Causal Modeling in Health Survey Studies

By: Min Qi Wang, PhD. James M. Eddy DEd., R. Carl Westerfield, PhD

Wang, M.Q., Eddy, J.M. & Westerfield, R.C. (1992). Causal modeling in health survey studies. *Health Values, 16, 4, 55-57.*

***** Note: Figures may be missing from this format of the document**

Article:

Many researchers are interested in providing causal interpretations of the statistical relationships they find in their one-shot health survey data. However, survey studies do not exert controls over independent variables while observing the dependent variable, as in experimental studies. Nor do they, as in longitudinal studies, define time order of variables (i.e., one event may occur prior to another). Survey studies are usually unable to isolate independent and dependent variables, making the causal interpretation of statistical relationships untenable. Nevertheless, these limitations should not discourage researchers from attempting the causal relationships by using a causal modeling approach. Causal modeling is a method that determines predictor variables that may have the potential to influence the criterion variable (the effects), and then analyzes the direct and indirect contributions made by each predictor variable to the effects. The two basic steps in causal modeling are: causal theorizing and causal analysis.

Causal Theorizing

Thinking causally about a research question and constructing an arrow diagram that reflects causal processes will likely provide mom explanatory power for the statistical relationships and generate additional insights into the research topic. In survey studies, causal theorizing poses a number of advantages. First, causal theorizing increases significant contributions to the discipline by developing theory-driven models and by testing models with a priori statistics (i.e., testing a set of causal assumptions the researcher has imposed). It reduces the chances of obtaining "piecemeal" results and making trivial or spurious contributions to the health discipline. Secondly, it identifies and includes a group of variables in the analysis based on a theoretical framework. Thirdly, causal theorizing forces the prediction ahead of time on why and how variables may relate directly or indirectly to each other, and then proceeds the statistical tests for the presence or absence of these relations.

Without causal theorizing, survey researchers tend to rely on available statistical models to treat their data. For example, they may enter all predictors in a stepwise regression equation. The resulting model does not necessarily reflect a best fit of the data for the population. Based "blindly" on statistical criterion in the stepwise procedure, relevant variables can be excluded while irrelevant variables can be included. Thus, this practice may increase the standard errors of all estimates without improving prediction.

Another statistical strategy researchers often adopt is to perform a series of bivariate analysis to test the relations between two variables. Though this method is easy to administer and results are understandable, it does not allow a systematic evaluation of the relationships among the variables; often providing researchers with only a partial picture of the results.

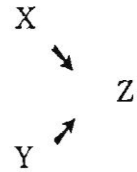
Causal Analysis

Causal analysis requires first the establishment of causal linkages among selected variables, and then the performance of a path analysis. These casual linkages could be established by several means including a theoretical framework, literature reviews, observations, or personal intuition. The path analysis helps to provide structure to the data analysis. To make the illustration simple, let's use three variables (X, Y, and Z) to define types of causal structure (Figure 1).

FIGURE 1

Types of Causal Structure with Explanation and Descriptive Examples

Diagram 1.

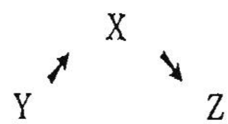


In Diagram 1, X is causally related to Z but not Y, and Y is causally related to Z but not X. For example, in this scenario let's assume that Z is the level of physical activity, X is the access to appropriate facilities, and Y is self-efficacy. You could establish the causal relationship that access to appropriate facilities (X) is casually related to level of physical activity (Z) and that self-efficacy (Y) is related to level of physical activity, but access to facilities (X) is not related to self-efficacy (Y).

Diagram 2. $X \rightarrow Y \rightarrow Z$

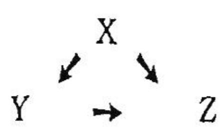
In Diagram 2, X is a direct cause of Y and an indirect cause of Z, mediated by Y. Child abuse provides a good example of this diagram. Y represents the abusive parent, Z the child, and X the parent of the abuser. Literature supports the notion that parents (Y) who abuse their children (Z) tended to have been abused by their parents (X).

Diagram 3.



In Diagram 3, X has a co-causal relationship between Y and Z. In this construct, we may say that X is self-efficacy related to the ability to develop a personal fitness regimen, Y is initiation of activity, and Z is relapse prevention skills. Therefore, self-efficacy (X) would be a causal factor in both program initiation (Y) and relapse prevention (Z) related to physical activity.

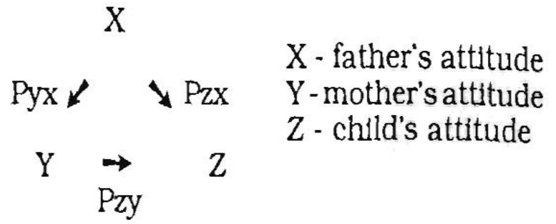
Diagram 4.



In Diagram 4, Z is causally dependent on both X and Y. plus an indirect cause of X via X's impact on Y. The following example In the article clarifies this relationship.

We will use a simple three-variable model with data from one of our pilot studies to illustrate path analysis. It is hypothesized that the parents' alcohol consumption attitudes are the major influence on the child's alcohol attitude. Previous work in the area of alcohol drinking also suggested that husbands' attitudes influence their wives, and further influence their children through the mediation of wives. This relationship is depicted in a path diagram (Figure 2).

FIGURE 2
Path Diagram with Three Variables



In Figure 2, as the arrows point, the child's attitude (Z) is likely affected by the father's and the mother's attitudes. The path coefficients represent the degree of influence (Pzx for the father and Pzy for the mother). To estimate the path coefficients, the following regression analysis needs to be conducted:

$$Z = Pzx\ X + Pzy\ Y$$

Here the dependent variable Z is the child's attitude. X and Y represent the father's and mother's attitudes, while Pzx and Pzy are the standardized regression coefficients referred to as beta weights. The SPSS command may be:

REGRESSION VARIABLES = Z X Y/ DEPENDENT = Z/METHOD = ENTER X Y

By the same token, the father's direct influence on the mother (Pyx) is estimated by the regression:

$Y = Pyx\ X$ and the SPSS command:

REGRESSION VARIABLES = Y X/ DEPENDENT = Y/METHOD = ENTER X

Three path coefficients were obtained from the above two regression analyses:

$$Pyx = .59$$

$$Pzx = .45$$

$$Pzy = .30$$

Like the regression coefficients, these path coefficients suggest that a unit change of one variable will produce an accompanying change of another variable. For example, a unit change of the father's attitude will produce a change of .45 in the child's attitude. The results tell us that the father's direct influence on the child is greater than the direct influence of the mother (.45 vs .30). In addition, according to the model, the father also indirectly influences the child via his impact on the mother and the degree of this indirect influence is estimated by the product of Pyx and Pzy. Thus, the father's total influence on the child is:

$$Pzx + Pzy * Pyx$$

Here Pzx (.45) represents the father's direct influence and the product of Pyx * Pzy (.59 x .30 = .18) represents the indirect influence on the child. Consequently, the father's total influence on the child is .45 + .18 = .63. That is to say, if the father's attitude changes one unit, the accompanying change in the child's is not .45, but .63. This example shows how path analysis enables us to determine the direct and substantively meaningful indirect effect of predictors.

For purpose of comparison, a stepwise regression model was conducted to determine the influences of the mother and father variables on child's attitude with the following SPSS commands:

REGRESSION VARIABLES = Z X Y/ DEPENDENT = Z/ METHOD = STEPWISE X Y

The results: the mother variable was excluded and only the father variable was retained with a beta weight of .63 which is the same as the total father influence obtained from path analysis. According to the results, we would conclude that the father, not the mother, significantly influenced the child. The truth is, however, the degree of the father influence (beta weight of .63) may consist of direct and indirect components. Obviously, the stepwise regression does not enable us to decompose the results.

When a bivariate analysis was conducted, with two separate regressions to predict the child's attitude - one with the father as the predictor and another with the mother as the predictor:

$$Z = B X$$

$$Z = B Y$$

It turned out that both regression models were significant. The beta weights (B) were .63 and .56 for the father and the mother, respectively. Obviously these findings provided only a partial picture of the true results without taking into account the indirect effect of the father over the child. Nor does the beta weight (.56) of the mother reflect the influence from the father.

It is worth noting that path analysis is not a procedure for demonstrating causality. It only estimates the magnitude of the linkages between variables and uses these estimates to provide information about the underlying causal structure imposed by the researcher. Like any statistics, path analysis also requires assumptions. Readers should keep in mind that seeking causal structure may not always be possible and imposing causal structure could lead to misinterpretations of data. Nevertheless, it is our belief that the risk is worth taking. The causal modeling approach helps us make clearer statement of problems, formulate more testable hypothesis, and provide greater insight into understanding the phenomenon, and thus, advance knowledge in the health discipline.