posttest ratios in the ordinary way, using these geometric mean pretest ratios

$$\bar{R}' = \frac{0.429}{0.567} = 0.757$$

3. Take natural logarithms of all adjusted posttest ratios. Table V lists these for the example data of Table I. In Table V, R_{p1}, and R_{p2} are the pretest ratios; $\bar{R}_G$ is the geometric means of the pretest ratios for each subject; R_t is the posttest ratio; R′ is the adjusted ratio; and R'_L is the natural logarithm of R′. Note the differences between the R′ values in Table IV and those in Table V. These are due to the different method of computing the mean pretest ratio.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

Table VI
ANOVA of R'_L Data of Table V
(Method C)

Source of Variation	DF	SS	MS	F
Due to side treated	1	0.0090	0.0090	0.106
Error	10	0.8470	0.0847	
Total	11	0.8560		

Most are very similar, but as can be seen, the difference increases as a function of the disagreement between the two pretest ratios. In experiments in which more than two pretest ratios are used, or with larger sets of subjects, such substantial disagreements will be more numerous.

4. Using the log-transformed adjusted ratios R'_L, carry out the ANOVA of Table VI (separating "side treated" and error). This procedure is equivalent to doing a t-test for side treated (the square root of F is t in this case), but is done as a convenient method for obtaining an error estimate free of any sides effect.

5. The standard error of the mean of the logs of the adjusted ratios is obtained from the ems in Table VI. Confidence limit are then calculated about the mean value of R_L, antilogs taken, and the per cent reduction and its 95% confidence limits obtained.

$$\text{Mean of logs of adjusted ratios} = \overline{R'_L} = -0.5097$$

$$\text{Standard error of mean} = \sqrt{\frac{0.0847}{12}} = 0.0840$$

$$t_{0.05} \text{ at } 10 \text{ df} = 2.228$$

$$CL_{0.95} = \overline{R'_L} \pm \bar{S}_x\, t_{0.05}$$
$$= -0.5097 \pm (0.0840)\,(2.228)$$
$$= -0.3226 \text{ to } -0.6969$$

$$\text{antilog } (-0.6969) = 0.498 \text{ (lower 95\% CL for ratio)}$$
$$\text{antilog } (-0.5097) = 0.601 \text{ (mean ratio)}$$
$$\text{antilog } (-0.3226) = 0.724 \text{ (upper 95\% CL for ratio)}$$

$$PR_1 = (1\text{-}0.724)100 = 27.60 \text{ (lower 95\% CL for PR)}$$
$$\overline{PR} = (1\text{-}0.601)100 = 39.90 \text{ (mean per cent reduction)}$$
$$PR_2 = (1\text{-}0.498)100 = 50.20 \text{ (upper 95\% CL for PR)}$$

Summary

Methods A, B, and C were applied to 5 sets of two-sample antiperspirant data involving varying numbers of subjects, following the procedures just illustrated. The results are summarized in Table VII.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

Table VII
Comparison of Results of Analysis of Two-Sample Data Sets Using Methods A, B, and C

Data Set Number	Statistic	Method A (SSEM)	Method B (RM)	Method C (RM)
1. (used in examples)	Number of subjects	12	12	12
	Mean PR	46.01	38.27	39.90
	CL_{95} about PR	38.4 to 52.7	28.5 to 48.0	27.6 to 50.2
	Width of CL_{95}	14.3	19.5	22.6
2.	Number of subjects	36	36	36
	Mean PR	20.51	15.03	16.61
	CL_{95} about PR	13.6 to 26.8	9.3 to 20.8	11.0 to 21.9
	Width of CL_{95}	13.2	11.5	10.9
3.[a]	Number of subjects	36	36	36
	Mean PR	44.14	39.91	43.10
	CL_{95} about PR	36.1 to 51.2	33.6 to 46.3	35.7 to 49.6
	Width of CL_{95}	15.1	12.7	13.9
4.	Number of subjects	8	8	8
	Mean PR	19.61	21.60	24.56
	CL_{95} about PR	−11.7 to 42.1	1.6 to 41.6	3.7 to 40.9
	Width of CL_{95}	53.8	40.0	37.2
5.	Number of subjects	12	12	12
	Mean PR	14.07	14.19	15.45
	CL_{95} about PR	1.3 to 25.2	3.2 to 25.2	3.8 to 25.7
	Width of CL_{95}	23.9	28.4	21.9

[a]This set showed a moderate sides effect in methods A and C. The sides effects in the other sets, before correction with ratios, appeared to be quite small.

Miscellaneous Related Topics

Sympathetic Effect

An hypothesis, which has been advanced occassionally, suggests that an effective antiperspirant applied to one axilla might stimulate additional sweating in the opposite (control) axilla. An argument in favor of the use of the RM, in spite of the disadvantages suggested in the foregoing discussion, might be the assumption that it will correct for this "cork effect" because it uses pretest control values, which are obviously unaffected by the antiperspirant treatment. We made calculations from the 5 sets of data summarized in Table VII, using the SSEM method, but with pretreatment instead of concurrent controls. Of course, this procedure introduces substantial additional variation,

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

but with these data we have found no evidence for the effect. If it existed, the expectation would be that the point estimates of per cent reduction obtained with the regular SSEM method (concurrent controls) would be biased in an upward direction (inflated), while the use of pretest controls should give lower PR values. If the cork effect were small or nonexistent, on the other hand, the two methods should vary in their outcomes in a random manner with, of course, larger errors when *a priori* controls are used. Our results on these sets of data do not show any consistency which could be considered significant; there were 4 cases in which the *a priori* controls yielded higher per cent reductions than those obtained with concurrent controls and one in which they were lower. We therefore believe that any cork effect in the data—if it exists at all—must be small.

Other Models

The discussion and examples in this paper have been confined to experiments in which only two "treatments" were considered: antiperspirant and control. However, it is possible to use three or even a greater number of treatments. In fact, at this laboratory the usual evaluation is a three-sample procedure in which a control and two different antiperspirants are tested. The ordinary SSEM procedure is used, but the design, instead of being a crossover, is a form of balanced incomplete blocks structure known as a Youden square. A schematic picture of such a design is shown in Fig. 2. The analysis is substantially more complex than that of the crossover, but is easily and quickly done with even small computer facilities and can be done in a few hours with a desk calculator. A similar design can be used for 5 samples. This is illustrated in Fig. 3.

Finally, there is a model for the two-sample test which may be superior in some ways to the crossover we have recommended. We are just beginning to

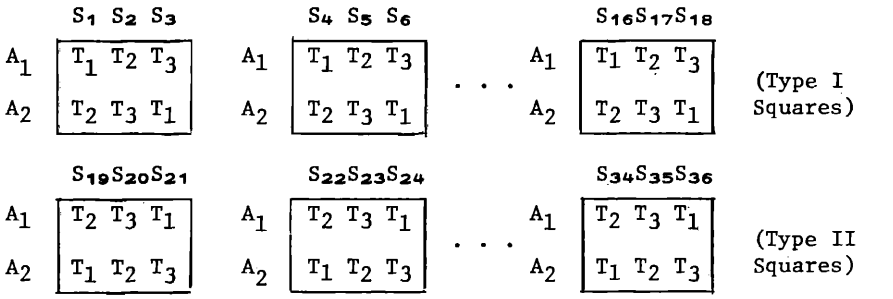

Figure 2. Three-sample test design (2 x 3 Youden squares) (illustrated for 36 subjects): A_i represents axillae, where i equals 1 or 2; T_j represents treatments (including control), where $j = 1, 2$, or 3; and S_k represents subjects, where $k = 1, 2 \ldots 36$. (Appropriate randomization not shown.)

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

(Type I Squares)

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
A_1	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
A_2	T_2	T_5	T_4	T_1	T_3	T_3	T_4	T_2	T_5	T_1

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
A_1	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
A_2	T_2	T_5	T_4	T_1	T_3	T_3	T_4	T_2	T_5	T_1

(Type II Squares)

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
A_1	T_2	T_5	T_4	T_1	T_3	T_3	T_4	T_2	T_5	T_1
A_2	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}	S_{40}
A_1	T_2	T_5	T_4	T_1	T_3	T_3	T_4	T_2	T_5	T_1
A_2	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5

Figure 3. Five-sample test design (2 x 10 Youden squares) (illustrated for 40 Subjects). A_i represents axillae, where i equals 1 or 2; T_j represents treatments (including control), where $j = 1, 2, \ldots 5$; and S_k represents subjects, where $k = 1, 2 \ldots 40$. (Appropriate randomization not shown.)

experiment with it, although it represents a well-known statistical design. It can be characterized as a 2^2 factorial design in which one degree of freedom is confounded with pairs of subjects. The laboratory protocol (it is a two-sample design) would be almost the same as the regular crossover of Fig. 1, but the error structure is more explicit and it will be possible to use it not only for the usual evaluations but also as a means of detecting treatments x sides interactions.

Conclusions and Discussion

General

In evaluating either the SSEM or the RM, the central questions are these: (1) does the method give an undistorted estimate of the true (population) per cent reduction; (2) if so, is it satisfactorily precise; and (3) how easy is it to use?

The SSEM analysis removes the average sides and subjects effects from error. The width of the confidence limits it produces is a function of the number of subjects used in the experiment, as well as the basic error associated with the protocol. The per cent reduction estimate can be shown to be unbiased, and the precision of the estimate is as good as can be obtained with a given number of subjects, using the crossover model.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

The RM gives estimates of per cent reduction which have been modified by the *a priori* mean ratio obtained at the time of the pretest measurements. If this mean ratio were constant, it would not alter the per cent reduction obtained relative to that of the SSEM. To the extent that this is not true, the per cent reduction will be different than that given by the SSEM. Since it has been shown that the ratios can be quite variable, it follows that the per cent reduction estimates obtained may be substantially altered. This difference in the estimates is obviously a function of the degree of variability of the pretest ratios. Under these circumstances, the precision of the estimate becomes irrelevant even though it is a function of the number of pretest determinations as well as the factors operating in the SSEM. A questionable estimate is not improved by reducing its variability.

It is possible, however, to utilize pretest data in a statistically valid manner to adjust posttest results by the use of an analysis of covariance, and this procedure might be shown to give more accurate, and precise estimates of per cent reduction than those given by the SSEM, if the pretest data are sufficiently well-correlated with the posttest ratio. The mathematics of such an analysis, however, differ from those of the ratio adjustment procedure, and it is impossible to predict results without experimentation. When time permits, we plan to investigate this matter. Meanwhile we recommend the SSEM over the RM as the scientifically valid method, as well as the one of choice on the bases of simplicity and low cost.

Acknowledgments

We wish to thank Dr. John Mandel for statistical advice which has been helpful in clarifying our thinking regarding a number of points. In addition, Mr. Giles Crane has contributed statistical assistance, and Miss Linda Timberlake performed a large number of statistical analyses in preparing the experiments. Finally, we would like to express our gratitude to the Editor and reviewers of this Journal for their careful, complete, and very constructive comments.

(Received November 12, 1974)

References

(1) W. G. Fredell and R. R. Read, Antiperspirant-axillary method of determining effectiveness, *Proc. Sci. Sect. Toilet Goods Ass.*, 23–7 (1951).
(2) W. G. Fredell and J. Longfellow, Report on evaluating antiperspirant and deodorant products, *J. Soc. Cosmet. Chem.*, **9**, 108-11 (1958).
(3) E. W. Daley, Antispirant testing: A comparison of two methods, *Proc. Sci. Sect. Toilet Goods Ass.*, 1–6 (1958).
(4) P. A. Majors and John E. Wild, The evaluation of antiperspirant efficacy–influence of certain variables, *J. Soc. Cosmet. Chem.*, **25**, 139–2 (1974).
(5) W. M. Wooding, The interpretation of gravimetric axillar antiperspirant data, *Proc. Joint Conf. Cosmet. Sci*, 91–105 (1968).
(6) N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, Inc., New York, 1966, chap. 3.
(7) F. J. Anscombe and J. W. Tukey, The examination and analysis of residuals, *Technometrics*, **5**, 141–60 (May 1963).

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)

(8) M. S. Bartlett, The use of transformations, *Biometrics*, **3**, 39–52 (March 1947).
(9) C. Eisenhart, The assumptions underlying the analysis of variance, *Biometrics*, **3**, 1–4 (March 1947).
(10) L. S. Nelson, Analysis of Residuals, *Nat. Convention Trans.*, Amer. Soc. Quality Control, 111–8 (1959).
(11) W. M. Wooding, The computation and use of residuals in the analysis of experimental data, *J. Qual. Technol.*, **1**, 175–88 (July 1969).
(12) G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 6th ed., Iowa State University Press, Ames, Iowa, 1967, p. 8–9.

Purchased for the exclusive use of nofirst nolast (unknown)
From: SCC Media Library & Resource Center (library.scconline.org)