Psychological Methods
2001, Vol. 6, No. 2, 115–134

1082-989X/01/$5.00 DOI: 10.1037//1082-989X.6.2.115

Estimating and Testing Mediation and Moderation in Within-Subject Designs

Charles M. Judd
University of Colorado at Boulder

David A. Kenny
University of Connecticut

Gary H. McClelland
University of Colorado at Boulder

Analyses designed to detect mediation and moderation of treatment effects are increasingly prevalent in research in psychology. The mediation question concerns the processes that produce a treatment effect. The moderation question concerns factors that affect the magnitude of that effect. Although analytic procedures have been reasonably well worked out in the case in which the treatment varies between participants, no systematic procedures for examining mediation and moderation have been developed in the case in which the treatment varies within participants. The authors present an analytic approach to these issues using ordinary least squares estimation.

The issues of mediation and moderation have received considerable attention in recent years in both basic and applied research (Baron & Kenny, 1986; James & Brett, 1984; Judd & Kenny, 1981b; MacKinnon & Dwyer, 1993). In addition to knowing whether a particular intervention has an effect, the researcher typically wants to know about factors that affect the magnitude of that effect (i.e., moderation) and mechanisms that produce the effect (i.e., mediation). Such knowledge helps in both theory development and intervention application.

To illustrate the difference between mediation and moderation, consider a design in which a researcher is interested in whether students who are taught with a new curriculum (the treatment condition) show higher performance on a subsequent standardized test than students taught under the old curriculum (the control condition). Assuming that a performance difference is found, one might plausibly hypothesize different mediating mechanisms for this effect. The new curriculum might increase students' interest in the subject matter; it might cause students to study harder outside of class; or it might convey the material more clearly. These are alternative reasons why the performance difference is found, that is, alternative mediators of the treatment effect. The researcher might also be interested in factors that affect the magnitude of the difference between performance following the old curriculum and performance following the new one. That difference might be larger or smaller for different types of students or in different types of classrooms or when taught by different kinds of teachers. All of these then are potential moderators of the treatment effect.

It is possible that the same variable may serve as both a mediator and a moderator. For instance, study time might serve both roles. First, as a mediator, the new curriculum might lead to higher performance because it causes students to study more. Second, as a moderator, the treatment might be especially effective for students who spend more time studying.

Procedures for assessing mediation and moderation have been relatively well worked out through ordinary least squares regression and analysis of variance procedures. Mediation is assessed through a four-step procedure (Baron & Kenny, 1986; Judd & Kenny,

Charles M. Judd and Gary H. McClelland, Department of Psychology, University of Colorado at Boulder; David A. Kenny, Department of Psychology, University of Connecticut.

Preparation of this article was partially supported by National Institute of Mental Health Grants R01 MH45049 and R01 MH51964.

Correspondence concerning this article should be addressed to Charles M. Judd, Department of Psychology, University of Colorado, Boulder, Colorado 80309-0345. Electronic mail may be sent to charles.judd@colorado.edu.

1981b) that involves testing whether the treatment effect is reduced in magnitude when the mediator is controlled. The test of moderation involves a test of the statistical interaction between the treatment variable and the moderator, examining whether a variable that is a product of the two predicts the outcome variable over and above the treatment and moderator (Aiken & West, 1991; McClelland & Judd, 1993).

These analytic procedures, however, apply only to the situation in which the treatment variable varies between experimental units, with some units or participants receiving one treatment and others receiving another. However, treatment comparisons often involve within-subject comparisons, examining every experimental unit or participant in each treatment condition. For example, as a means of evaluating the effects of mood differences on judgment, the same participants might be observed both in a positive mood condition and in a negative one. Or to evaluate a new remedy for chronic migraine headaches, the same participant might one day take a placebo pill and the next day the new remedy pill (or vice versa). A design in which the independent variable varies within participants rather than between them is particularly useful when the effects of the independent variable are relatively transitory and the independent variable is easily varied. Such a design can often result in a dramatic increase in statistical power relative to between-subjects designs, because variance due to stable individual differences contributes to the error variance in testing between-subjects effects but not to the error variance in within-subject analyses. As a result, within-subject designs often involve smaller samples than those typically used in between-subjects designs. Surprisingly, there is very little work that has formally considered issues of mediation and moderation in within-subject designs (for an exception, see Judd, McClelland, & Smith, 1996).

In the analysis of variance literature, a variety of authors have dealt with the use of covariates in within-subject designs (for instance, Bock, 1975; Huitema, 1980; Khattree & Naik, 1995; Myers, 1979; Winer, Brown, & Michels, 1991), and some of this literature is relevant.[1] But this literature has not explicitly addressed the conditions under which covariates or, more generally, concomitant variables may be considered to be either mediators or moderators of a within-subject treatment effect. In addition, this literature has largely confined itself to concomitant variables that are measured only once (i.e., between-subject variables in within-subject designs) rather than dealing with concomitant variables that vary within participants (but see Khattree & Naik, 1995). Certainly, recent literature on multilevel modeling and hierarchical linear modeling (Bryk & Raudenbush, 1992; Kenny, Kashy, & Bolger, 1998; Snijders & Bosker, 1999) has considered such time-varying concomitant variables, but again this literature has not explicitly addressed the conditions under which these variables serve mediating and moderating roles.

Interestingly, in spite of an absence of a formal literature on within-subject mediation and moderation, there are several empirical articles that have attempted to evaluate mediational models with within-subject treatment variables (e.g., Fiske, Kenny, & Taylor, 1982; Neuberg, 1989). In general, these articles have explored mediation by analyzing difference scores in ordinary least squares regression models. But there has been no formal work that has developed the rationale for the use of such an approach.

The purpose of this article is to provide a formal basis for estimating and testing mediation and moderation effects in within-subject designs, wherein the observations under one level of the treatment variable are not independent of observations under the other level(s).[2] Although much of what we present could be developed within the context of multilevel modeling or hierarchical linear modeling procedures, we made the decision to explore these issues in the context of ordinary least squares regression for a number of reasons. First, many researchers who are interested in questions of mediation and moderation are likely to be much more conversant with analysis of variance and

[1] Traditionally, in the analysis of variance literature, covariates have been seen as variables assumed to be uncorrelated with the between-subjects treatment variable. They are included in an analysis of covariance to increase statistical power. Because this assumption is not made for potential mediators and moderators, we refer to them as concomitant variables rather than as covariates.

[2] Although we generally refer to the designs we are considering as within-subject designs, in fact the procedures that we develop are widely applicable whenever observations are linked or nonindependent of each other because they come from the same participant, the same family, the same classroom, the same setting, or whatever induces dependence. In general, the designs considered are designs in which the factor that causes the dependence, and under which observations are nested, is crossed with the independent or treatment variable of interest.

regression procedures that use ordinary least squares estimation than with multilevel modeling procedures. Second, within-subject designs typically involve relatively few participants. As a result, the large-sample assumptions underlying some estimation procedures used in multilevel modeling may be untenable. Third, the detailed examination of the within-subject mediation and moderation models we offer holds independent of what statistical methods are used to estimate the parameters of those models. The analyses that we outline involve the use of difference scores in regression models, thus providing a more formal basis for the procedures that researchers have occasionally used on an ad hoc basis.

The within-subject case requires us to address a set of issues that are typically not addressed in the between-subject case. Doing so will make clear certain assumptions underlying the definition and assessment of mediation and moderation, both in within-subject and between-subject designs.

The Basic Model

We start with a relatively simple research design in which there are two treatment conditions being compared. At a later point, we generalize our approach to more complex designs involving multiple levels of a within-subject treatment variable. In the two-condition case, each unit (e.g., participant, respondent, or case) is observed twice, once in each condition, and these observations are denoted Y_1 and Y_2.[3] The treatment effect is equal to the difference in the population between the mean of Y_1 and the mean of Y_2. We assume, for now, that there are no confounding factors between the observations, so the difference between the means is an unbiased estimate of the true treatment effect. Accordingly, following the model of causal inference developed by Rubin (1974, 1978) and Holland (1986, 1988; see also West, Biesanz, & Pitts, 2000), we assume both temporal stability (i.e., the value of an observation does not depend on when the treatment is delivered) and causal transience (i.e., the effect of a treatment or of a prior measurement does not persist over time). These are strong assumptions that may not be met in many applications of within-subject designs in psychological research.[4]

We occasionally illustrate our discussion by referring to a hypothetical research design involving a drug versus placebo treatment comparison. In the design, participants who chronically suffer from migraine headaches are given an experimental drug one day and a placebo pill another. The variables Y_1 and Y_2 are measures of pain taken after the administration of the placebo or drug. In a later section, we present hypothetical data from such a design and provide illustrative analyses of these data.

We start by considering the simple case without a potential mediator or moderator. We present more detail here than necessary to provide a basis for later generalization. Then we introduce a concomitant variable that may serve as a mediator or moderator. We first consider the case of a stable concomitant variable, assumed to represent a stable individual-differences attribute or other variable that varies between units but not across treatment levels. Finally, we consider the case of a varying concomitant variable, assumed to vary within units across treatment levels as well as between units.

Although we discuss estimation, in general we define models in terms of population parameters. These are designated with Greek letters, whereas their estimates are in the Roman alphabet.

Case 1: No Concomitant Variable

Treatment effects in the case of between-subjects experimental designs are estimated by asking whether the mean value of Y differs between units exposed to one treatment and those exposed to the other. In the within-subject case, however, we have Y_1 and Y_2 measured on each unit, because each unit is observed in each treatment condition. Formally, each response is modeled as

$$\begin{aligned} Y_{1i} &= \alpha_{10} + \varepsilon_{1i}, \\ Y_{2i} &= \alpha_{20} + \varepsilon_{2i}. \end{aligned} \tag{1}$$

In these equations, Y_{1i} and Y_{2i} are the Y measurements in each treatment condition for the ith participant. The αs are expected values of each Y, and the εs are errors or residuals for each observed Y value. According to these equations, we are assuming that all observations within a treatment condition are equal, except for er-

[3] In the text, we exclude the subscripts that designate unit or participant. These are given in the equations, however.

[4] In many within-subject designs in psychology, researchers may use a Latin square design, counterbalancing the order of treatment exposure across participants. This is commonly thought to yield unbiased estimates of the treatment effect when temporal stability cannot be assumed. At a later point in this article, we examine the implications of relaxing this temporal stability assumption in greater detail.

ror, to a constant. In other words, the true scores are all equal within each condition. We assume that the errors or residuals (εs) share a multivariate normal distribution, each with a mean of zero and a common variance, and, given dependence, a nonzero covariance between residuals from the same unit (typically positive).

If $\alpha_{10} = \alpha_{20}$, then there is no treatment difference on average, although there may be differences between the Ys for individual units as a result of ε_{1i} not equaling ε_{2i}. Assuming that $\alpha_{10} \neq \alpha_{20}$, the treatment difference for each unit equals the difference between the preceding two equations:

$$Y_{Di} = Y_{2i} - Y_{1i} = (\alpha_{20} - \alpha_{10}) + (\varepsilon_{2i} - \varepsilon_{1i}). \quad (2)$$

It is assumed that the treatment difference, except for error, is constant for all units. The mean difference

$$\mu_{YD} = \mu_{Y2} - \mu_{Y1} = \alpha_{20} - \alpha_{10} \quad (3)$$

is the treatment effect. In a sample, it is estimated as the difference between the sample mean of Y_{1i} and the sample mean of Y_{2i}, both computed across the participants in the sample.

A test of whether this mean treatment effect differs from zero is provided equivalently by a paired- or dependent-samples t test, a one-factor repeated measures analysis of variance, or a single-sample t test testing whether the mean difference score differs from zero. In our example, the mean pain difference between the two experimental conditions estimates the overall treatment effect, and each participant is assumed to manifest this same pain difference, except for error.

Case 2: A Stable Concomitant Variable

Instead of assuming that the true scores are constant within each conditions, it may be more realistic to assume that they are related to one or more concomitant variables. We first consider the case of a single stable concomitant variable, X, assumed to be an individual-differences attribute or some other variable that does not vary over the time frame of the experiment. In the context of our example, this variable might be the participant's age or gender.

The Ys are now modeled as a function of this concomitant variable:

$$\begin{aligned} Y_{1i} &= \beta_{10} + \beta_{11}X_i + \varepsilon_{1i}, \\ Y_{2i} &= \beta_{20} + \beta_{21}X_i + \varepsilon_{2i}. \end{aligned} \quad (4)$$

In other words, the X variable contributes to some extent to each individual's two Y scores. We consider two cases depending on whether the two within-condition slopes are equal to each other, that is, whether or not $\beta_{11} = \beta_{21}$.

Parallel functions. If $\beta_{11} = \beta_{21}$ in Equation 4, then the two linear functions for the Ys, one in each of the two conditions, are parallel. This situation is depicted in Figure 1 with two participants, P and Q. P_1 and Q_1 are the data points for these participants in Condition 1, and P_2 and Q_2 are their data points in Condition 2. Accordingly, in this and subsequent figures, both Y_1 and Y_2 are graphed on the common ordinate axis. Because each participant has only a single X value, both data points for a given participant have exactly the same location on the x-axis. In addition, with only two data points in each condition, we assume no error, so that all data points lie exactly on the two linear functions. The bottom line, labeled Y_1:X, represents the linear function in Condition 1; the Y_2:X line is the function in Condition 2. The treatment effect for each participant is then the vertical distance between the two data points for each participant, represented by the two vertical dashed lines: between P_1 and P_2 and between Q_1 and Q_2. Importantly, because of the equal slopes in the two conditions, these two within-subject treatment effects are equal to each other. With X centered in the figure so that the mean X score for the two participants equals zero, the solid vertical line represents the treatment effect on aver-

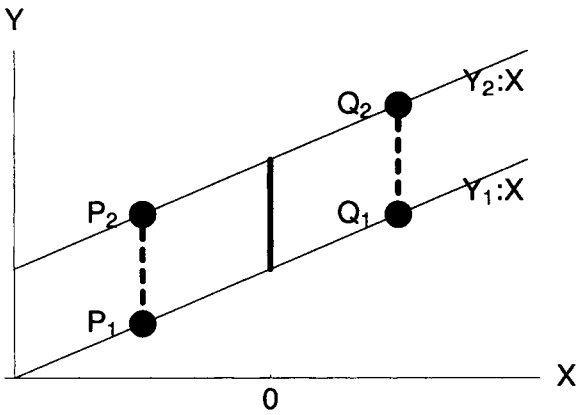

Figure 1. Treatment effect (dashed vertical line) unrelated to a stable concomitant variable. P and Q are two units measured on the concomitant X variable and on Y_1 and Y_2 in Treatments 1 and 2. The treatment effect (solid vertical line) is the distance between the two lines relating the two Y scores, respectively, to the unvarying concomitant variable X. The parallel slopes indicate a constant treatment effect that does not depend on X.

age. And note once again that this is equal to the vertical distance between the two linear functions.

Nonparallel functions. Now consider what happens if the two within-condition slopes are not equal to each other, that is, $\beta_{11} \neq \beta_{21}$. This situation, again with two participants, *P* and *Q*, is depicted in Figure 2. Here the slope of the Y_2:*X* linear function is more positive than the slope for the Y_1:*X* linear function. In other words, *X* relates to *Y* in Condition 2 more strongly than it does in Condition 1. Again, the treatment difference for each participant is represented by the dashed vertical line between the two *Y*:*X* functions, at that participant's particular *X* value. Given the slopes in the figure, participant *P* has a smaller *X* value than participant *Q*, and so the treatment difference for *P* is smaller than that for *Q*. Again, with *X* centered, the mean *X* equals zero, and the average treatment difference is represented by the solid vertical line at this point on the *x*-axis. It necessarily equals the average of the two within-subject treatment differences.

Figures 1 and 2 differ in that the within-condition slopes are equal in the former and unequal in the latter. The situation in Figure 2 represents moderation of the within-subject treatment effect. With unequal within-condition slopes, the relationship between *Y* and *X* depends on condition. Equivalently, the magnitude of the treatment effect—that is, the vertical distance between the two functions—depends on the value of *X*.

Algebraically, the situation is captured by taking the difference between the two within-condition functions in Equation 4. For each unit, the difference between the two observations, one in Condition 2 and one in Condition 1, equals

$$Y_{Di} = Y_{2i} - Y_{1i} = (\beta_{20} - \beta_{10}) + (\beta_{21} - \beta_{11})X_i + (\varepsilon_{2i} - \varepsilon_{1i}), \quad (5)$$

which represents the treatment effect for each unit (plus error). There are two components to this treatment effect, one that is common to all units, $\beta_{20} - \beta_{10}$, and one that varies between units, $(\beta_{21} - \beta_{11})X_i$. If the two within-condition slopes are equal to each other, $\beta_{11} = \beta_{21}$, then the second component of the treatment effect equals zero. Hence, the treatment difference is the same for all units. On the other hand, if the two within-condition slopes are unequal, then the treatment effect depends on the value of X_i. This defines moderation. The magnitude of the treatment effect depends on X_i. Equivalently, the *Y*:*X* slopes are different in the two conditions.

Estimation. Moderation due to a stable *X* is estimated by regressing the *Y* difference on *X*, thus estimating the model in Equation 5:

$$Y_{Di} = b_0 + b_1X_i + e_{YDi}. \quad (6)$$

The slope for *X* in this linear regression will exactly equal the difference in the two slopes if separate regression models are estimated within each condition. Accordingly, a test of whether this slope differs from zero provides a statistical test of moderation. In other words, moderation of the within-subject treatment difference due to a stable *X* variable is indicated if *X* predicts the *Y* difference.

For ease of interpretation, it is helpful in estimation to center *X* by deviating each unit's *X* score from the mean of *X*, as we did in Figures 1 and 2. If *X* has been centered, then the estimated *Y* difference is given by

$$Y_{Di} = b_0 + b_1(X_i - \bar{X}) + e_{YDi}, \quad (7)$$

and the estimated mean difference equals

$$\bar{Y}_D = \bar{Y}_2 - \bar{Y}_1 = b_0 + b_1(\bar{X} - \bar{X}) \quad (8)$$

or

$$\bar{Y}_D = \bar{Y}_2 - \bar{Y}_1 = b_0, \quad (9)$$

which equals the estimated treatment effect for a unit with an average value of *X*. Thus, with centered *X*, the estimated intercept equals the average treatment effect.[5]

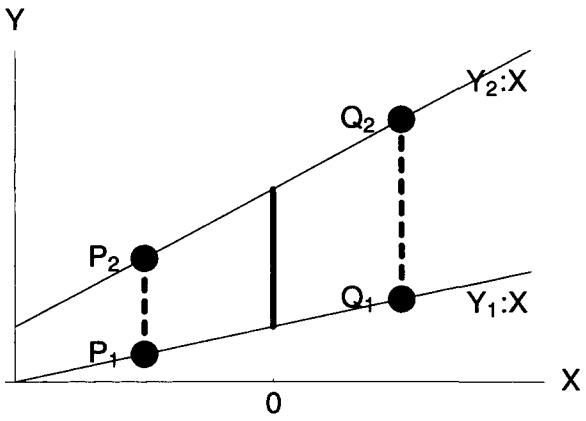

Figure 2. Treatment effect (dashed vertical line) moderated by a stable concomitant variable. *P* and *Q* and the regression lines are as in Figure 1. The nonparallel slopes indicate a varying treatment effect that depends on the concomitant variable *X*.

[5] At first, there may appear to be an inconsistency between the −1, 1 coding for the within-subject difference here

In the between-subjects case, moderation is tested by computing a product term between a dichotomous treatment variable and the moderator and testing whether that product variable predicts the outcome once both the treatment variable and the moderator have been partialed. In other words, in the between-subjects case, moderation is shown by the familiar practice of testing whether an interaction, computed as a partialed product variable, predicts the outcome. Given this context, moderation in the within-subject case may seem a bit strange, because we are not forming a product variable and so one might wonder whether an interaction is being tested or not. In fact, the model given in Equation 5 does imply an interaction, because X predicts the treatment difference for each unit only if $\beta_{21} \neq \beta_{11}$, that is, only if the effect of X on Y depends on whether the observation is in Treatment Condition 1 or Treatment Condition 2.

In the context of our example, suppose the concomitant variable was the participant's age. If the within-condition relationships between pain judgments and age are different, then moderation of the within-subject treatment effect by age is indicated. An assessment of whether such moderation is present would be provided by testing whether age is a predictor of each participant's difference between his or her pain ratings in the two conditions. If age does predict the pain difference, then the magnitude of the treatment difference depends on the participant's age. Equivalently, the relationship between pain and age is different in the two conditions.

Case 3: A Varying Concomitant Variable

We now allow there to be variation in X within participants across the two treatment conditions. Just as we have two measures of Y for each participant (Y_1 in one treatment condition and Y_2 in the other), we now have X_1 in the first treatment condition and X_2 in the other, and these X scores, measured after the administration of each treatment, may be influenced by the treatment variable. As we did for Y, we make the strong assumptions from Rubin and Holland of temporal stability and causal transience in these X measures. Thus, any mean difference in X between the two treatment conditions is assumed to be a causal effect of treatment. In addition, we assume that the X variables are causally prior to the Y variables in each treatment condition.

Just as we did in the case of the Y variables, we can develop a model for the X variables:

$$\begin{aligned} X_{1i} &= \lambda_{10} + \varepsilon_{1i}, \\ X_{2i} &= \lambda_{20} + \varepsilon_{2i}. \end{aligned} \tag{10}$$

From these equations, we can derive the treatment difference for each unit:

$$X_{Di} = X_{2i} - X_{1i} = (\lambda_{20} - \lambda_{10}) + (\varepsilon_{2i} - \varepsilon_{1i}). \tag{11}$$

And the mean treatment effect in X equals

$$\mu_{XD} = \mu_{X2} - \mu_{X1} = \lambda_{20} - \lambda_{10}. \tag{12}$$

In the context of our example, if the Y variables are measures of subjective pain on each day, following the administration of the experimental drug or the placebo, X might be the degree to which a certain hormone, presumed to be released by the administration of the experimental drug, is found in the blood system on each of the 2 days.

Each of the Y variables can be modeled as a function of the X variable measured in the same conditions:[6]

$$\begin{aligned} Y_{1i} &= \delta_{10} + \delta_{11} X_{1i} + \varepsilon_{1i}, \\ Y_{2i} &= \delta_{20} + \delta_{22} X_{2i} + \varepsilon_{2i}. \end{aligned} \tag{13}$$

These two equations are very similar in form to those we presented for the case of the stable concomitant

and the −.5 and .5 contrast coding recommended by Judd, McClelland, and Culhane (1995) and West, Aiken, and Krull (1996) to estimate the treatment difference in the between-subjects case. However, any apparent inconsistency is entirely due to the difference between slope and intercept estimates. If we were to compute a regression equation for each participant to estimate each participant's treatment effect as a slope, it would equal $(\sum \lambda_j Y_j)/\sum \lambda_j^2 = (-.5Y_1 + .5Y_2)/(.25 + .25) = Y_2 - Y_1$. This of course parallels the between-subjects regression for which the slope, using the −.5 and .5 codes, equals $\bar{Y}_2 - \bar{Y}_1$. Note also that the intercept in a within-subject regression would equal the average of the two observations.

[6] As already stated, we are assuming that X in each condition is causally prior to Y in each condition. Thus, in modeling the X effects, we did not include Y variables as predictors. However, we do include X in modeling the Y variables. In addition, because of our assumptions of temporal stability and causal transience, we can assume that in the population the X in one condition exerts no causal effect on the Y in the other. In other words, each Y is a function of the X from that same condition alone and is uninfluenced by the X measured in the other condition. At a later point at which we consider additional complexities, we examine designs in which this assumption cannot reasonably be made.

variable, the difference being that the X predictor variable varies between the two treatment conditions; that is, there are two Xs, not one. Again, interpretation depends on whether the within-condition slopes are equal, that is, whether $\delta_{11} = \delta_{22}$. We consider each case in turn.

Parallel functions. If $\delta_{11} = \delta_{22}$, then the two linear functions in Conditions 1 and 2 are parallel. Figure 3 presents a graph of these two parallel slopes. The top sloping line is the function in Condition 2 relating Y_2 to X_2. The lower line graphs the function in Condition 1. As in the earlier two figures, both Ys are plotted on the common ordinate. Unlike the earlier figures, where there was only one X, now both Xs are plotted on the common abscissa. Again, the important assumption in this figure is that the two slopes are identical, that is, $\delta_{11} = \delta_{22}$.

We have plotted the two observations from a single unit in Figure 3, participant Q, with observation Q_1 from the first condition and observation Q_2 from the second. Both observations lie perfectly on the two functions; accordingly, for purposes of illustration, we are assuming that their Y scores are measured without error.[7] Note that the two observations from this unit not only are at two different locations on the y-axis or ordinate axis (corresponding to their two different Y scores), but they also occupy two different locations on the x-axis, because we are not presuming that X_{1i} equals X_{2i}. The total treatment effect for this unit equals the difference between Y_{2i} and Y_{1i} (the vertical dashed line). But note that these two different Y values, one in Condition 1 and one in Condition 2, are associated with different X values (and the difference between the two Xs is represented by the horizontal dashed line). Part of the difference in the two Y values for this participant derives from the fact that higher X values in both conditions lead to higher Y values, and the X value for the observation in Condition 1 is lower than the X value for the observation from this unit in Condition 2. In other words, of the total treatment difference between Y_{2i} and Y_{1i} for this unit, part of it is due to the X difference between the two conditions, and part of it is a treatment difference over and above that. The treatment difference that persists over and above the X difference in the two conditions is indicated by the solid vertical line between the two lines, located above the participant's mean X value.

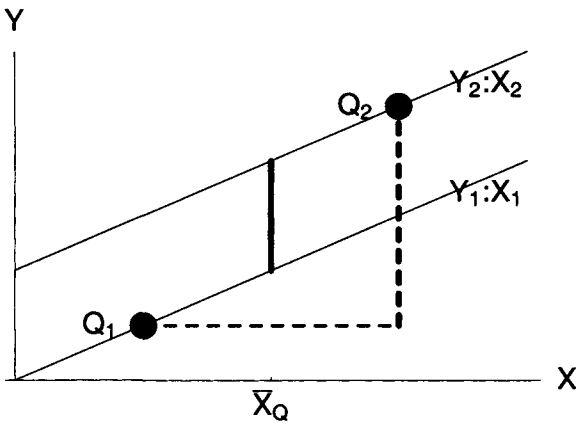

Figure 3. Treatment effect adjusted within units by changes in the concomitant variable. Q is a unit measured twice on both the outcome variable Y and the concomitant variable X after Treatments 1 and 2. The total treatment effect is the difference between Y_1 and Y_2 (the dashed vertical line). The adjusted treatment effect (the solid vertical line) is the distance between the two lines relating each Y score, respectively, to its concomitant X variable at the mean value of X for that unit.

The Y difference for any unit (e.g., unit Q in Figure 3) equals the difference between the two functions in Equation 13:

$$Y_{Di} = Y_{2i} - Y_{1i} = (\delta_{20} - \delta_{10}) + \delta_{22}X_{2i} - \delta_{11}X_{1i} + (\varepsilon_{2i} - \varepsilon_{1i}). \quad (14)$$

Assuming parallel slopes in the two conditions—that is, $\delta_{11} = \delta_{22}$—this reduces to

$$Y_{Di} = Y_{2i} - Y_{1i} = (\delta_{20} - \delta_{10}) + \delta_{11}(X_{2i} - X_{1i}) + (\varepsilon_{2i} - \varepsilon_{1i}), \quad (15)$$

where δ_{11} is the common slope in the two conditions. The total treatment effect is then equal to the sum of two different components: $(\delta_{20} - \delta_{10})$ and δ_{11} $(X_{2i} - X_{1i})$. The first of these is the vertical distance between the two lines in the figure, at the mean level of X for the unit. The second of these two components represents the part of the treatment difference in the Ys that is attributable to the X difference for that unit. In essence, the vertical distance between the two lines, at the value of the mean X for the unit, represents the treatment effect in the Ys over and above or adjusting for the X difference between the two conditions for that unit. It represents an adjusted treatment difference for the unit, adjusting for the within-unit differences between the two Xs.

[7] This is only an assumption of convenience to better illustrate the situation for this particular unit in the figure. In general, there is assumed to be residual variation in all observations.

With multiple units, interpretation depends on whether or not there is a treatment effect for X on average. For simplicity, we first consider the case of no X treatment effect. Figure 4 presents two different participants, P and Q. As in Figure 3, unit Q has a positive $X_{2i} - X_{1i}$ difference. Given the positive relationship between X and Y in the two conditions, this means that Qs adjusted treatment difference, over and above his X difference, is smaller than his total or unadjusted treatment difference. That is, the vertical distance between the two lines at his mean X value is smaller than the total difference between his Y values. On the other hand, the adjustment for participant P runs in exactly the opposite direction; her $X_{2i} - X_{1i}$ difference is negative, whereas her total treatment effect, $Y_{2i} - Y_{1i}$, is slightly positive. As a result, participant P's adjusted treatment difference is larger than the unadjusted difference.

For there to be no average treatment effect in X, the degree of the adjustments must be exactly the same, except for sign, for P and Q. That is, the $X_{2i} - X_{1i}$ differences are equal, but of opposite signs, for the two. Hence, the net adjustment collapsing across both units is exactly zero. Accordingly, the mean X-adjusted treatment effect, portrayed by the length of the solid vertical line between the two functions at the centered value of zero on the X axis, exactly equals the mean of the two unadjusted treatment effects for the two units. Although there is individual adjustment

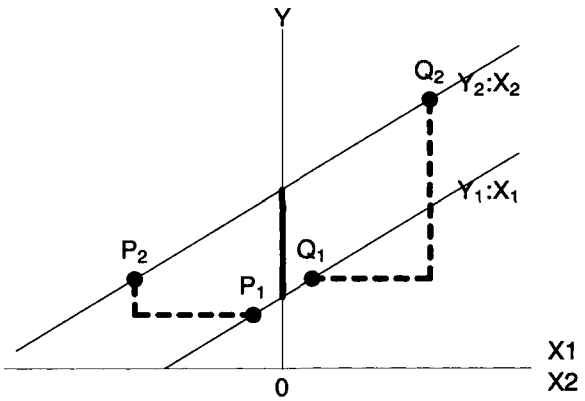

Figure 4. Treatment effect adjusted within units by changes in the concomitant X without between-units mediation. P and Q are units measured on both X and Y in Treatments 1 and 2. On average across the two units, the means of X_1 and X_2 are equal; they both can be centered at zero without loss of generality. The average treatment effect adjusted within-unit (the solid vertical line) equals the average of the two unadjusted treatment effects (the two dashed vertical lines).

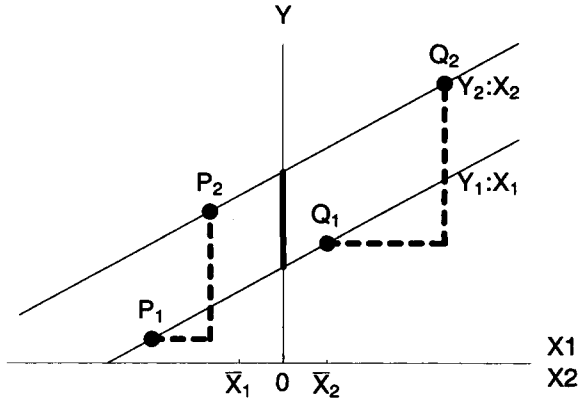

Figure 5. Treatment effect adjusted within units by changes in the concomitant X with between-units mediation. P and Q, the regression lines, and the within-unit adjustments are as in Figure 4. The difference in the means of X_1 and X_2 implies that the treatment effect adjusting for the average change in the concomitant variable (the solid vertical line) is smaller than the average of the two unadjusted treatment effects (the two dashed vertical lines).

of the Y treatment difference within each unit, there is no aggregate adjustment in the magnitude of the treatment difference when X is controlled.

Finally, consider Figure 5, again with two units, P and Q. In this case, both units have a positive $X_{2i} - X_{1i}$ difference. Hence, on average across the two units, there is a treatment difference in X as well as in Y. Accordingly, the two $\bar{X}$s are at different locations on the joint abscissa (which has been centered at their joint mean). Given a positive relationship between X and Y in each condition, the adjustment in the magnitude of the treatment effect is such that the total treatment effect for each unit, $Y_{2i} - Y_{1i}$, is larger than the treatment effect adjusted for the X difference. As a result, the average adjusted treatment difference, represented by the solid vertical line at the joint mean of the two Xs, is smaller than the average unadjusted treatment effect.

The situation that is graphically displayed in Figure 5 represents mediation. It is captured by the fact that (a) there are, on average, treatment differences in both X and Y; (b) X and Y are related in each condition (we assume that X has been scaled to have a positive relationship with Y); and (c) the treatment effects in Y and X are in the same direction (assuming the scaling of X just specified).[8] In terms of our example, treat-

[8] It is possible that the mean X difference between the two conditions is in the opposite direction from the mean Y

ment differences in hormone levels in the blood would be a mediator of treatment differences in pain judgments if (a) pain judgments are related to blood hormone levels in each condition (scaled so that the relationship is positive) and (b) there are mean treatment differences in both variables in the same direction. In this case, the adjusted treatment effect in pain judgments would be smaller than the unadjusted one, once those effects are estimated controlling for blood hormone levels. The specifics of the estimation are discussed later.

Nonparallel functions. Let us return to the two equations relating Y and X in each condition (Equation 13): $Y_{1i} = \delta_{10} + \delta_{11}X_{1i} + \varepsilon_{1i}$ and $Y_{2i} = \delta_{20} + \delta_{22}X_{2i} + \varepsilon_{2i}$. Their difference tells us about the treatment difference in Y for each unit (Equation 14):

$$Y_{Di} = Y_{2i} - Y_{1i} = (\delta_{20} - \delta_{10}) + \delta_{22}X_{2i} - \delta_{11}X_{1i} + (\varepsilon_{2i} - \varepsilon_{1i}).$$

In our discussion so far, we have been assuming that the two slopes are homogenous, that is, $\delta_{11} = \delta_{22}$. Now, however, we want to examine the consequences if these two within-condition slopes are not equal to each other, $\delta_{11} \neq \delta_{22}$. To examine this situation, for pedagogical purposes we initially make the unrealistic assumption that there are no treatment differences in X, either on average or at the level of individual units, that is, $X_{1i} = X_{2i}$. Then the treatment difference in Y specified in Equation 14 becomes

$$Y_{Di} = Y_{2i} - Y_{1i} = (\delta_{20} - \delta_{10}) + (\delta_{22} - \delta_{11})X_i + (\varepsilon_{2i} - \varepsilon_{1i}), \quad (16)$$

where X_i is either X_{1i} or X_{2i}, as they are assumed to be equal to each other. This model is identical to Equation 5 when we were considering the case of the stable concomitant variable. It captures what we earlier defined as moderation of the within-subject treatment effect and was graphically depicted in Figure 2. Accordingly, when the within-condition slopes are unequal, moderation of the treatment effect is once again indicated. Moderation, as shown earlier, has two equivalent definitions: (a) The X:Y relationship depends on condition, and (b) the magnitude of the treatment effect depends on the value of X (in this case, assuming one common value).

We now relax the assumption that $X_{1i} = X_{2i}$, continuing to allow unequal within-condition slopes, so that the model as specified in Equation 14 holds:

$$Y_{Di} = Y_{2i} - Y_{1i} = (\delta_{20} - \delta_{10}) + \delta_{22}X_{2i} - \delta_{11}X_{1i} + (\varepsilon_{2i} - \varepsilon_{1i}).$$

It is useful to reexpress this equation in terms of the corresponding difference and the sum of the two X variables. This amounts to a 45° rotation of the two original x-axes. The preceding difference equation can be equivalently expressed as[9]

$$Y_{Di} = (\delta_{20} - \delta_{10}) + \frac{\delta_{22} - \delta_{11}}{2}X_{Si} + \frac{\delta_{11} + \delta_{22}}{2}X_{Di} + (\varepsilon_{2i} - \varepsilon_{1i}), \quad (17)$$

where $X_{Si} = X_{1i} + X_{2i}$ and $X_{Di} = X_{2i} - X_{1i}$. In this full model, allowing unequal within-condition slopes as well as X differences within participants, there are three different components to each unit's treatment effect in Y, ignoring error.

First, there is a component that is common to all units, $\delta_{20} - \delta_{10}$. Its magnitude is unaffected by individual differences in the Xs.

Second, there is a component of each unit's treatment effect that varies as a function of the sum of each unit's two X scores, $[(\delta_{22} - \delta_{11})/2]X_{Si}$. Its contribution to the unit's treatment effect will be large depending on the magnitude of the condition differences in the Y:X slopes. Assuming that X is scaled to have a positive within-condition relationship with Y and assum-

difference, even given the scaling of X to have a positive relationship with Y in each condition. In such a case, we might talk about inconsistent mediation, wherein the direction of the indirect or mediated effect is opposite from the direction of the unmediated effect. In this case, the average adjusted treatment difference in Y, over and above the treatment difference in X, would actually be larger than the total treatment difference in Y. Although inconsistent mediation is possible, empirically it seems less likely than the case of consistent mediation described in the text.

[9] Any orthogonal basis can be transformed by an orthogonal matrix to another orthogonal basis (Cliff, 1987, pp. 320–326). Solving

$$\begin{pmatrix} X_S \\ X_D \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

for X_1 and X_2 and then substituting these values into the original difference equations yields the reexpression. Such transformations have long been used for mathematical convenience but often have substantive importance in factor analysis and elsewhere. For a substantive example using the 45° rotation to the sum and difference, see Coombs, Coombs, and McClelland (1975).

ing that $\delta_{22} > \delta_{11}$, larger treatment differences will be found for units having larger average (or sum of) X values across the two conditions. This second component reflects moderation: With different average X values, we get different treatment effects. Equivalently, the relationships between Y and X are not homogeneous in the two conditions.

The final component of each unit's treatment effect is that due to the difference between the two Xs, $[(\delta_{22} + \delta_{11})/2]X_{Di}$. It will be large whenever the sum of the two within-condition slopes is large and when there are large differences in the unit's X values. This component reflects the individual adjustment due to differences between X_{2i} and X_{1i} for each unit. When there is a large X difference in the same direction as the Y difference, and X and Y are highly related, then part of the individual's treatment effect in Y is due to the individual's treatment effect in X. And if, on average, across units, there is a mean treatment effect in X, then mediation is indicated.

Figure 6 captures the full complexity, again for two perfectly measured participants, P and Q. First, there is moderation of the treatment effect according to each unit's average or typical X value in the two conditions. This is revealed by the unequal slopes between the two conditions, with the Y:X slope being steeper in Condition 2 than in Condition 1. Thus, X relates to Y more strongly in the second condition than the first, and this means that the magnitude of the treatment difference for each unit, represented by the dashed vertical line for each unit, increases as the unit's average X value increases. Thus, unit Q shows a larger treatment effect than does unit P, over and above any within-unit adjustment due to differences in the unit's two X values.

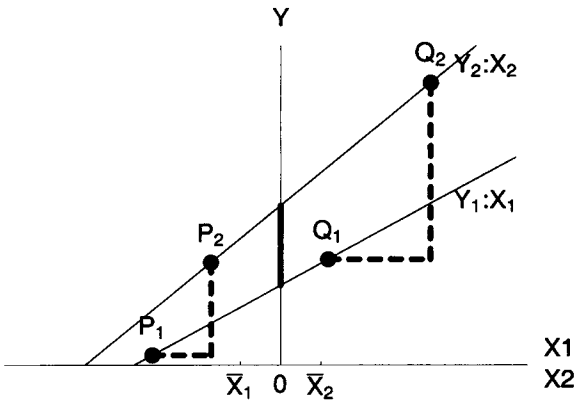

Figure 6. Between-units mediation of treatment effect with moderation of treatment effect due to concomitant X. All points and lines are as in Figure 5, except that the nonparallel slopes indicate a varying adjusted treatment effect depending on each unit's average X score. Mediation is shown by the fact that the average adjusted treatment effect (the solid vertical line) is smaller than the average of the two unadjusted treatment effects (the two dashed vertical lines). Moderation is shown by the fact that the adjusted treatment effect depends on the average of X_1 and X_2 for a given unit.

Second, each unit has different X values in the two different conditions, and there are positive relationships between Y and X in both conditions. Hence, part of the Y difference for each unit is due to the X difference between conditions. Adjusting for the X difference within each unit leaves each unit's adjusted treatment difference (i.e., the vertical distance between the two lines at each unit's mean X). In addition, as in Figure 5, averaging across the two units, there is an average treatment effect in X, such that the mean X value in Condition 2 is higher than the mean X value in Condition 1. Given that the treatment effect in Y is in the same direction and that the within-condition relationships between Y and X are positive, then on average, across units, there is mediation due to X: The overall treatment effect in Y is partly due to the treatment difference in X and is reduced when the X difference is controlled.

As in previous figures, we have centered the x-axis so that the joint mean of X_{1i} and X_{2i} equals zero. The solid vertical line at this point represents the average adjusted treatment difference, over and above any X difference, averaged across the two units and allowing for moderation. Note that it is smaller than the average of the two unadjusted treatment effects.

Estimation. Having discussed this full model, we can now turn to estimation. We need to estimate three different models. First, Equations 2 and 11 should be estimated, testing for overall treatment effects in both Y and X. Second, the full model portrayed in Equation 14 should be estimated. However, the equivalent formulation, given in Equation 17, is much more amenable to interpretation. Accordingly, one can estimate a single regression model in which the difference in the two Y scores for each unit is regressed on two predictors, the sum of that unit's X scores and the difference in that unit's X scores (with the Y and X differences taken in the same direction):

$$Y_{Di} = d_0 + d_1X_{Si} + d_2X_{Di} + e_{YDi}. \qquad (18)$$

The regression coefficients in this model provide estimates of the parameters in Equation 17. Thus, d_0 estimates the component of the treatment difference that is constant across all units, $\delta_{20} - \delta_{10}$. For most

applications, we recommend centering the X_{Si} predictor in the model, as we did in Figure 6. If X_{Si} has been centered, then the estimated intercept, d_0, estimates the mean treatment effect over and above any mediation of that treatment effect due to X. In addition, this estimate is at the mean value of X_{Si}, allowing moderation as a function of the magnitude of X_{Si}. If there is an overall treatment effect in both Y and X, but the intercept in Equation 18 is not different from zero (given X_{Si} centered), then complete mediation is indicated.

The regression coefficient for the sum predictor variable, d_1, estimates the difference in the two within-condition slopes, $(\delta_{22} - \delta_{11})/2$. Accordingly, it estimates the degree to which the treatment effect in Y is moderated by a unit's average X score. Assuming that X has been scaled to relate positively to Y in each condition, then a positive value for this regression coefficient indicates larger treatment effects for units with larger average X scores.

Finally, the regression coefficient for the difference predictor variable, d_2, estimates the degree to which each unit's treatment difference in Y is due to that unit's X difference. It represents the within-unit adjustment in the magnitude of the Y difference due to the X difference. If this regression coefficient is significant (in a positive direction, given the scaling of X), and if there is a mean difference in X between the two conditions, then mediation of the treatment effect in Y is indicated.

In summary, first, the slope d_1 for the sum X_{Si} ($= X_{1i} + X_{2i}$) estimates $(\delta_{22} - \delta_{11})/2$, half of the difference of the within-condition slopes for the concomitant variable X, and indicates moderation of the treatment effect in Y depending on the level of the X sum. Second, the slope d_2 for the difference X_{Di} ($= X_{2i} - X_{1i}$) estimates $(\delta_{22} + \delta_{11})/2$, the average of the two within-condition slopes for the concomitant variable X, and indicates mediation of the treatment effect in Y, assuming a mean difference between the two Xs. Third, the intercept d_0 estimates $\delta_{20} - \delta_{10}$, which is (assuming X_{Si} centered[10]) (a) the average treatment effect over and above mediation, if d_2 differs from zero, and (b) that treatment effect for a unit with an average sum of the X values, if d_1 differs from zero.

Illustration

Our example has been a hypothetical study involving a drug treatment versus placebo comparison, examining effects of these treatments on chronic pain judgments. We analyze hypothetical data to illustrate our proposed analysis strategy.

Table 1
Data Used in Example

Y_1	Y_2	A	H_1	H_2
73	61	59	37	33
57	55	60	30	28
57	61	56	30	36
67	49	69	31	30
80	75	67	37	35
56	60	51	33	34
72	73	56	38	35
81	65	60	43	29
61	59	58	33	31
67	48	61	20	17
74	64	58	43	41
70	55	64	34	27
83	68	58	41	39
62	61	60	35	30
49	55	56	32	32
74	79	63	35	37
73	60	61	36	38
46	51	59	25	24
60	55	54	26	25
93	71	62	46	39

Data from this example are presented in Tables 1 and 2. Variables Y_1 and Y_2 are the pain judgment scores for each participant in the two conditions (Y_1 in the placebo condition and Y_2 in the drug condition, both measures taken on a hypothetical 100-point scale in which higher numbers indicate greater levels of pain). Three other variables are measured. The first, A, is the participant's age, assumed to be unchanging for all participants during the course of the study. The second and third, H_1 and H_2, are measures of levels of the hypothetical hormone in the blood of each participant, measured after the administration of either the placebo or the drug. The hormone variables have been scaled so that lower numbers indicate higher levels of

[10] It is a useful default practice to mean-center predictor variables in moderator models so that zero will be a meaningful value (Aiken & West, 1991). However, we specifically do not do so for the difference predictor because its uncentered zero value is meaningful; it represents no difference in X between the two treatment conditions. The intercept then equals the treatment effect in Y at the average level of the Xs (assuming that the X sum is centered) and when the X difference equals zero (i.e., acting as if there were no treatment difference in X).

Table 2
Correlations, Means, and Standard Deviations Used in Example

Variable	1	2	3	4	5
1. Y_1	—				
2. Y_2	.64	—			
3. A	.42	.06	—		
4. H_1	.74	.70	.10	—	
5. H_2	.48	.71	−.05	.78	—
M	67.75	61.25	59.60	34.25	32.00
SD	11.89	8.58	4.21	6.41	5.96

the hormone in the blood, and thus they are scaled to have positive relationships with the pain variables: Higher levels of hormone in each condition coincide with lower levels of pain. These variables are all measured for 20 hypothetical participants.[11]

The overall effect of treatment on pain judgments is estimated by the mean difference between Y_2 and Y_1, which equals −6.50. A test of the null hypothesis that the true difference is zero yields $t(19) = -3.15$, $p < .01$. Hence, in these data there is a statistically significant effect of the drug relative to the placebo, with lower levels of pain reported in the drug condition. Next, we estimate this treatment effect considering age as the concomitant variable and then considering hormone levels as the concomitant variables.

Regressing each pain score on age yields $\hat{Y}_{1i} = -3.39 + 1.19\,A_i$ and $\hat{Y}_{2i} = 53.64 + 0.13\,A_i$. In the first equation, A is a marginally significant predictor of Y_1, $t(18) = 1.98$, $p < .10$; in the second equation, the slope for A is not significantly different from zero, $t(18) = 0.27$, $p > .50$.

When the pain difference, Y_D $(Y_2 - Y_1)$, is regressed on A, the following equation results: $\hat{Y}_{Di} = 57.03 - 1.06\,A$. Here the slope for A is significantly different from zero, $t(18) = 2.36$, $p < .05$. It equals the difference in the two slopes from the separate Y_2 and Y_1 equations just described (i.e., $-1.06 = .13 - 1.19$). Thus, the test of whether this slope differs from zero shows equivalently that the slope for A in the Y_2 equation differs from the slope for A in the Y_1 equation. As a result, we can conclude that there is a significant Age × Drug Condition interaction, meaning that age (A) relates more strongly to pain judgments (older participants report greater pain) in the placebo condition than in the drug condition. Equivalently, we can say that age moderates the treatment difference, with larger differences between the drug and placebo conditions for older participants; that is, older participants have more negative pain differences between Condition 2 (drug) and Condition 1 (placebo).

The interpretation of the parameter estimates in this model is made simpler if age is centered from its mean ($A' = A - \bar{A}$): $\hat{Y}_{Di} = -6.50 - 1.06\,A'$. In this equation, the intercept equals the predicted treatment difference at the mean age (i.e., $A' = 0$), and it is identical to the estimated treatment difference when age was not included as a predictor. A test of whether this treatment difference differs from zero yields $t(18) = 3.50$, $p < .01$. Although this test of the mean treatment effect has one fewer degrees of freedom than the test of the treatment difference when age was not included as a predictor, it provides a more powerful test of the overall treatment effect because age is strongly related to the treatment difference.

Turning to the two hormone variables, we want to know whether the hormone levels in the two conditions both mediate and moderate the treatment difference in pain judgments. Initially, we regress each pain measure on the hormone measure from the same condition: $\hat{Y}_{1i} = 20.51 + 1.38\,H_{1i}$ and $\hat{Y}_{2i} = 28.54 + 1.02\,H_{2i}$. As expected in the present research context, hormone levels in each condition are significantly related to pain levels, with higher hormone levels associated with lower judgments of pain. (Remember that lower scores on Y indicate less pain and lower scores on H indicate higher hormone levels.)

When the pain difference is regressed on the hormone sum (H_S) and difference (H_D), the following estimate results:[12] $\hat{Y}_{Di} = 0.55 - 0.07H_{Si} + 1.22H_{Di}$. In this model, the hormone level difference is a significant predictor of the pain difference, $t(17) = 2.66$, $p < .05$, but the hormone sum is not. To aid interpre-

[11] Both types of concomitant variables (age and hormone level) could be included in one analysis. For simplicity of exposition, we treat them separately, just as we have done throughout.

[12] Given Equation 16 presented earlier, it may come as a surprise that the intercept and slopes in this estimated model are not exact functions of the intercepts and slopes in the two equations estimated in each treatment condition. The algebra underlying Equation 16 presumes that the two crossed slopes (effect of H in one condition on P in the other) are exactly zero, which they will be in the population if the assumptions that we made in the earlier section hold. Even though in the present data those crossed effects are not significant, they do not exactly equal zero. We discuss the role of these crossed slopes in a later section devoted to complications.

tation, as previously explained, we reestimate this model after centering the sum variable: $\hat{Y}_{Di} = -3.77 - 0.07\ H'_{Si} + 1.22\ H_{Di}$, where H'_S is the centered sum.

Differences in hormone levels can be said to mediate the treatment difference in pain if two conditions are met. First, there must be a condition difference in hormone levels that is in the same direction as the condition difference in pain (assuming that hormone levels have been scaled to have a positive relationship with pain judgments). The mean hormone level in the drug condition is 32.00; in the placebo condition, it is 34.25. The difference between these two is significant, $t(19) = 2.45, p < .05$. Second, the hormone level difference must be predictive of the pain difference, as we have shown it to be.

The intercept in the preceding model in which the sum predictor has been centered, −3.77, estimates the mean difference in pain judgments between the two treatment conditions over and above the hormone difference, or acting as if the hormone difference were zero. Thus, it represents the residual treatment difference over and above mediation. In this case, that residual difference remains marginally significantly different from zero ($p < .10$). Nevertheless, it is necessarily less than the total or unmediated treatment difference because of the fact that the hormone difference is a significant predictor in this model and there is a mean hormone difference between the two conditions.

In conclusion, with these hypothetical data, we have evidence of a significant treatment difference in pain judgments, with the age of the participant moderating that difference. Older participants show a larger treatment effect. In addition, there is a significant treatment effect on levels of the blood hormone that serves as a mediator of the treatment difference in pain. Over and above the treatment difference in hormone levels, the residual pain difference between the two treatment conditions is marginally significant. There is no evidence that average hormone levels of the participant moderate the treatment effect.

Elaborations and Complications

Having laid out the underlying logic and estimation of mediation and moderation in within-subject designs, we now turn to two different complicating issues. The first is that many within-subject treatment comparisons involve independent variables having more than two levels. The second is that within-subject comparisons are frequently conducted in nonexperimental situations, wherein the assumptions underlying causal inference laid out by Rubin (1974, 1978) and Holland (1986, 1988) are very unlikely to be met.

Within-Subject Independent Variables With More Than Two Levels

The models developed earlier for the case of two repeated measures readily generalize to studies with three or more levels of a within-subject independent variable. For example, a study of a treatment effect might include two different drugs being evaluated as well as a placebo control condition. Or a study of mood effects might involve comparisons among a positive, negative, and neutral mood induced in participants at three different times. And in these cases there might also be a single concomitant variable measuring a stable characteristic or a varying concomitant variable measured in each treatment condition. With less detail, we show the generalization from two to three levels of the within-subject treatment variable and describe the general principles that also apply to more than three repeated measures.

First consider the case of three repeated measures without concomitant variables:

$$\begin{aligned} Y_{1i} &= \alpha_{10} + \varepsilon_{1i}, \\ Y_{2i} &= \alpha_{20} + \varepsilon_{2i}, \\ Y_{3i} &= \alpha_{30} + \varepsilon_{3i}. \end{aligned} \tag{19}$$

For illustrative purposes, we assume that Y_{1i} and Y_{2i} are measures of the dependent variable after administration of one of two drugs, whereas Y_{3i} comes from the placebo control condition.

Procedures for assessing the treatment effect for two repeated measures generalize to orthogonal contrasts for three or more repeated measures. For example, the researcher might want to compare the two drug conditions with the placebo control condition $(1,1,-2)$[13] to determine whether there is an overall effect of the drugs. Then he or she might want to compare the two drug conditions with each other $(1,-1,0)$. We use these two contrasts for illustrative purposes; many other pairs of orthogonal contrasts are, of course, also possible. To obtain the treatment

[13] We represent contrasts with their coefficients. The contrast $(1,1,-2)$ represents $Y_{1i} + Y_{2i} - 2Y_{3i}$, and the contrast $(1,-1,0)$ represents $Y_{1i} - Y_{2i}$.

difference coded by each contrast, we take the corresponding differences among the *Y*s:

$$\begin{aligned} Y_{D1i} &= Y_{1i} + Y_{2i} - 2Y_{3i} \\ &= (\alpha_{10} + \alpha_{20} - 2\alpha_{30}) + (\varepsilon_{1i} + \varepsilon_{2i} - 2\varepsilon_{3i}), \\ Y_{D2i} &= Y_{1i} - Y_{2i} = (\alpha_{10} - \alpha_{20}) + (\varepsilon_{1i} - \varepsilon_{2i}). \end{aligned} \tag{20}$$

In other words, the two treatment differences of interest, except for error, are constant for all of the observational units. The mean treatment effects according to these two contrasts are then

$$\begin{aligned} \mu_{YD1} &= \alpha_{10} + \alpha_{20} - 2\alpha_{30}, \\ \mu_{YD2} &= \alpha_{10} - \alpha_{20}. \end{aligned} \tag{21}$$

Next consider the case of a single stable concomitant variable. That is,

$$\begin{aligned} Y_{1i} &= \beta_{10} + \beta_{11}X_i + \varepsilon_{1i}, \\ Y_{2i} &= \beta_{20} + \beta_{21}X_i + \varepsilon_{2i}, \\ Y_{3i} &= \beta_{30} + \beta_{31}X_i + \varepsilon_{3i}. \end{aligned} \tag{22}$$

Then the contrast differences are given by

$$\begin{aligned} Y_{D1i} &= Y_{1i} + Y_{2i} - 2Y_{3i} \\ &= (\beta_{10} + \beta_{20} - 2\beta_{30}) + (\beta_{11} + \beta_{21} - 2\beta_{31})X_i \\ &\quad + (\varepsilon_{1i} + \varepsilon_{2i} - 2\varepsilon_{3i}), \\ Y_{D2i} &= Y_{1i} - Y_{2i} \\ &= (\beta_{10} - \beta_{20}) + (\beta_{11} - \beta_{21})X_i + (\varepsilon_{1i} - \varepsilon_{2i}). \end{aligned} \tag{23}$$

Now treatment differences according to each contrast have two components (other than error). First, there is a constant effect for all units (represented by the intercept differences in the preceding equations). Second, the magnitude of each contrast difference depends on the unit's *X* value. Just as was the case when the independent variable had only two levels, if the relationships between *Y* and *X* are homogeneous within conditions (i.e., $\beta_{11} = \beta_{21} = \beta_{31}$), then the contrast effects do not depend on or are not moderated by *X*. In this case, the estimates of the contrast effects are the same whether or not *X* is controlled. However, if the relationships are not homogeneous, then the contrast effects are moderated by *X*. The average contrast effects are

$$\begin{aligned} \mu_{YD1} &= \mu_{Y1} + \mu_{Y2} - 2\mu_{Y3} \\ &= (\beta_{10} + \beta_{20} - 2\beta_{30}) + (\beta_{11} + \beta_{21} - 2\beta_{31})\mu_X, \\ \mu_{YD2} &= \mu_{Y1} - \mu_{Y2} = (\beta_{10} - \beta_{20}) + (\beta_{11} - \beta_{21})\mu_X. \end{aligned} \tag{24}$$

Estimation is accomplished by regressing each contrast difference on the *X* variable. A significant regression coefficient associated with *X* indicates that the relationship between *X* and *Y* varies across the levels of treatment specified in the *Y* contrast. Equivalently, *X* moderates the magnitude of the treatment effect specified in the contrast, such that larger or smaller contrast differences are associated with larger or smaller *X* scores. In addition, if *X* has been centered, the intercept estimates the contrast effect at the average level of *X*.

Just as we did in the case of two treatment levels, we might also have an *X* variable whose values vary across the treatment conditions. If we assume, as we did in the case of two treatments, that this *X* variable manifests stability and causal transience, then treatment differences in it are assumed to be causal effects, as they are in the outcome variable. In this case, we can model each *Y* as a function of the *X* measured in that condition:

$$\begin{aligned} Y_{1i} &= \delta_{10} + \delta_{11}X_{1i} + \varepsilon_{1i}, \\ Y_{2i} &= \delta_{20} + \delta_{22}X_{2i} + \varepsilon_{2i}, \\ Y_{3i} &= \delta_{30} + \delta_{33}X_{3i} + \varepsilon_{3i}. \end{aligned} \tag{25}$$

Given these models, we can calculate the differences due to each of the contrasts of interest:

$$\begin{aligned} Y_{D1i} &= Y_{1i} + Y_{2i} - 2Y_{3i} \\ &= (\delta_{10} + \delta_{20} - 2\delta_{30}) + \delta_{11}X_{1i} + \delta_{22}X_{2i} - 2\delta_{33}X_{3i} \\ &\quad + (\varepsilon_{1i} + \varepsilon_{2i} - 2\varepsilon_{3i}) \\ Y_{D2i} &= Y_{1i} - Y_{2i} \\ &= (\delta_{10} - \delta_{20}) + \delta_{11}X_{1i} - \delta_{22}X_{2i} + (\varepsilon_{1i} - \varepsilon_{2i}). \end{aligned} \tag{26}$$

As in the case of two treatment levels, these models can be equivalently reexpressed as two models, with three predictors in each model: the sum of the *X*s, the difference in the *X*s according to the first contrast weights (1,1,–2), and the difference in the *X*s according to the second contrast weights (1,–1,0):

$$\begin{aligned} Y_{D1i} &= Y_{1i} + Y_{2i} - 2Y_{3i} \\ &= (\delta_{10} + \delta_{20} - 2\delta_{30}) + \left(\frac{\delta_{11} + \delta_{22} - 2\delta_{33}}{3}\right)X_{Si} \\ &\quad + \left(\frac{\delta_{11} + \delta_{22} + 2\delta_{33}}{6}\right)X_{D1i} + \left(\frac{\delta_{11} - \delta_{22}}{2}\right)X_{D2i} \\ &\quad + (\varepsilon_{1i} + \varepsilon_{2i} - 2\varepsilon_{3i}) \\ Y_{D2i} &= Y_{1i} - Y_{2i} \\ &= (\delta_{10} - \delta_{20}) + \left(\frac{\delta_{11} - \delta_{22}}{3}\right)X_{Si} \\ &\quad + \left(\frac{\delta_{11} - \delta_{22}}{6}\right)X_{D1i} + \left(\frac{\delta_{11} + \delta_{22}}{2}\right)X_{D2i} + (\varepsilon_{1i} \\ &\quad - \varepsilon_{2i}). \end{aligned} \tag{27}$$

These unwieldy expressions reduce considerably if we assume that all within-condition relationships between X and Y are homogeneous, that is $\delta_{11} = \delta_{22} = \delta_{33}$:

$$\begin{aligned} Y_{D1i} &= Y_{1i} + Y_{2i} - 2Y_{3i} \\ &= (\delta_{10} + \delta_{20} - 2\delta_{30}) + \left(\frac{\delta_{11} + \delta_{22} + 2\delta_{33}}{6}\right) X_{D1i} \\ &\quad + (\varepsilon_{1i} + \varepsilon_{2i} - 2\varepsilon_{3i}) \qquad (28) \\ Y_{D2i} &= Y_{1i} - Y_{2i} \\ &= (\delta_{10} - \delta_{20}) + \left(\frac{\delta_{11} + \delta_{22}}{2}\right) X_{D2i} + (\varepsilon_{1i} - \varepsilon_{2i}). \end{aligned}$$

In other words, homogeneity implies that the contrasts among the Ys are a function not of the sum of the Xs but only of the contrast among them that corresponds to the contrast among the Ys that is being modeled. As a result, whenever the sum of the Xs is found to be a significant predictor of a Y contrast, moderation of that treatment contrast by the Xs is implied. If there is a mean contrast difference for both the Xs and the Ys, then the X contrast being predictive of the Y contrast implies mediation. In this case, part of the overall contrast difference in the Ys can be attributed to the overall contrast difference in the Xs.

The preceding conclusions generalize to situations having more than three levels of a within-subject treatment variable. Estimation in such cases involves defining a series of orthogonal contrast scores among both the Ys and the Xs. Then the Y contrasts are regressed on the sum of the Xs and the full set of corresponding contrast differences among the Xs. Moderation of a given contrast difference is indicated if the sum of the Xs is predictive of that given Y contrast. Mediation is indicated if the corresponding X contrast is significant, having already shown mean differences in both the Y and X contrasts of interest.

Complications When Assumptions Are Relaxed

Following the lead of Rubin (1974, 1978) and Holland (1986, 1988), we have made the strong assumptions of temporal stability and causal transience. In fact, within-subject experimental designs are frequently used in psychological research wherein these assumptions are unlikely. There is considerable consensus that such designs should not be used in cases in which causal transience is violated (i.e., treatment effects carry over and may affect observations at later points in time). However, the assumption of temporal stability (that responses are stable over time in the absence of treatment differences) is frequently not tenable. Instead, to overcome the threats to internal validity that violations of this assumption imply (e.g., history, maturation, regression toward the mean, instrumentation, and testing threats; Campbell & Stanley, 1963; Judd & Kenny, 1981a), alternative orders of treatment administration are typically used, and participants are randomly assigned to these alternative orders.

In still other cases, within-subject analyses are frequently used in cases in which causal inference is clearly impossible; that is, "treatments" are in no sense under the control of the experimenter, and experimental manipulation of their levels is not conceivable. Consider a researcher who is interested in gender differences in subjective happiness. He or she recruits married couples for the research and asks the spouses about their subjective happiness. To assess gender differences in happiness in these couples, it is entirely appropriate to conduct a repeated measures analysis of variance or paired samples t test to ask whether the mean happiness of the men is different from the mean happiness of the women. But clearly the Rubin and Holland assumptions make no sense in this context. Although we can ask whether the mean difference is different from zero, causal inference to gender is impossible.

And yet, it is entirely possible in such a situation that one might be interested in knowing whether some concomitant variable (stable across the two individuals in the couple, e.g., length of time they have been married) moderates the within-couple gender difference in subjective happiness. Perhaps couples who have been married longer are less likely to show gender differences in subjective happiness. Certainly, one can ask about this within-couple moderation using the procedures developed earlier, even though we would not want to refer to it as moderation of a "treatment effect."

Similarly, we could focus on some other concomitant variable that possibly varies between gender within a couple (such as overall mental health) and ask whether differences in mental health "mediate" the gender difference in subjective happiness. In this context, it seems that we should be particularly cautious in using the term *mediation* because there seems to be an implicit assumption in this term that an "effect" is "due to" the mediator. But certainly one can reframe the question, avoiding the term *mediation,* and conduct analyses to ask whether the gender difference in subjective happiness is reduced once we control for gender differences in mental health. In

light of this discussion, it seems appropriate to consider the ways in which the models that we have outlined need to be modified once we move outside of the experimental context.

In general, analyses either without a concomitant variable or with a single stable concomitant variable are identical to the models presented earlier. That is, the overall mean difference in Y between the two "treatment" levels (e.g., male vs. female spouses) is estimated as the mean difference in Y. In addition, that difference can be regressed on some concomitant variable, X, that is stable across the two "treatment levels" (e.g., length of marriage of the couple) to determine whether it moderates the magnitude of the difference in Y. If it does, then that means that X relates to Y differently in the two "treatment levels." In terms of our example, it might be the case that female spouses report, on average, significantly less subjective happiness than the male spouses. But this difference may be less pronounced among couples who have been married longer than among more recently married couples. This would imply that length of marriage relates differently to the men's self-reports of subjective happiness than to women's self-reports.

In the case of an X variable that varies across "treatment levels," however, the models necessarily become a bit more complicated, because it may be the case that the X in one "treatment level" exerts effects on the Y in the other level. Again, consider our example. Suppose we were interested in each spouse's overall mental health status as both a potential "mediator" and moderator of the spousal difference in subjective happiness. It seems likely that each spouse's mental health not only influences his or her own subjective happiness but also the subjective happiness of his or her marriage partner. In the previous models with a varying X, we could make the assumption that such crossed effects were zero. Now we no longer can. Accordingly, each Y must be modeled as a function not only of X in the same "treatment level" but also of X in the other "level":

$$\begin{aligned} Y_{1i} &= \delta_{10} + \delta_{11}X_{1i} + \delta_{12}X_{2i} + \varepsilon_{1i}, \\ Y_{2i} &= \delta_{20} + \delta_{21}X_{1i} + \delta_{22}X_{2i} + \varepsilon_{2i}. \end{aligned} \tag{29}$$

In these models, δ_{11} and δ_{22} represent the within-level effects, whereas δ_{12} and δ_{21} represent the crossed effects.

Given these models, the model for the difference between the two Ys is more complicated:

$$\begin{aligned} Y_{Di} &= Y_{2i} - Y_{1i} \\ &= (\delta_{20} - \delta_{10}) + (\delta_{21} - \delta_{11})X_{1i} + (\delta_{22} \\ &\quad - \delta_{12})X_{2i} + (\varepsilon_{2i} - \varepsilon_{1i}). \end{aligned} \tag{30}$$

So too is the model in which the difference in the Ys is modeled as a function of the sum of the Xs and the difference in the Xs:

$$\begin{aligned} Y_{Di} &= Y_{2i} - Y_{1i} \\ &= (\delta_{20} - \delta_{10}) + \frac{\delta_{21} + \delta_{22} - \delta_{11} - \delta_{12}}{2} X_{Si} \\ &\quad + \frac{\delta_{11} + \delta_{22} - \delta_{21} - \delta_{12}}{2} X_{Di} + (\varepsilon_{2i} - \varepsilon_{1i}). \end{aligned} \tag{31}$$

Now the effect of the sum of the Xs on the Y difference will be large to the extent that the two within-level effects differ and to the extent that the two crossed effects differ (in the same direction). The effect of the X difference on the Y difference will be large to the extent that the crossed effects are smaller than the two within effects (assuming all δs are positive).

The estimation of "mediation" and moderation can still be accomplished by regressing the Y difference on the X sum and X difference. If X serves as a moderator, then the X sum should predict the Y difference. Equivalently, moderation implies that within and crossed effects of the X in one "treatment level" are different from the within and crossed effects in the other. If the X difference predicts the Y difference, then X can be said to "mediate" the Y difference, assuming that there is a meaningful mean difference in X as well as in Y. Finally, if the X sum has been centered as a predictor, then the intercept in this regression model estimates the mean residual "treatment effect" in Y when the "treatment effect" in X equals zero, that is, over and above any difference in X. To clarify the nature of the within and crossed effects in this context, we also recommend estimating the separate Y models, regressing each Y on the two Xs (one from the same "treatment level" and one from the other).

In terms of the example that we have been developing, the female–male difference in subjective happiness within each couple would be regressed on the centered sum of the two mental health scores and the female–male difference in the two mental health scores. If greater within-couple differences in happiness are associated with lower average mental health scores for the two individuals, then the sum predictor would have a significant regression coefficient. This

would imply moderation. If greater within-couple differences in happiness are associated with greater within-couple differences in mental health status, and if on average there was a female–male difference in mental health status in the same direction as the subjective happiness mean difference, then mental health status might be said to "mediate" the subjective happiness difference. That is, part of the gender difference in happiness might be attributable to the gender difference in mental health status. This interpretation, of course, depends heavily on an implicit causal model suggesting that mental health status is somehow causally prior to subjective happiness. In addition, we again issue the warning here that talking about "mediation" in this situation, where we clearly cannot talk about causal effects of a treatment variable, is highly suspect. At best, perhaps, we could argue that gender differences are found in both mental health and subjective happiness and that, when the former difference is controlled, the magnitude of the latter one is diminished. In the absence of clear conditions that establish the causal effect of a treatment variable, "mediational" analyses can ultimately do nothing to help make the causal arguments that are implicit in the terminology used.[14]

Summary and Conclusion

In this concluding section, we have three main goals. First, we want to summarize the analytic recommendations we have made for assessing mediation and moderation in within-subject designs. Second, in addition to the complications discussed in the previous section, we want to mention some caveats and concerns that need to be raised whenever the analyses we have outlined are conducted. Finally, we conclude with some general comments about ways in which the assessment of mediation and moderation in within-subject designs permits us to think more clearly about these processes and how they are assessed in between-subjects designs.

Summary of Analyses

The assessment of overall treatment effects in within-subject designs involves estimating the mean difference between the two treatment conditions in Y, the outcome variable. To assess whether the magnitude of this treatment effect is moderated by some stable concomitant variable, X, the Y difference is regressed on that concomitant variable. If the magnitude of the treatment effect depends on X, then X will be predictive of the Y difference. Equivalently, this means that X relates to Y more strongly in one of the treatment conditions than the other. In the estimated model, if X has been centered, then the intercept will equal the overall or average treatment effect, equivalent to the estimated treatment effect assessed independently of moderation.

When the concomitant variable is affected by the treatment, then it may serve as a mediator of the treatment effect if it is causally prior to Y. In this case, X is measured in each treatment condition, and there is on average a mean difference in X between the two conditions. This concomitant variable, measured in each treatment condition, may also serve as a moderator of the treatment effect, with treatment effects of different magnitudes depending on a participant's total or average X score.

To assess both mediation and moderation due to a concomitant variable that varies between treatment conditions, one regresses the Y difference on both the X sum and the X difference. Assuming that there is an overall treatment effect on X and that the X difference predicts the Y difference, mediation of the treatment effect in Y by X is indicated (assuming that X and Y are scaled to have a positive relationship and the treatment effects in Y and X are in the same direction). If the X sum is predictive of the Y difference, then X also serves as a moderator of the treatment effect and, equivalently, X relates to Y differently in the two treatment conditions. Finally, if the X sum has been centered, then the intercept in this regression equation will equal the magnitude of the residual treatment difference in Y, over and above mediation due to X and at the mean value of X. In other words, it will equal the portion of the mean treatment effect that is not mediated through X.

In cases in which the treatment variable has more than two levels, this analytic approach can be accom-

[14] Given the content of the example we have been pursuing, it makes sense to link the "mediation" and moderation model we have been developing with the actor–partner effect model that Kenny and colleagues have developed (Kashy & Kenny, 2000; Kenny, 1996). That model is in fact given by Equation 29. In our model, wherein the difference is regressed on the sum and the difference, the regression coefficient for the difference estimates the extent to which the partner effects are smaller than the actor effects. The coefficient for the sum estimates the extent to which the effects on one target (both actor effect and partner effect) are greater than the effects on the other target.

plished by computing contrast differences in both *Y* and *X*. One then focuses on the analysis of individual *Y* contrasts among the treatment conditions, regressing these on the *X* sum as well as the set of *X* contrast differences. Mediation of a particular contrast difference in the *Y*s is indicated if there is a parallel *X* difference and if the *Y* contrast difference is predicted by that same *X* contrast difference.

Our approach, as just summarized, relies exclusively on estimates provided by ordinary least squares regression models. Estimation can be conducted through any general purpose multiple regression program. As we noted in the introduction, much of what we have presented could also have been developed within the context of multilevel modeling, using programs that rely on other estimation strategies (e.g., HLM or Proc Mixed in SAS). We chose the ordinary least squares approach simply because most psychological researchers are much more familiar with regression and analysis of variance procedures than with multilevel modeling.

Further Concerns

In the initial sections of this article, we made a series of assumptions when we laid out the basic model. These included that any treatment differences in *Y* and *X* were causal in nature and were estimated without bias. An additional assumption is important if, in fact, one wishes to use the proposed analyses to argue for mediation. It must be the case that *X* is causally prior to *Y;* that is, changes in *X* induced by the treatment cause changes in *Y* between the two treatment conditions. Because *X* is a measured variable rather than one that is manipulated in the designs we have been considering, this assumption cannot be critically examined without additional research in which *X* is actually manipulated and its effects on *Y* assessed experimentally.

Implicitly, we have also been assuming that both *Y* and *X* are measured without error. Random errors of measurement, particularly in the concomitant *X* variable, will result in biased assessments of both mediation and moderation. In the case of moderation due to a single stable *X,* measurement error in *X* will attenuate the estimate of moderation. In the case of a concomitant *X* that varies between the two treatment conditions, the situation is more complicated because both mediation and moderation are assessed in the same model. Random errors of measurement in this case can lead to either an underestimation or an overestimation of these effects. All of the analyses that we have outlined could be conducted within a structural equation approach that allows for multiple indicators and controls for errors of measurement. Such an approach would need to model the means of the variables as well as their variances and covariances (see Kline, 1998, pp. 293–307).

Concluding Comments

Assessment of mediation and moderation in within-subject designs suggests some refinement in our thinking about these processes in both these designs and designs in which the treatment variable varies between participants. We conclude with two such thoughts.

First, the within-subject analysis with a varying *X* that can serve as both a moderator and a mediator is conducted by assessing whether the magnitude of the treatment difference in *Y* depends on two different linear functions of the *X*s: their sum and their difference. Moderation is indicated if the treatment difference in *Y* depends on the *X* sum. Mediation is indicated if the treatment difference in *Y* depends on the *X* difference, assuming overall treatment effects in both *X* and *Y*. Thus, in a deeper sense, both moderation and mediation suggest that the magnitude of the *Y* treatment difference varies and is predictable from some function of the *X*s. Although we are convinced of the utility of thinking about moderation and mediation as fundamentally different processes, they both imply treatment effects varying in magnitude in predictable ways as a function of a concomitant variable.

Second, our analyses suggest that when *X* varies between treatment conditions and *Y* is a function of that *X* in each condition, the full model necessitates including both the *X* sum and the *X* difference as predictors. To assess the impact of one of these without the other also included in the model is to risk biased assessments of either mediation or moderation.[15] The same presumably should apply in between-subjects assessments of mediation and moderation. Furthermore, in cases in which crossed effects may exist (i.e., the *X* in one treatment condition affects *Y* in the other), analyses in between-subject de-

[15] If the *X*s in the two conditions have equal variances, it can be shown that their sum and difference are uncorrelated (Kenny, 1979, p. 21). Accordingly, in this equal variance case, there will be no bias if estimation includes only the sum or only the difference as predictor.

signs may be misleading because such crossed effects cannot be estimated. In this sense, one can think of within-subject designs as containing considerably more information than between-subjects designs. Or, equivalently, one can think of between-subjects designs as within-subject designs with half of the data from each participant randomly missing.

As with any analytic approach to data, the models we have outlined should be applied only after careful consideration of their appropriateness and of the assumptions that underlie them. Mediation and moderation of treatment effects are clearly important analytic issues to be addressed in both between-subjects and within-subject designs. And we hope to have clarified methods in the latter case. But these analyses make sense only if the many assumptions that underlie them can realistically be made.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

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

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cliff, N. (1987). *Analyzing multivariate data.* San Diego, CA: Harcourt Brace Jovanovich.

Coombs, L. C., Coombs, C. H., & McClelland, G. H. (1975). Preference scales for number and sex of children. *Population Studies, 29,* 273–298.

Fiske, S. T., Kenny, D. B., & Taylor, S. E. (1982). Structural models for the mediation of salience effects on attribution. *Journal of Experimental Social Psychology, 18,* 105–127.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81,* 945–960.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models. In G. C. Clogg (Ed.), *Sociological methodology 1988* (pp. 449–484). Washington, DC: American Sociological Association.

Huitema, B. E. (1980). *The analysis of covariance and alternatives.* New York: Wiley Interscience.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69,* 307–321.

Judd, C. M., & Kenny, D. A. (1981a). *Estimating the effects of social interventions.* New York: Cambridge University Press.

Judd, C. M., & Kenny, D. A. (1981b). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602–619.

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology, 46,* 433–465.

Judd, C. M., McClelland, G. H., & Smith, E. (1996). Testing treatment by covariate interactions when treatment varies within subjects. *Psychological Methods, 1,* 366–378.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). Cambridge, England: Cambridge University Press.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley Interscience.

Kenny, D. A. (1996). Models of interdependence in dyadic research. *Journal of Social and Personal Relationships, 13,* 279–294.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–268). Boston: McGraw-Hill.

Khattree, R., & Naik, D. N. (1995). *Applied multivariate statistics with SAS software.* Cary, NC: SAS Institute.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17,* 144–158.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon.

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology, 56,* 374–386.

Rubin, D. B. (1974). Estimating causal effects of treatments

in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6,* 34–58.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64,* 1–48.

West, S. G., Biesanz, J. E., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). Cambridge, England: Cambridge University Press.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Received February 3, 1999
Revision received January 5, 2001
Accepted January 8, 2001 ■