Journal of Counseling Psychology
1987, Vol. 34, No. 4, 383-392

0022-0167/87/$00.75

Analysis of Covariance: Its Model and Use in Psychological Research

Andrew C. Porter and Stephen W. Raudenbush
College of Education
Michigan State University

Since its development 50 years ago, the analysis of covariance (ANCOVA) has become a standard tool for data analysis in psychological research. Nevertheless, it is common for researchers to underestimate both the benefits of the technique and its potential for misuse. In this article, we consider the two major ways in which psychologists have used the technique: for increasing the precision of estimation in randomized experiments and for seeking to remove bias in nonrandomized studies. In each case, the ANCOVA is compared with analytic alternatives. The argument emphasizes the benefits of using the technique in randomized experiments and warns of the dangers of using it in nonrandomized studies.

The analysis of covariance (ANCOVA) is a statistical procedure developed and popularized by Sir Ronald Fisher (1948) over 50 years ago. In his words, it "combines the advantages and reconciles the requirements of the two very widely applicable procedures known as regression and analysis of variance" (Fisher, 1948, p. 281). Despite a long and distinguished history of use in data analysis and particular prominence in the social science research literature, ANCOVA remains an often misunderstood and misused technique.

Fisher (1948) originally developed ANCOVA as a method for reducing error variance in randomized experiments, thereby increasing both the statistical power of hypothesis tests and the precision in estimating effects. However, social scientists have at least as frequently used the method to provide statistical control in nonrandomized (quasi) experiments. Perhaps the fact that ANCOVA offers these two distinctly different potential benefits—increasing precision in randomized experiments and reducing bias in nonrandomized studies—has contributed to confusion over its use. In our view, social scientists have underutilized the procedure for the purpose of improving the precision of estimates and overutilized it for the purpose of attempting to reduce bias in nonrandomized studies.

In the following discussion, we give the statistical model for ANCOVA and then consider each of its major uses. Throughout, we make reference to potential pitfalls often found in the empirically based literature in which ANCOVA has been used. Generally, we argue for a more balanced consideration of what the procedure has to offer to psychological research. The benefits of adjustment in nonrandomized experiments are fewer than are suggested in the literature, but the need for better precision of estimates is greater.

The Model

The ANCOVA model includes three types of variables. Independent variables represent the contrasts that motivated the study, typically including a treatment–control contrast (or alternative types of treatment contrasts) and other independent variables to test for possible interactions with treatment conditions (e.g., the effect of treatment might depend on age or some other characteristic of the subject). The dependent variable represents a hypothesized outcome of treatment. The ANCOVA model allows for only one dependent variable but can be generalized to a multivariate model. Covariables are included in the model so that their relation to the dependent variable can be statistically controlled. They represent nuisance variables not directly under investigation. In ANCOVA, hypotheses are about relations between independent variables and the dependent variable while the covariables are held constant.

Like analysis of variance (ANOVA), ANCOVA extends to a wide variety of models implied by factorial designs (including crossed and nested factors). The primary conceptual issues and principal applications, however, can be understood within the context of a simple, single classification model. Specifically, we consider a study that compares treatment and control groups and in which pre- and posttreatment measures are available for each subject. The formal model, then, for a simple analysis of covariance is

$$Y_{ij} = \mu' + \alpha'_j + \beta_{Y \cdot X}(X_{ij} - \bar{X}_{..}) + e'_{ij},$$

where

Y_{ij} represents the dependent variable (e.g., a posttest) for the ith subject in the jth group;

X_{ij} represents a covariable (e.g., a pretest);

μ' is a grand mean, across all subjects and groups;

α'_j is called a *treatment effect* and is the distance of the mean of treatment group j from the grand mean;

$\beta_{Y \cdot X}$ is a parameter representing the slope of the regression line for Y on X; and

e'_{ij} is the final term in any linear model.

What remain hidden in this model are the definitions of μ' and α'_j and e'_{ij}. By rewriting the model as

$$Y_{ij} - \beta_{Y \cdot X}(X_{ij} - \bar{X}_{..}) = \mu' + \alpha'_j + e'_{ij},$$

Correspondence concerning this article should be addressed to Andrew C. Porter, Institute for Research for Teaching, College of Education, Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

it becomes clear that μ', α'_j, and e'_{ij} cannot be the same as when similar notation is used to describe parallel terms in a one-way ANOVA. We have used primes to make this distinction.

Because interest invariably focuses on α'_j, the treatment effect, we concentrate there:

$$\alpha'_j = \alpha_{Y_j} - \beta_{Y.X}(\bar{X}_j - \bar{X}_.),$$

or in words, an ANCOVA treatment effect is an ANOVA treatment effect adjusted by the product of two terms. The first term, $\beta_{Y.X}$, is the slope of the regression line of Y, the dependent variable on X, the covariable (sometimes called the *concomitant variable*). The second term, $\bar{X}_j - \bar{X}_.$, is the distance of the jth group's average on the covariable from the grand mean of the covariable. With one exception, this second term is identical to an ANOVA treatment effect in which the covariable is the dependent variable. The exception is an important one, however, and worth a brief digression.

The ANCOVA was developed with X, the covariable, fixed, just as is the case for any regression model (i.e., values of X are specified a priori, and subjects who have those values of X are then selected). Probably because social science research rarely if ever has designs in which covariables are fixed (or, for that matter, regression equations in which predictors are fixed), this distinction between fixed and random has been glossed over. The term $(X_{ij} - \bar{X}_.)$, indicating that results are conditioned on fixed values of X, is used interchangeably with the term $(X_{ij} - \mu_{X.})$, which suggests that X is a random variable with expectation $\mu_{X.}$ across replications. Fortunately, at least for applications on randomized experiments, ANCOVA estimation procedures and hypothesis tests are as valid for random covariables as they are for fixed covariables (e.g., DeGracie & Fuller, 1972; Porter, 1967), with the only loss being some precision in estimation (or statistical power in hypothesis testing).

Conceptually, then, the ANCOVA treatment effect can be represented as

$$\alpha'_j = \alpha_{Y_j} - \beta_{Y.X}\alpha_{X_j},$$

or in words, an ANCOVA effect is an ANOVA effect adjusted by subtracting the covariable effect, α_{X_j}, weighted by $\beta_{Y.X}$, the regression slope of Y on X. Definitions for other terms in the ANCOVA model take a parallel form.

Model Assumptions

To complete discussion of the model, we must consider the assumptions on which the estimation and hypothesis-testing procedures rest. First, standard ANOVA assumptions are made on the model's residuals; the e'_{ij}s are distributed normally, with equal variance, and are independent of each other. Because the e'_{ij}s are residuals from regression lines defined separately for each treatment condition (i.e., separate regression lines for each of the j groups), these assumptions are really made on conditional distributions of Y given X. Second, as implied by the model's having a single regression parameter, $\beta_{Y.X}$, it is assumed that when regressing Y on X separately for each treatment group, the regression lines all have the same slope (i.e., they are either collinear or parallel). Third, it is assumed that the relation between X and Y is linear. Sometimes a fourth assumption is specified, that the covariable X be free from errors of measurement (i.e., that X be observed with perfect reliability).

The consequences of violating ANCOVA assumptions have been thoroughly investigated (e.g., Glass, Peckham, & Sanders, 1972). Generally, the procedure, like the ANOVA, is robust with respect to violations of the assumptions of normality and equal variance (especially for balanced, equal-sample-size designs). Neither procedure is robust with respect to violations of independence (e.g., design constraints including grouping variables must be included in the statistical model).

The assumptions of equal slopes across treatment groups and a linear relation between X and Y can also be violated without invalidating the statistical estimation and hypothesis-testing procedures, but for these assumptions there is a more important problem than accuracy in hypothesis testing. if treatment-group regression lines are not parallel, there is an interaction between the covariable X and the independent variable defined by treatment conditions. This means that the effect of the treatment depends on each subject's value on the covariable. This can be seen in Figure 1, where, depending on the value of X selected, the distance between the two treatment conditions changes (i.e., larger effects favoring TX_1 for low values of X over TX_1 for high values of X). Thus, if the equal-slopes assumption is violated, the researcher needs to describe and interpret the interaction effects in the data. Main effects for treatment are then of little interest.

The assumption about the reliability of X is addressed in some detail later. Briefly, the assumption is not of concern when randomized experiments are analyzed but is one of many difficulties in analyzing nonrandomized studies.

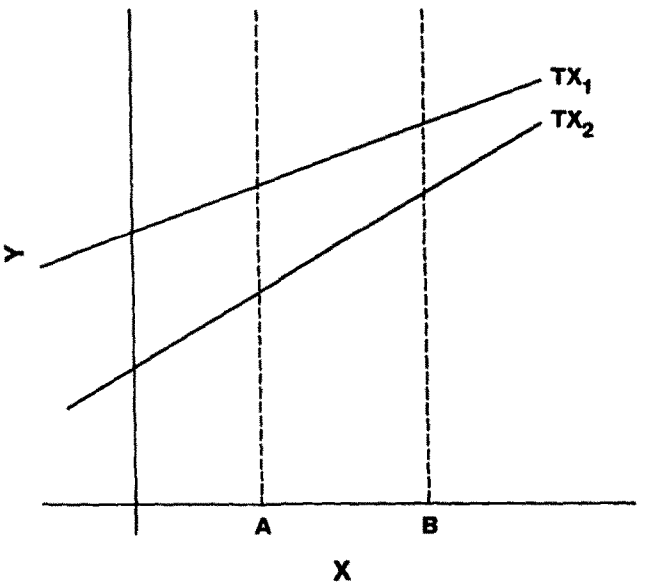

Figure 1. Nonparallel regression lines with two treatment groups TX_1 and TX_2. Notice that the magnitude of the treatment effect diminishes as X increases from A to B.

Choosing a Covariable

To improve precision, a covariable must be linearly related to the outcome. Often a pretest of the dependent variable is a good choice for a covariable, but the procedure is appropriate for use with any quantitative variable that meets the assumptions of the model. When two covariables are used, both should be strongly related to the outcome and weakly related to each other. It is crucial that the covariable be unaffected by the treatment. Thus only variables that are measured prior to treatment implementation or variables that are unarguably impervious to change by the treatment should be considered as covariables. (More will be said on this later in the discussion on adjustments.) Finally, there should not be an interaction between the covariables and the experimental conditions under investigation. The model requires common regression across treatments.

ANCOVA and Statistical Power

Lack of statistical power plagues social science research, and counseling psychology research is no exception to this rule. Put simply, poor statistical power means that a study that aims to discover the effectiveness of a new counseling approach will fail to do so with high probability—even when this new approach works.

Consider a study in which 30 patients who received a new therapy were compared with 30 other patients who received a control therapy. All 60 patients were assessed on a relevant criterion, and an ANOVA was computed to test the significance of the mean difference between treatments at the .05 level of significance. (The ANOVA yielded the same results as a t test in the two-group case.) Even if the true effect of the treatment was large, e.g., as large as one half of the standard deviation, the probability of correctly rejecting the no-treatment-effect hypothesis was less than .50. Poor statistical power can result in a highly effective treatment's being dismissed as a failure.

Statistical power depends on several factors: the size of the sample, the degree of variability in the dependent variable, the choice of research design and method of statistical analysis, the significance level chosen by the researcher, and the magnitude of the treatment effect. The researcher has little control over some of these factors. Choice of significance level is constrained by tradition; most psychological journals do not report significance levels greater than .05. The size of experimental effect is what is being investigated. Of course, large effects are desired, and interventions are designed with that in mind, but the discovery of the experimental effect is a goal of a study, not a part of a study's design. Similarly, once populations have been selected and dependent variables chosen, the degree of variability in the dependent variable is no longer under control. Thus, the researcher is primarily limited to controlling statistical power through sample size and choice of design and analysis procedures.

Sample sizes are limited by budget and time constraints; unavailability of subjects or trained therapists also conspires to limit sample sizes. However, choice of experimental design, of statistical analysis, or of both can increase statistical power, sometimes dramatically, without much extra expense. ANCOVA is often the most effective means of doing so.

How ANCOVA works. Once the significance level, sample size, and magnitude of treatment effect are fixed, the power of a statistical test depends on the precision with which treatment effects are estimated. The larger the sampling variance of estimated treatment effects, the weaker is statistical power. Technically, the precision of an experimental comparison is the inverse of the sampling variance of an estimated effect. For example, the sampling variance of a comparison between two group means, where both groups have equal sample size, n, is given by

$$2\sigma^2/n,$$

where σ^2 is the variance of the dependent variable after the treatment effect (i.e., the residual or error variation) is accounted for. Clearly, this sampling variance is large when n is small or σ^2 is large; thus, the two strategies for increasing statistical power mentioned above are to increase n, which costs money, and to reduce σ^2, which often costs little or nothing.

ANCOVA can reduce σ^2, thereby improving precision and statistical power. Figure 2 illustrates this idea conceptually. Without ANCOVA, ANOVA partitions the total variation on the dependent variable into two pieces: variation explained by the treatment variable (a typically small fraction) and residual, or unexplained, variation (σ^2). With ANCOVA, the total variation is partitioned into three pieces: variation explained by the treatment variable, variation explained by the covariable, and residual variation. If the covariable has a strong correlation with the outcome, the residual variation will be small, and statistical power will be substantially improved.

Again, consider a two-group study for which there are n subjects per group and the goal is to estimate the mean difference between the two groups. Now let the symbol ρ_{XY} denote the correlation between the covariable and the outcome. The sampling variance of an ANCOVA adjusted mean difference is

$$2\sigma^2(1 - \rho_{XY}^2)/n + k,$$

with

$$k = \sigma^2(1 - \rho_{XY}^2)(\bar{X}_{.1} - \bar{X}_{.2})^2/\Sigma\Sigma\,(X_{ij} - \bar{X}_{..})^2,$$

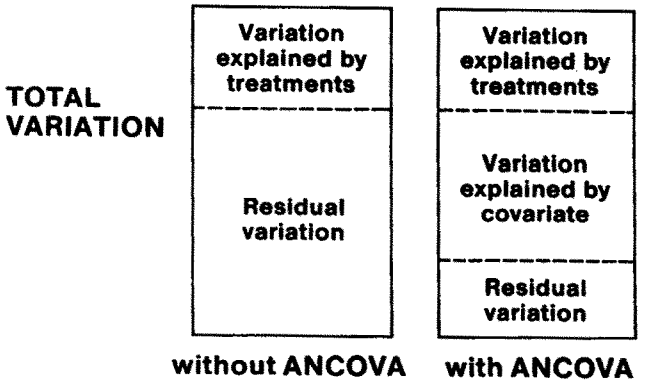

Figure 2. Partitioning of variation of the dependent variable with and without analysis of covariance.

where X_{ij} denotes the covariable and $\bar{X}_{.1}$ and $\bar{X}_{.2}$ denote group means on the covariable.[1] In randomized experiments, $\bar{X}_{.1} - \bar{X}_{.2}$ will be small, so that k will be near zero.

A comparison of the sampling variance for ANOVA effects with the sampling variance for ANCOVA effects makes the role of the correlation between X and Y clear. As ρ_{XY} increases, the sampling variance of ANCOVA decreases (and statistical power increases). For example, using ANCOVA rather than ANOVA when $\rho_{XY} = .7$ essentially cuts the sampling variance in half for randomized experiments.

To illustrate this idea, we analyze a simple, hypothetical data set by using ANOVA and ANCOVA. The data set, provided in Table 1, has two groups, each with 10 cases, and includes measurements on a premeasure, which is the covariable (X) and a postmeasure, which is the dependent variable (Y). For simplicity, group variances are the same; the correlation between X and Y is .80 in both groups, and pretest means X_1 and X_2 are identical. This last condition is not farfetched, because group means on a premeasure are likely to be similar in a randomized study.

As shown in Table 2, when the data are analyzed by means of ANOVA, the treatment effect is nonsignificant, $F(1, 18) = 2.25$. However, ANCOVA yields a significant treatment effect, $F(1, 17) = 5.90$, $p < .05$. In both cases, the treatment effect estimate is the same: Experimental subjects gained 1 point more on average than controls. Each mean square treatment is 5. The difference between the two analyses arises from the smaller error variation for ANCOVA. Whereas 89% of the total variation was left unexplained by ANOVA, only 32% was unexplained by ANCOVA.

Table 2 also shows a significant effect of the covariate, $F(1, 17) = 30.22$, $p < .001$. This effect is hardly surprising, because the experimenter chooses the covariate because of its expected strong relation with the outcome. Therefore, the test of the effect of the covariate is commonly omitted from research reports.

Although use of ANCOVA to improve statistical power is not as prominent in the research literature or counseling psychology as might be desirable, there are illustrations of its use. Some recent examples worth considering are provided by Jacobson (1984), Kazdin and Mascitelli (1984), and Piper, Debbane, and Bienvenu (1984).

Table 1
Hypothetical Data

	Group 1: Experimental		Group 2: Control	
Case	X	Y	X	Y
1	8	9	6	7
2	10	10	8	8
3	6	8	10	9
4	9	11	7	6
5	7	7	9	10
6	7	7	8	8
7	10	10	10	9
8	6	8	6	7
9	9	11	9	10
10	8	9	7	6
M	8	9	8	8
SD	1.49	1.49	1.49	1.49

Table 2
ANOVA and ANCOVA Results for Hypothetical Data

Results	*df*	*SS*	*MS*	*F*
ANOVA				
Treatment group	1	5	5.60	2.25
Within groups (error)	18	40	2.22	
Total	9	45		
ANCOVA				
Covariate	1	25.6	25.60	30.22**
Treatment	1	5	5.00	5.90*
Error	17	14.4	.85	
Total	19	45		

Proportion of variation explained

Results	*SS*	Proportion of total
ANOVA		
Treatment groups	5	.11
Error	40	.89
Total	45	1.00
ANCOVA		
Covariate	25.6	.57
Treatment groups	5	.11
Error	14.4	.32
Total	45	1.00

Note. ANOVA = analysis of variance; ANCOVA = analysis of covariance; *SS* = sum of squares; *MS* = mean square.
* $p < .05$. ** $p < .001$.

How much does analysis of covariance help? Earlier we noted that an investigator using the ANOVA to test the significance of a difference between a treatment group and a control group with 30 subjects each would fail to find a significant difference at the .05 level more than half the time—even though the treatment produced an average gain of half a standard deviation (a large effect). To achieve power of .80 by using ANOVA, an investigator would need 84 subjects per group. Now suppose that a covariable that correlated .8 with the dependent variable was used in ANCOVA. Then, only 33 subjects per group would be needed to achieve a power of .80 at the .05 significance level. In this case, the use of a covariable is roughly equivalent to adding 50 subjects to each group in terms of added power. If the correlation between covariable and outcome were .60, 57 subjects per group would be needed to achieve a power of .80 at the same significance level, equivalent to adding 25 subjects per group in terms of the added power. The benefit of the covariable diminishes sharply as its correlation with the outcome falls below about .40 (e.g., Cox, 1957).

Comparison With Other Strategies for Increasing Power

ANCOVA is just one strategy for using ancillary information about the subjects of an experiment to improve precision and, therefore, power. Other common strategies include the use of

[1] This expression is for a fixed covariable, which is the classical model most covered in the literature. Although the expression is not exactly correct for a random covariable, the expression is simply the expectation over all samples. Fortunately, there is no difference in the correct computational procedures between a fixed and a random covariable.

gain scores and blocking variables. How does ANCOVA compare with these strategies for increasing statistical power?

Gain scores. A common strategy for increasing power is to administer a premeasure and to compute posttest–pretest gain score or difference score for each subject. These gain scores are then analyzed by means of ANOVA. For illustrations in the counseling psychology literature, see Hobbs, Walle, and Caldwell (1984), Sanchez-Craig, Annis, Bonet, and McDonald (1984), and Scovern et al. (1980).

With gain scores, the mean difference between two groups is then

$$\mu_{Y_1} - \mu_{Y_2} - (\mu_{X_1} - \mu_{X_2}),$$

where μ_{X_1} and μ_{X_2} are pretest means and μ_{Y_1} and μ_{Y_2} are posttest means for Groups 1 and 2, respectively. Because the comparable adjusted mean difference for ANCOVA in the two-group case is

$$\mu_{Y_1} - \mu_{Y_2} - \beta_{Y.X}(\mu_{X_1} - \mu_{X_2}),$$

where $\beta_{Y.X}$ is the pre–post regression coefficient (common within groups), it is clear that gain-score analysis estimates the same treatment effects as does ANCOVA when $\beta_{Y.X} = 1$. Actually, in randomized experiments, these two procedures estimate the same treatment effect even when $\beta_{Y.X} \neq 1$ because $(\mu_{X_1} - \mu_{X_2}) = 0$ given random assignment.

Gain scores are one specific case of a more general analysis strategy called *index of response.* Let us define Z_{ij} as an index of response:

$$Z_{ij} = Y_{ij} - kX_{ij},$$

where Y and X are as before and k is any constant. With this index of response as the dependent variable, the population mean difference between two treatment groups becomes

$$\mu_{Z_1} - \mu_{Z_2} = \mu_{Y_1} - \mu_{Y_2} - k(\mu_{X_1} - \mu_{X_2}).$$

Clearly, when $k = 1$, the index of response is also a gain score.

The following question arises: What choice of k will minimize the sampling variance of mean differences, thus maximizing statistical power? It can be shown that when $k = \beta_{Y.X}$, the sampling variance of the above mean difference for Z is minimized, thus maximizing precision and statistical power (Cox, 1957). This minimum sampling variance, or lower bound, is

$$\sigma_Y^2(1 - \rho_{XY}^2).$$

One might suggest, therefore, using

$$Z_{ij} = Y_{ij} - \beta_{Y.X}X_{ij}$$

as the index of response. The problem is that $\beta_{Y.X}$, the population slope for predicting Y from X, is not known. In using ANCOVA, the investigator estimates $\beta_{Y.X}$ from the data. The penalty is that the estimates will vary from sample to sample, so that estimating $\beta_{Y.X}$ increases the sampling variance above the lower bound. Also, estimating $\beta_{Y.X}$ costs a degree of freedom. Nevertheless, the error variance for ANCOVA is typically small, compared with that obtained by using gain scores (i.e., setting $k = 1.0$). Because in psychological research, $\beta_{Y.X}$ will typically be quite different from 1.0, the use of gain scores seldom yields acceptable precision. Further, forming an index of response by using a sample estimate of the slope (i.e., $k = \hat{\beta}_{Y.X}$) is not a correct procedure and yields standard errors of mean difference that are too small (and inflated Type I error rates for hypothesis tests).

Another advantage of ANCOVA over gain-score analysis is its flexibility. Because administration of premeasures is often impossible or undesirable, gain-score analysis may be impossible. However, any prior information about the subjects may be used as a covariable, subject to certain restrictions addressed above. (See the Choosing a Covariable section.) Also, multiple covariates may be used. Generally, however, little precision is gained through adding additional covariables because when several covariables are strongly related to the outcome, they tend to also be related to each other. Hence, after the best covariate (the one most strongly related to the outcome) is chosen, additional covariates may contribute little fresh information about the outcome.

Blocking. Another alternative to ANCOVA when the improvement of precision is sought is the randomized blocks design. (In the counseling psychology literature, see Foy, Nunn, & Rychatarik, 1984.)

Briefly, a randomized blocks design involves the creation of groups of subjects believed to be similar on the dependent variable and then, for each group, the random assignment of subjects to treatment conditions. Repeated measures designs are a special case of randomized block designs; subjects constitute "blocks of observations" to be assigned in random order to treatments. Like ANCOVA, randomized blocks designs partition total variation on the dependent variable, removing between-blocks variability from the error variation. The procedure is effective, then, to the extent that a researcher can create blocks with large mean differences on the dependent variable (and small variance within each block).

When a quantitative variable that could also be used as a covariable in ANCOVA, is used to form blocks, comparisons between the two procedures in terms of precision can be made. A common rule of thumb is to prefer blocking when the quantitative variable has a low correlation with the dependent variable, $\rho_{XY} < .4$, and to prefer ANCOVA when the correlation is high, $\rho_{XY} > .8$ (Cox, 1957). Because psychological researchers typically have information on one or more quantitative variables that correlate substantially with the dependent variable (e.g., pretests of the dependent variable), ANCOVA is a favored approach. Still, blocking does not require quantitative information (as does ANCOVA), nor must the relation between information used to form blocks and the dependent variable be linear. In this sense, blocking is a somewhat more flexible approach than is ANCOVA. A limitation of blocking, however, is that the blocking variable must be chosen in advance and built into the design, whereas covariables may be chosen somewhat opportunistically after the experiment.

Adjustment

As was seen earlier, effects estimated in the ANCOVA model can be thought of as differences between unadjusted effects

on the dependent variable and a weighted unadjusted effect on the covariable (where the weight is the regression coefficient of Y or X),

$$\alpha_j' = \alpha_{Y_j} - \beta_{Y.X}\alpha_{X_j}.$$

When subjects are randomly assigned to groups, the long-run expectation is that each treatment group starts out equal to each other treatment group on any dimension that might be measured. Therefore, the α_{X_j}s, the unadjusted effects on the covariable, are zero. As a result, when subjects are randomly assigned to groups,

$$\alpha_j' = \alpha_{Y_j}.$$

In words, ANCOVA adjusted effects are equal to ANOVA unadjusted effects.

Thus, from the point of view of interpreting ANCOVA treatment effects, in randomized experiments the interpretation remains the same as for ANOVA. In this regard, however, two points are important. First, in the above discussion, a random covariable (rather than a fixed covariable) has been assumed. In psychological research, covariables are invariably random because subjects are identified and then values of the covariables are observed. Second, even though adjusted effects are equal to unadjusted effects for randomized experiments, the estimates of these two sets of effects will generally not be equal. Thus,

$$\hat{\alpha}_j' = \hat{\alpha}_{Y_j} - \hat{\beta}_{Y.X}\hat{\alpha}_{X_j},$$

but $\hat{\alpha}_{X_j}$ will generally take on nonzero values, making

$$\hat{\alpha}_j' \neq \hat{\alpha}_{Y_j}.$$

These chance differences between ANCOVA estimated effects and ANOVA estimated effects of the same data set are just that, chance differences. They have no bearing whatsoever on the substantive interpretations to be made of results. Adjusted means should be reported, because they are the most precise estimates for forming treatment contrasts and because they are consistent with the model on which the tests of hypotheses rest.

For randomized experiments, then, the sole motivation for using ANCOVA rather than ANOVA is a concern for statistical precision, as discussed in the preceding section. The hypotheses tested and the population effects estimated are identical for these two models, even if their sample estimates are not. But what about uses of ANCOVA in nonrandomized experiments?

A great many social science researchers seek to investigate the possibility of causal relations by using data drawn from other than randomized experiments. Much of their work relies on studies of natural variation in which, for a single sample, several variables of interest—some of which are hypothesized as causal agents, some of which may be hypothesized as intervening variables, and some of which represent outcomes of interest—are observed simultaneously. For instance, Bragg (1979) studied the relation between the causal attributions college students make for their loneliness, their coping responses to loneliness, and their level of depression. When the data for such studies are gathered at a single point in time, the investigator must impose a temporal ordering of constructs based on theory. Moreover, many demographic and personality variables may confound key causal inferences. ANCOVA is not generally appropriate for analyzing such data. Rather, researchers look to path analysis models and structural regression models in efforts to estimate causal relations (e.g., Blalock, 1971; Cook & Campbell, 1979; Duncan, 1975; Joreskog, 1974).

But sometimes social scientists who wish to estimate effects of interventions are unable to conduct a randomized experiment and must settle instead for a nonrandomized experiment. Perhaps the single most common analysis of nonrandomized experiments is ANCOVA. The hope is that differences among treatment conditions that are present because of a failure to randomly assign subjects will be (a) reflected in the observations on one or more covariables and (b) appropriately removed from dependent variable effects through ANCOVA adjustments. In order to understand the potential of ANCOVA for satisfying these two conditions, it is helpful to have a way to think about data from nonrandomized experiments. Again, the primary conceptual issues can be seen within the limited context of a design that contrasts one treatment group and one control group and in which the covariable is a pretest and the dependent variable is a posttest.

The question of what constitutes the proper adjustment necessary to compensate for nonrandom assignment in such studies has often been referred to as the problem of measuring change (Harris, 1963; Linn, 1981). This is because the researcher wishes to estimate the differential amount of growth (change) on the dependent variable that can correctly be attributed to differing effects of treatment conditions. In randomized experiments, the control group serves as the estimate of no growth due to treatment.

In Figure 3, a hypothetical set of population data are represented for a randomized experiment. The solid curve A → B represents growth on the dependent variable dimension (averaged across all subjects) for a no-treatment control

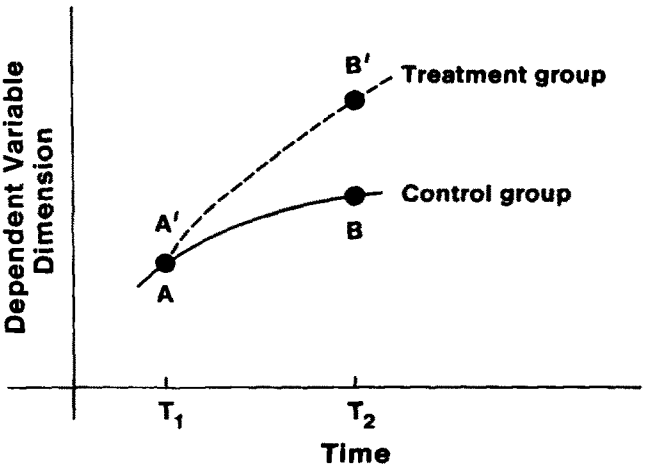

Figure 3. Hypothetical growth curves for treatment and control groups in a randomized study.

population over the period of the experiment, time T_1 to time T_2. The dashed curve A′ → B′ represents growth on the dependent variable dimension (again averaged across subjects) for a treatment population. Had there been no treatment effect, A → B and A′ → B′ would have been collinear. Thus, the difference B′ − B represents the amount of growth attributable to the treatment (i.e., the correct solution to the problem of measuring change). When Figure 3 is translated into the terms of ANOVA and ANCOVA, it can be seen that either procedure provides an unbiased estimate of B′ − B.

Observations at T_1 represent the covariable, X, and observations at T_2 represent the dependent variable, Y. For ANCOVA,

$$\alpha'_j = \alpha_{Y_j} - \beta_{Y.X}\alpha_{X_j}.$$

For the treatment group, this is

$$\alpha'_T = B' - \frac{B + B'}{2} - \beta_{Y.X}\left(A' - \frac{A + A'}{2}\right)$$

Because A = A′,

$$\alpha'_T = B' - \frac{B + B'}{2},$$

or in words, half of the distance from B to B′, which is the typical deviation form for defining treatment effects in linear models.

For ANOVA,

$$\alpha_j = \mu_{Y_j} - \mu_{Y.}$$
$$= B' - \frac{B + B'}{2}.$$

As stated before, the two procedures estimate the same treatment parameters, and, as can be seen in Figure 3, these treatment parameters are conceptually correct for randomized experiments.

In nonrandomized experiments, researchers wish to estimate the same treatment effects that would have been estimated had a randomized experiment been performed. Again, the problem of measuring change can be represented by growth curves for two treatment conditions and one outcome variable. In Figure 4, curve A → B represents growth on the dependent variable dimension for a no-treatment control group over the period of the experiment, time T_1 to time T_2, and the dashed line A′ → B′ represents growth for the treatment group. Curve A′ → C represents the growth of the treatment group if there had been no intervention (or no effect of the intervention). Thus, B′ − C represents the treatment effect that would have been estimated had a randomized experiment been done, B′ − B represents the treatment effect that an ANOVA would estimate, and therefore, C − B is the appropriate adjustment for ANCOVA.[2] The key question for evaluating the usefulness of ANCOVA with nonrandomized experiments becomes, under what conditions will the ANCOVA adjustment equal C − B?

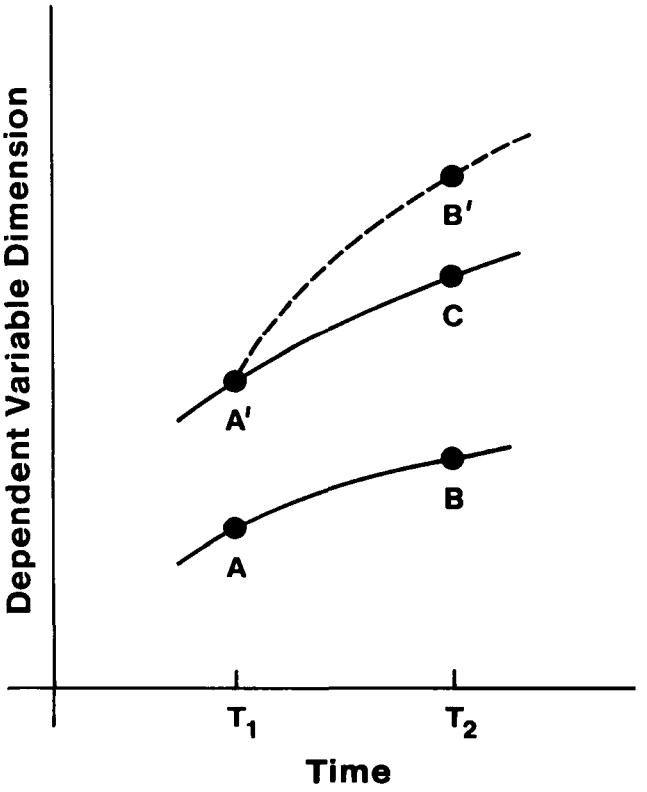

Figure 4. Hypothetical growth curves for treatment and control groups in a nonrandomized study.

Consider again an ANCOVA treatment effect

$$\alpha'_j = \alpha_{Y_j} - \beta_{Y.X}\alpha_{X_j}.$$

As shown previously, an ANCOVA adjusted treatment effect is

$$\alpha'_j = \frac{B' - B}{2} - \beta_{Y.X}\frac{A' - A}{2}.$$

We also know that the desired value of α'_j is (B′ − C)/2. Thus, the conditions under which ANCOVA will provide the appropriate adjustment are found by solving the following expression for $\beta_{Y.X}$:

$$\frac{B' - C}{2} = \frac{B' - B}{2} - \beta_{Y.X}\frac{A' - A}{2},$$

or

$$\beta_{Y.X} = \frac{C - B}{A' - A}.$$

Unfortunately, in an application of ANCOVA, only points A,

[2] Actually, because linear model effects are always deviations from a grand mean rather than differences between comparison groups this statement is only conceptual, not algebraic.

A′, B, and B′ are directly estimable. Point C represents the hypothetical mean of the treatment group if the treatment had not been given. But, of course, the treatment always is given. The point is that a researcher can never be sure whether α_j' is the right value because distance C − B can be estimated only by doing a different study, one involving random assignment of subjects to experimental conditions.

There is a great deal of literature on the problem of measuring change in nonrandomized experiments and a great many myths and some mystery to go with it (e.g., Elashoff, 1969; Evans & Anastasio, 1968; Lord, 1967). Despite everyone's hopes to the contrary, the problem remains unsolved and almost certainly unsolvable if data are collected at only two points in time, as is minimally required for ANCOVA (i.e., one covariate measured prior to the study and one dependent variable measured after the treatment or treatments). Recent methodological work is focused on issues of measurement, design, and analysis in nonrandomized studies based on more than two time points (Bryk & Raudenbush, 1987; Rogosa, Brand, & Zimowski, 1982). The inescapable uncertainties associated with two-time-point data and nonrandom assignment may be considerably reduced by researchers designing multiple-time-point research and attending to these issues.

Some insight into the promise of ANCOVA for use in nonrandomized experiments can be gained through further consideration of $\beta_{Y.X}$ and of various forms of growth models. The slope parameter is, perhaps, easier thought about within the context of ANCOVA if one remembers that

$$\beta_{Y.X} = \rho_{XY}\,\sigma_Y/\sigma_X.$$

With this in mind and with the types of growth curves depicted in Figure 4 as a framework, one can consider some special types of data. One possibility would be for growth curves to remain parallel over time in the absence of treatment effects. For example, in Figure 4, curves A′ → C and A → B could have been drawn parallel to each other. In such a case, (C − B)/(A′ − A) = 1.0. For ANCOVA to provide the appropriate adjustment, $\beta_{Y.X}$ must equal 1.0. It is possible to imagine that at least for studies of short duration, the variance of the pretest and the variance of the posttest might be equal (i.e., $\sigma_X = \sigma_Y$), but surely errors of measurement would attenuate the value of the correlation between pretest and posttest (i.e., $\rho_{XY} < 1.0$). Thus, for parallel growth, only under rare circumstances (when attenuation of the correlation coefficient is offset by an increase in variance) can $\beta_{Y.X}$ be expected to take the desired value of 1.0.

Do not be misled. We are not saying that parallel growth curves are typical of psychological data, and therefore, we are not saying that the desired value of $\beta_{Y.X}$ is typically 1.0. If that were the case, a more straightforward and more statistically powerful approach would be to form gain scores and use them as the dependent variable in ANOVA (Porter & Chibucos, 1974).

One of the most popular assumptions about growth is called the *fan spread hypothesis* (e.g., Bryk & Weisberg, 1977; Campbell & Erlebacher, 1970). For data to conform to the fan spread hypothesis, growth in the absence of a treatment effect (and in the terms of Figure 4) must satisfy

$$\frac{\text{C} - \text{B}}{\text{A} - \text{A}'} = \frac{\sigma_Y}{\sigma_X},$$

where σ_Y and σ_X are the common within-group standard deviations for Y and X. In words, the ratio of posttest mean differences to pretest mean differences must equal the ratio of posttest standard deviation to pretest standard deviation. This would be true if, for example, each group growth curve were the average of individual growth curves, all of the same form but with differing rates of change, and there were no errors of measurement in the data (Blumberg & Porter, 1983). The definition of the fan spread hypothesis is nearly identical to the condition identified earlier as being necessary for ANCOVA to provide the appropriate adjustment. What is missing in the fan spread formulation that is in the ANCOVA conditions is ρ_{XY}. In short, if $\rho_{XY} = 1.0$, then

$$\beta_{Y.X} = \rho_{XY}\,\frac{\sigma_Y}{\sigma_X} = \frac{\text{C} - \text{B}}{\text{A} - \text{A}'}.$$

This is identical to the problem identified for parallel growth (in fact, parallel growth is a special case of the fan spread hypothesis) and is often called the *fallible covariable problem* (because errors of measurement in the covariable are the cause of difficulty).

It can be shown that the slope of the regression line defined on observed variables (containing errors of measurement) is not equal to the slope of the regression line defined on the latent true variables (containing no errors of measurement). In fact,

$$\beta_{\text{T}} = \beta_{Y.X}/\rho_{XX},$$

where β_{T} denotes the slope on the latent true variables, $\beta_{Y.X}$ is as before, and ρ_{XX} is the reliability of the X variable (covariable). If we go back to the definition of $\beta_{Y.X}$,

$$\beta_{\text{T}} = (\rho_{XY}/\rho_{XX})\,(\sigma_Y/\sigma_X).$$

If the correlation between the covariable and dependent variable is 1.0 when corrected for attenuation (and if the reliability of X and Y are equal so that ρ_{XX} does correct for attenuation),

$$\beta_{\text{T}} = \sigma_Y/\sigma_X.$$

Thus, when the covariable is without error of measurement, ANCOVA provides the appropriate adjustment for data conforming to the fan spread hypothesis. When the covariable is fallible, this will not be the case, and some other procedures have been suggested (e.g., see Porter & Chibucos, 1974, for discussions of estimated true scores ANCOVA and true residualized gains).

For some time, the literature on the problem of measuring change in nonrandomized experiments centered on such issues as correcting for errors of measurement in the covariable. The real issue, however, is the one with which we started this brief discussion. ANCOVA provides the appropriate adjustment only under a very limited set of conditions. Worse, the re-

searcher is never in a position to know whether data satisfy those conditions. In short, ANCOVA cannot be counted on to estimate the right effects or test the correct hypothesis when it is used to analyze nonrandomized experiments. The best solution to this difficulty is to avoid nonrandomized experiments, conduct randomized experiments instead, and use ANCOVA as a procedure for improving the precision with which treatment effects are estimated.

Using Outcome Variables as Covariables

We said earlier that the covariable or covariables should either be observed prior to treatment or be known to be unaffected by the treatments. From the preceding discussion on adjustment, the reason for this requirement should be clear. If the covariable is measured after the treatment and if the treatment affects the covariable, then ANCOVA will, by adjusting on the covariable, remove some or all of the treatment effect for the dependent variable.

Sometimes, however, a design includes two or more dependent variables. Theory suggests that some of the dependent variables are affected directly by the treatment, whereas other dependent variables are affected only indirectly through intervening dependent variables. Alternatively, a researcher may wish to know whether the treatment effect on one dependent variable is independent of the treatment effect on other dependent variables. ANCOVA has been used as one approach to answer such questions. The idea is straightforward: by using some dependent variables as covariables while analyzing other dependent variables, researchers' use of ANCOVA serves to estimate or test hypotheses about partial treatment effects (Cochran, 1957). When used in this way, the ANCOVA tests of hypotheses are sometimes called *step down Fs*.

Just as we have shown that the ANCOVA is suspect when used to adjust for nonrandom assignment in experiments, the use of ANCOVA to separate treatment effects among dependent variables is equally problematic. Interpretation of ANCOVA adjusted treatment effects when the covariables are also outcome variables should be performed with extreme caution and with the assistance of clear theory.

Precision Under Nonrandom Assignment

Earlier we examined ANCOVA as a tool for increasing precision in the context of experiments with random assignment of subjects to groups. When random assignment is not possible, methodological attention understandably tends to focus more on sources of bias than on imprecision. But the need for precision in nonrandomized experiments, though perhaps overshadowed by the need to avoid danger of bias, remains a concern.

ANCOVA is a less useful means of increasing precision in nonrandomized experiments than in randomized experiments. Mathematically, the point will be clear if we reexamine the expression for error variance of a two-group comparison using ANCOVA. That expression included a term k, where

$$k = \sigma^2 (1 - \rho^2_{XY})(\bar{X}_{.1} - \bar{X}_{.2})/\Sigma\Sigma (X_{ij} - \bar{X}_{..})^2.$$

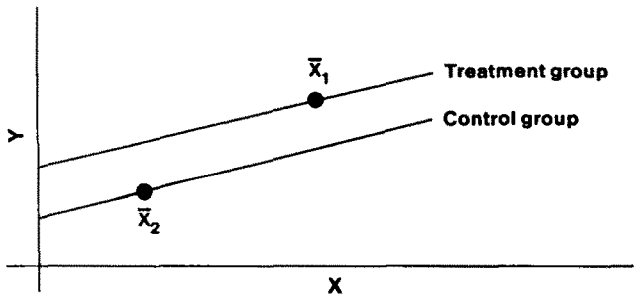

Figure 5. Population regression lines for two treatment groups.

Clearly, as $X_{.1} - X_{.2}$ increases, so does the sampling variance for ANCOVA effects. Because comparison groups in nonrandomized experiments can be quite different in their mean values on the covariable, ANCOVA has substantially poorer precision for nonrandomized experiments than for randomized experiments.

The relatively poor precision of ANCOVA when used in nonrandomized experiments should not necessarily be thought of as bad and as to be avoided if possible. Conceptually, ANCOVA treatment comparisons (contrasts of adjusted means) can be thought of as differences on the dependent variable axis, Y, between parallel treatment group regression lines (see Figure 5). For nonrandomized experiments in which comparison groups differ substantially on the covariable, X, even small sampling fluctuation in the estimates of regression line slopes causes major changes in the distances among the treatment group regression lines. In short, the nature of the ANCOVA adjustment is highly dependent on the sampling fluctuation in estimates of $\beta_{Y.X}$. A major problem with gain scores or other indexes of response is their failure to inflate appropriately the error variance of treatment effect estimates when groups differ substantially on the premeasure. Thus, in nonrandomized experiments, gain scores may provide a false sense of precision, testing with great statistical power incorrectly defined effects.

Summary

The ANCOVA is not a new statistical procedure. Most of the issues involved in its appropriate use have been known for some time. In this regard, an especially useful set of references is an entire issue of *Biometrics* devoted to ANCOVA (see especially articles by Cochran, 1957, and by Smith, 1957). Nevertheless, the rather straightforward application of ANCOVA for purposes of improving statistical power (or precision of estimates) is too seldom found in social science research. The highly problematic use of ANCOVA to adjust for the effects of nonrandom assignment in experiments continues, all too often without appropriate caveats in the interpretation of results.

The purpose of this article has been to sketch how ANCOVA works, both in reducing error variance and in adjusting effects. While keeping the general nature of counseling psychology research in mind and by taking into account the various alternatives to ANCOVA, we conclude that most randomized

experiments would profit from having a covariable and use of ANCOVA. For nonrandomized experiments, however, ANCOVA is not a good approach. If random assignment is simply not possible, the next best approach appears to be to take multiple measures over time in a way that allows the modeling of growth separately from the effects of treatments.

References

Blalock, H. M., Jr. (1971). *Causal models in the social sciences.* Chicago: Aldine.

Blumberg, C., & Porter, A. C. (1983). Analyzing quasi-experiments: Some implications of assuming continuous growth models. *The Journal of Experimental Education, 51,* 150–159.

Bragg, M. A. (1979). *A comparative study of loneliness and depression.* Unpublished doctoral dissertation, University of California, Los Angeles.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158.

Bryk, A. S., & Weisberg, H. I. (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin, 84,* 950–962.

Campbell, D. T., & Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate, Vol. 3. The disadvantaged child* (pp. 185–210). New York: Brunner/Mazel.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13,* 261–281.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation.* Chicago: Rand McNally.

Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika, 44,* 150–158.

DeGracie, J. S., & Fuller, W. A. (1972). Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. *Journal of the American Statistical Association, 67,* 930–937.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6,* 383–401.

Evans, S. H., & Anastasio, E. J. (1968). Misuse of analysis of covariance when treatment effect and covariance are confounded. *Psychological Bulletin, 69,* 225–234.

Fisher, R. A. (1948). *Statistical methods for research workers* (10th ed.). New York: Hafner.

Foy, D. W., Nunn, L. B., & Rychatarik, R. G. (1984). Broad spectrum behavioral treatment for chronic alcoholics: Effects of training controlled drinking skills. *Journal of Consulting and Clinical Psychology, 52,* 218–230.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research, 42,* 237–288.

Harris, C. W. (1963). *Problems in measuring change.* Madison: University of Wisconsin.

Hobbs, S. A., Walle, D. L., & Caldwell, H. S. (1984). Maternal evaluation of social reinforcement and time out: Effects of brief parent training. *Journal of Consulting and Clinical Psychology, 52,* 130–136.

Jacobson, N. S. (1984). A component analysis of behavioral marital therapy: The relative effectiveness of behavioral exchange and communication/problem-solving training. *Journal of Consulting and Clinical Psychology, 52,* 268–279.

Joreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1–56). San Francisco: W. H. Freeman.

Kazdin, A. E., & Mascitelli, S. (1984). Covert and overt rehearsal and homework practice in developing assertiveness. *Journal of Consulting and Clinical Psychology, 50,* 250–258.

Linn, R. L. (1981). Measuring pretest–posttest performance changes. *Education evaluation methodology: The state of the art.* Baltimore, MD: Johns Hopkins University Press.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68,* 304–305.

Piper, W. E., Debbane, E. G., & Bienvenu, J. P. (1984). A comparative study of four forms of psychotherapy. *Journal of Consulting and Clinical Psychology, 52,* 268–279.

Porter, A. C. (1967). *The effects of using fallible variables in the analysis of covariance.* Unpublished doctoral dissertation, University of Wisconsin, Madison.

Porter, A. C., & Chibucos, T. R. (1974). Selecting analysis strategies. In G. Borich (Ed.), *Evaluating educational programs and products* (pp. 415–464). Englewood Cliffs, NJ: Educational Technology Publications.

Rogosa, D. R., Brand, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 728–748.

Sanchez-Craig, M., Annis, H. M., Bonet, H. R., & McDonald, K. R. (1984). Random assignment of abstinence and controlled drinking: Evaluation of a cognitive behavioral program for problem drinkers. *Journal of Consulting and Clinical Psychology, 52,* 390–403.

Scovern, A. W., Bukstel, L. H., Kilmann, P. R., Laval, R. A., Busemeyer, J., & Smith, V. (1980). Effects of parent counseling on the family system. *Journal of Counseling Psychology, 27,* 268–275.

Smith, H. F. (1957). Interpretations of adjusted treatment means and regressions in analysis of covariance. *Biometrics, 13,* 281–308.

Received April 11, 1986
Revision received August 27, 1986
Accepted August 27, 1986 ■