MODERATING AND MEDIATING EFFECTS IN CAUSAL MODELS

Jin-Sun Kim, RN, PhD
Judy Kaye, ARNP, PhD
Lore K. Wright, PhD, RN, CS, FAAN
Medical College of Georgia, School of Nursing,
Augusta, Georgia, USA

This article explains causal relationships in conceptual models of mental health phenomena. Direct, moderating, mediating, and reciprocal effects among variables are defined, appropriate statistical analyses are described, and the correct interpretations of moderating versus mediating effects are discussed. Examples are provided that will help the reader to distinguish between moderating and mediating effects.

Conceptual models help us to understand complex phenomena, and they are often used to guide nursing research. Conceptual models typically depict causal relationships among several variables for the purpose of explaining when, how, and why human phenomena occur (Lindley & Walker, 1993). The inclusion of several variables to explain mental health phenomena is especially appropriate because rarely is one factor the sole cause of mental health problems such as depression, lack of motivation, or ineffective coping.

In a model, causal relationships between variables may depict direct effects, moderating effects, mediating effects, and, as yet less common in nursing research, reciprocal effects. Direct effects between independent and dependent variables do not usually present statistical and interpretative problems. However, the terms moderating and mediating effects in

causal models can be problematic. The frequent interchangeable use of these terms has led to inappropriate analytical strategies and misinterpretation of causal relationships (Baron & Kenney, 1986). Conversely, reciprocal effects between two dependent variables are usually handled appropriately by knowledgeable researchers using sophisticated statistical techniques with structural equation analysis. The purpose of this article is to explain theoretical differences between moderator and mediator variables and describe appropriate statistical strategies for both conditions.

MODERATING EFFECTS

A moderator variable specifies when or under what conditions a predictor variable influences a dependent variable (Baron & Kenny, 1986; Holmbeck, 1997). A moderator variable may reduce or enhance the direction of the relationship between a predictor variable and a dependent variable, or it may even change the direction of the relationship between the two variables from positive to negative or visa versa (Lindley & Walker, 1993).

A moderator variable can be considered when the relationship between a predictor variable and a dependent variable is strong, but most often it is considered when there is an unexpectedly weak or inconsistent relationship between a predictor and a dependent variable (Baron & Kenny, 1986; Holmbeck, 1997; Lindley & Walker, 1993). The moderating effect is typically expressed as an interaction between predictor and moderator variable (Aldwin, 1994; Baron & Kenny, 1986; Holmbeck, 1997).

For example, a researcher may hypothesize that depression in family caregivers (the dependent variable or y) is predicted by two independent variables: the care recipients' severity of impairment (x1) and dysfunctional family dynamics (x2). When depression (y) is regressed on "impairment" (x1) and on "dysfunctional family" (x2), the researcher finds only a weak relationship between impairment and depression, but the relationship between dysfunctional family and depression is significant. Because the relationship between impairment and depression had been expected to be significant, the researcher formulates a new hypothesis, namely that dysfunctional family dynamics moderate, or in this case, enhance the effect of care recipients' impairment on caregiver depression. A third independent variable is, therefore, introduced, that is, the interaction term of impairment × dysfunctional family (x3). This new variable (x3) tests whether dysfunctional family (x2) is a moderator variable. When x3 is added to the regression analysis and found to have a significant relationship with depression, the moderator hypothesis is supported (Figure 1).
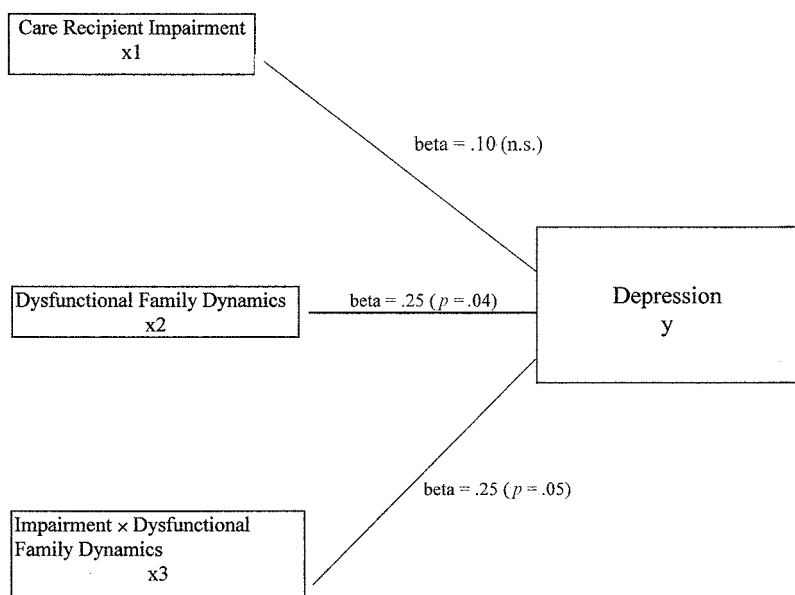
FIGURE 1. The effect of care recipients' impairment severity on caregiver depression moderated by dysfunctional family relationships.

The interpretation of this finding is that the effect of impairment on depression depends on, or is conditional on, the level (or severity) of dysfunctional family dynamics. If the interaction term had not been introduced, an incorrect conclusion would have been that care recipients' impairment severity is not a predictor of depression in caregivers. Conversely, if the interaction term was found to be insignificant, the moderator hypothesis would not be supported. The correct interpretation in that situation would be that dysfunctional family dynamics are significant predictors of depression but the care recipients' impairment characteristics are not.

It should be noted that moderator variables are always at the same level as predictor variables in regard to their roles as causal variables. This means that in any model they are antecedent or exogenous to dependent variables.

Statistical Strategies for Testing Moderating Effects

Moderator variables can be at the interval, continuous, or ratio level, and they can also be categorical (Baron & Kenny, 1986; Lindley & Walker, 1993). Depending on the type (level) of the moderator variable, different statistical analyses are used to measure and test the differential

effects. The statistical tests are multiple regression analyses, structural equation modeling (SEM), and analysis of variance (ANOVA; Baron & Kenny, 1986; Holmbeck, 1997).

Multiple Regression Analyses

When predictor and moderator variables are interval or continuous, multiple regression analyses are used for testing moderating effects. Most commonly, researchers assume that a continuous moderator variable alters the relationship between the independent and dependent variables in a linear function (Baron & Kenny, 1986). Using the example shown in Figure 1 in which both moderator and dependent variable are continuous, the following statistical analyses would be appropriate. First, the predictor variables "impairment" and "dysfunctional family" are entered into the regression equation to test their main effects. This is followed by the interaction term which is generated by multiplying the predictor by the moderator (impairment × dysfunctional family). Depending on the researcher's conceptual framework, the main effects can be entered into the equation in hierarchical, stepwise, or simultaneous methods (Cohen & Cohen, 1983). Although the main effects may be entered in any order, they must be entered first and before the interaction term (impairment × dysfunctional family) is introduced at a separate step. If the change in R^2 (ΔR^2) for the interaction term is statistically significant, it is said to have a moderating effect, and the moderator hypothesis is supported (Aldwin, 1994; Baron & Kenny, 1986; Holmbeck, 1997). Table 1 shows hypothetical data for the above discussed example. Note that the change in R^2 (.02) is significant at the $p = 0.05$ level.

Interpretations of statistically significant interactions require several steps. First, the researcher calculates low, medium, and high levels for "impairment" and also for "dysfunction" which are usually defined as the mean -1 standard deviation (SD) for low levels, the mean for median levels, and the mean $+1$ SD for high levels. Simple regression equations are then solved for each level of the moderator. (For a more detailed

TABLE 1. Caregiver Depression Regressed on Care Recipient Impairment and Dysfunctional Family Dynamics

Predictor variables	R^2 change	Significance of change	Cumulative R^2
Care recipient impairment	.04	.35	.04
Dysfunctional family dynamics	.08	.04	.12
Impairment × dysfunction	.02	.05	.14
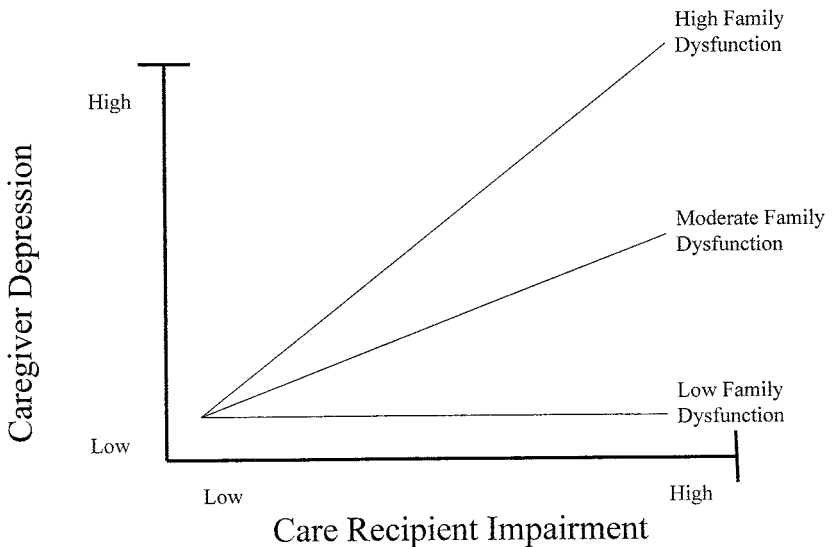
FIGURE 2. Enhancing effect of dysfunctional family dynamics on caregiver depression.

discussion on these regression equations refer to Aldwin, 1994). The obtained regression lines for high, medium, and low values of the moderator variable are then plotted to determine whether there is a buffering, enhancing, or situation-specific effect (Aldwin, 1994; Cohen & Cohen, 1983; Holmbeck, 1997). Figure 2 shows that when the level of dysfunctional family dynamics is high, the stronger the relationship between care recipient impairment and caregiver depression is. This represents an enhancing effect of the moderator variable. Figure 2 also shows that when dysfunctional family dynamics are low (bottom line), there is no relationship between recipients' impairment level and caregiver depression.

Researchers must also be aware of the problem of multicollinearity that may result when the variables being multiplied to generate the interaction term (impairment × dysfunctional family in the above example) are highly correlated with each other. Multicollinearity causes "bouncing betas" in which the direction of the beta terms can shift from previously positive to negative relationships or vice versa (Cohen, 1978). For example, the previously identified positive relationship between dysfunctional family dynamics and caregiver depression could become negative if care recipients' impairment ($x1$) and dysfunctional family relations ($x2$) were highly correlated.

Multicollinearity may be reduced by centering continuous predictor and moderator variables. This is accomplished by subtracting the sample mean from the respective variable, thereby obtaining a centered deviation score with a mean of zero. Centering the beta terms reduces the magnitude of the correlations between the independent variables, thus reducing multicollinearity (Aldwin, 1994).

Structural Equation Modeling (SEM)

Although multiple regression analyses are useful to test moderating effects, SEM, which is based on maximum likelihood analysis, should be used if any of the following conditions exist: (1) the model is nonrecursive, (2) the model has correlated residuals, or (3) the model has multiple indicator variables for unobserved (or latent) variables (Pedhazur, 1982). Nonrecursive models, that is, models with reciprocal relationships, cannot be analyzed with regression analysis. However, with SEM it is possible to separate out the confounding aspects of reciprocal effects (Biddle & Marlin, 1987; Peyrot, 1996). SEM also makes allowances for errors in measurements in the statistical model. Measurement errors are important because they can attenuate the relationship between two variables (Baron & Kenny, 1986; Peyrot, 1996). Moreover, SEM is capable of generating solutions for models in which unobserved variables (known as "constructs" or "latent" variables) are measured by multiple indicators (Biddle & Marlin, 1987; Mason-Hawkes & Holm, 1989; Pedhazur, 1982; Peyrot, 1996).

Although a number of SEM statistical programs have been developed, the two most well known are LISREL VIII (Joreskog & Sorbom, 1993) and EQS (Bentler, 1995). It is highly recommended that researchers who want to use these programs obtain advanced statistical education and analysis training.

Analysis of Variance (ANOVA)

When both the predictor and moderator variable are dichotomous (categorical), 2×2 ANOVA (also called two-way ANOVA) is used for testing moderating effects. For example, the researcher may want to know if the results for two different types of therapy for caregiver depression, that is, group therapy versus individual therapy (coded A/B) are influenced by the presence or absence of reported agitated behavior in care recipients (coded no/yes). The two-way ANOVA will provide main effects for the predictor variables (treatment A and treatment B). Next, ANOVA provides the effect for the interaction between the predictors (treatment A and B $\times$ agitation). If the interaction term is statistically
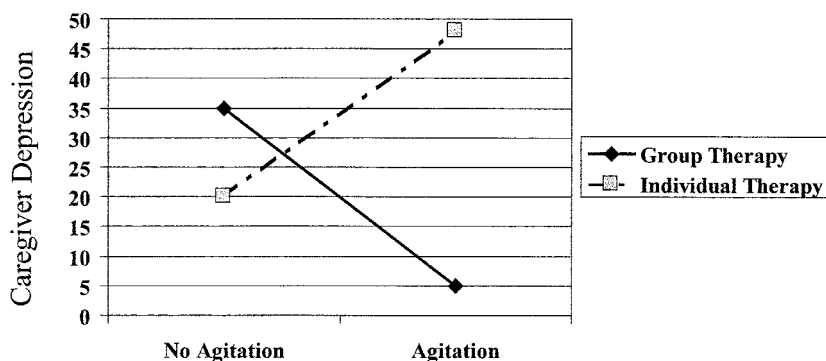
FIGURE 3. Interaction between type of therapy for caregiver depression in the presence or absence of care recipient agitation.

significant, the moderator hypothesis is supported. Sample means for each situation are then used to plot and visually demonstrate the interaction, that is, sample mean depression scores for group therapy in the presence or absence of agitated behavior and sample mean depression scores for individual therapy in the presence or absence of agitated behavior (Polit, 1996). As shown in Figure 3, the plotted graph may reveal a crossed interaction. This would indicate that caregiver depression is lowest for those who received group therapy (the solid line) and reported agitation in their care recipient; but group therapy was less effective for caregivers who did not report agitation. Figure 3 also shows that depression is higher for those caregivers who received individual therapy and reported agitation, but depression is lower when there is no report of agitation.

MEDIATING EFFECT

A mediator variable represents an intervening variable or, stated differently, a mechanism through which an independent variable is able to influence a dependent variable (Baron & Kenny, 1986; Peyrot, 1996). A mediator explains how or why a relationship exists between the predictor and dependent variable, and a mediator is often an attribute or an intrinsic characteristic of individuals (Holmbeck, 1997; Lindley & Walker, 1993; Peyrot, 1996). In contrast to moderating effects which are most commonly introduced when there is an unexpected weak relationship between predictor and dependent variable, there must be a significant relationship between the predictor and the dependent variable before testing for a mediating effect (Baron & Kenny, 1986).

For example, researchers have found that women undergoing breast cancer treatment who use active forms of coping live longer than women who use more passive coping styles (Greene & Larson, 1991; Greer & Morris, 1975; Morris, Greer, Pettingale, & Watson, 1981). A possible interpretation of this finding would be that active forms of coping directly affect longevity. In actuality, active coping involves health behaviors such as exercising, good nutrition, and adherence to medications, all of which facilitate the prescribed treatment regime, and hence the treatment regime affects the person's longevity. Active coping may not have a direct effect on longevity, but only a *mediating* effect by facilitating other things that do directly affect longevity (Aldwin, 1994). In the example described above, the mediator specifies the mechanism by which the given effect of the treatment regime occurs; the mediator indirectly affects health outcomes (Aldwin, 1994; Baron & Kenny, 1986). As shown in Figure 4, a mediator is a third variable that falls in the causal pathway between the predictor variable and the dependent variable.

The model in Figure 4 assumes a three-variable system. First, a direct and significant relationship between "treatment regime" (A) and "longevity" (C) is established (see Figure 4a). After introducing the mediator variable "coping" (B), the path between A and C becomes nonsignificant (Figure 4b). However, A now influences the mediator B,
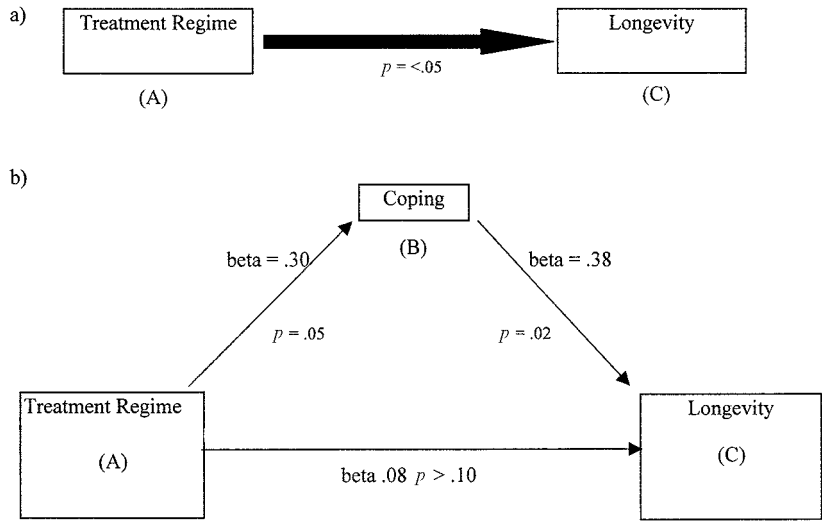
FIGURE 4. a) A significant direct relationship between A and C is established. b) Introduction of coping (B) as a mediator shows that the relationship between A and C is no longer significant.

and B influences C. This can be interpreted as the treatment regime having an indirect effect on longevity through coping. In a "trimmed" model, the path between A and C could be deleted. It should be noted that statistical textbooks (e.g., Polit, 1996; Tabachnick & Fidell, 1996) tend to use the terms mediated and indirect effects interchangeably.

Statistical Strategies for Testing Mediating Effects

Mediating effects can be investigated through path analysis and SEM strategies. Path analysis is a series of regression equations that track out the direct and indirect pathways between predictor and dependent (or outcome) variables. SEM strategies are based on maximum likelihood analysis and, as previously noted, are necessary if the model includes unobserved (i.e., unmeasured) variables that are represented by several indicator variables, or if the model includes reciprocal effects or has correlated residuals.

Multiple Regression Analyses

Prior to using path analytic regression techniques, Pearson correlations among variables in the model are examined. Using the example in Figure 4, the predictor A (treatment regime) must be significantly associated with the dependent variable C (longevity) and with the mediator B (coping). Furthermore, the mediator B (coping) must be significantly associated with the dependent variable C (longevity). After significant correlations have been established, three multiple regression analyses are performed (Cohen & Cohen, 1983).

In the first regression, the significance of the path from A to B is examined. In the second regression, the significance of the path A to the dependent variable C is examined. Finally, the significance of the path B to C is examined in the third regression by using A and B as predictors of C. In the third equation simultaneous entry, rather than hierarchical entry, is used. Simultaneous entry allows for controlling the effect of A while the effect of B on C is examined, and controlling the effect of B while the effect of A on C is examined (Baron & Kenny, 1986; Holmbeck, 1997). The results are then compared, that is, the relative effect of A on C (when B is controlled in the third equation) to the effect of A on C (when B is not controlled in the second equation). The sequence of these regression equations is summarized in Table 2.

If the path A to C in the third equation is reduced to zero, it provides strong evidence for a single, dominant mediator. If the residual path A to C is not zero, it indicates that multiple mediating factors may be operating. Many nursing phenomena have multiple causes, and to

TABLE 2. Sequence of Regression Analyses to Establish a Mediating Effect

1) Coping (B) regressed on treatment regime (A)
2) Longevity (C) regressed on treatment regime (A)
3) Longevity (C) regressed on treatment regime (A) and coping (B) simultaneously
4) Compare Equation 3 (coping controlled) with Equation 2 (coping not controlled)
5) Mediating effect established if A to C is nonsignificant in third equation

find a path reduced to zero is unlikely. In reality, therefore, nurse researchers will look for mediators that significantly decrease path A to C instead of trying to eliminate the relationship between the independent and dependent variables altogether (Baron & Kenny, 1986). The degree to which the effect is reduced (i.e., the change in the regression coefficient in Equation 3 versus the regression coefficient in Equation 2) indicates how powerful the mediator is (Baron & Kenny, 1986; Holmbeck, 1997).

Structural Equation Modeling (SEM)

When the model to be tested has multiple indicators for unobserved (latent) variables, SEM is used for testing mediating effects (Holmbeck, 1997). Assuming a model with a latent predictor variable (A), a hypothesized latent mediator variable (B), and a latent outcome variable (C), factor analytic techniques are generally used first to generate the latent variable constructs. Once this has been accomplished, the mediating effect is tested by assessing the fit of the A to B to C model. Interpretation is similar to the regression approach. If there is a mediating effect, the A to C path is reduced to nonsignificance when B (the mediator variable) is in the model. However, if the path from A to C remains significant even when B is in the model, a mediating effect cannot be assumed.

To arrive at these interpretations with SEM strategies, the overall model is tested under two conditions: (1) when the A to C path is constrained to zero, and (2) when the A to C path is not constrained. (Constraining a path means that the path is not estimated.) The computer print-out will provide a modification index for the constrained path that can guide the researcher in deciding whether the path should be included or deleted. A modification index value of <2 generally means that adding the path would not significantly improve the overall fit of the model. The two models (with and without the A to C path) can also be compared by calculating a nested chi-square statistic (Munro, 1997). The difference should be significant to indicate that one model is a better fit than the other; the model with the smaller chi-square is chosen because it indicates the better fit (Munro, 1997).

If a mediating effect has been demonstrated and the researcher has deleted the A to C path, the chosen model can then be evaluated with a chi-square significance test and one or several goodness of fit indices. Correct interpretation of these tests is important. When testing one, that is, the chosen model, the chi-square compares whether the actual data (the measures obtained by the researcher) and the theoretical structure of the model differ from each other; therefore, chi-square test in this situation should be insignificant to indicate that data and theory are congruent (Munro, 1997). Because chi-square tests can be influenced by sample size, the goodness-of-fit index (GFI) and the adjusted-goodness-of-fit index (AGFI; adjusted for the degrees of freedom in the model) are used. The two tests are independent of sample size. A good fit of the model to the data is indicated by a value of .90 or greater (Polit, 1996).

It should be noted that while the terms mediating and indirect effects are used interchangeably in path analysis, Holmbeck (1997) has discussed important distinctions between mediating and indirect effects when using SEM. For example, Capaldi, Crosby, and Clark (1996) utilized SEM to detect an indirect effect even though significance of a direct effect between the predictor (A) and criterion (C) had not been established. The researchers did find, however, that the relationship between the predictor (A) and mediator (B) and the relationship between the mediator (B) and dependent variable (C) were significant. In this situation it would be correct to state that the researchers tested the significance and found support for an indirect pathway. It would be incorrect to state that a mediating pathway had been tested.

CONCLUSION

Moderators and mediators serve different functions in causal models, and accurate interpretation of these functions is important to mental health clinicians and researchers. Table 3 summarizes the major points of this discussion.

A moderator variable specifies when or under what conditions a predictor variable influences the dependent variable; a mediator explains how or why a relationship exists between the predictor and dependent variable. Moderators are most often introduced when the relationship between predictor and dependent variables is unexpectedly weak; mediator variables are introduced only after a significant relationship between a predictor and a dependent variable has been established. Moderators are always at the same level as predictor variables, whereas mediator variables lie between a predictor and a dependent variable. These distinctions are important if nurses are to accurately depict causal relationships

TABLE 3. Contrasting Moderator and Mediator Variables in Causal Models

	Moderator variables (see Figure 1)	Mediator variables (see Figure 4)
Why used	To establish when/under what conditions a predictor variable influences a dependent variable	To establish how or why a relationship between a predictor variable and a dependent variable exists
Position in model	Always at the level of predictor variables: x_2 same level as x_1	Lies between predictor (A) and dependent variable (C): (A- - - - -B- - - - -C)
Type of variable	Interval, continuous, ratio, or categorical	Interval, continuous, or ratio
Statistical significance	If interaction between independent variable and moderator variable is significant (x_1 multiplied by x_2)	If path A to C becomes *in*significant *after* introduction of B and A to B and B to C are significant

in conceptual models. Appropriate analytic and statistical strategies not only prevent erroneous interpretations but can lead to exciting findings and new understandings of complex phenomena.

REFERENCES

Aldwin, C. M. (1994). *Stress, coping, and development: An integrative perspective.* New York: Guilford.

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.

Bentler, P. M. (1995). *EQS: Structural equations program manual.* Encino, CA: Multivariate Software.

Biddle, B. J., & Marlin, M. M. (1987). Causality, confirmation, credulity, and structural equation modeling. *Child Development, 58*, 4–17.

Capaldi, D. M., Crosby, L., & Clark, S. (1996). The prediction of aggression in young adult intimate relationships from aggression in the family of origin: A mediational model. *Proceedings of the Society for Research on Adolescence.* 6th annual meeting, March 7–10, 1996. Boston, MA.

Cohen, J. (1978). Partial products *are* interactions; partialed powers *are* curve components. *Psychological Bulletin, 85*, 858–866.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Greene, A. L., & Larson, R. W. (1991). Variation in stress reactivity during adolescence. In E. M. Cummings, A. L. Greene, & K. H. Karraker (Eds.), *Life-span developmental psychology: Perspectives on stress and coping* (pp. 195–209). Hillsdale, NJ: Erlbaum.

Greer, S., & Morris, T. (1975). Psychological attitudes of women who develop breast cancer. *Journal of Psychosomatic Research, 19*, 147–153.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 4*, 599–610.

Joreskog, K., & Sorbom, D. (1993). *LISREL VIII*. Chicago: Scientific Software.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Old Tappan, NJ: Englewood Cliffs.

Lindley, P., & Walker, S. N. (1993). Theoretical and methodological differentiation of moderation and mediation. *Nursing Research, 42*, 276–279.

Mason-Hawkes, J., & Holm, K. (1989). Causal modeling: A comparison of path analysis and LISREL. *Nursing Research, 38*, 312–314.

Morris, T., Greer, S., Pettingale, K. W., & Watson, M. (1981). Patterns of expression of anger and their psychological correlates in women with breast cancer. *Journal of Psychosomatic Research, 25*, 111–117.

Munro, B. H. (1997). *Statistical methods for health care research* (3rd ed.). Philadelphia, PA: Lippincott.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Peyrot, M. (1996). Causal analysis: Theory and application. *Journal of Pediatric Psychology, 21*, 3–24.

Polit, D. F. (1996). *Data analysis and statistics for nursing research*. Stanford, CT: Appleton & Lange.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins College Publishers.