

SPSS and SAS Programming for the Testing of Mediation Models

William N. Dudley ▼ Jose G. Benuzillo
Mineh S. Carrico

- **Background:** Mediation modeling can explain the nature of the relation among three or more variables. In addition, it can be used to show how a variable mediates the relation between levels of intervention and outcome. The Sobel test, developed in 1990, provides a statistical method for determining the influence of a mediator on an intervention or outcome. Although interactive Web-based and stand-alone methods exist for computing the Sobel test, SPSS and SAS programs that automatically run the required regression analyses and computations increase the accessibility of mediation modeling to nursing researchers.
- **Objectives:** To illustrate the utility of the Sobel test and to make this programming available to the *Nursing Research* audience in both SAS and SPSS.
- **Methods:** The history, logic, and technical aspects of mediation testing are introduced. The syntax files *sobel.sps* and *sobel.sas*, created to automate the computation of the regression analysis and test statistic, are available from the corresponding author.
- **Results:** The reported programming allows the user to complete mediation testing with the user's own data in a single-step fashion. A technical manual included with the programming provides instruction on program use and interpretation of the output.
- **Conclusion:** Mediation modeling is a useful tool for describing the relation between three or more variables. Programming and manuals for using this model are made available.
- **Key Words:** mediation • mediation modeling • regression analysis • Sobel test • statistical analysis • statistical models

The early work on mediation and moderation by Baron and Kenny (1986) provided behavioral researchers with a powerful analytical tool: mediation modeling. This type of statistical modeling has recently gained attention in nursing research literature. The research of Bennett (2000) and Bennett, Stewart, Kayser-Jones, and Glaser (2002) provided an application for mediation modeling. The case for using mediation modeling in cancer nursing to test models that explain why cancer patients often

experience clusters of symptoms was made by Dudley, Beck, and Barsevick (2003).

A variable is defined as a mediating variable if it accounts for the relationship between two other variables (Baron & Kenny, 1986; MacKinnon, 1994).

Mediation is often useful for conceptualizing the relation among three or more variables as a way to explain the nature of the relation. For instance, Williamson and Schulz (1995) showed that the relation

between pain and depression in patients with cancer was attributable to the mediating variable, function. In their study, patients in pain experienced a decline in function as measured by their ability to accomplish activities of daily living, and this decline led to increased depression. Thus function, or in this case a decrease in function, was a mediating variable in that it accounted for the relation between pain and depression.

This finding has important implications in that it opens the possibility of nonpharmacologic approaches to easing pain-related depression (Williamson & Schulz, 1995). For example, interventions to improve a patient's ability to function within the limitations imposed by pain (e.g., patient education or restructuring of the environment) may alleviate depression.

Mediation Modeling to Test Intervention Efficacy

Mediation analyses also may be used to test the effect of an intervention.

William N. Dudley, PhD, is Research Professor and Director of Applied Statistics at the Emma Eccles Jones Nursing Research Center, College of Nursing, University of Utah.

Jose G. Benuzillo, MPP, is a Research Analyst at the Huntsman Cancer Institute, University of Utah.

Mineh A. S. Carrico, BA, BS, is a graduate from the College of Nursing, University of Utah.

Most interventions are aimed at modifying behaviors by changing intervening variables (MacKinnon, 1994). With this type of modeling, the mediating variable mediates the relation between the levels of intervention (e.g., treatment and control) and the outcome. Lewis, Marcus, Pate, and Dunn (2002) reviewed studies that used this approach. In one study, Miller, Trost, and Brown (2002) investigated the impact of their intervention on the physical activity of mothers with young children and discovered that the effects of the intervention were not significant except when mediated by self-efficacy and partner support. Unlike traditional methods of analyzing the relation between intervention and outcome, which merely indicate to what extent change has occurred, mediation modeling provides insight into how or why such changes occurred. This, in turn, provides a structure for designing more powerful interventions and for testing the effect of these interventions on targeted outcomes. As such, mediation modeling may prove to be an important tool for advancing behavioral science.

Statistical Testing for Mediation Effects

The statistical methods for testing mediation have progressed from the initial testing of regression models (Baron & Kenny, 1986) to more elaborate methods developed more recently (Sobel, 1990) and popularized through the work of MacKinnon (1994), MacKinnon and Dwyer (1993), and MacKinnon, Lockwood, Hoffman, West, and Sheets (2002). Although the SAS and SPSS programming described in this report automate the computational aspects of testing for mediation effects, the use of these programs without sufficient understanding of mediation modeling could lead to an inaccurate interpretation of results. An in-depth review of mediation modeling (Baron & Kenny, 1986; Bennett, 2000; Lindley & Walker, 1993; MacKinnon, 1994; MacKinnon & Dwyer, 1993) and structural equation modeling (Short & Hennessy, 1994) literature is recommended before attempts to use or interpret these tests.

Interventions to improve a patient's ability to function within the limitations imposed by pain may alleviate depression

Traditional Mediation Testing

Baron and Kenny (1986) proposed testing mediation through three regression analyses. The decision to accept or reject a mediation model is based on the results of these analyses. This testing can be best discussed with the aid of a simple model (Figure 1).

The Baron and Kenny (1986) model tests four paths in three regression analyses. The primary predictor of the outcome is referred to as the independent variable.

1. It must be shown that the independent variable (pain) is associated with the outcome (depression). This is a test of path c.
2. To test path a, it must be shown that the independent variable is associated with the mediator (function).
3. In a model with both the independent variable and the mediator predicting the outcome, it

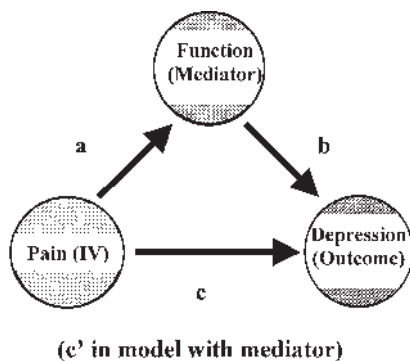


FIGURE 1. A conceptual model of mediation.

should be verified that the mediator is associated with the outcome (path b), and that in this multiple regression model, the path between the independent variable and the outcome (known as path c' in this two predictor model) is reduced to zero (total mediation) or reduced by a significant amount (partial mediation).

This modeling has several benefits. It can be conducted with standard statistical software and does not require specialized programming abilities. In addition, the technique does not require large samples. Work by Walters and Williamson (1999), Williamson and Schulz (1995), and Williamson, Walters, and Shaffer (2002) has involved fewer than 100 respondents.

There are some weaknesses in this approach. The hypothesis testing is reminiscent of hypothesis testing in which critical values are found in a table and compared with computed values in that either full or partial mediation is established. That is, information about the strength of the mediation is not gained, as with significance testing, nor is there a measure for the proportion of the independent variable's effect on the outcome through the mediation pathway. In addition, mediation modeling has shown that path c' is rarely reduced to zero. More often, c' may remain statistically significant, but is reduced by a significant amount. That is, c-c' is significantly greater than zero. However, this test requires an intervening test of the difference between c and c'. With large samples, even trivial differences may be statistically different, and this may raise the question of how much of the independent variable's effect was mediated through the pathway a*b.

Recent Research

More recent methodologic work by Sobel (1990) and MacKinnon (1994) has provided statistical techniques that build on the Baron and Kenny (1986) model, but provide a more precise picture of the mediation. The first advance was the provision of a statistical test for the mediation path.

MacKinnon and Dwyer (1993) showed that the test of the mediation path can be expressed as the "Sobel test,"

$$t_{a*b} = \frac{a*b}{se_{ab}}$$

where:

$$se_{a*b} = \sqrt{(a^2 * seb^2) + (b^2 * sea^2)}$$

In this formula, *a* and *b* refer to the unstandardized regression coefficients, with *seb* and *sea* referring to the standard errors of these coefficients. Both coefficients *a* and *b* and their standard errors, *sea* and *seb*, are readily available from the statistical output. In addition to the statistical test of the hypothesis showing that this pathway is not equal to zero, the newer methods allow for the amount of explained variance accounted for by the mediation to be computed. This is computed simply as:

$$= \frac{a*b}{(a*b) + c'}$$

or alternatively as:

$$= \frac{c}{c - c'}$$

As in the previous formulas, *a* and *b* are the unstandardized path coefficients, and *c'* is the unstandardized path between the independent variable and the outcome in the full model (the model with both the independent variable and the mediator). It is important to keep in mind that this is the portion of the independent variable's effect on the outcome accounted for by the mediation. Thus, in a simple linear model in which the independent variable is the sole predictor with an *R*² of 0.2 and in which the mediation accounts for half of this, it can be concluded that the mediation accounts for 10% of the variance in the outcome, not 50%.

Programming Information

Although the work by Preacher and Leonardelli (2001) provides an interactive Web-based and stand-alone method for computing the Sobel test, it is useful to have a self-contained program that will run the required regression analyses and compute the test statistic automatically. Such a



program helps to reduce errors in hand computation and provides consistency in the analyses from one study to the next. Two such programs have been developed, one each in SPSS and SAS, that run the Sobel test and the Goodman I and II tests (Goodman, 1960) and compute the amount of explained variance attributable to the mediation path.

The syntax files *sobel.sps* and *sobel.sas* were created on SPSS version 11.0 (SPSS, Chicago, Ill) and SAS version 8.02 (SAS Institute, Cary, NC), respectively. The SPSS syntax should run on earlier versions with no adaptations. The SAS program may require modification to run with older versions of SAS. The two syntax files are similar in logic and structure. Both open the data set to be analyzed as the first step and run the regression analyses as the second step. One difference is that the SPSS program runs three regression analyses, each representing a different model, which results in three data sets that contain the results. These files are concatenated into one data set. On the other hand, the SAS program runs one regression analysis with the three different models specified in the same procedure.

In both SPSS and SAS, the next step calculates the Sobel test equation with its corresponding *p* value, the percentage of the total effect that is mediated, and the ratio of the indirect to the direct effect. Finally, the results are presented in a table or report that can be saved or printed.

Program Availability

The programs, example data, and manuals are available for free download at <http://www.nurs.utah.edu/faculty/dudley/index.html>. Users who encounter any difficulty should contact the first author at William.Dudley@nurs.utah.edu.

Mediation modeling is a statistical technique that can be useful for understanding the relations among several variables and for testing the mechanisms by which an intervention leads to behavioral change. This article provides some background on the use of mediation, highlights the potential for the use of mediation modeling in nursing research, and makes available two programs that can be used to automate mediation testing. ▀

Accepted for publication July 10, 2003.

This work was supported partially by a grant from the National Institute for Nursing Research (R01 NR04573).

Corresponding author: William N. Dudley, PhD, University of Utah, College of Nursing, 10 South 2000 East, Salt Lake City, Utah 84112-5880 (e-mail: William.Dudley@nurs.utah.edu).

References

- 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(6), 1173-1182.
- Bennett, J. A. (2000). Mediator and moderator variables in nursing research: Conceptual and statistical differences. *Research in Nursing and Health*, 23(5), 415-420.
- Bennett, J. A., Stewart, A. L., Kayser-Jones, J., & Glaser, D. (2002). The mediating effect of pain and fatigue on level of functioning in older adults. *Nursing Research*, 51(4), 254-265.
- Dudley, W. N., Beck, S., & Barsevick, A. (2003). *Mediation: Concepts, statistical tests, and application to cancer fatigue*. Paper presented at the 7th National Conference on Cancer Nursing Research, San Diego, CA.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55(292), 708-713.
- Lewis, B. A., Marcus, B. H., Pate, R. R., & Dunn, A. L. (2002). Psychosocial mediators of physical activity behavior among adults and children. *American*

- Journal of Preventive Medicine*, 23(Suppl. 2), 26-35.
- Lindley, P., & Walker, S. N. (1993). Theoretical and methodological differentiation of moderation and mediation. *Nursing Research*, 42, 276-279.
- MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention research. *NIDA Research Monograph*, 139, 127-153.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144-158.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83-104.
- Miller, Y. D., Trost, S. G., & Brown, W. J. (2002). Mediators of physical activity behavior change among women with young children. *American Journal of Preventive Medicine*, 23(Suppl. 2), 98-103.
- Preacher, K. J., & Leonardelli, G. L. (2001). Calculation for the Sobel test: An interactive calculation tool for mediation tests. Retrieved May 30, 2003 from <http://quantrm2.psy.ohio-state.edu/kris/sobel/sobel.htm>.
- Short, L. M., & Hennessy, M. (1994). Using structural equations to estimate effects of behavioral interventions. *Structural Equation Modeling*, 1(1), 68-81.
- Sobel, M. E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika*, 55(3), 495-515.
- Walters, A. S., & Williamson, G. M. (1999). The role of activity restriction in the association between pain and depression: A study of pediatric patients with chronic pain. *Children's Health Care*, 28(1), 33-50.
- Williamson, G. M., & Schulz, R. (1995). Activity restriction mediates the association between pain and depressed affect: A study of younger and older adult cancer patients. *Psychology and Aging*, 10(3), 369-378.
- Williamson, G. M., Walters, A. S., & Shaffer, D. R. (2002). Caregiver models of self and others, coping, and depression: Predictors of depression in children with chronic pain. *Health Psychology*, 21(4), 405-410.