This article proposes and evaluates a method to test for mediation in multilevel data sets formed when an intervention administered to intact groups is designed to produce change in individual mediator and outcome variables. Simulated data of this form were used to compare ordinary least squares (OLS) and two multilevel estimators of the mediated effect. OLS and multilevel standard error approximations were also evaluated and recommendations given for optimal estimator choice. These methods were applied to data from an existing substance use intervention to show the impact multilevel mediation modeling can have on the conclusions drawn from real-world evaluation studies.

MULTILEVEL MEDIATION MODELING IN GROUP-BASED INTERVENTION STUDIES

JENNIFER L. KRULL
DAVID P. MacKINNON
Arizona State University

Most prevention programs are based on theory of individual behavior. Consequently, the success of a prevention program is measured by its effect on individuals. Often, however, individuals are clustered within intact groups, and for practical reasons, the prevention intervention is randomized and administered at the group level. The groups involved in such a design may include classes or schools in educational contexts, hospitals or other treatment sites in clinical contexts, companies or offices in organizational contexts, and neighborhoods, counties, states, or countries in geographical contexts. When intact groups are assigned to conditions and observations are made on individuals within these groups, a multilevel data set is formed. The independent variable, assignment to intervention or control condition, is a group level variable, and the dependent variable is measured at the individual

level. Overcoming the difficulties introduced by the structure of multilevel data has been a focus of much research during the past 20 years. Generally accepted methods of dealing with such data have been developed (Bryk and Raudenbush 1992; Goldstein 1995), and applications are now appearing in the applied research literature. However, several topics remain unresolved. The purpose of this article is to describe methods to assess mediated effects in multilevel data.

TRADITIONAL ANALYSES OF MULTILEVEL DATA

Traditional analytical techniques, such as ordinary least squares (OLS) regression, are typically limited to single level data. Applying traditional techniques to multilevel data requires restructuring the data set to eliminate one level of the hierarchy. The group level data can be disaggregated and analyzed as if all variables were measured at the individual level, or the individual level data can be aggregated and the analysis conducted at the group level (de Leeuw 1992; de Leeuw and Kreft 1986). Both of these approaches involve potential drawbacks.

In the disaggregation approach, all individuals within a group are assigned an identical score on the group level variable as if this were an observation unique to the individual. Any error associated with this group measure will be identical for all group members. Additionally, the individuals within a given group may be more similar to each other than they are to members of other groups on any number of unmeasured variables. Intraclass correlation (ICC) is a measure of the extent of this within-group homogeneity (Haggard 1958, 6). Common error associated with the group level predictor and the unmeasured similarities among group members will result in correlated error terms among the individuals within a group and nonzero ICC. Because traditional analysis techniques typically assume independent and identically distributed error terms, employing such methods to analyze multilevel data violates assumptions, producing inefficient estimates and, more importantly, biased standard errors and overly liberal inferences (Hanushek and Jackson 1977, 145-146, 156-157; Moulton 1986; Scott and Holt 1982). With even small ICCs, the actual probability of making a Type I error can be much higher than the nominal alpha level of the test (Barcikowski 1981; Scariano and Davenport 1987; Walsh 1947), particularly when group sizes are large.

The alternative to the disaggregation approach, aggregating all individual level data to the group level, is also quite limiting. When individual level data are aggregated by computing group means, the number of observations in the analysis is reduced from the number of individuals to the number of groups,

decreasing the power of tests to detect effects of one variable on another. Also, discarding individual observations eliminates the ability to predict individual level variation, which may comprise the majority of the total variation present in the data (de Leeuw 1992). Moreover, in many intervention studies, the individual level effects are of primary interest, and the results of an aggregated group level analysis cannot necessarily be used to draw inferences at the individual level. Indeed, individual and group level analyses of the same data may indicate relationships that differ in both magnitude and direction (Robinson 1950), and making individual level inferences based on group level relationships has come to be known as the "ecological fallacy" (Pedhazur 1982, 529). Although clear evidence of the ecological fallacy has not yet been observed in the analysis of randomized interventions, researchers must be careful to confine all statements based on the results of an analysis to the particular level at which the analysis was conducted.

MULTILEVEL MODELING

Multilevel data can be analyzed without artificially restructuring the data by employing random coefficient multilevel models. These models can simultaneously examine effects of both individual and group level variables on an individual level outcome. Moreover, the correlated errors and nonzero ICC inherent in grouped data are appropriately incorporated in multilevel models, giving accurate standard error estimates and inferences.

A multilevel model may be conceptualized as a series of models at the different data levels. At the lowest level of the hierarchy, a linear model is constructed relating individual level predictors to the individual level outcome. Consider, for example, a situation with an individual level outcome Y_{ij} and a single individual level predictor X_{ij}. The double subscript for these variables indicates that the observations are unique for each individual i within each group j. The relationship between the two variables is modeled with a linear equation within each group.

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}.$$

A key feature of multilevel models is that the coefficients associated with the intercept and the X slope in this individual level equation can be allowed to vary across groups. The subscripts for the β coefficients in this lowest level equation indicate that they can differ for each group j. The observed coefficient values are viewed as realizations drawn from distributions of possible values of intercepts and slopes. These coefficient values, then, can be used as

outcome variables in group level equations, which can be predicted by group level variables W_j.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}.$$

Forming prediction equations at each of the different levels of the hierarchy in this manner accommodates individual and family level predictors in a single multilevel analysis. The specification of error terms at both the individual (ε) and group (u) levels allows multilevel models to appropriately model the error in grouped data (i.e., nonzero ICC).

The series of level-specific models is combined algebraically into a single model for the estimation of parameters, allowing the estimation of fixed effects (γ parameters) without actually estimating the individual level model within each group. In addition to these estimates of fixed effects, multilevel analyses also produce estimates of the variances and covariances of the ε and u error terms, known as the variance components. More detailed discussion of multilevel procedures can be found in Bryk and Raudenbush (1992), Goldstein (1995), Kreft and de Leeuw (1998), and Longford (1993).

Many researchers have recognized the appropriateness of multilevel modeling for group-based interventions and have used this analysis technique to examine intervention effects. Seltzer (1994) illustrated how multilevel models can incorporate the dependency present in grouped data and simultaneously examine the effects of variables at the individual and the group level in the evaluation of a mathematics curriculum. Other researchers have used multilevel modeling to evaluate school-based interventions on substance use behaviors, knowledge, and attitudes, both in terms of overall program effects and in interaction with student and school level characteristics (DeVries et al. 1994; Hedeker, Gibbons, and Flay 1994; Kreft 1994, 1997; Murray et al. 1992).

MEDIATION IN SINGLE LEVEL MODELS

In addition to determining whether an intervention has its desired effect on the outcome, many researchers are also interested in identifying the processes that bring about these effects. Mediational analysis is a method for examining such processes. Components of an intervention program are designed to change specific mediational constructs, which are believed to cause changes in the outcome variable (MacKinnon and Dwyer 1993).

Research examining the mechanisms by which prevention programs achieve their effects have begun to appear in the evaluation of drug use prevention (MacKinnon et al. 1991; Donaldson, Graham, and Hansen 1994) and reemployment (Vinokur and Schul 1997) programs. For example, the Adolescents Training and Learning to Avoid Steroids (ATLAS) program, which will be discussed in greater detail in a later section of this article, is an intervention designed to directly change a number of mediators (e.g., peer influences) that are believed to ultimately affect steroid use. Figure 1 illustrates such a mediational model in which an intervention (X) affects a mediator (M), which in turn affects the outcome (Y).

In models involving only a single level of data, point estimates of the mediated effect can be calculated in two ways (MacKinnon, Warsi, and Dwyer 1995). Each of the two methods requires the estimation of two regression equations. The first of these methods involves estimating

$$Y = k + cX + \varepsilon \tag{1}$$

and

$$Y = k + c'X + bM + \varepsilon. \tag{2}$$

Equation 1 estimates the effect of the intervention X on the outcome measure Y. The c coefficient in this equation represents the total effect of the intervention on the outcome, without taking the mediator into account. The second equation estimates the simultaneous effects of both the intervention X and the mediator M on the outcome. In this equation, the c' coefficient represents the effect of the intervention on the outcome, with the effect of the mediator statistically removed. The difference between the c and c' coefficients associated with the intervention variable X in the two equations is an estimate of the mediated effect (Judd and Kenny 1981). This $c - c'$ difference measures the extent to which the mediator in Equation 2 accounts for the relationship between the intervention and outcome assessed in Equation 1.

The second method for calculating point estimates of mediated effects requires estimating the two equations

$$M = k + aX + \varepsilon \tag{3}$$

and

$$Y = k + c'X + bM + \varepsilon. \tag{2}$$

In Equation 3, the a coefficient is an estimate of the effect of the intervention on the mediator. The second equation is identical to Equation 2 from the first
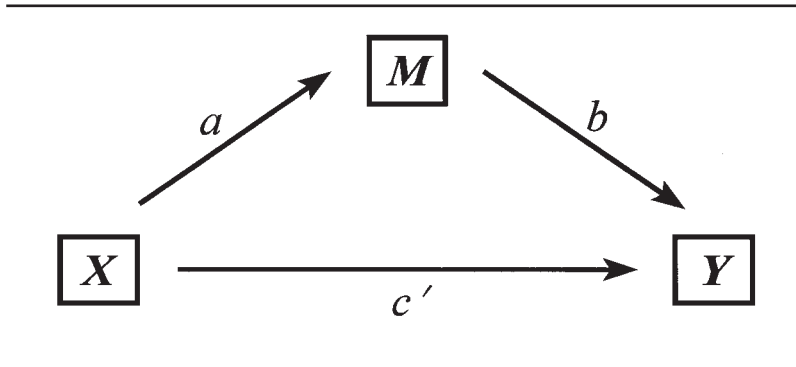
Figure 1:   A Single Mediator Model

method for calculating point estimates of the mediated effect. However, in this case, the focus is on the b coefficient, the effect of the mediator on the outcome variable, rather than on the c' coefficient, as in the previous method. The product of the two estimated coefficients, ab, is an estimate of the mediated effect. This method makes explicit that the estimate of the mediated effect depends on the extent to which the intervention affects the mediator and the extent to which the mediator affects the outcome. The two methods of calculation, $c - c'$ and ab, produce algebraically equivalent estimates for single level models (MacKinnon, Warsi, and Dwyer 1995).

The ab point estimate of the mediated effect calculated by the second method can be considered the product of two random variables (Sobel 1982), and methods for determining the variance of a product of random variables can provide estimates for the standard error of the mediated effect, which can be used in forming interval estimates and in conducting significance tests. These methods are based on large sample theory and are thus approximate for small samples. As outlined in MacKinnon, Warsi, and Dwyer (1995), estimates of the standard error of the mediated effect can be calculated in a number of different ways, including (a) first-order Taylor series expansion or the multivariate delta method (Sobel 1982; 1986):

$$s_{ab} = \sqrt{s_a^2 b^2 + s_b^2 a^2} \, ;$$

(b) second-order Taylor series expansion or exact variance under the condition of independence (Goodman 1960; Mood, Graybill, and Boes 1974):

$$s_{ab} = \sqrt{s_a^2 b^2 + s_b^2 a^2 + s_a^2 s_b^2} \, ;$$

or (c) the estimate of unbiased variance (Goodman 1960):

$$s_{ab} = \sqrt{s_a^2 b^2 + s_b^2 a^2 - s_a^2 s_b^2}.$$

In models involving only a single level of data, each of these standard error estimates is calculated using estimates of a and b and their associated standard errors s_a and s_b obtained from OLS estimation of Equations 2 and 3.

MULTILEVEL MEDIATION

In prevention intervention studies, the mediational effects that are of particular interest are those that explain the relationship between the intervention program and the outcome measure. When the program is randomized and administered at the group level and the outcome measure is assessed at the individual level, multilevel mediational analysis is most appropriate. Each of the three equations described above in the context of a single level analysis can be reformulated as a multilevel model. Table 1 shows the single level expressions for each of the three equations and the multilevel expression equivalent to each.

Equation 1 shows the OLS and multilevel equivalent specifications for determining the effect of the intervention X on the outcome measure Y, ignoring (for the moment) the mediator. Although the OLS equation assumes that X and Y are at the same level, the multilevel specification clearly shows that Y is an individual level outcome, and the intervention X is a group level predictor. Also, although the OLS specification contains only a single error term ε, the multilevel version contains error terms as both levels, ε and u, to appropriately model the error in the grouped data. Equations 2 and 3 are similar: The multilevel equations include the individual level Y and M variables in individual level equations, while the group level X predictor appears in group level equations. Multilevel Equations 2 and 3 also include error terms at both levels to accommodate the dependency in grouped data. To parallel the single level OLS analyses as closely as possible, only the intercept terms of the multilevel models are treated as random, and the b coefficient (the slope associated with the effect of the mediator on the outcome) in the multilevel Equation 2 is fixed.

Although Table 1 shows that each OLS equation necessary for mediational analysis can be reformulated as a multilevel model, some of the relationships found in single level mediational models do not hold for multilevel mediational models. Specifically, the two different point estimates of the mediated effect, $c - c'$ and ab, are algebraically equivalent in single level

TABLE 1:   OLS and Multilevel Expressions for the Three Mediational Analysis Equations

	OLS	Multilevel
Equation 1	$Y = k + cX + \varepsilon$	Individual Level: $Y_{ij} = \beta_{0j} + \varepsilon_{ij}$ Group Level: $\beta_{0j} = \gamma_{00} + cX_j + u_{0j}$
Equation 2	$Y = k + c'X + bM + \varepsilon$	Individual Level: $Y_{ij} = \beta_{0j} + bM_{ij} + \varepsilon_{ij}$ Group Level: $\beta_{0j} = \gamma_{00} + c'X_j + u_{0j}$
Equation 3	$M = k + aX + \varepsilon$	Individual Level: $M_{ij} = \beta_{0j} + \varepsilon_{ij}$ Group Level: $\beta_{0j} = \gamma_{00} + a\_X_j + u_{0j}$

NOTE: OLS = ordinary least squares.

models, but the equivalence does not hold in the multilevel framework. Correspondingly, the basic decomposition of the total effect into a direct effect and an indirect (mediated) effect does not hold for multilevel models. Estimates of direct and indirect effects will not necessarily sum to the estimate of the total effect if estimates of the different effects are generated using different multilevel equations. The reason for these discrepancies lies in the nature of multilevel estimation.

Procedures for estimating multilevel models differ from standard OLS estimation in that multilevel estimation is iterative in nature. The fixed effects and variance component parameters of the model are estimated in separate steps, using current estimates of one parameter type in updating estimates of the other. The estimates of the fixed effects in a multilevel model, given estimates of the variance components, are generated via a generalized least squares expression, $(X\_VX)^{-1}X\_VY$. This differs from the expression used to generate OLS estimates, $(X\_X)^{-1}X\_Y$, only in the use of a weighting matrix V. The elements that make up this V matrix are functions of the variance component estimates. Estimating Equations 1, 2, and 3 to perform a mediational analysis requires estimating 3 different multilevel models, and the variance components and the V matrices produced by each estimation process will differ. Consider, for example, the ε_{ij} terms in multilevel expressions of Equation 1 and Equation 3. In Equation 1, this term represents the residual error in the outcome measure after adjusting for program participation. In Equation 3, this term represents the residual error in the mediator after adjusting for program participation. Because the two ε_{ij} terms are associated with different measures, these errors will, of course, be different, and the V matrices constructed using estimates of the corresponding variance components will differ as well. Unless the three V matrices are identical, the algebraic equivalencies of effect decomposition and mediational analysis will not hold.

As with the three equations for estimating mediated effects, single level methods for estimating the standard error of the mediated effect also can be reformulated for multilevel models. This straightforward adaptation simply requires computing the first-order Taylor series estimator, the second-order Taylor series estimator, or the unbiased estimator using estimates of a and b and their standard errors that are generated by the iterative estimation of a multilevel model in place of those typically generated through OLS analysis of a single level model. However, because these are large sample approximations, the adequacy of their performance in small samples requires study.

The purpose of this article is to examine the performance of multilevel mediational analysis in a simulated data set and in an existing prevention intervention data set. The simulation study was designed to answer several specific questions. First, given that multilevel estimates of the mediated effect produced using the $c - c'$ and ab methods will not be algebraically equivalent, how discrepant are the values likely to be? Second, is there any reason to prefer one estimate over the other? And finally, how well do the various large-sample-based approximations of the standard error of the mediated effect in multilevel data perform in small samples? Following this, OLS and multilevel mediational analyses of an existing prevention intervention data set are used to illustrate the differences between the two types of analyses in a real-world application. These comparisons allow the researcher to determine whether the improvements made through the use of multilevel mediational analysis warrant the additional complexity in model specification and estimation inherent in the multilevel framework.

SIMULATION STUDY

The SAS programming language was used to generate simulated multilevel data sets reflecting the form of an intervention study with randomization at the group level. Data were created according to the single mediator model depicted in Figure 2. The RANNOR function within SAS was used to create a dichotomous independent variable, X, representing assignment to intervention or control conditions. The other variables in the model, the mediator M and the outcome Y, were created as individual level variables according to the relationships specified in the single mediator model shown in Figure 2.

Two aspects of sample size, the number of groups and group size, were systematically varied in simulating the data for this study. Simulated data sets were composed of 10, 20, 30, 50, 100, or 200 groups. These groups were of
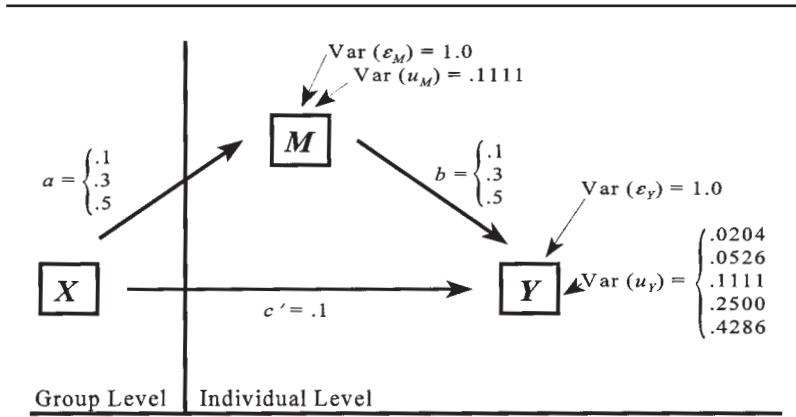
Figure 2:   Model Form and Parameter Values for Multilevel Simulated Data

either small or moderate size. For the small group size conditions, half of the groups were composed of 5 individuals, the other half of 10 individuals. For the moderate group size conditions, half of the groups were made up of 20 individuals, the other half of 30 individuals.

In the simulated model, the parameters a (representing the effect of the intervention on the mediator) and b (representing the effect of the mediator on the outcome) could take on three different values: .1, .3, and .5. True a and b values were set equal in each model, and the c' parameter (representing the direct effect of the intervention on the outcome) was set to a constant value of .1. These values of a and b are among those used in previous simulation studies of mediational models (e.g., MacKinnon, Warsi, and Dwyer 1995). Moreover, the true values of the mediated effect ab (.01, .09, and .25) reflect those observed in intervention studies. For example, these values correspond quite well to mediated effect values found in the analysis of a substance use intervention (MacKinnon, Goldberg, Clarke, et al. 1999), in which single mediator analyses of 18 possible mediators for each of three different dependent variables revealed significant ($p < .05$) point estimates of mediated effects ranging in magnitude from .01 to .27, with an average magnitude of .07 ($SD = .06$).

The proportion of residual variance in the simulation outcome variable due to between-group variability was systematically adjusted by changing the variability of the group error terms, u_{0j}, relative to the constant variability of the individual error terms, ε_{ij}, which was set equal to 1.0. Values were chosen

so that the proportion of residual variance due to between-group variability (residual ICC), calculated as Var (u_{0j}) / (Var $[\varepsilon_{ij}]$ + Var $[u_{0j}]$), was equal to .02, .05, .10, .20, or .30. These values were chosen to be representative of actual proportions that might be observed in intervention studies in a number of areas, including psychology, education, and family studies. The proportion of residual variance in the mediator due to between-group variability was held constant at .10.

All possible combinations of the 6 values for the number of groups in the analysis, the 2 group sizes, the 3 true values of *ab*, and the 5 values for the proportion of residual variability at the between-group level were examined. A total of 500 replications were conducted for each of the 180 conditions, producing 90,000 simulated data sets for analysis.

Each data set was analyzed using the three equations necessary for mediational analysis discussed previously, in both the OLS (single level) and multilevel frameworks. These analyses were carried out using the PROC REG and PROC MIXED routines of the SAS system for the single level and multilevel analyses, respectively. First-order Taylor series approximations were used to estimate standard errors of both single level and multilevel mediated effects.

The resultant data were examined to determine (a) the extent of the differences between the $c - c'$ and *ab* multilevel estimates of the mediated effect, (b) the relative bias and efficiency of the OLS and two multilevel estimators of the mediated effect, and (c) the relative bias of estimates of the standard error of the mediated effect. Relative bias (RB) was calculated using the equation

$$RB = \frac{\hat{w} - w}{w},$$

and efficiency was measured by calculating mean squared error (MSE) using the equation

$$MSE = (\hat{w} - w)^2,$$

where $\hat{w}$ is an estimate of the mediated effect or its standard error, and w is the true parameter value or an approximation of the true value when the precise value is not known. True values of the mediated effect were the $ab = .01, .09,$ or .25 values used in creation of the simulated data set. Empirical standard errors, calculated as the standard deviation of mediated effect estimates across the 500 replications in a given condition, were used as true value approximations in calculating the relative bias of the standard error estimates.

Following the recommendation of Hauck and Anderson (1984) that simulation studies be analyzed using the same analytical tools as any other

experimental study, factorial analyses of variance were conducted for each of the outcome measures, using number of groups, group size, true value of the mediated effect, and residual ICC of the *Y* variable as predictors. In those analyses involving comparisons among estimator types, estimation technique was treated as a repeated-measures factor, with single degree of freedom contrasts specified to focus on the comparisons of interest. Due to the large number of observations, standard probability values were not particularly useful cutoffs for distinguishing between meaningful and inconsequential effects. Therefore, an effect size value of .01 or greater, calculated via

$$\eta^2 = \frac{SS_{effect}}{SS_{error} + SS_{effect}}$$

was used in most cases to identify important effects. This particular expression for η^2 is used to estimate the effect size associated with a given effect independent of other effects included in the model. Using this expression, the effect sizes are not directly interpretable as percentages of total variance, because effect sizes of the various effects in a given design may total more than 1.00 (Tabachnick and Fidell 1989, 55).

SIMULATION RESULTS

DISCREPANCY BETWEEN $C - C$↲ AND *AB* MULTILEVEL ESTIMATES OF THE MEDIATED EFFECT

Table 2 shows the average discrepancy between $c - c'$ and *ab* estimates of the mediated effect produced using multilevel modeling techniques. For all conditions, the average discrepancy was essentially zero. An ANOVA predicting this discrepancy from group size, number of groups, true value of the mediated effect, and residual ICC of the outcome variable revealed no significant main effects or interactions.

Also shown in Table 2 is the average absolute value of the discrepancy between the $c - c'$ and *ab* estimates of the mediated effect. ANOVA results showed that the size of this absolute discrepancy was a function of group size, $F(1, 89820) = 8994.58$, $p < .0001$, $\eta^2 = .09$; number of groups, $F(5, 89820) = 3264.27$, $p < .0001$, $\eta^2 = .15$; true value of the mediated effect, $F(2, 89820) = 4140.30$, $p < .0001$, $\eta^2 = .08$; and the two-way interactions among these variables—for group size and number of groups, $F(5, 89820) = 850.45$, $p < .0001$, $\eta^2 = .05$; for number of groups and true value of the mediated effect,

TABLE 2:   Average (ave) and Absolute (abs) Differences Between *c* – *c*⌐ and *ab* Estimates of the Mediated Effect by Number of Groups, Group Size, and True Value of the Mediated Effect (true med)

			Number of Groups					
Group Size	True Med	Difference	10	20	30	50	100	200
5 to 10	.01	Ave	−.0002	.0000	.0001	.0000	−.0000	.0000
		Abs	.0047	.0025	.0017	.0011	.0007	.0004
	.09	Ave	−.0000	−.0002	−.0000	−.0000	−.0000	−.0000
		Abs	.0108	.0064	.0047	.0031	.0019	.0012
	.25	Ave	−.0002	.0003	.0004	−.0001	−.0001	.0000
		Abs	.0157	.0099	.0074	.0050	.0031	.0020
20 to 30	.01	Ave	.0000	−.0000	−.0000	−.0000	−.0000	−.0000
		Abs	.0013	.0007	.0005	.0004	.0002	.0002
	.09	Ave	−.0000	−.0001	−.0000	−.0000	.0000	−.0000
		Abs	.0035	.0022	.0016	.0011	.0007	.0005
	.25	Ave	−.0003	−.0000	−.0000	.0001	−.0000	.0000
		Abs	.0055	.0031	.0023	.0016	.0011	.0007

$F(10, 89820) = 310.42, p < .0001, \eta^2 = .03$; for group size and true value of the mediated effect, $F(2, 89820) = 965.22, p < .0001, \eta^2 = .02$. The main effects indicated that the absolute size of the discrepancy between the mediated effect estimates was larger in conditions with smaller numbers of groups, smaller group sizes, and larger true values of the mediated effect. Moreover, these effects combined interactively so that the effects were stronger in combination. However, although larger mediated effect values were associated with larger absolute discrepancies, it should be pointed out that these larger discrepancies represent proportionally smaller effects. For example, with 10 groups of 5 to 10 individuals, the average absolute discrepancy was .0047 for a true mediated effect equal to .01, and this value was more than three times as large, .0157, for a true mediated effect equal to .25. However, this discrepancy is nearly half (47%) of the size of the true mediated effect value in the .01 case, while the discrepancy is less than 7% of the size of the true mediated effect value in the .25 case.

RELATIVE BIAS AND MSE OF OLS AND MULTILEVEL
ESTIMATES OF THE MEDIATED EFFECT

Table 3 presents the relative bias in point estimates of the mediated effect produced by OLS and by the two multilevel estimators. There is essentially no bias in any of these estimators. There were no significant differences

TABLE 3:   Relative Bias in OLS and Multilevel *c* − *c*⌐ and *ab* Estimates of the Mediated Effect by Number of Groups, Group Size, and True Value of the Mediated Effect (true med)

Group Size	True Med	Estimate	Number of Groups					
			10	20	30	50	100	200
5 to 10	.01	OLS	−.0731	−.0141	.0176	−.0437	.0438	.0108
		c − *c*′	−.0670	.0025	.0177	−.0385	.0369	.0097
		ab	−.0492	−.0001	.0107	−.0387	.0371	.0091
	.09	OLS	−.0295	−.0018	.0089	−.0242	−.0176	.0051
		c − *c*′	−.0298	−.0007	.0154	−.0255	−.0178	.0054
		ab	−.0295	.0017	.0153	−.0253	−.0175	.0056
	.25	OLS	.0282	.0020	.0056	.0124	−.0002	−.0014
		c − *c*′	.0334	.0021	.0092	.0102	.0001	−.0023
		ab	.0342	.0010	.0076	.0104	.0005	−.0025
20 to 30	.01	OLS	.0591	.0246	−.0037	−.0381	−.0007	.0007
		c − *c*′	.0297	.0341	−.0057	−.0307	−.0036	.0033
		ab	.0271	.0363	−.0047	−.0305	−.0034	.0039
	.09	OLS	.0185	−.0079	.0045	.0111	.0011	−.0035
		c − *c*′	.0214	−.0052	.0050	.0104	−.0004	−.0033
		ab	.0216	−.0043	.0051	.0103	−.0006	−.0033
	.25	OLS	−.0068	.0030	.0049	−.0014	−.0024	−.0042
		c − *c*′	−.0084	.0043	.0058	−.0025	−.0023	−.0043
		ab	−.0072	.0043	.0059	−.0028	−.0023	−.0043

NOTE: OLS = ordinary least squares.

between the relative bias in the OLS estimator and the relative bias in the multilevel estimators, or between the *c* − *c*′ and *ab* multilevel estimators. Moreover, there were no significant differences in bias associated with the various combinations of group size, number of groups, true value of the mediated effect, and level of intraclass correlation examined in this study.

Table 4 presents the MSE values associated with the OLS and multilevel point estimates of the mediated effect. The two multilevel estimators were very slightly more efficient (i.e., had lower MSE values) than the OLS estimator, but this effect accounted for only a very small portion of the variance in MSE, $F(1, 89820) = 111.42$, $p < .0001$, $\eta^2 = .001$. Similarly, the *ab* multilevel estimator of the mediated effect was very slightly more efficient than the $c − c'$ estimator, but this effect accounted for an even smaller portion of the variance, $F(1, 89820) = 7.59$, $p < .01$, $\eta^2 = .00008$. Greater efficiency (smaller MSE) was found in conditions with larger numbers of groups, $F(5, 89820) = 1917.60$, $p < .0001$, $\eta^2 = .10$; larger group sizes, $F(1, 89820) = 929.58$, $p < .0001$, $\eta^2 = .01$; and smaller true values of the mediated effect, $F(2, 89820) = 3295.17$, $p < .0001$, $\eta^2 = .07$. Significant two-way interactions between

TABLE 4: Mean Squared Error (MSE) of OLS and Multilevel $c - c'$ and *ab* Estimates of the Mediated Effect by Number of Groups, Group Size, and True Value of the Mediated Effect (true med)

Group Size	True Med	Estimate	Number of Groups					
			10	20	30	50	100	200
5 to 10	.01	OLS	.003289	.001066	.000601	.000327	.000143	.000062
		$c - c'$	.003180	.001008	.000550	.000304	.000137	.000058
		ab	.003150	.000995	.000545	.000300	.000136	.000058
	.09	OLS	.014200	.006045	.003933	.002192	.001013	.000557
		$c - c'$	.013640	.005843	.003850	.002148	.000986	.000542
		ab	.013549	.005835	.003846	.002139	.000976	.000537
	.25	OLS	.036614	.016001	.010293	.005988	.002903	.001454
		$c - c'$	.036643	.015340	.010153	.005767	.002832	.001397
		ab	.036298	.015354	.010063	.005774	.002812	.001394
20 to 30	.01	OLS	.001283	.000471	.000278	.000141	.000073	.000034
		$c - c'$	.001140	.000430	.000254	.000134	.000067	.000033
		ab	.001139	.000430	.000253	.000133	.000068	.000033
	.09	OLS	.007240	.003215	.002107	.001307	.000601	.000306
		$c - c'$	.007007	.003049	.002018	.001250	.000579	.000213
		ab	.007001	.003040	.002009	.001248	.000577	.000292
	.25	OLS	.019204	.009471	.006027	.003489	.001704	.000848
		$c - c'$	.018223	.009043	.005794	.003355	.001629	.000806
		ab	.018232	.009039	.005771	.003352	.001620	.000804

NOTE: OLS = ordinary least squares.

number of groups and group size, $F(5, 89820) = 216.87$, $p < .0001$, $\eta^2 = .01$, and between number of groups and the true value of the mediated effect, $F(10, 89820) = 590.80$, $p < .0001$, $\eta^2 = .06$, indicated that these effects were even more pronounced in combination. All these effects operated similarly for each of the three point estimators.

COMPARISONS BETWEEN OLS AND MULTILEVEL ESTIMATES OF STANDARD ERROR

Table 5 presents the empirical standard error (the standard deviation of parameter estimates over the replications in a given condition) of the mediated effect and the relative bias in standard error estimates calculated using first-order Taylor series, second-order Taylor series, and unbiased estimators based on both OLS and multilevel analyses.

This table shows that the empirical standard errors associated with the OLS analyses were slightly larger than those associated with the multilevel

analyses for all combinations of mediated effect value, number of groups, and group size. This indicates that the OLS analyses were, as expected, less efficient than the multilevel analyses. The table also shows that the OLS-based estimates of the standard error of the mediated effect were substantially downwardly biased. This bias was most apparent for the larger group size (20-30 observations), in which estimates were typically 35% to 40% smaller than the empirical standard error value. Downward biases of 18% to 20% were common in the smaller group size (5-10 observations) condition.

Overall, multilevel-based estimates of the standard error showed considerably less bias than the OLS-based estimates (although a notable exception occurred in the condition with small mediated-effect value, small number of groups, and small group size). The mean contrast between the OLS and multilevel estimates of the standard error was significant, $F(1, 87125) > 99999.99$, $p < .0001$, $\eta^2 = .35$. The effect of group size was also significant for this contrast, $F(1, 87125) = 37898.56$, $p < .0001$, $\eta^2 = .30$, indicating that the OLS underestimation of the standard error was more substantial when group size was large.

COMPARISONS AMONG MULTILEVEL ESTIMATES OF THE STANDARD ERROR OF THE MEDIATED EFFECT

Separate ANOVAs were performed on relative bias values for each of the three multilevel standard error estimators. No effects of group size, number of groups, true value of the mediated effect, or level of intraclass correlation were found for the first-order Taylor series estimator (all η^2s $< .01$). A significant main effect of true value of the mediated effect, $F(2, 89820) = 1485.00$, $p < .0001$, $\eta^2 = .03$, and a significant interaction between number of groups and the true value of the mediated effect, $F(10, 89820) = 142.14$, $p < .0001$, $\eta^2 = .02$, revealed that the second-order Taylor series estimator tended to overestimate the empirical standard error when the true value of the mediated effect was small, especially when the number of groups in the analysis was also small. Finally, examination of a weak three-way interaction among number of groups, group size, and the true value of the mediated effect, $F(10, 87667) = 42.97$, $p < .0001$, $\eta^2 = .005$, suggested that the unbiased estimator tended to underestimate the empirical standard error when the number of groups was small, except when small numbers of groups (10-30), small group sizes (5-10), and small values of the true mediated effect (.01) combined, resulting in overestimation.

These patterns of overestimation and underestimation have clear implications in the selection of the optimal estimator for use with mediated effects of

TABLE 5: **Empirical Standard Errors (*SE*) and Relative Bias in Estimated Standard Errors of the Mediated Effect For OLS and Multi-level Analyses by Number of Groups, Group Size, and True Value of the Mediated Effect (true med)**

Group Size	True Med		OLS Number of Groups						Multilevel Number of Groups					
			10	20	30	50	100	200	10	20	30	50	100	200
5 to 10	.01	Empirical *SE*	.0573	.0326	.0244	.0180	.0119	.0079	.0561	.0315	.0233	.0173	.0116	.0076
		First-order estimate (%)	−7.14	−10.93	−14.38	−17.78	−20.46	−19.53	7.52	4.25	3.50	0.32	−1.30	2.03
		Second-order estimate (%)	10.32	2.27	−3.52	−10.51	−16.42	−17.12	32.05	22.96	18.73	10.51	4.15	5.01
		Unbiased estimate (%)	−7.54	−13.24	−17.56	−21.30	−24.19	−21.46	7.92	1.30	−0.98	−4.94	−6.51	−0.95
	.09	Empirical *SE*	.1188	.0777	.0627	.0467	.0318	.0235	.1161	.0763	.0620	.0462	.0312	.0232
		First-order estimate (%)	−19.48	−20.56	−21.24	−20.55	−18.03	−22.28	−0.70	−1.19	−1.98	−0.92	3.69	−1.60
		Second-order estimate (%)	−14.79	−18.11	−19.62	−19.54	−17.51	−22.04	5.62	2.04	0.13	0.39	4.37	−1.27
		Unbiased estimate (%)	−22.89	−22.92	−22.90	−21.56	−18.55	−22.53	−5.46	−4.56	−4.15	−2.25	3.01	−1.92
	.25	Empirical *SE*	.1911	.1265	.1015	.0771	.0538	.0381	.1901	.1239	.1003	.0758	.0530	.0373
		First-order estimate (%)	−21.54	−21.10	−20.67	−19.76	−19.86	−20.25	−3.06	−0.58	−0.69	0.67	1.29	1.37
		Second-order estimate (%)	−19.79	−20.21	−20.07	−19.40	−19.68	−20.16	−0.81	−0.59	0.09	1.15	1.53	1.49
		Unbiased estimate (%)	−23.35	−22.01	−21.27	−20.13	−20.05	−20.34	−5.39	−1.78	−1.47	0.19	1.05	1.25

20 to 30	.01	Empirical *SE*	.0357	.0216	.0166	.0119	.0086	.0059	.0337	.0207	.0159	.0115	.0082	.0057
		First-order estimate (%)	−34.41	−36.82	−39.15	−38.59	−42.03	−41.70	−2.15	−1.10	−0.72	2.96	0.35	0.21
		Second-order estimate (%)	−28.47	−33.01	−36.27	−36.66	−41.09	−41.22	11.07	7.09	5.22	6.75	2.21	1.14
		Unbiased estimate (%)	−36.19	−39.31	−41.30	−40.40	−42.96	−42.19	−6.27	−6.36	−5.95	−1.04	−1.55	−0.73
	.09	Empirical *SE*	.0851	.0567	.0459	.0361	.0245	.0175	.0836	.0552	.0448	.0353	.0240	.0171
		First-order estimate (%)	−41.70	−41.75	−42.14	−43.74	−41.87	−42.68	−3.62	0.75	0.34	−2.08	1.49	0.36
		Second-order estimate (%)	−40.65	−41.21	−41.78	−43.53	−41.76	−42.63	−1.64	1.79	1.03	−1.67	1.70	0.46
		Unbiased estimate (%)	−42.78	−42.29	−42.51	−43.95	−41.98	−42.74	−5.66	−.30	−.36	−2.49	−1.28	−0.26
	.25	Empirical *SE*	.1383	.0974	.0776	.0590	.0412	.0291	.1348	.0951	.0759	.0579	.0402	.0283
		First-order estimate (%)	−42.49	−44.42	−43.63	−43.15	−42.72	−42.85	−1.32	−3.49	−2.04	−0.91	0.79	0.69
		Second-order estimate (%)	−42.10	.44.23	−43.50	−43.08	−42.68	−42.83	−0.59	−3.13	−1.79	−0.76	0.87	0.73
		Unbiased estimate (%)	−42.89	−44.61	−43.56	−43.23	−42.76	−42.87	−2.06	−3.85	−2.28	−1.06	0.72	0.66

NOTE: OLS = ordinary least squares.

different magnitudes. For the smallest mediated effect in this study ($ab = .01$), the first-order Taylor series or the unbiased estimator would be preferred, because the second-order Taylor series estimator tended to overestimate the standard error, often substantially. With a slightly larger true mediated effect ($ab = .09$), the first-order Taylor series estimator would be recommended when the number of groups and group size were small, because the second-order estimator still tended to overestimate, and the unbiased estimator tended to underestimate the standard error in these conditions. However, when both number of groups and group size were relatively large, the unbiased estimator had a slight advantage over the first-order Taylor series. For the largest mediated effect in this study ($ab = .25$), the second-order estimate of the standard error would be preferred when the number of groups in the analysis was small, because this estimate was somewhat less biased than the other two under these conditions. However, when the number of groups in the analysis was large (100 or more groups), the unbiased estimator was the least biased of the three. It should be noted, however, that the differences among the various estimators were quite small in these conditions.

In most cases, however, researchers who wish to use multilevel mediational analyses will not know in advance the size of the mediated effect they wish to test. Although number of groups, and to some extent, group size, are controllable design features, the true size of the mediated effect is a feature of the data-generating process and is not directly under the researcher's control. Moreover, a particular intervention study may involve a number of mediated effects of different sizes. Therefore, it may not be possible to base decisions about the best standard error estimator on the effect size specific information in Table 5. Therefore, Table 6 presents the relative bias in the multilevel-based first-order, second-order, and unbiased estimators of the standard error collapsed across the various values of the mediated effect. This table may be more useful in situations in which effect sizes are unknown or variable. Table 6 shows that the second-order estimator tended to overestimate the standard error, and this bias was quite substantial for conditions with small numbers of groups and small group sizes. For conditions with 50 or fewer groups, the first-order estimator had the least bias; for conditions with 100 groups or more, the unbiased estimator had the least bias.

ATLAS DATA

To illustrate the extent of the differences between single level and multi-level mediation analyses in an actual evaluation study, these methods were

TABLE 6:   Relative Bias in Multilevel Estimated Standard Errors of the Mediated
             Effect by Number of Groups and Group Size (%)

Group Size	SE Estimate	Number of Groups					
		10	20	30	50	100	200
5 to 10	First order	1.25	0.83	0.28	0.02	1.23	0.60
	Second order	12.29	8.53	6.32	4.01	3.35	1.74
	Unbiased	−1.86	−1.89	−2.26	−2.27	−0.80	−0.54
20 to 30	First order	−2.37	−1.28	−0.81	−0.01	0.88	0.42
	Second order	2.95	1.92	1.49	1.44	1.59	0.78
	Unbiased	−4.60	−3.46	−2.85	−1.53	0.15	0.06

applied to data from the evaluation trial of the ATLAS program. The ATLAS program is a team-based educational intervention program designed to ultimately decrease steroid use among high school football players. Thirty-one teams from different high schools were randomly assigned to program or control conditions. Details of the prevention program can be found in Goldberg et al. (1996). The program was designed to achieve its effects through changing a number of mediators believed to be related to the outcome measure. These targeted mediators included knowledge constructs, peer and non-peer influence measures, attitudes, and individual player characteristics. Subjects were measured prior to (pretest) and immediately following the 7-week intervention period (posttest).

Although the ultimate object of the ATLAS program was the reduction of steroid use, more immediate aims included lessening intentions to use steroids, lowering risk factors, increasing healthy behaviors and attitudes, and presenting alternatives to steroid use, including improved sports nutrition and strength training self-efficacy (Goldberg et al. 1996). The analyses reported here concentrate on the strength training self-efficacy outcome variable. This measure was constructed as the sum of six items (e.g., I know how to train with weights to get as strong and as quick as possible), each scored from 1 to 5, with lower values representing lower levels of this construct. Cronbach's α for this measure was .86 at pretest (MacKinnon, Goldberg, Lapin, et al. 1999).

The mediator variables included in these analyses were selected because preliminary analyses had shown them to be simultaneously significant ($p <$ .05) or marginally significant ($p < .10$) mediators of the relationship between the intervention program and strength training self-efficacy in a single level mediational analysis of the ATLAS data (MacKinnon, Goldberg, Clarke, et al. 1999). These nine mediators were,

1. knowledge of steroid effects (18 items, pretest α = .86),
2. perceived coach tolerance of steroid use (3 items, pretest α = .59),
3. peers as an information source (3 items, pretest α = .85),
4. team as an information source (3 items, pretest α = .75),
5. ability to turn down offers of drugs (4 items, pretest α = .88),
6. perceived severity of steroid effects (3 items, pretest α = .82),
7. reasons against using steroids (14 items, pretest α= .84),
8. perception of athletic competence (4 items, pretest α = .86), and
9. self-esteem (4 items, pretest α = .87).

Details of these measures are reported in Goldberg et al. (1996), MacKinnon, Goldberg, Clarke, et al. (1999), and MacKinnon, Goldberg, Lapin, et al. (1999).

Analyses were conducted estimating each of the three mediation equations in both the single level (OLS) and multilevel frameworks. Pretest measures of the mediators and the outcome variable were used as covariates in each of these three equations. The use of a common set of covariates across the three equations is necessary to preserve the $c - c' = ab$ equivalence in single level models. Using different sets of covariates in the various equations would confound the nonequivalence due to multilevel estimation techniques with that due to differences in covariate sets. Therefore, a common set of covariates, pretest level of the outcome variable and pretest measures of all mediators included in a model, were used in estimating each of the three meditation equations.

Two kinds of mediator models are reported here. First, each mediated relationship was examined in a separate analysis. These single mediator models were estimated for each of the nine mediators individually, using both OLS and multilevel estimation methods. Second, all nine mediators were combined into a multiple mediator model, which examined the entire set of mediated effects simultaneously, again using both OLS and multilevel estimation methods. In all models, standard error approximations were calculated using the first-order Taylor series expression.

SINGLE MEDIATOR MODELS

Table 7 presents the point estimates and standard errors of the mediated effect for both OLS and multilevel analyses. The first four columns of the table pertain to the single mediator models. For each of these nine models, the standard errors associated with the multilevel estimates of the mediated effect were larger than those associated with the OLS estimates. The multilevel standard errors were, on average, 33% larger, and the largest increase

TABLE 7: OLS and Multilevel Estimates and Standard Errors of ATLAS Mediated Effects in Single and Multiple Mediator Models

| | Single Mediator Models | | | | Multiple Mediator Model | | |
| | OLS | Multilevel | | | OLS | Multilevel | |
Mediator	$c - c_J = ab$ Estimate	$c - c_J$ Estimate	ab Estimate	% change in SE	ab Estimate	ab Estimate	% change in SE
Knowledge of steroid effects	.1576*** (.0227)	.1499*** (.0253)	.1479*** (.0253)	11.45	.0570*** (.0160)	.0553*** (.0161)	0.63
Perceived coach tolerance of steroid use	.1256*** (.0237)	.1218** (.0376)	.1182** (.0376)	58.65	.0318** (.0103)	.0301* (.0129)	25.24
Perceived severity of steroid effects	.1336*** (.0212)	.1256*** (.0282)	.1248*** (.0282)	33.02	.0472*** (.0128)	.0482*** (.0146)	14.06
Peers as an information source	.2305*** (.0276)	.2289*** (.0336)	.2272*** (.0336)	21.74	.1289*** (.0209)	.1351*** (.0239)	14.35
Team as an information source	.2414*** (.0296)	.2429*** (.0529)	.2407*** (.0529)	78.72	.0964*** (.0181)	.0995*** (.0277)	53.04
Perceived athletic competence	.1097*** (.0237)	.0967** (.0313)	.0976** (.0313)	32.07	.0340** (.0107)	.0336** (.0124)	15.89
Self-esteem	.0984*** (.0223)	.0982*** (.0286)	.0935** (.0286)	28.25	.0183* (.0080)	.0169 (.0087)	8.75
Ability to turn down offers of drugs	.0836*** (.0200)	.0853*** (.0248)	.0826*** (.0248)	24.00	.0155* (.0072)	.0149 (.0080)	11.11
Reasons against using steroids	.0447*** (.0123)	.0416** (.0137)	.0417** (.0137)	11.38	.0013 (.0059)	.0009 (.0057)	−3.39

NOTE: OLS = ordinary least squares; ATLAS = Adolescents Training and Learning to Avoid Steroids. Standard errors are in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$.

was 79% for the mediated effect involving team as an information source. Given the magnitude of the effects in these single mediator models, however, the substantial increases in standard errors had only minor effects on the significance tests of the mediated effects. For four of the mediator variables, perceived coach tolerance of steroid use, perceived athletic competence, self-esteem, and reasons against using steroids, mediated effects that were significant at the $p < .001$ level in OLS analyses were significant only at the $p < .01$ level in multilevel analyses. In all four of these cases, this minor change in significance level could be attributed to both a larger standard error and a smaller estimate of the mediated effect. For only one of these four mediator variables, reasons against using steroids, would the mediated effect still have had a lower significance level if the magnitude of the multilevel estimate had been equal to that of the OLS estimate.

MULTIPLE MEDIATOR MODEL

The last three columns of Table 7 pertain to the multiple mediator model, in which all nine mediated effects were examined simultaneously. No $c - c'$ estimates are provided in this section of the table. In a multiple mediator model, $c - c'$ estimates the total mediated effect for the model, rather than the individual contribution of any particular mediator. Therefore, only *ab* estimates from both OLS and multilevel analyses are reported. Each of the *ab* estimates represents the unique effect of a given mediator over and above those of the other eight mediators. Thus, these effects would be expected to be considerably smaller than those in the single mediator models. Indeed, the table shows that all nine effects were smaller in the multiple mediator model than in the single mediator models. However, with the exception of the reasons against using steroids mediator, all still significantly ($p < .05$) mediated the relationship between program and outcome according to the OLS analysis.

The standard errors associated with the mediated effects were again larger in the multilevel analyses than in the OLS analyses (except for the nonsignificant reasons against using steroids mediated effect, which was very slightly smaller), reflecting the downward bias in the OLS case shown in the simulation study. These differences were not quite as dramatic as those in the single mediator models: Multilevel standard errors were 0.63% to 53.04% larger than OLS standard errors, with an average increase of 15.52% across the nine mediators. However, the differences in significance levels and the inferences that could be made on the basis of these were more apparent in the multiple mediator model. The perceived coach tolerance of steroid use mediator was

significant at the .01 level in the OLS analysis but achieved only a .05 level of significance in the multilevel analysis. More tellingly, both the self-esteem and the ability to turn down offers of drugs mediators, significant at the .05 level in the OLS analysis, were nonsignificant in the multilevel analysis. Again, this change can be attributed both to larger multilevel standard errors and smaller multilevel estimates of the mediated effects. However, the change in the significance level of the perceived coach tolerance of steroid use variable and the nonsignificance of the ability to turn down offers of drugs variable would still hold even if the multilevel and OLS estimates were of equal magnitude.

DISCUSSION

Now let us readdress the questions that the simulation study and ATLAS example were designed to answer. The first of these concerned the discrepancy between the $c - c'$ and ab estimates of the mediated effect in multilevel models. That two different estimates of the same quantity are not algebraically equivalent may initially be unsettling, but the results of the simulation study show that on average, the discrepancy between the two is equal to zero. Moreover, the absolute value of the discrepancy becomes vanishingly small as the number of groups and group size increase, implying that for very large samples, the two estimates would be equivalent. This small sample inequality, then, seems unlikely to be problematic.

The second question asks whether there is any reason to prefer one of these two estimators over the other. Both the $c - c'$ and ab multilevel estimators provide good estimates of the mediated effect. Both estimators are essentially unbiased, as is the OLS estimator. Both multilevel estimators are slightly more efficient than the OLS estimator, and the ab multilevel estimate is slightly more efficient than the $c - c'$ estimate. However, both of these effects account for only a small portion of the variance in MSE. Rather than base a choice between the multilevel ab and $c - c'$ estimators on the nearly identical relative bias and efficiency results, a stronger argument might be made based on the usefulness of the estimators in different types of mediational models. Although both $c - c'$ and ab may be calculated as estimates of the mediated effect in a single mediator model, the two quantities are interpreted quite differently in a multilevel mediator model. In such a model, $c - c'$ estimates the total mediated effect, whereas ab estimates a single unique mediated effect. So, although ab estimates can be summed to provide an estimate of the total mediated effect, there is no way in which to apportion

the $c - c'$ estimate to determine the unique contributions of individual mediators. Therefore, the *ab* estimator provides a greater amount of information than the $c - c'$ estimator in models involving more than one mediator and generally may be preferred for that reason.

Comparisons between the standard error approximations calculated using OLS and multilevel estimates show a substantial downward bias for the OLS estimates and considerably less bias among the multilevel estimates. This bias in OLS estimates in grouped data is widely known and, in fact, is a primary rationale for multilevel modeling. Among the multilevel estimators, the second-order Taylor series approximation tends to overestimate the standard error, and the unbiased estimator tends to underestimate the standard error in many small sample conditions. Similar patterns are found in single level mediation studies as well (MacKinnon, Warsi, and Dwyer 1995), which would suggest that the biases are characteristic of small sample mediational analyses in general, rather than the result of multilevel estimation. Given these patterns, and in the absence of detailed information about the true value of the mediated effect, it is recommended that the multilevel first-order Taylor series approximation be used to estimate the standard error of the mediated effect in analyses with 50 groups or less. The unbiased estimator may provide a slight advantage in analyses with 100 groups or more.

The ATLAS example shows how different conclusions about the importance of particular mediators targeted by interventions can be reached, based on OLS and multilevel mediational analyses. The standard errors generated in the multilevel solution can be substantially larger than those from the OLS solution, and statistical theory and simulation results show that this difference is due to downward bias in the OLS standard errors. When the mediated effects are relatively small, as the unique effects of particular mediators may be in a multiple mediator model, using unbiased multilevel standard errors may result in making different inferences and in drawing different conclusions than an OLS analysis. The increased accuracy that results from correcting for the OLS bias in grouped data seems sufficient to justify the increased complexity in model specification and estimation involved in multilevel modeling.

REFERENCES

Barcikowski, R. S. 1981. Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics* 6:267-85.

Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

de Leeuw, J. 1992. Series editor's introduction to Hierarchical Linear Models. In *Hierarchical linear models: Applications and data analysis methods*, edited by A. S. Bryk and S. W. Raudenbush, xiii-xvi. Newbury Park, CA: Sage.

de Leeuw, J., and I. Kreft. 1986. Random coefficient methods for multilevel analysis. *Journal of Educational Statistics* 11:57-85.

DeVries, H., E. Backbier, M. Dijkstra, G. Van Brenkelen, G. Parcel, and G. Kok. 1994. A Dutch social influence smoking prevention approach for vocational school students. *Health Education Research* 9:365-74.

Donaldson, S. I., J. W. Graham, and W. B. Hansen. 1994. Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. *Journal of Behavioral Medicine* 17:1-22.

Goldberg, L., D. Elliot, G. N. Clarke, D. P. MacKinnon, E. Moe, L. Zoref, C. Green, S. L. Wolf, E. Greffrath, D. J. Miller, and A. Lapin. 1996. Effects of a multidimensional anabolic steroid prevention intervention: The Adolescents Training and Learning to Avoid Steroids (ATLAS) program. *Journal of the American Medical Association* 20:1555-62.

Goldstein, H. I. 1995. *Multilevel statistical models*. London: Edward Arnold.

Goodman, L. A. 1960. On the exact variance of products. *Journal of the American Statistical Association* 55:708-13.

Haggard, E. A. 1958. *Intraclass correlation and the analysis of variance*. New York: Dryden.

Hanushek, E. A., and J. E. Jackson. 1977. *Statistical methods for social scientists*. New York: Academic Press.

Hauck, W. W., and S. Anderson. 1984. A survey regarding the reporting of simulation studies. *American Statistician* 38:214-6.

Hedeker, D., R. D. Gibbons, and B. R. Flay. 1994. Random effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology* 62:757-65.

Judd, C. M., and D. A. Kenny. 1981. Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review* 5:602-19.

Kreft, I.G.G. 1994. Multilevel models for hierarchically nested data: Potential applications in substance abuse prevention research. In *Advances in Data Analysis for Prevention Intervention Research* (NIDA Research Monograph No. 142), edited by L. Collins and L. Seitz, 140-84. Rockville, MD: US Department of Health and Human Services.

Kreft, I.G.G. 1997. The interactive effect of alcohol prevention programs in high school classes: An illustration of item homogeneity scaling and multilevel analysis techniques. In *The Science of prevention: Methodological advances from alcohol and substance abuse research*, edited by K. J. Bryant, M. Windle, and S. G. West, 251-77. Washington, DC: American Psychological Association.

Kreft, I., and J. de Leeuw. 1998. *Introducing multilevel modeling*. London: Sage.

Longford, N. 1993. *Random coefficient models*. Oxford, UK: Clarendon.

MacKinnon, D. P., and J. H. Dwyer. 1993. Estimating mediated effects in prevention studies. *Evaluation Review* 17:144-58.

MacKinnon, D. P., L. Goldberg, G. N. Clarke, D. L. Elliot, J. Cheong, A. Lapin, and J. L. Krull. 1999. Mediating mechanisms in a program to reduce intentions to use steroids, improve exercise self-efficacy, and dietary behavior. Manuscript submitted for publication.

MacKinnon, D. P., L. Goldberg, A. Lapin, G. N. Clarke, D. L. Elliot, and E. Moe. 1999. Psychometric properties of an adolescent health behavior and anabolic steroid questionnaire: The

Adolescents Training and Learning to Avoid Steroids (ATLAS) Project. Manuscript in preparation.

MacKinnon, D. P., C. A. Johnson, M. A. Pentz, J. H. Dwyer, W. B. Hansen, B. R. Flay, and E. Y. Wang. 1991. Mediating mechanisms in a school-based drug prevention program: First-year effects of the Midwestern Prevention Project. *Health Psychology* 10:164-72.

MacKinnon, D. P., G. Warsi, and J. H. Dwyer. 1995. A simulation study of mediated effect measures. *Multivariate Behavioral Research* 30:41-62.

Mood, A., F. A. Graybill, and D. C. Boes. 1974. *Introduction to the theory of statistics*. New York: McGraw-Hill.

Moulton, B. R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32:385-97.

Murray, D. M., C. L. Perry, G. Griffin, K. C. Harty, D. R. Jacobs, L. Schmid, K. Daly, and U. Pallonen. 1992. Results from a statewide approach to adolescent tobacco use prevention. *Preventive Medicine* 21:449-72.

Pedhazur, E. J. 1982. *Multiple regression in behavioral research: Explanation and prediction*. Fort Worth, TX: Holt, Rinehart & Winston.

Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15:351-7.

Scariano, S. M., and J. M. Davenport. 1987. The effects of violations of independence assumptions in one-way ANOVA. *The American Statistician* 41:123-9.

Scott, A. J., and D. Holt. 1982. The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association* 77:848-54.

Seltzer, M. H. 1994. Studying variation in program success: A multilevel modeling approach. *Evaluation Review* 18:342-61.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. In *Sociological methodology*, edited by S. Leinhardt, 290-312. Washington, DC: American Sociological Association.

Sobel, M. E. 1986. Some new results on indirect effects and their standard errors in covariance structure models. *Sociological methodology*, edited by N. Tuma, 159-86. Washington, DC: American Sociological Association.

Tabachnick, B. G., and L. S. Fidell. 1989. *Using multivariate statistics*, 2d ed. New York: HarperCollins.

Vinokur, A. D., and Y. Schul. 1997. Mastery and inoculation against setbacks as active ingredients in the JOBS intervention for the unemployed. *Journal of Consulting and Clinical Psychology* 65:867-77.

Walsh, J. E. 1947. Concerning the effect of the intraclass correlation on certain significance tests. *Annals of Mathematical Statistics* 18:88-96.

Jennifer L. Krull is completing postdoctoral work in the Program for Prevention Research at Arizona State University. She has accepted an assistant professor position at the University of Missouri–Columbia to begin Fall 1999. Her current research interests center on the extension and application of multilevel modeling techniques to psychological data.

David P. MacKinnon is an associate professor in the psychology department at Arizona State University. His interests include statistical methods in prevention research and applied psychological research.