This article presents the rationale and procedures for conducting a process analysis in evaluation research. Such an analysis attempts to identify the process that mediates the effects of some treatment, by estimating the parameters of a causal chain between the treatment and some outcome variable. Two different procedures for estimating mediation are discussed. In addition we present procedures for examining whether a treatment exerts its effects, in part, by altering the mediating process that produces the outcome. Finally, the benefits of process analysis in evaluation research are underlined.

PROCESS ANALYSIS

Estimating Mediation in Treatment Evaluations

CHARLES M. JUDD
Harvard University
DAVID A. KENNY
University of Connecticut

Social interventions or treatments may or may not exert the effects that we expect of them. To assess whether the desired effects are obtained, evaluation research is typically conducted using any of a variety of different research designs. In randomized experiments, where subjects are randomly assigned to treatment conditions, it is a relatively straightforward task to test for treatment effects. In quasi-experimental designs, where the assignment variable is other than a random one, treatment effects are much more difficult to estimate, since the assignment variable can be expected to relate to the outcome variable even in the absence of treatment effects (Judd and Kenny, 1981).

AUTHORS' NOTE: *A version of this article was presented at the conference "Health Promotion in Youth: Problems and Perspectives," organized by the Institute for Social Medicine and Epidemiology, Federal Health Ministry, Federal Republic of Germany, April 1980. We sincerely thank the sponsors of that conference. Support for this article*

EVALUATION REVIEW, Vol. 5 No. 5, October 1981 602-619

In such cases, techniques such as analysis of covariance and change score analysis have been used to estimate treatment effects, although each of these procedures has serious drawbacks (Campbell and Erlebacher, 1970; Cronbach and Furby, 1970; Kenny, 1975).

When a treatment is judged effective, regardless of the procedure used to estimate its effects, it is usually informative to examine the *mediating process* that produces those effects.[1] Such a process analysis is an attempt to specify the causal chain responsible for the observed treatment effects. With a process analysis one asks not *whether* a treatment produced the desired effects, but rather *how* it did so. Our purpose in these pages is to examine the procedures and benefits of conducting a process analysis. What we have to say has been influenced by others who have studied issues of treatment implementation (e.g., Boruch and Gómez, 1979) or of "evaluability assessment" (e.g., Wholey et al., 1975). Nevertheless, our notions of process analysis and the procedures behind it are not subsumed under these other topics, as will be shown.

A process analysis is of value in evaluation research for three reasons. First, by specifying and examining the causal mechanisms that produce some outcomes, we gain knowledge about the genesis of the outcome behavior of interest. Through a process model we not only examine treatment effects, but we also build and test a theory regarding the more general causal mechanisms responsible for the outcome behavior. Second, once we have a theoretical causal model for the outcome behavior, it becomes much easier to generalize the results of the research to other settings, populations, and treatments. In other words, once we understand how treatment effects are produced, we can better predict whether they will be produced in other contexts. Finally, if we know the process, then we also know the variables that have a very direct impact upon the outcome of interest. With such knowledge more efficient treatments may be designed.

To clarify the procedures and problems in conducting a process analysis, we will use a particular applied research example throughout this article. During the early 1970s a large-scale health intervention study, the Stanford Heart Disease Prevention Project (Maccoby and Farquhar, 1975) was conducted to reduce the risk of heart disease

was also provided by the National Science Foundation grants BNS 8005737 to the first author and BNS 7826672 to the second. We also thank Reuben Baron, Jon Krosnick, and Michael Milburn for helpful comments on an earlier version.

Treatment ⟶ Knowledge ⟶ Behavior ⟶ Outcome

Figure 1: A Process Model of Treatment Effects Produced by the Stanford Heart Disease Prevention Project

through mass media and personal instruction interventions. Among the outcome measures focused upon as indicators of heart disease risk were levels of serum cholesterol and triglyceride in the blood. Individuals participating in the research were either in a treated or control group depending upon their residence. Two towns were exposed to the mass media treatment while a third was not. Data were collected before the treatment initiation in 1971, and again in 1972, 1973, and 1974.

Since individuals were not randomly assigned to treatment conditions, various adjustment strategies were used to demonstrate treatment effects upon the physiological outcomes of interest. These analyses, summarized by Milburn (1978), demonstrated small but quite significant treatment effects. In the treatment towns, the mass media campaign and personal instructions led to relatively lower levels of cholesterol and triglyceride.

A process analysis of treatment effects begins with a theoretical model that takes the form of a causal chain, linking the treatment at one end with the outcome variable at the other end. This model is derived from theory and prior research. Ideally it should be constructed prior to the collection of the research data since the ease with which the causal links can be hypothesized may affect the choice of outcome variables (Wholey et al., 1975). In our example, the hypothesized causal chain might be as follows: We might suspect that exposure to the mass media campaign led to increased knowledge about dietary factors associated with heart disease risk. In turn, this increased knowledge might be expected to lead to changes in dietary practices or behaviors that in turn are ultimately responsible for the changes in the physiological outcome variables that were observed. The mediating causal chain that we are hypothesizing is presented in Figure 1. Depending on the precision of our theory, this causal chain could be stretched out quite far. Many, many different mediating constructs could be inserted, for instance, between knowledge and behavior, specifying the process that produces behavior change from knowledge acquisition. Perhaps the one important mediating variable we have omitted from this process model is

exposure to the treatment. Exposure is clearly caused by the treatment variable but is not identical to it, since it is certain that not all residents of the treatment towns received equal exposure to the mass media campaign. The inclusion of this exposure mediating variable points immediately to one of the benefits of a process model: In such a model we can allow for differences in program implementation, arising either from administrative imperfections in the program or from differences between individuals in receptivity to it (Boruch and Gómez, 1979).

In order to demonstrate mediation, or, alternatively, in order to validate the hypothesized process model, the researcher must present evidence for the following three conclusions:

Conclusion I. The treatment affects the outcome variable. Without treatment effects, it makes little sense to speak of a causal process mediating them. Occasionally, something resembling a process analysis is attempted as a fallback analysis strategy when treatment effects are not found. Such an analysis, however, should not be called a process analysis, as we are defining the term, since there are no treatment effects to be mediated. This is not to suggest however, that such an analysis, in the absence of treatment effects, is without merit. Examining "mediating" variables in the absence of treatment effects may be informative in understanding why the treatment was ineffective. For instance, it may reveal that there was little variability in exposure to the treatment, even though the research design called for both treated and control groups. It is also possible, although unlikely, that the effect of the treatment variable, controlling for the mediating process, is cancelled out by that mediating process.

Conclusion II. Each variable in the causal chain affects the variable that follows it in the chain, when all variables prior to it, including the treatment, are controlled. In the example of Figure 1, the treatment affects knowledge, knowledge affects behavior when the treatment is controlled, and behavior affects the outcome variable when both the treatment and knowledge variables are controlled.

Conclusion III. The treatment exerts no effect upon the outcome when the mediating variables are controlled. This conclusion is necessary to establish that the hypothesized mediating process is *the* sufficient mediating process. Partial mediation of the treatment effect can be demonstrated by presenting evidence for only the first and second conclusions. The third conclusion does more than just establish mediation. It says in addition that the hypothesized chain accounts for all of the relationship between the treatment and the outcome.

TECHNIQUES AND PROBLEMS IN ASSESSING MEDIATION

Each of the three conclusions for demonstrating mediation can, in certain circumstances, be assessed using regression analysis to estimate

the causal parameters of the hypothesized causal chain. If all variables are transformed so their means equal zero and their variances equal unity, such an analysis is known as path analysis (Duncan, 1975; Kenny, 1979).

To demonstrate the first criterion—that the treatment affects the outcome—the outcome variable should be regressed upon the treatment, controlling for the variable or variables that determined assignment (Judd and Kenny, 1981). While treatment effects must be demonstrated in order to claim mediation, the causal parameters in a long process model may be fairly substantial even though the overall effect of the treatment on the outcome variable may seem quite small. In a causal chain, the total effect of the treatment equals the product of all mediating path coefficients. Hence, in Figure 1, if each of the three mediating path coefficients (beta coefficients) were .40, the treatment effect coefficient would only equal .064 (i.e., $.40^3$).

To reach the two other conclusions for establishing mediation, each mediating variable is regressed upon all the variables that precede it in the causal chain, and likewise the outcome variable is regressed on the treatment and all mediating variables. Evidence for the second conclusion is gathered from the regression coefficients for all variables that immediately precede the various dependent variables. The second conclusion can be reached if most, or ideally all, these coefficients are significantly different from zero. To illustrate this from the example of Figure 1, knowledge would be regressed upon the treatment variable. Behavior is regressed upon treatment and knowledge. And the outcome is regressed on all three. In each of these equations, evidence for the second conclusion depends upon the significance of the regression coefficient for the variable that immediately precedes the dependent variable in the hypothesized causal chain. Thus, the treatment should affect the knowledge variable; kowledge should affect behavior when the treatment is controlled; and behavior should affect the outcome when both other variables are controlled. In all of these regression equations, it is wise to control for either the variable that determined treatment assignment or a proxy for it, in order to reduce bias. In randomized experiments, assignment is randomly determined and hence does not need to be controlled.

The third conclusion can be reached in this series of regression equations if the regression coefficient for the treatment is essentially zero in the equation where the outcome is regressed on all the prior variables. If this is true and yet the treatment has a significant effect when the mediating variables are *not* controlled, we can have some confidence that the hypothesized process model is accurate. If in fact the

hypothesized model is a simple chain, as in Figure 1, where prior variables in the chain exert their effects only through mediating variables, then in each of the regression equations there should be only a single significant predictor, other than the assignment variable, if one is included. This single significant predictor should be the variable that immediately precedes the dependent variable in the causal chain.

While this path-analytic approach to assessing mediation seems relatively straightforward, there are at least two problems in using the techniques outlined in the preceding paragraphs. First, the power or precision of the statistical tests used to test the mediational hypothesis may be quite low. Second, the coefficients generated by this analysis may be biased because of measurement error, omitted variable problems, or reciprocal causation. As Judd and Kenny (1981) note, "We should recognize a mediational analysis for what it really is: a correlational analysis." As such it is likely to yield biased estimates of causal parameters.

The problem of relatively low statistical power in a mediational analysis arises because substantial correlations are expected between variables that occur close together in the causal chain. If this is true, then the estimate of their effects upon a later dependent variable will be collinear. As is well known, collinearity increases the standard error of regression coefficients. Hence we may conclude that a particular variable in a causal chain exerts no effect upon the next variable simply because of poor power.

Another source of low power in a path-analytic mediational analysis arises from the inevitable presence of measurement error in the variables. One of the effects of measurement error in a dependent variable is to reduce the chance of finding significant predictors of it. Since all variables (except the treatment) in a process chain are treated as dependent variables at some point in the analysis, measurement error in any of them decreases statistical power. Other effects of measurement error are discussed later in this article.

In addition to problems of statistical power, a mediational analysis may also yield biased estimates because of omitted variables that cause both the outcome and one or more of the mediating variables. If variables that affect the outcome and that also are causes of the included mediating variables are not controlled in the analysis, biased estimates of the mediation process will result, *even when a randomized experimental research design has been used.* To illustrate this bias, a hypothetical example is presented in Figure 2. Both of the path diagrams in Figure 2 were computed from the same correlations in Table 1. In this sample, we are assuming that the omitted variable (O) is

A. Model with Bias due to Omitted Variable (O)

B. Unbiased Model

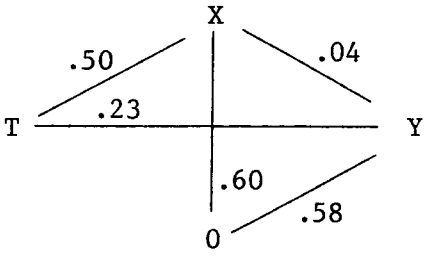

Figure 2: Biased and Unbiased Mediational Models Based on the Correlations in Table 1

TABLE 1
Hypothetical Correlation Matrix for Treatment (T), Mediating (X), Outcome (Y) and Omitted (O) Variables

		T	*X*	*Y*	*O*
Treatment	(T)	1.00			
Mediating	(X)	.50	1.00		
Outcome	(Y)	.25	.50	1.00	
Omitted	(O)	.00	.60	.60	1.00

a relatively stable cause of both the outcome (Y) and the hypothesized mediating variable (X). For instance, O might be education. We assume for illustrative purposes that its effects upon both Y and X are substantial. We are also assuming that the research employed a randomized experimental design, hence treatment (T) is uncorrelated with the omitted variable (O).

In model A of Figure 2, the mediational coefficients are estimated for the model that does not include the omitted variable. In this model, all criteria we have earlier discussed are met. In particular, T has no independent effect on Y once X is controlled, even though the treatment effect ignoring X is substantial (i.e., .25). We would conclude from this model that X totally mediates the effect of T on Y. In model B of Figure 2, the omitted cause (O) is included in the regressions and the causal coefficients are reestimated. Here we see that when O is controlled, the role of X as a mediator is sharply reduced. The treatment continues to exert a very substantial effect on Y even with X controlled. In fact the magnitude of this unmediated effect is only slightly less than the original overall treatment effect. Hence, important causes of both the outcome and mediating variables can lead to substantial bias in estimating mediation unless they are controlled in the analysis, even if the treatment variable is manipulated.

We have already discussed the role of measurement error in reducing the power of the mediational analysis. In addition, measurement error can also lead to biased estimates of the causal parameters. As has been shown elsewhere (e.g., Duncan, 1975), measurement error in a predictor causes us to underestimate that predictor's causal effect. In essence, this bias is the same as that arising from omitted variables, since the true score construct that is measured with error can be seen as an omitted variable.

Bias can also result when multiple regression is used to estimate mediation if the mediating variables reciprocally or mutually affect each other. For instance, in Figure 1, it might be that knowledge causes behavior and, in addition, that different behaviors are instrumental in procuring different information. As others have discussed in some detail (e.g., Duncan, 1975; James and Singh, 1978), other procedures than regression analysis must be used to estimate causal coefficients in models with reciprocal causation.

A more general approach to structural equation modeling may be used as an alternative to regression analysis for estimating a mediational model's causal parameters. This general approach, which employs the

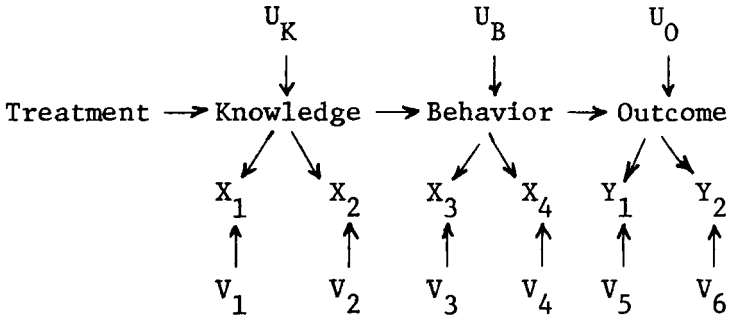

Figure 3: A Multiple Indicator Process Model

computer program LISREL IV (Joereskog and Soerbom, 1978), generates maximum likelihood estimates of all parameters simultaneously under the assumption of multivariate normality. The parameters are estimated by an iterative procedure that minimizes a weighted discrepancy between the observed variance-covariance matrix of all variables and the matrix predicted by the estimated causal model. Assuming that the observed variance-covariance matrix contains sufficient information, the procedure generates a χ^2 goodness-of-fit statistic for testing whether the observed data are consistent with the hypothesized model.

This general structural equation estimation procedure has a number of advantages over the regression-based approach previously discussed. First, by using the overall χ^2-test, the conclusions necessary for mediation can be assessed simultaneously. For instance, to assess the third conclusion, that the treatment exerts no independent effects upon the outcome over and above those mediated via the process model, we can specify a model in which such unmediated effects are not present. If the χ^2-test reveals that the data are consistent with such a model, then we have support for the third conclusion.

Second, if multiple indicators of mediating constructs and the outcome are available, causal parameters among the constructs themselves can be estimated, thus reducing or even eliminating the biasing effects of measurement error. Such a model is illustrated in Figure 3 where two indicators of each of the constructs included in Figure 1 are used, with the exception of the treatment variable. In a model like this at least two indicators are needed for each construct. Two measures (x_1, x_2) of health and dietary knowledge are used as indicators of the mediating construct. Likewise two indicators of both the behavior and outcome

constructs are used. The two indicators of the outcome might be the measurements of serum cholesterol and triglyceride levels already mentioned. Included in the model are the disturbances to the indicators (V_1 to V_6) which represent both random and systematic variance that is unexplained by the constructs. Disturbances to the dependent or endogenous constructs are also included (U_K, U_B, and U_O).

While models incorporating multiple indicators eliminate bias from measurement error, other sources of bias may continue to affect the estimated parameters. For instance, an omitted cause of both knowledge and the outcome will result in bias. To overcome this threat we need to control for such omitted causes in the analysis. In essence, what we might like to do is to control for all the other causes of the outcome that are also causes of the mediating variables. One approach for doing this, very similar to analysis strategies used to reduce bias in the estimate of treatment effects in the nonequivalent control group design, is to include longitudinal data in the model. Specifically we might include the mediational and outcome constructs assessed at a point in time prior to the delivery of the treatment. We then might construct a model such as that illustrated in Figure 4 and estimate the coefficients of the model using LISREL. Here again we are assuming a randomized experimental research design, so that treatment is not related to any of the pretreatment measures. In this model, we are reducing bias in the estimation of the mediational process by controlling for pretreatment differences on all mediating and outcome variables. In essence we are using an analysis of covariance adjustment to remove bias just as it is used in the nonequivalent control group design. The success of this strategy depends on meeting two assumptions besides the usual assumptions of ANCOVA. First of all, the pretreatment constructs must be assessed without error in order to adequately control for them. Second, as in the use of covariance analysis in the nonequivalent control group design, we are assuming that the effects of all omitted variables that cause two or more of the Time 2 variables are mediated through the Time 1 variables (Judd and Kenny, 1981).

The first assumption, that concerning measurement error in the Time 1 constructs, is eased considerably through the use of multiple indicators of all three constructs at both time points. For instance, for the outcome construct we might use the measures of both serum cholesterol and triglyceride at both time points as indicators of the unmeasured outcome. Each of these indicators can be expected to reflect both true score variance as well as error variance. To the extent that error in an indicator is systematic as well as random, it is likely to be correlated with error in the same indicator at other points in time. Hence, in multiple

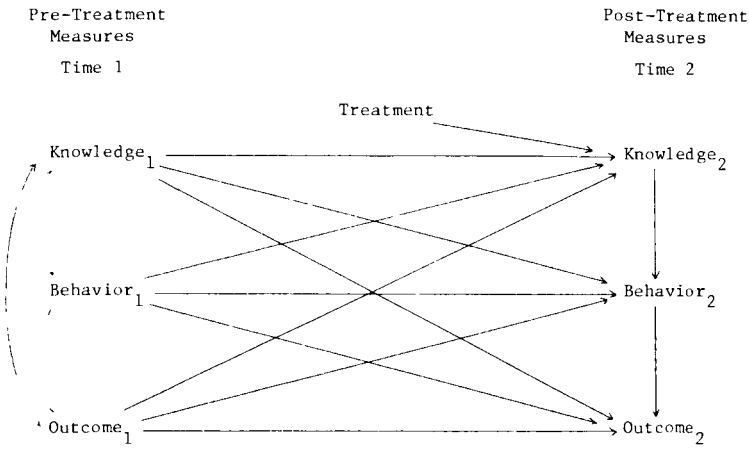

Figure 4: A Longitudinal Process Model

indicator longitudinal models we need to allow errors to indicators to be correlated over time (Wheaton et al., 1977). To solve for the coefficients of models that allow errors to indicators to be correlated over time at least three indicators of the various constructs are needed. An example of such a model with a single mediating construct (knowledge) and with three indicators of both the mediating and outcome constructs at each time point is presented in Figure 5. Only a single mediating construct has been included here in order to make the figure simpler once the correlations between the errors to indicators are included. In this figure, the indicators to the mediating knowledge construct are labeled X_{ij}, where i refers to the specific measure and j refers to time point. The disturbances to these indicators are designated as U_{ij}. The indicators of the latent outcome construct are labeled Y_{ij} and their disturbances are designated as V_{ij}. The W_k are disturbances to the two latent endogenous or dependent constructs.

As should be apparent, longitudinal process models with multiple indicators can become quite complex. Likewise, their estimation under the LISREL procedure may seem to be a formidable task. We will not discuss the estimation strategy here in more detail, as that is well covered in more formal treatments of the LISREL procedure (i.e., (Joereskog and Soerbom, 1978, 1979). We should only note that the most formidable part of estimating such models may be to determine whether the model is identified, that is, whether the observed variance-covariance matrix contains sufficient information to estimate a given

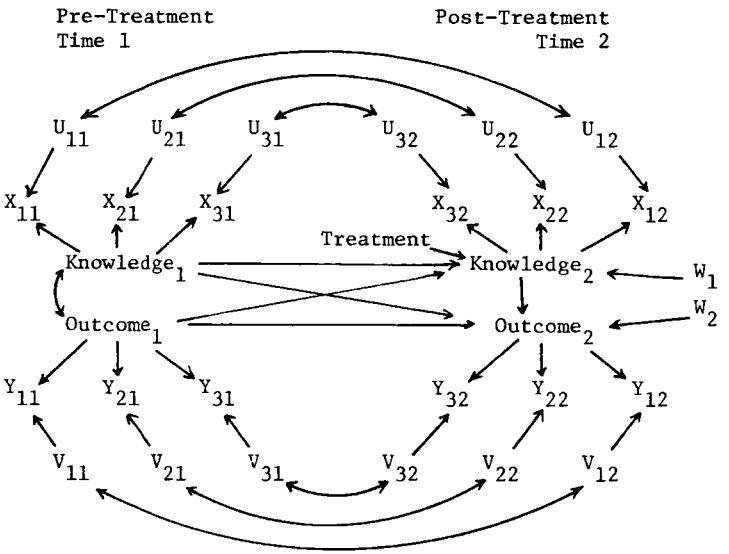

Figure 5: A Longitudinal Multiple Indicator Process Model

model's parameters. It is for this reason that three indicators of each construct were necessary in the longitudinal model of Figure 5. Again, we refer the reader to a more formal treatment of the identification issue (Long, 1976).

While the estimation of longitudinal multiple indicator process models is complex, it is also likely to be quite rewarding, since only through such an analysis can we glimpse the process whereby treatment effects are produced. Without knowledge of this process, generalizing treatment effects may be difficult. In addition, by knowing the mediating process we may design a new treatment to produce the same results at less cost.

ESTIMATING TREATMENT EFFECTS ON THE MEDIATION PROCESS

So far in our discussion, we have developed models in which a treatment affects a chain of mediating constructs which in turn affect or produce the outcome. We have implicitly assumed that the mediating process is similar in both the treatment and control conditions. That is, for both treated and untreated subjects, we have assumed that the causal

coefficients of the mediating process are of roughly equal magnitude. It may be, however, that the treatment produces its effects in part by altering the process that normally produces the outcome. Our example illustrates this possibility. We might suspect that, in the absence of the mass media and personal instruction treatment, dietary behaviors result from habit, neglect, or happenstance rather than from knowledge about what foods are nutritious or health promoting. The treatment may produce changes in dietary practices in part by making that behavior more contingent on rational planning based upon knowledge about the nutritional values of foods. In other words, the treatment may induce changes in the outcome measure by altering the mediational process. Dietary practices may become more knowledge-based and they may be less dependent upon past dietary practices.

In statistical terms, we are suggesting that the magnitude of the mediational parameters varies by treatment, or alternatively that the treatment and the mediating variables interact in producing the desired outcome. Such interactions may be quite informative about how it is that the treatment produces its effects. Looking for these treatment by mediation variable interactions is therefore an important component of a process analysis of treatment effects. Both a regression based analysis as well as the maximum likelihood estimation procedure can be used to conduct this interactive analysis.

Under the multiple regression model, we would like to test the treatment by mediating variable interactions at each step in the analysis. To explain the details of this analysis, we will refer once again to the simple causal chain model in Figure 1. In the noninteractive analysis for this chain, each mediating variable and the outcome are regressed upon all variables that precede them in the causal chain, once overall treatment effects have been established upon the outcome.

Once mediation has been established (or even, in fact, if it has not), we can examine whether the treatment exerts its effects, in part, by altering the causal parameters of the process model. To do this, the treatment by mediating variable product terms are added to the regression equations. In the model of Figure 1, the treatment by knowledge product term is included in the regression of behavior on knowledge and treatment. Likewise the treatment by knowledge and treatment by behavior product terms are included in the final regression equation that predicts the outcome. If the regression coefficients for these various product terms are significantly different from zero, then we have evidence that the strength of the mediating process is different in the treatment and control groups. If the treatment variable had been coded as a "dummy"

variable with treatment coded as "1" and control as "0," then the regression coefficient for the treatment by behavior product term in the final equation estimates the difference in the causal effect of behavior on outcome in the two experimental groups (Cohen and Cohen, 1975).

Suppose we had already established that the treatment was effective and that our hypothesized process model was generally accurate. If in addition we established through the interactive analysis that the effect of knowledge on behavior was greater in the treatment group than among subjects in the control condition, we could build a strong case for the conclusion that the treatment exerts its effects, in part, by encouraging dietary decisions that are based upon health knowledge. In addition, we might then suspect that dietary practices in the treatment condition may be less based upon habit or past practices. That is, if longitudinal data were available to us, we might expect more stability in pretreatment to posttreatment dietary behavior among the control subjects than among the treatment subjects. To test this hypothesis, a treatment by pretreatment behavior product term might be included in the regression equation used to predict the posttreatment behavior variable.

Models incorporating treatment interactions, either treatment by mediating variable or treatment by pretreatment variable product terms, can also be tested quite efficiently using the maximum likelihood estimation procedure of Joereskog and Soerbom, (1978). Under this estimation procedure, such interactions are tested not through the creation of product terms, as in the regression based analysis, but rather by simultaneously deriving the parameters of the model in both the treatment and control groups. LISREL IV permits the estimation of causal parameters simultaneously in multiple groups. In addition, restrictions can be placed upon those parameters between groups. For instance a restriction can be made that a given causal coefficient be equal in the groups. The estimated models are then tested for their simultaneous fit with the variance-covariance matrices from the groups. In the case of a process analysis in treatment evaluations, the two groups across which restrictions upon the causal parameters would be placed are the treatment and control groups.

There is one modification that would have to be made in the models discussed so far if this simultaneous multiple group estimation procedure is employed. Within the treatment group and within the control group there will be no variance in the treatment variable. Hence in separate models for the two groups, this variable must be omitted. For this reason, overall treatment effects and the validity of the process model should be established through the techniques already discussed

prior to estimating models in the two groups simultaneously. Only after we have support for the validity of the mediating process should we test whether the parameters of that model differ in the two groups by examining between sample parameter restrictions.

In a recent article Judd and Milburn (1980) show that a hierarchy of models can be tested with the goal of identifying differences between samples in the causal parameters. At the highest level of this hierarchy, all parameters are constrained to be equal between the two groups. At the lowest level, no between group constraints are placed on the causal parameters. Once the parameters of these models are estimated, each of them will have a χ^2 statistic associated with it. The difference between these χ^2 value is itself a χ^2 and can be used to test whether the model with no between group constraints fits the data better than the model with all parameters constrained equal between groups. If the resulting difference χ^2 is significant, then the model with between group constraints cannot be accepted. Hence, treatment by mediating variables interactions are indicated, since we know that at least some of the parameters in the two groups differ.

In between the two extremes of the hierarchy, a series of models can be estimated in which some but not all of the parameters in the models are constrained to be equal between the treatment and the control groups. To illustrate this, suppose we were estimating the parameters of the longitudinal multiple indicator process model of Figure 5 simultaneously in the two groups (with treatment omitted). If we had determined that a model with all parameters equal between groups did not fit the data, we might then test a model in which we allowed the effect of posttreatment knowledge on posttreatment outcome to differ in the two groups, but forced all other parameters to be equal. If such a model fit the group matrices just as well as the model with no between group constraints, then we would have evidence that the only causal parameter in the mediational process that differs between the treated and untreated subjects is the causal link between posttreatment knowledge and posttreatment outcome. It is in fact likely that a number of the causal parameters in the two groups would differ, given the complexity of the longitudinal causal model in Figure 5. In order to detect the various treatment interactions that do exist in the data, various between group equality constraints should be relaxed with repeated simultaneous estimation. The choice of the parameters that are allowed to differ between groups should be dictated by theory. In addition, the LISREL procedure outputs the first derivatives of all parameters across the

iterations. These can be used to determine where equality constraints should be relaxed to yield the maximum improvement in fit (Soerbom, 1975).

Up to this point in this paper, we have not worried about the distinction between path or standardized causal coefficients and structural or unstandardized coefficients (Blalock, 1967). When it comes to comparing parameters between groups, however, it is necessary to make a distinction between the two. Standardized measures of association in two groups may differ because the variances of the variables differ in the two groups, even when structural coefficients in the raw or unstandardized metrics are equivalent (Duncan, 1975). Therefore, to adequately test for differences between the causal parameters in the treatment and control groups we need to estimate the structural coefficients in the raw or unstandardized metric. With multiple indicators of latent constructs, as in Figure 5, this seems to present problems since the variances of the latent constructs are unobserved. However, these unobserved variances can be estimated by setting the coefficient of one of the indicators on each latent construct equal to unity (Long, 1976; Kenny, 1979). As long as the coefficient for the same indicator of each construct is so fixed in the two groups, then comparisons between groups can be made of the resulting unstandardized structural parameters. A more detailed exposition of both the necessity and the procedure for comparing unstandardized structural parameters between groups in LISREL models is contained in Judd and Milburn (1980).

CONCLUSION

An analysis of the causal process that mediates treatment effects should be an important part of most evaluation studies. Once overall or gross treatment effects have been established, it is important to understand the process that has produced them. Such a process model enables us to have more confidence in our generalizations from the research. It may even help us design a more efficient or effective intervention. Finally, through a process analysis we gain basic knowledge about the causal mechanisms that produce socially significant outcomes.

We started this article with a discussion of three conclusions that should be reached to establish mediation of treatment effect. Two

techniques were then discussed for examining the hypothesized process model and for demonstrating its validity. The first technique was a multiple regression path analytic approach. Problems of power and bias in this approach were identified. The second technique used a simultaneous maximum likelihood estimation procedure (LISREL) to test the process model. Within this procedure, we recommended using multiple indicators of latent constructs to reduce bias from measurement error. We also recommended the inclusion of pretreatment longitudinal measures to further reduce bias in the estimates of the parameters of the process model.

In the second half of the article, we discussed procedures for examining whether the treatment exerts its effects, in part, by altering the process that produces the desired outcome. Again, the use of both a regression based analysis and the LISREL estimation procedure were discussed for examining the process differences between treatment and control groups.

In this article we hope that the benefits, procedures, and problems in establishing the process that produces treatment effects have been made clear. Identifying the mediating process in evaluation research is an important component in the evaluation of any treatment.

NOTE

1. Even when a program is judged ineffective, an examination of the hypothesized mediating process may also be quite informative. It may tell us, for instance, the reasons why no results were demonstrated. We return to this issue when we discuss the conclusions that constitute a process analysis.

REFERENCES

BLALOCK, H. M. (1967) "Path coefficients versus regression coefficients." Amer. J. of Sociology 72: 675-676.

BORUCH, R. F. and H. GOMEZ (1979) "Sensitivity, bias, and theory in impact evaluation," in L. Sechrest et al. (eds.) Evaluation Studies Review Annual, Volume 4. Beverly Hills: Sage.

CAMPBELL, D. T. and A. ERLEBACHER (1970) "How regression in quasi-experimental evaluation can mistakenly make compensatory education harmful," in J. Hellmuth (ed.) The Disadvantaged Child, Volume III: Compensatory Education: A National Debate. New York: Brunner-Mazel.

COHEN, J. and P. COHEN (1975) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.

CRONBACH, L. J. and L. FURBY (1970) "How we should measure 'change'—or should we?" Psych. Bull. 74: 68-80.

DUNCAN, O. D. (1975) Introduction to Structural Equation Models. New York: Academic Press.

JAMES, L. R. and B. K. SINGH (1978) "An introduction to the logic, assumptions and basic analytic procedures of two stage least squares." Psych. Bull. 85: 1104-1123.

JOERESKOG, K. G. and D. SOERBOM (1979) Advances in Factor Analysis. Cambridge, MA: Abt Associates.

——— (1978) LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources.

JUDD, C. M. and D. A. KENNY (1981) Estimating the Effects of Social Interventions. New York: Cambridge Univ. Press.

JUDD, C. M. and M. A. MILBURN (1980) "The structure of attitude systems in the general public: comparisons of a structural equation model." Amer. Soc. Rev. 45: 627-643.

KENNY, D. A. (1975) "A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design." Psych. Bull. 82: 345-362.

——— (1979) Correlation and Causality. New York: Wiley-Interscience.

LONG, J. S. (1976) "Estimation and hypothesis testing in linear models containing measurement error." Soc. Methods and Research 5: 157-206.

MACCOBY, N. and J. W. FARQUHAR (1975) "Communication for health: unselling heart disease." J. of Communication 25: 114-126.

MILBURN, M. A. (1978) "Process analysis of mass media campaigns." Ph.D. dissertation, Harvard University.

SOERBOM, D. (1975) "Detection of correlated errors in longitudinal data." British J. of Mathematical and Statistical Psychology 28: 138-151.

WHEATON, B., B. MUTHEN, D. F. ALWIN, and G. F. SUMMERS (1977) "Assessing reliability and stability in panel models," in D. R. Heise (ed.) Sociological Methodology 1977. San Francisco: Jossey-Bass.

WHOLEY, J. S., J. N. NAY, J. W. SCANLON, and R. E. SCHMIDT (1975) "Evaluation: when is it really needed?" Evaluation Magazine 2: 89-93.

Charles M. Judd is Associate Professor in the Department of Psychology and Social Relations, Harvard University. His research and teaching interests lie in social research methods and in public opinions and attitudes.

David A. Kenny is Associate Professor in the Department of Psychology, University of Connecticut. He received his Ph.D. from Northwestern University in 1972. His interests are study of social interaction, nonexperimental methods, and evaluation research.