J. Soc. Cosmet. Chem., 44, 13–21 (January/February 1993)

A comparison of antiperspirant data analysis methods

FRANK H. DIETRICH II, JAMES P. BOWMAN, BARBARA M. FATH, and JOHN E. WILD, *Hill Top Research, Inc., Cincinnati, OH 45147 (F.H.D., J.P.B., B.M.F., J.E.W.), and Northern Kentucky University, Highland Heights, KY 41076 (F.H.D.).*

Received January 22, 1992.

Synopsis

The purpose of an antiperspirant study may be to estimate the efficacy of a product, to compare two or more products, or to provide support for an advertising claim. In any of these cases, various statistical methods are available to analyze the data. In this paper, we will compare three methods of estimating the efficacy of a single antiperspirant product. Over the years Hill Top Research, Inc. has collected vast amounts of antiperspirant data, and this large base of historical information will be utilized for comparing the statistical methods of interest. This investigation suggests that one of the methods is preferable to the others.

INTRODUCTION

Evaluation of human perspiration evolved from simply viewing the amount of sweat produced during a set time period to collecting and weighing the actual sweat in the axillae. Collecting and weighing axillary perspiration through the use of absorbent pads has been used since the early 1950s and with minor modifications has remained fairly consistent for over 40 years. Analyses have also evolved over time.

The analysis of the early visual data estimated reduction in the amount of perspiration seen. One form of analysis of the data collected with absorbent pads is done using treated axilla over untreated axilla ratios, adjusting the posttreatment ratios by a pretreatment ratio (1). As early as 1964 the use of analysis of variance to test for differences between treatments was employed (2). Adaptation of the analysis of variance (i.e., analysis of covariance, etc.) and other methods have also been used to analyze antiperspirant data.

In this paper, we will consider experiments with the objective of estimating the percent reduction obtained following use of an antiperspirant test product. That is, we will consider experiments designed to estimate the average percent reduction in perspiration that could be achieved by consumers using an antiperspirant.

ANTIPERSPIRANT TERMINOLOGY

Terminology that is typically used to define the various stages of an antiperspirant study is as follows:

CONDITIONING

Each subject abstains from the use of axillary antiperspirants and uses only a deodorant product in the axillae for a period of 17 or more days (3) to effect washout of any previously used antiperspirant formulations.

BASELINE

Baseline sweat volumes and baseline axillary ratios are determined by sweat collections taken after the conditioning period, but before applying the antiperspirant test product.

TREATMENT

A series of supervised applications of the test formulation to the axillary areas is made during this period. The effect of treatment on axillary sweating is evaluated by sweat collections made at specific intervals following supervised applications.

All antiperspirant studies have conditioning and treatment stages; however, some do not utilize a baseline evaluation.

Sweating of the test panelists is induced by having the panelists sit in a room maintained at 100°F ± 2°F and at a relative humidity of 35% ± 5%. During the first 40 minutes of the sweat stimulation period, the panelists hold unweighed pads of Webril (nonwoven cotton padding fabric) in their axillae. This preliminary warm-up period is followed by two successive 20-minute collection periods, during which the panelists hold weighed Webril pads in the axillae. These pads are placed in tightly capped polystyrene vials and weighed before and after use.

STATISTICAL METHODS

For each of the following methods, the source data for the statistical analysis consist of milligrams of sweat collected from right and left axillae of panelists who have participated in antiperspirant studies. Treatments consist of one half of the panelists receiving a placebo, Sample X, on the right and the antiperspirant of interest, Sample Y, on the left, and one half of the panelists receiving the treatment in the opposite right-left order. The axilla that has been identified to receive the antiperspirant of interest, Sample Y, is designated as Axilla Y, and the other axilla as Axilla X.

ADJUSTED RATIO METHOD

For this method of analysis, pretreatment Y-to-X ratios are determined for each subject using baseline sweat collections. This ratio is calculated for each subject by

$$\text{Pretreatment ratio} = \frac{\text{Baseline milligrams of sweat from Axilla Y}}{\text{Baseline milligrams of sweat from Axilla X}}.$$

Posttreatment ratios are similarly calculated by

$$\text{Posttreatment ratio} = \frac{\text{Milligrams of sweat after treatment from Axilla Y}}{\text{Milligrams of sweat after treatment from Axilla X}}.$$

The data that is actually analyzed is called the adjusted treatment ratio, and is calculated by

$$\text{Adjusted treatment ratio} = \frac{\text{Posttreatment ratio}}{\text{Pretreatment ratio}}.$$

The mean of the adjusted treatment ratios is employed to find a point estimate of the mean percent reduction in sweating achieved by consumers. This value is given by

$$\text{Estimate of mean percent reduction in sweating} = (1 - \text{mean of adjusted ratio}) \times 100.$$

To obtain an interval estimate of the mean percent reduction in sweating, first the adjusted treatment ratios are used to calculate a small sample (Student's t) confidence interval. These values are then subtracted from one and multiplied by 100 (4).

WOODING-FINKELSTEIN METHOD

For this analysis, no baseline measurements are used. For each subject, the posttreatment milligrams of sweat for Axilla Y and Axilla X are transformed by calculating the natural logarithm of each. The means of the transformed data for Axilla Y and Axilla X are calculated and denoted by $\overline{Y}_{log}$ and $\overline{X}_{log}$, respectively. A point estimate of the percent reduction is calculated using the antilogs of these values as follows:

$$\text{Estimate of mean percent reduction} = [1 - \text{Antilog}\,(\overline{Y}_{log})/\text{Antilog}\,(\overline{X}_{log})] \times 100.$$

To find an interval estimate of the mean percent reduction, two steps are required. First, following an analysis of variance using the transformed data, a confidence interval is calculated using a small sample (Student's t) method. The endpoints of this interval are then exponentiated to transform them back to the percent reduction scale (5).

DIRECT METHOD

Again, no baseline measurements are used in this analysis. Posttreatment individual percent reductions are determined for each subject by employing the posttreatment ratios (defined for the ratio method). This is accomplished by

$$\text{Individual percent reductions} = (1 - \text{posttreatment ratio}) \times 100 =$$

$$\frac{(\text{posttreatment Axilla X milligrams}) - (\text{posttreatment Axilla Y milligrams})}{(\text{posttreatment Axilla X milligrams})} \times 100.$$

The mean of the individual percent reductions is used as a point estimate of the mean percent reduction in sweating for all consumers. To obtain an interval estimate of the mean percent reduction, a confidence interval is calculated from the individual percent

Table I
Estimates of Mean Percent Reductions

Study	No. of panelists	Method		
		ARM	WFM	DM
1	29	−1.16	3.21	−3.67
2	30	9.65	14.59	8.81
3	30	19.49	20.60	18.04
4	30	19.98	19.54	15.95
5	10	22.72	21.72	19.16
6	30	22.77	21.77	18.34
7	52	27.68	29.65	26.64
8	29	32.96	35.51	32.24
9	30	41.96	48.70	44.30
10	30	42.05	42.93	38.92
11	30	43.56	45.72	42.65
12	33	45.28	47.13	43.99
13	32	50.30	55.42	53.31
14	15	56.55	58.10	54.17
15	15	59.92	65.96	62.49

reductions[1]. The calculation of this interval is obtained by a method used in deodorant efficacy studies (6).

RESULTS

For any particular antiperspirant study of interest, the adjusted ratio method (ARM), the Wooding-Finkelstein method (WFM), and the direct method (DM) will generally produce slightly different point estimates of the percent reduction in sweating for that study. To demonstrate how much different the estimates typically are for the three methods, we applied the methods to fifteen recent antiperspirant studies conducted at Hill Top Research[2]. The posttreatment data analyzed were the one-hour collection taken after the third application[3]. In Table I we present the three point estimates of percent reduction for each of the fifteen antiperspirant studies.

If you examine the results of Study 8, which included 29 subjects, you will see that the WFM produced the largest estimate of percent reduction, 35.51%. The ARM is next at 32.96%, and the DM is smallest at 32.24%. Although this exact pattern does not exist for every study, the overall trend is similar. In fact, in all 15 studies the percent reduction produced by the WFM is larger than the corresponding estimate produced by

[1] Depending on the number of subjects sampled, either a Student's t or a large sample Z interval might be found. To assure validity, we would recommend sampling over thirty subjects and using the large sample procedure.

[2] When selecting the fifteen antiperspirant studies to be analyzed, we made sure they covered a wide range of efficacies. This was the only criterion used to select the studies, and no studies were eliminated because of lack of support for our conclusions.

[3] This collection is accepted as one that is appropriate to use when estimating the efficacy of an antiperspirant.

the DM. The average of the estimates of percent reduction using the WFM are 3.68 larger than the average of the estimates using the DM. In 12 of the 15 studies, the percent reduction produced by the ARM is larger than the corresponding estimate produced by the DM. The ARM produced estimates that averaged 1.22 more than those produced by the DM. There are theoretical reasons that explain the differences in the estimates shown by these empirical results.

A basic statistical result is that the arithmetic mean of a sample is an unbiased estimator[4]. Since the DM uses the mean of the sample of percent reductions in sweating as an estimator, this method will provide an unbiased estimator, one that tends (on the average) to be neither larger nor smaller than the true efficacy of the antiperspirant being tested. It can be mathematically shown that for any given antiperspirant study, the estimated percent reduction produced by the WFM will always be larger than the corresponding estimate produced by the DM[5]. Thus the WFM will tend to overestimate the efficacy of an antiperspirant.

The estimate produced by the ARM uses the average of ratios. It is for this reason that in Table I the estimates for percent reduction for the ARM tend to be slightly larger than those produced by the DM[6].

When comparing estimators, an unbiased estimator is generally preferred to one that is biased. For this reason, the DM is preferable to both the ARM and the WFM. A second point that is often considered when comparing methods of estimation is the variability associated with the estimators[7]. This can be thought of as a measure of precision, and may be assessed by examining the widths of the confidence intervals estimates.

In Table II we present 95% confidence intervals calculated by the WFM, ARM, and DM for the fifteen antiperspirant studies introduced in Table I.

The average width of the ARM confidence intervals is 2.52 less than the average width of the DM confidence intervals. Thus it appears that the ARM estimates are less variable than the DM estimates, and for the same sample sizes, this tends to be true. However, baseline measurements must be collected to obtain the ARM estimates, while this is not necessary for the DM to be used. If baseline measurements were not collected, these resources could be used to collect more posttreatment measurements. Thus, for similar expenditures, the ARM and DM would produce confidence intervals of comparable widths.

The average width of the WFM intervals is 0.22 less than the average width of the DM intervals. These empirical results suggest there is little difference in the variability of the two estimators.

Finally, for any statistical procedure to be used, it should be statistically valid. By the validity of the procedure we mean certain conditions (assumptions) must be met for any

[4] In this context, an unbiased estimator will *on the average* be equal to the value it is estimating.

[5] This is due to the relationship between the geometric mean (calculated for the WFM) and the arithmetic mean (calculated for the DM).

[6] When the sample average ratio of two variables is used to estimate the ratio of the true averages of the two variables, the estimate is generally not unbiased.

[7] A comparison of the variability of different estimators is of most interest when comparing different unbiased estimators.

Table II
95% Confidence Intervals for Mean Percent Reductions

	Method					
	ARM		WFM		DM	
Study	Interval	Width	Interval	Width	Interval	Width
1	−12.78, 10.46	23.24	−12.96, 17.06	30.02	−17.55, 10.21	27.76
2	−0.35, 19.65	20.00	1.94, 25.60	23.66	−3.04, 20.66	23.70
3	12.48, 26.50	14.02	12.29, 28.13	15.84	10.39, 25.69	15.30
4	13.62, 26.34	12.72	10.38, 27.76	17.38	6.57, 25.33	18.76
5	10.80, 34.64	23.84	5.05, 35.46	30.41	4.81, 33.51	28.70
6	15.11, 30.43	15.32	13.37, 29.36	15.99	10.63, 26.05	15.42
7	20.87, 34.49	13.62	23.58, 35.24	11.66	20.54, 32.74	12.20
8	25.58, 40.34	14.76	27.85, 42.35	14.50	24.60, 39.88	15.28
9	33.94, 49.98	16.04	40.32, 55.90	15.58	35.38, 53.22	17.84
10	35.72, 48.38	12.66	34.43, 50.32	15.89	30.29, 47.55	17.26
11	38.11, 49.01	10.90	37.83, 52.61	14.78	36.13, 49.17	13.04
12	39.83, 50.73	10.90	40.19, 53.27	13.08	37.24, 50.74	13.50
13	44.19, 56.41	12.22	50.47, 59.89	9.42	48.48, 58.14	9.66
14	49.17, 63.93	14.76	45.97, 67.51	21.54	44.04, 64.30	20.26
15	50.96, 68.88	17.92	56.04, 73.65	17.61	51.47, 73.51	22.04

statistical method to "work properly." These conditions involve how the data are collected, the type of data, the distribution of the data, etc. If the appropriate conditions are not met, a statistical method may yield unreliable results.

It has been noted that the distribution of adjusted ratios should have a normal distribution for the ARM to be valid. An examination of the distribution of approximately 5000 adjusted ratios calculated from vast amounts of historical data (Figure 1) shows that this distribution is definitely not symmetric. Thus the distribution is not normal. This means that for small samples the ARM is not appropriate. However, when more than thirty subjects are used in antiperspirant studies, the t-statistic is approximately valid and the mentioned criticism is no longer a worry.

The WFM performs an analysis of variance to analyze antiperspirant studies, and the data must again be normally distributed. Since the milligrams of sweat collected are not normally distributed (Figure 2), the WFM analyzes log-transformed data. While the transformed data are more nearly normally distributed than the original milligrams of sweat collected, the transformed data are not exactly normal; it is somewhat nonsymmetric (Figure 3). Thus the WFM is approximately theoretically valid.

The DM analyzes the collection of individual percent reductions in sweating for each subject. The distribution of percent reductions is not normal (Figure 4). Due to the nonsymmetry of this distribution, we recommend that over thirty panelists be used to provide percent reductions; in this manner the DM would be approximately theoretically valid.

DISCUSSION

The adjusted ratio method and direct method agree quite well, with the adjusted ratio

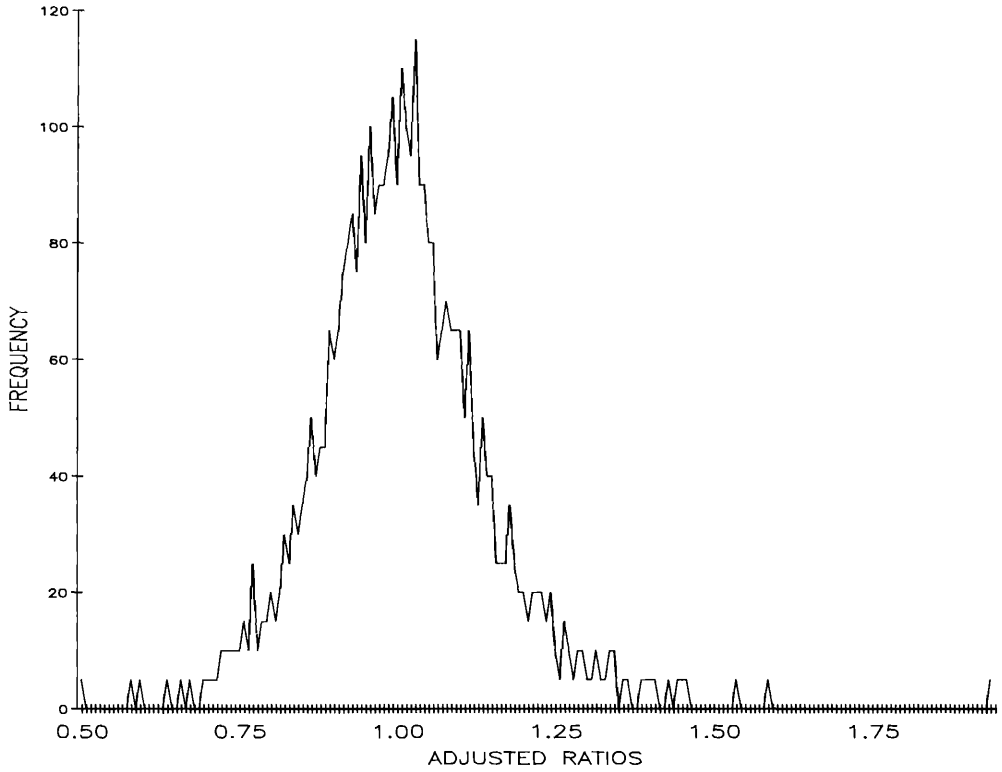

Figure 1. Graph of approximately 5000 adjusted ratios.

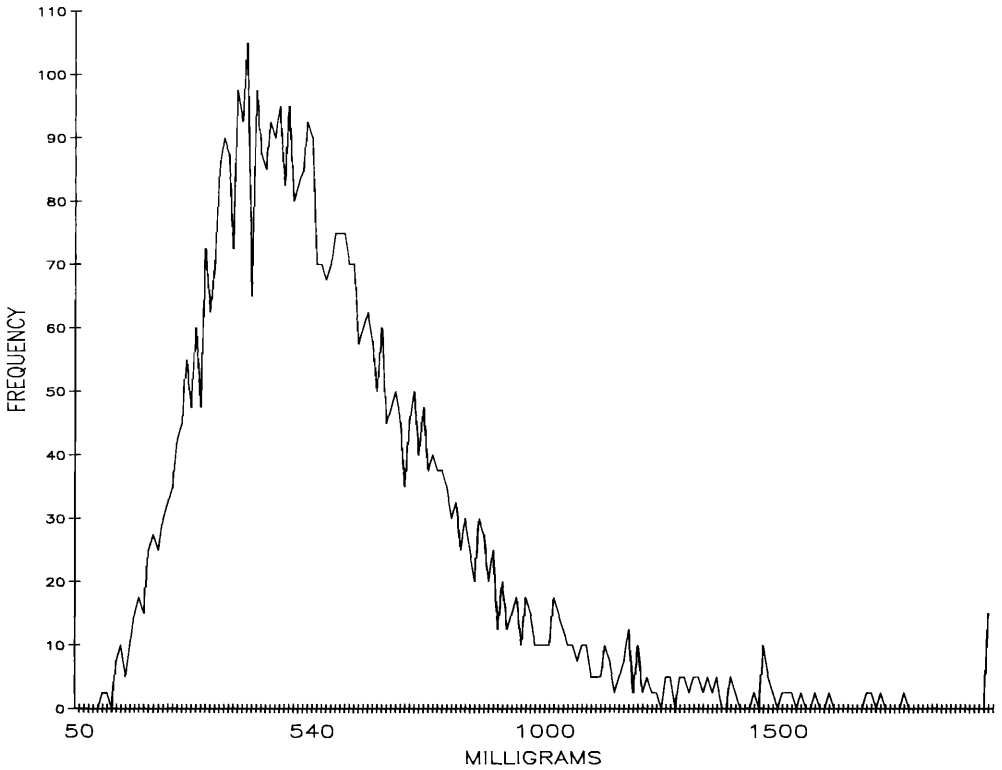

Figure 2. Graph of approximately 5000 milligram values.

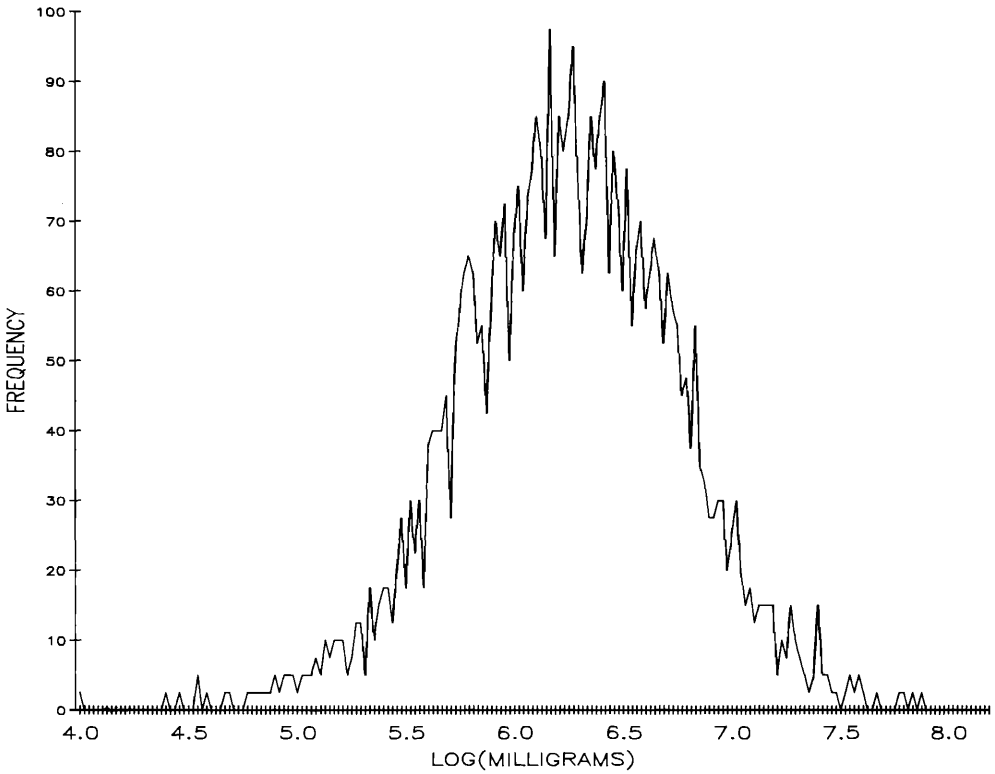

Figure 3. Graph of approximately 5000 natural logarithms of milligram data.

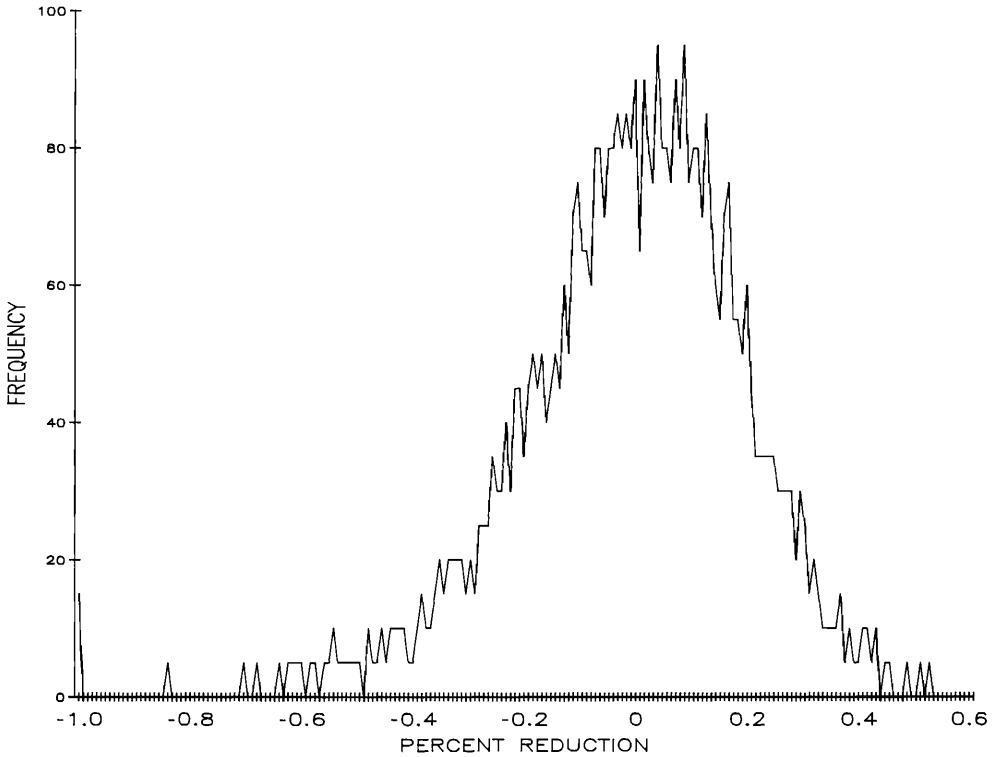

Figure 4. Graph of approximately 5000 percent reductions.

method tending to have slightly higher estimates than the direct method. The Wooding-Finkelstein method always produces higher estimates of percent reduction than the direct method and almost always produces higher estimates than the adjusted ratio method. These statements are supported by empirical results as well as theoretical considerations.

Since the direct method produces statistically unbiased estimates, while the other two methods do not, the direct method is preferred. Also, there is little in the way of increased precision (width of confidence intervals) to recommend one of the three methods over the other two.

Other methods of analysis such as analysis of covariance, analysis of log-transformed adjusted ratios, and non-parametric methods have also been used by experimenters to analyze antiperspirant data. Since the first two of these methods use transformed data, they will provide biased estimators, as the Wooding-Finkelstein method does. Although a non-parametric technique might be valid, it is well known that non-parametric methods are less powerful than parametric methods. Thus the direct method would be preferred.

Finally, we want to emphasize that the results of this paper are in reference to experiments in which the objective is to estimate the percent reduction of an antiperspirant. For studies with other objectives, such as testing which of two or more antiperspirants has the greater (or greatest) efficacy, it is as yet to be determined what, if any, statistical analysis is most appropriate.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Cindy Yablok and Sally Burkhard for their assistance in the preparation of this manuscript.

REFERENCES

(1) W. G. Fredall and R. R. Read, Antiperspirant-axillary method of determining effectiveness, *Proc. Sci. Sect., Toilet Goods Assoc.*, **15**, 23–27 (1951).
(2) W. M. Wooding *et al.*, Statistical evaluation of quantitative antiperspirant data. I., *J. Soc. Cosmet. Chem.*, **15**, 579–592 (1964).
(3) F. B. Carabello, Guidelines for the clinical study of antiperspirant and deodorant efficacy, *Cosmet. Toiletr.*, **95**, 33 (1980).
(4) P. A. Majors and John E. Wild, The evaluation of antiperspirant efficacy—Influence of certain variables, *J. Soc. Cosmet. Chem.*, **25**, 139–152 (1974).
(5) W. M. Wooding and P. Finklestein, A critical comparison of two procedures for antiperspirant evaluation, *J. Soc. Cosmet. Chem.*, **26**, 255–275 (1974).
(6) ASTM Committee E-18, "Standard Practice for the Sensory Evaluation of Axillary Deodorancy," in *Annual Book of ASTM Standards*, Vol. 15.07 (1988).