Draft chapter for:
Handbook of Causal Analysis for Social Research, S. Morgan (Ed.) (Springer 2012).

TECHNICAL REPORT
R-393
January 2012

EIGHT MYTHS ABOUT CAUSALITY AND STRUCTURAL EQUATION MODELS

Kenneth A. Bollen, University of North Carolina at Chapel Hill
Judea Pearl, University of California at Los Angeles

EIGHT MYTHS ABOUT CAUSALITY AND STRUCTURAL EQUATION MODELS

Social scientists' interest in causal effects is as old as the social sciences. Attention to the philosophical underpinnings and the methodological challenges of analyzing causality has waxed and waned. Other authors in this volume trace the history of the concept of causality in the social sciences and we leave this task to their skilled hands. But we do note that we are at a time when there is a renaissance, if not a revolution in the methodology of causal inference, and structural equation models play a major role in this renaissance.

Our emphasis in this chapter is on causality and structural equation models (SEMs). If nothing else, the pervasiveness of SEMs justifies such a focus. SEM applications are published in numerous substantive journals. Methodological developments on SEMs regularly appear in journals such as *Sociological Methods & Research, Psychometrika, Sociological Methodology, Multivariate Behavioral Research, Psychological Methods,* and *Structural Equation Modeling*, not to mention the econometrics literature. Over 3,000 subscribers belong to SEMNET, a listserv devoted to SEMs. Thus interest in SEMs is high and continues to grow.

Discussions of causality in SEMs are hardly in proportion to their widespread use. Indeed, criticisms of using SEM in analysis of causes are more frequent than explanations of the role causality in SEMs. Misunderstandings of SEMs are evident in many of these. Some suggest that there is only one true way to attack causality and that way excludes SEMs. Others claim that SEMs are equivalent to regression analysis or that SEM methodology is incompatible with intervention analysis or the potential outcome framework. On the other hand, there are valid concerns that arise from more thoughtful literature that deserve more discussion. We will address both the distortions and the insights from critics in our chapter.

Our paper is organized by presenting eight myths about causality and SEMs in the hope that this will lead to a more accurate understanding. More specifically, the eight myths are: (1) SEMs aim to establish causal relations from associations, (2) SEMs and regression are essentially equivalent, (3) No causation without manipulation, (4) SEMs are not equipped to handle nonlinear causal relationships, (5) A potential outcome framework is more principled than SEMs, (6) SEMs are not applicable to experiments with randomized treatments, (7) Mediation analysis in SEMs is inherently noncausal, and (8) SEMs do not test any major part of the theory against the data.

In the next section we provide the model and assumptions of SEM. The primary section on the eight myths follows and we end with our conclusion section.

MODEL AND ASSUMPTIONS OF SEMs

Numerous scholars across several disciplines are responsible for the development of

and popularization of SEMs. Blalock (1960, 1961, 1962, 1969), Duncan (1966, 1975), Jöreskog (1969; 1970, 1973), and Goldberger (1972; Goldberger & Duncan, 1973) were prominent among these in the wave of developments in the 1960s and 1970s. But looking back further and if forced to list just one name for the origins of SEMs, Sewall Wright (1918, 1921, 1934), the developer of path analysis, would be a good choice.

Over time this model has evolved in several directions. Perhaps the most popular general SEM that takes account of measurement error in observed variables is the LISREL model proposed by Jöreskog & Sörbom (1978). This model simplifies if measurement error is negligible as we will illustrate below. But for now we present the general model so as to be more inclusive in the type of structural equations that we can handle. We also note that this model is linear in the parameters and assumes that the coefficients are constant over individuals. Later when we address the myth that SEMs cannot incorporate nonlinearity, we will present a more general nonparametric form of SEMs which relaxes these assumptions. But to keep things simpler, we now stay with the widely used linear SEM with constant coefficients.

This SEM consists of two major parts. The first is a set of equations that give the causal relations between the substantive variables of interest, also called "latent variables", because they are often inaccessible to direct measurement (Bollen, 2002). Self-esteem, depression, social capital, and socioeconomic status are just a few of the numerous variables that are theoretically important but are not currently measured without substantial measurement error. The latent variable model gives the causal relationships between these variables in the absence of measurement error. It is[1]

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_\eta + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i \qquad (1)$$

The second part of the model ties the observed variables or measures to the substantive latent variables in a two equation measurement model of

$$\mathbf{y}_i = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \qquad (2)$$

$$\mathbf{x}_i = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi}_i + \boldsymbol{\delta}_i \qquad (3)$$

In these equations, the subscript of *i* stands for the *i*th case, $\boldsymbol{\eta}_i$ is the vector of latent endogenous variables, $\boldsymbol{\alpha}_\eta$ is the vector of intercepts, $\mathbf{B}$ is the matrix of coefficients that give the expected effect[2] of the $\boldsymbol{\eta}_i$ on $\boldsymbol{\eta}_i$ where its main diagonal is zero[3], $\boldsymbol{\xi}_i$ is the vector of latent exogenous variables, $\boldsymbol{\Gamma}$ is the matrix of coefficients that give the

[1] The notation slightly departs from the LISREL notation in its representation of intercepts.

[2] The expected effect refers to the expected value of the effect of one η on another.

[3] This rules out a variable with a direct effect on itself.

expected effects of ξ_i on η_i, and ζ_i is the vector of equation disturbances that consists of all other influences of η_i that are not included in the equation. The latent variable model assumes that the mean of the disturbances is zero [E(ζ_i)=**0**] and that the disturbances are uncorrelated with the latent exogenous variables [COV(ζ_i, ξ_i)=**0**]. If on reflection a researcher's knowledge suggests a violation of this latter assumption, then the offending variable is not exogenous and should be included as an endogenous latent variable in the model.

The covariance matrix of ξ_i is **Φ** and the covariance matrix of ζ_i is **Ψ**. The researcher determines whether these elements are freely estimated or are constrained to zero or some other value.

In the measurement model $\mathbf{y}_i$ is the vector of indicators of η_i, α_y is the vector of intercepts, Λ_y is the factor loading matrix that gives the expected effects of η_i on $\mathbf{y}_i$, and ε_i is the vector of unique factors (or disturbances) that consists of all the other influences on $\mathbf{y}_i$ that are not part of η_i. The $\mathbf{x}_i$ is the vector of indicators of ξ_i, α_x is the vector of intercepts, Λ_x is the factor loading matrix that gives the expected effects of ξ_i on $\mathbf{x}_i$, and δ_i is the vector of unique factors (or disturbances) that consists of all the other influences on $\mathbf{x}_i$ that are not part of ξ_i. The measurement model assumes that the means of disturbances (unique factors) [E(ε_i), E(δ_i)] are zero and that the different disturbances are uncorrelated with each other and with the latent exogenous variables [i.e., COV(ε_i, ξ_i),COV(δ_i, ξ_i),COV(ε_i, ζ_i),COV(δ_i, ζ_i) are all zero]. Each of these assumptions requires thoughtful evaluation. Those that are violated will require a respecification of the model to incorporate the covariance. The covariance matrix for δ_i is $\mathbf{\Theta_\delta}$ and the covariance matrix for ε_i is $\mathbf{\Theta_\varepsilon}$. The researcher must decide whether these elements are fixed to zero or some other constraint or are freely estimated.

The SEM explicitly recognizes that the substantive variables represented in η_i and ξ_i are likely measured with error and possibly measured by multiple indicators. Therefore the preceding separate specification links the observed variables that serve as indicators to their corresponding latent variables. Indicators influenced by single or multiple latent variables are easy to accommodate. Researchers can include correlated disturbances from the latent variable or measurement model by freely estimating the respective matrix entries in the covariance matrices of these disturbances mentioned above (i.e., **Ψ** , $\mathbf{\Theta_\delta}$, $\mathbf{\Theta_\varepsilon}$). If it happens that an observed variable has negligible measurement error, it is easy to represent this by setting the observed variable and latent variable equal (e.g., $x_{3i} = \xi_{3i}$).

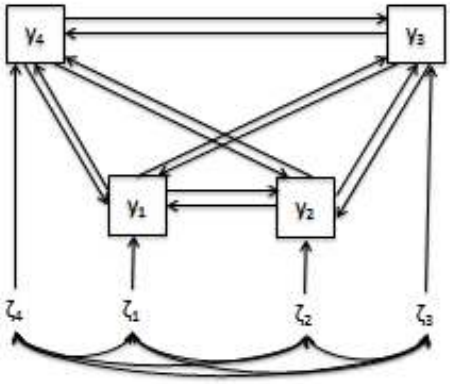
Now we focus on the "Structural" in Structural Equation Models. By structural we mean that the researcher incorporates causal assumptions as part of the model. To illustrate

this, we retreat from the general latent variable structural equation model presented above and make the simplifying assumption that all variables are measured without error. Formally, this means that the measurement model becomes $\mathbf{y}_i = \boldsymbol{\eta}_i$ and $\mathbf{x}_i = \boldsymbol{\xi}_i$. This permits us to replace the latent variables with the observed variables and our latent variable model becomes the well-known simultaneous equation model of

$$\mathbf{y}_i = \boldsymbol{\alpha}_\eta + \mathbf{B}\mathbf{y}_i + \boldsymbol{\Gamma}_i\mathbf{x}_i + \boldsymbol{\zeta}_i \qquad (4)$$

We can distinguish weak and strong causal assumptions. Strong causal assumptions are ones that assume that parameters take specific values. For instance, a claim that one variable has no causal effect on another variable is a strong assumption encoded by setting the coefficient to zero or if one assumes that two disturbances are uncorrelated, then we have another strong assumption that the covariance equals zero. A weak causal assumption excludes some values for a parameter, but permits a range of values. A researcher who includes an arrow between two variables usually makes the causal assumption of a nonzero effect, but if no further restrictions are made then this permits an infinite variety of values (other than zero) and this represents a weak causal assumption. The causal assumption is more restrictive if the researcher restricts the coefficient to be positive, but the causal assumption still permits an infinite range of positive values and is a weaker causal assumption than specifying a specific value such as zero.

Figure 1



[FIGURE 1 ABOUT HERE]

To further explain the nature of causal assumptions, consider the special case of the

simultaneous equations where there are four *y* variables as in Figure 1. In this path diagram the boxes represent observed variables. Single headed straight arrows represent the effect of the variable at the base of the arrow on the variable at the head of the arrow. The two-headed curved arrows connecting the disturbances symbolize possible association among the disturbances. Each disturbance contains all of the variables that influence the corresponding *y* variable but that are not included in the model. The curved arrow connecting the disturbances means that these omitted variables are correlated. The equations that correspond to the path diagram are

$$\begin{aligned} y_1 &= \alpha_1 + \beta_{12} y_2 + \beta_{13} y_3 + \beta_{14} y_4 + \zeta_1 \\ y_2 &= \alpha_2 + \beta_{21} y_1 + \beta_{23} y_3 + \beta_{24} y_4 + \zeta_2 \\ y_3 &= \alpha_3 + \beta_{31} y_1 + \beta_{32} y_2 + \beta_{34} y_4 + \zeta_3 \\ y_4 &= \alpha_4 + \beta_{41} y_1 + \beta_{42} y_2 + \beta_{43} y_3 + \zeta_4 \end{aligned} \qquad (5)$$

with $COV(\zeta_j, \zeta_k) \neq 0 \; for \; j,k$.

As a linear simultaneous equation system, the model in Figure 1 and equation (5) assumes linear relationships, the absence of measurement error, and incorporates only weak causal assumptions that all coefficients and covariances among disturbances are nonzero. All other values of the coefficients and covariances are allowed. Other than assuming nonzero coefficients and covariances, this model represents near total ignorance about the data-generating process. Needless to say this model is underidentified in the sense that none of the structural coefficients is estimable from the data

A researcher who possesses causal knowledge of the domain may express this knowledge by bringing stronger causal assumptions to the model and draw their logical consequences. Or a researcher who wants to examine the implications of or plausibility of a set of causal assumptions can impose them on the model and test their compatibility with the data. The two strongest types of causal assumptions are: (1) imposing zero coefficients and (2) imposing zero covariances. For instance, consider the models in Figure 2.

[FIGURE 2 ABOUT HERE]

Figure 2

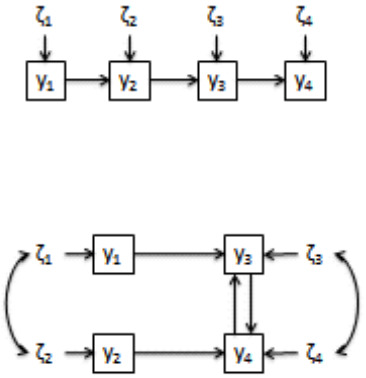


Figure 2a is the same as Figure 1 with the addition of the following strong causal assumptions:

$$\beta_{12} = \beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{31} = \beta_{34} = \beta_{41} = \beta_{42} = 0 \text{ , } C(\zeta_j, \zeta_k) = 0 \; for\, all\; j,k \quad (6)$$

This is a causal chain model. The strong causal assumptions include forcing nine coefficients to zero and setting all disturbance covariances to zero. The weak causal assumptions are that the coefficients and covariances remaining in the model are nonzero. The resulting model differs from that of Figure 1 in two fundamental ways. First, it has testable implications and, second, it allows all of the remaining structural coefficients to be estimable from the data (i.e., identifiable). The set of testable implications of a model as well as the set of identifiable parameters can be systematically identified from the diagram (although some exceptions exist). The ability to systematize these two readings has contributed substantially to our understanding of the causal interpretation of SEM, as well as causal reasoning in general.

Figure 2b shows what results from Figure 1 when imposing a different set of causal assumptions on the coefficients and disturbance covariances. The causal assumptions of Figure 2b are

$$\begin{aligned} &\beta_{12} = \beta_{13} = \beta_{14} = \beta_{21} = \beta_{23} = \beta_{24} = \beta_{32} = \beta_{41} = 0 \text{ , } \\ &C(\zeta_1, \zeta_3) = C(\zeta_1, \zeta_4) = C(\zeta_2, \zeta_3) = C(\zeta_2, \zeta_4) = 0 \end{aligned} \quad (7)$$

The model in Figure 2b has eight strong causal assumptions on the coefficients that are

set to zero and four strong causal assumptions about disturbance covariances set to zero. It can be shown that this model has no testable implications for the strong causal assumptions yet all parameters are identified. The weak causal assumptions of nonzero values for those coefficients and covariances that remain in the model can be tested, given that the strong assumptions hold, but are less informative than the zero coefficient and covariance restrictions present in Figure 2a.

In Figures 1 and 2 we treated only models of observed variables in simultaneous equations. Suppose we stay with the same four *y* variables, but consider them measures of latent variables. The measurement model equation of

$$\mathbf{y}_i = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \qquad (8)$$

covers factor analysis models.

[FIGURE 3 ABOUT HERE]

Figure 3

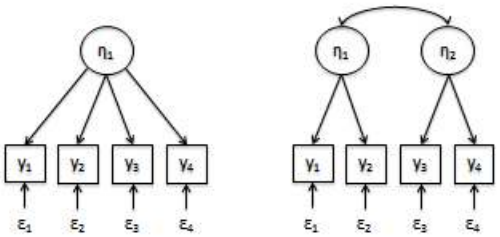


Figure 3 contains two hypothetical measurement models for the four *y* variables that we have used for our illustrations. In the path diagram the ovals or circles represent the latent variables. As stated above these are variables that are part of our theory, but not in our data set. As in the previous path diagrams, the observed variables are in boxes, single-headed arrows stand for direct causal effects, two-headed arrows (often curved) signify sources of associations between the connected variables, though the reasons for their associations are not specified in the model. It could be that they have direct causal influence on each other; that some third set of variables not part of the model influence both; or there could be some other unspecified mechanism (preferential selection) leading them to be associated. The model only says that they are associated and not why. Disturbances ("unique factors") are included in the model not enclosed in circles or boxes. These are the ε s in the diagram. Given that they could be considered as latent variables, they are sometimes enclosed by circles or ovals, though we do not do so here.

In Figure 3a our causal assumptions are that none of the indicators (*y*s) has direct effects on each other and all covariances of disturbances are zero. In other words, the model assumes that a single latent variable (η) explains all the association among the measures (*y*s). In addition, the model assumes that causal influences run from the latent variable to the indicators and that none of the indicators has a causal effect on the latent variable. The weak causal assumptions are that the coefficients (i.e., "factor loadings") in the model are nonzero. Similarly, the strong causal assumptions of Figure 3b are that none of the indicators (*y*s) has direct effects on each other and all covariances of disturbances are zero. But in addition, it assumes that η_1 has zero effect on y_3 and y_4 and that η_2 has zero effect on y_1 and y_2. It also assumes that two correlated latent variables are responsible for any association among the four indicators. It assumes that all causal influences run from the latent variable to the indicators and none in the reverse direction. The weak causal assumptions are that the coefficients and covariance of the latent variables are nonzero.

Imposing different causal assumptions lead to different causal models as illustrated by our examples. The causal assumptions derive from prior studies, research design, scientific judgment, or other justifying sources. In a minority of cases, the causal assumptions are well-supported and widely accepted (e.g., a variable at time 2 cannot cause a variable at time 1). But there are few situations where all causal relations are without challenge.

More typically the causal assumptions are less established, though they should be defensible and consistent with the current state of knowledge. The analysis is done under the perspective of "what if these causal assumptions were true." These latter analyses are useful because there are often ways of testing the whole or parts of the model that can be helpful in rejecting one or more of these causal assumptions and thereby reveals flaws in specification. Of course, passing these tests does not prove the causal assumptions validity, but it lends credibility to them. If we repeatedly test the model in diverse data sets and find good matches to the data then the causal assumptions further gain in their credibility. In addition, when there are competing causal models, equally compatible with the data, an analyst can compare their performances under experimental conditions to see which are best. We will have more to say about testing these causal assumptions later under the myth that SEMs do not permit any testing of these assumptions.

A second reason that the models resulting from causal assumption are valuable is that they enable an estimate of the coefficients (as well as variances, and covariances) that are important for guiding policies. For instance, Figure 2a allows for y_1 having a direct effect on y_2, but it does not specify its magnitude. With SEM estimation, and with the help of the strong assumptions, we can quantify the magnitude of this effect and of other estimated parameters and thus evaluate (albeit provisionally) the merits of interventional policies that depend on this effect.

This ability to quantify effects is available even in a saturated model (as in Figure 2b) when it is not possible to test any of the strong causal assumptions, nor any combination thereof. In such cases, the quantified effects are still useful for policy evaluation, though they are predicated on the validity of modeling assumptions that received no scrutiny by the data.

The traditional path diagram as well as the graphical model notation that we discuss later, makes the causal assumptions of the model clear through the absence of certain arrows, and certain curved arcs (double-headed arrows). The equation forms of these models are equally capable of making these causal assumptions clear, but can be more complicated to interpret and to analyze, especially in their nonparametric form.

EIGHT MYTHS:

In the previous section we presented the model, notation, and causal assumptions for SEM as well as the role of identification, model testing and advice to policy making. A great deal of misinformation on SEMs and causality appears in a variety of publications. Rather than trying to address all such inaccuracies we highlight eight that are fairly frequent and widespread. The remaining part of this section is organized around these myths.

Myth #1 SEMs aim to establish causal relations from associations.

This misunderstanding is striking both in its longevity and in its reach. In essence the critique states that developers and users of SEMs are under the mistaken impression that SEMs can convert associations and partial associations among observed and/or latent variables into causal relations. The mistaken suggestion is that researchers developing or using SEM believe that if a model is estimated and it shows a significant coefficient, then that is sufficient to conclude that a significant causal influence exists between the two variables. Alternatively, a nonsignificant coefficient is sufficient to establish the lack of a causal relation.

As an illustration of these critiques, Guttman (1977, page 97) argues that sociologists using path analysis or causal analysis do so under the mistaken belief that they can use correlation to imply causation between variables. De Leeuw's (1985) influential review of four early SEM manuscripts and books (Long, 1983a, 1983b; Everitt, 1984; Saris and Stronkhorst, 1984) gives an illustration of this claim: "I think that the use of causal terminology in connection with linear structural models of the LISREL type means indulging in a modern, but nevertheless clearly recognizable, version of the 'post hoc ergo propter hoc' fallacy" (page 372). The 'post hoc ergo propter hoc' fallacy is "after this, therefore because of this" where association (with a temporal lag) is incorrectly used to justify a causality claim.

Freedman (1987) critiques recursive path models, a special case of SEM, suggesting

that researchers are assuming causal or structural effects based on associations: “Of course, it is impossible to tell just from data on the variables in it whether an equation is structural or merely an association. In the latter case, all we learn is that the conditional expectation of the response variable shows some connection to the explanatory variables, in the population being sampled” (page 103).

Baumrind (1983) bemoans the tendency of those using SEM to assume that associations lead to causal claims: “Since the publication of Kenny's (1979) book *Correlation and Causation*, there has been an explosion in the research literature of studies making causal inferences from correlational data in the absence of controlled experiments.” (page 1289).

If these distorted portrayals ended in the 1980s, there would be little need to mention them today. But they have not. Goldthorpe (2001, page 11) suggests “causal path analysis” is regarded as a “means of inferring causation directly from data…”. Freedman (2004) suggests that: “Many readers will ‘know’ that causal mechanisms can be inferred from nonexperimental data by running regressions” (page 268) and he asks readers to suspend this belief. Or look at Sobel (2008, page 114) who writes: “First, there is a putative cause *Z* prior in some sense to an outcome *Y*. Futhermore, *Z* and *Y* are associated (correlated). However, if the *Z* - *Y* association vanishes when a (set of) variable(s) *X* prior to *Z* is conditioned on (or in some accounts, if such a set exists), this is taken to mean that *Z* ‘does not cause' *Y*. The use of path analysis and structural equation models to make causal inferences is based on this idea. Granger causation (Geweke, 1984; Granger, 1969) extends this approach to time series.”

Other quotes and authors could be presented, but the clear impression created by them is that SEM users and developers are assuming that we can derive causality from complicated models of partial associations.

Is this true? To address this question, it is valuable to read papers or books that present SEMs to see what they actually say. Duncan (1966) was a key work introducing path analysis or SEMs into sociology and the social sciences. His abstract states: “Path analysis focuses on the problem of interpretation and does not purport to be a method for discovering causes” (page 1).

James, Mulaik, & Brett (1982) published a book devoted to causality in models and they were far from suggesting that mere association (or lack thereof) equals causality. A chapter of Bollen (1989, ch.3) on SEMs begins by saying that a SEM depends on causal *assumptions* and then goes on to examine the threats to and the consequences of violating causal assumptions. The chapter distinguishes the differences between model-data consistency versus model-reality consistency where the latter is essentially impossible to prove. A recent SEM text by Mulaik (2009, Ch.3) devotes a chapter to causation in SEM which deals with the meaning of and threats to establishing causality.

As we explained in the last section, Researchers do not derive causal relations from a

SEM. Rather the SEM incorporates the causal assumptions of the researcher. These assumptions derive from the research design, prior studies, scientific knowledge, logical arguments, temporal priorities, and other evidence that the researcher can marshal in support of them. The credibility of the SEM depends on the credibility of the causal assumptions in each application.

In closing this subsection it is useful to turn to Henry E. Niles, a critic of Wright's path analysis from 1922. He too suggested that path analysis was confusing associations with causation. Wright responded that he "never made the preposterous claim that the theory of path coefficients provides a general formula for the deduction of causal relations …" (Provine, 1986, pages 142-143). Rather as Wright (1921, page 557) had explained: "The method [of path analysis] depends on the combination of knowledge of the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypothesis in regard to them."

The debate from the preceding paragraph occurred 90 years ago. How is it possible that we have the same misunderstandings today?

We see several possible reasons. One is that the critics were unable to distinguish causal from statistical assumptions in SEM, or to detect the presence of the former. An equation from a SEM stripped of its causal assumptions appears identical to a regression equation and the assumptions of zero covariances among disturbance terms and covariates appeared to be statistical in nature. Accordingly, Pearl (2009, pp. 135-138) suggests that notational inadequacies and the hegemony of statistical thinking solely in terms of probability distributions and partial associations contributed to these misunderstandings. Furthermore, SEM researchers were not very effective in explicating both the causal assumptions that enter a model and the "logical consequences" of those assumptions, which Wright considered so essential. For example, many SEM authors would argue for the validity of the weak causal assumptions of nonzero coefficients instead of attending to the strong ones of zero coefficients or covariances. SEM defenders who highlighted the weak over the strong causal assumptions might have contributed to the critics' misunderstanding of the role of causal assumptions in SEM. The development of graphical (path) models, nonparametric structural equations, "do-calculus," and the logic of counterfactuals now makes the causal content of SEM formal, transparent and impossible to ignore by even its worse critics (Pearl, 2009, 2012).

Lest there be any doubt:

SEM does not aim to establish causal relations from associations.

Perhaps the best way to make it clear is to state formally and unambiguously what SEM does aim to establish. SEM is an inference engine that takes in two inputs, qualitative

causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions. Failure to fit the data casts doubt on the strong causal assumptions of zero coefficients or zero covariances and guides the researcher to diagnose, or repair the structural misspecifications. Fitting the data does not reject the causal assumptions, but makes them more plausible though these results need to be replicated and to withstand the criticisms of researchers who suggest other models for the same data.

Myth # 2 SEM and regression are essentially equivalent.

This second misunderstanding also is traced back to the origins of path analysis. In a biography of Wright, Provine (1986, page 147) states that Henry Wallace who corresponded with Wright "kept trying to see path coefficients in terms of well-known statistical concepts, including partial correlation and multiple regression. Wright kept trying to explain how and why path coefficients were different from the usual statistical concepts." More contemporary writings also present SEM as essentially the same as regression.

Consider Holland's (1995, page 54) comment on models: "I am speaking, of course, about the equation: $y = a + bx + \varepsilon$. What does it mean? The only meaning I have ever determined for such an equation is that it is a shorthand way of describing the conditional distribution of *y* given *x*. It says that the conditional expectation of *y* given *x*, $E(y|x)$, is $a + bx$ …)."

More recently the same perspective is expressed by Berk (2004, page 191): "However, the work of Judea Pearl, now summarized in a widely discussed book (Pearl, 2000), has made causal inference for structural equation models a very visible issue. Loosely stated, the claim is made that one can routinely do causal inference with regression analysis of observational data." In the same book, Berk (2004, page 196) says: "The language of Pearl and many others can obscure that, beneath all multiple equation models, there is only a set of conditional distributions. And all that the data analysis can do by itself is summarize key features of those conditional distributions. This is really no different from models using single equations. With multiple equations, additional complexity is just laid on top. Including some more equations per se does not bring the researcher any closer to cause and effect."

The gap between these critics and the actual writings on SEM is wide. The critics do not directly address the writings of those presenting SEM. For instance, Goldberger (1973, page 2) has a succinct description of the difference between a SEM and a regression: "In a structural equation model each equation represents a causal link rather than a mere empirical association. In a regression model, on the other hand, each equation represents the conditional mean of a dependent variable as a function of explanatory variables." Admittedly, Goldberger's quote emphasizes the weak causal

assumptions over the strong causal assumptions as distinguished by us earlier, but it does point to the semantic difference between the coefficients originating with a regression where no causal assumptions are made versus from a structural equation that makes strong and weak causal assumptions.

In light of the lingering confusion regarding regression and structural equations, it might be useful to directly focus on the difference with just a single covariate. Consider the simple regression equation:

$$Y_i = \alpha_y + \beta_{yx} X_i + \zeta_{yi}$$

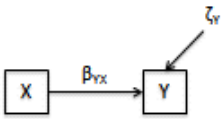
whose aim is to describe a line from which we can "best" predict Y_i from X_i. The slope is a regression coefficient. If prediction is the sole purpose of the equation, there is no reason that we could not write this equation as

$$X_i = \alpha_x + \beta_{xy} Y_i + \zeta_{xi}$$

where $\alpha_x = -\beta_{yx}^{-1}\alpha_y$, $\beta_{xy} = \beta_{yx}^{-1}$, and $\zeta_{xi} = -\beta_{yx}^{-1}\zeta_{yi}$ and use it to predict X from observations of Y. However, if the first equation, $Y_i = \alpha_y + \beta_{yx} X_i + \zeta_{yi}$, is a structural equation *then* β_{yx} is a structural coefficient that tells us the causal effect on Y_i for a one unit difference in X_i. With this interpretation in mind, a new structural equation will be needed to describe the effect of Y on X (if any); the equation $X_i = \alpha_x + \beta_{xy} Y_i + \zeta_{xi}$ (with $\beta_{xy} = \beta_{yx}^{-1}$) will not serve this purpose.

A similar confusion arises regarding the so called "error term" ζ. In regression analysis ζ stands for whatever deviation remains between Y and its prediction $\beta_{yx} X_i$. It is therefore a human made quantity, which depends on the goodness of our prediction. Not so in structural equations. There, the "error term" stands for substantive factors and an inherent stochastic element omitted from the analysis. Thus, whereas errors in regular regression equations are by definition orthogonal to the predictors, errors in structural equations may or may not be orthogonal, the status of which constitutes a causal assumption which requires careful substantive deliberation. It is those substantive considerations that endow SEM with causal knowledge, capable of offering policy-related conclusions (see Pearl, 2011b).

Figure 4



[FIGURE 4 ABOUT HERE]

The ambiguity in the nature of the equation is removed when a path diagram (graphical model) accompanies the equation or, when the equality sign is replaced by an assignment symbol :=, which is used often in programming languages to represent asymmetrical transfer of information, and here represents a process by which nature assigns values to the dependent variable in response to values taken by the independent variables.

In addition to judgments about the correlation of ζ_{yi} with X_i, the equation $Y_i = \alpha_y + \beta_{yx} X_i + \zeta_{yi}$, embodies three causal assumptions that the model builder should be prepared to defend:

1. linearity – a unit change from X=x to X=x+1 will result in the same increase of Y as a unit change from X = x' to X = x'+1.
2. exclusion – once we hold X constant, changes in all other variables (say Z) in the model will not affect Y. (This assumption applies when the model contains other equations. For instance, if we added an equation $X_i = \alpha_x + \beta_{xz} Z_i + \zeta_{xi}$ to the model in Figure 4, then changes in Z have no effect on Y once X is held constant.)
3. Homogeneity – every unit in the population has the same causal effect β_{yx}.

We can write the first two assumptions in the language of do-calculus as:

E(Y| do(x), do(z)) = α_y + β_{yx}x

which can be tested in controlled experiments. The third assumption is counterfactual, as it pertains to each individual unit in the population., and cannot therefore be tested at the population level.

In the path diagram of Fig. 4, the single-headed arrow from X_i to Y_i, the absence of an arrow from Y_i to X_i, and the lack of correlation of the disturbance with X_i clearly represent the causal assumptions of the model in a way that the algebraic equation does not. The causal assumptions can be challenged by researchers or in more complicated models the set of causal assumptions could prove inconsistent with the data and hence worthy of rejection. However, the claim that a structural equation and a regression equation are the same thing is a misunderstanding that was present nearly a century ago and has lingered to the current day, primarily because many critics are either unaware of the difference or find it extremely hard to accept a new type of assumptions cast in a language that is not part of standard statistics.

Myth #3 No causation without manipulation.

In an influential *JASA* article, Paul Holland (1986, page 959) wrote on causal inference, he discusses the counterfactual or potential outcome view on causality. Among other points, Holland (1986, page 959) states that some variables can be causes and others cannot: "The experimental model eliminates many things from being causes, and this is probably very good, since it gives more specificity to the meaning of the word cause. Donald Rubin and I once made up the motto

NO CAUSATION WITHOUT MANIPULATION

to emphasize the importance of this restriction."

Holland uses race and sex as examples of "attributes" that cannot be manipulated and therefore cannot be causes and explicitly criticized SEMs and path diagrams for allowing arrows to emanate from such attributes. In the extreme case of viewing manipulation as something done by humans only, we would reach absurd conclusions such as there was no causation before humans evolved on earth. Or we would conclude that the "moon does not cause the tides, tornadoes and hurricanes do not cause destruction to property, and so on" (Bollen, 1989, page 41). Numerous researchers have questioned whether such a restrictive view of causality is necessary. For instance, Glymour (1986), a philosopher, commenting on Holland's (1986) paper finds this an unnecessary restriction. Goldthorpe (2001, page 15) states: "The more fundamental difficulty is that, under the - highly anthropocentric – principle of 'no causation without manipulation', the recognition that can be given to the action of individuals as having causal force is in fact peculiarly limited."

Bhrolchain & Dyson (2007, page 3) critique this view from a demographic perspective: "Hence, in the main, the factors of leading interest to demographers cannot be shown to be causes through experimentation or intervention. To claim that this means they cannot be causes, however, is to imply that most social and demographic phenomena do not have causes—an indefensible position. Manipulability as an exclusive criterion is defective in the natural sciences also." Economists Angrist & Pischke (2009, page 113) also cast doubt on this restrictive definition of cause.

Pearl (2011) further shows that this restriction has led to harmful consequence by forcing investigators to compromise their research questions only to avoid the manipulability restriction. The essential ingredient of causation, as argued in Pearl (2009, p. 361) is responsiveness, namely, the ability of some variables to respond to variations in other variables, regardless of how those variations came about.

Despite this and contrary to some critics, the restriction of "no causation without manipulation" is not incompatible with SEM. A SEM specification incorporates the causal assumptions of the researcher. If a researcher believes that causality is not possible for "attributes" such as "race" and "gender," then the SEM model of this researcher should treat those attributes as exogenous variables and avoid asking any query regarding their "effects."[4] Alternatively, if a researcher believes that such attributes can serve as causes, then such attributes can act as ordinary variables in the SEM, without restrictions on queries that can be asked.

Myth # 4 The potential outcome framework is more principled than SEMs.

The difficulties many statisticians had in accommodating or even expressing causal assumptions have led them to reject Sewell Wright's ideas of path analysis as well as the SEMs adapted by econometricians and social scientists in the 1950s to 1970s. Instead, statisticians found refuge in Fisher's invention of randomized trials (Fisher, 1931), where the only assumptions needed were those concerning the nature of randomization, and required no mathematical machinery for cause-effect analysis. Many statisticians clung to this paradigm as long as they could and, later on, when mathematical analysis of causal relations became necessary, they developed the Neyman-Rubin "potential outcome" (PO) notation (Rubin, 1974) and continued to oppose structural equations as a threat to principled science (Rubin, 2004, 2009, 2010; Sobel, 2008). The essential difference between the SEM and PO frameworks is that the former encodes causal knowledge in the form of functional relationships among ordinary variables, observable as well as latent, while the latter encodes such knowledge in the form of statistical relationships among hypothetical (or counterfactual) variables, whose value is determined only after a manipulative treatment is enacted For example, to encode the causal assumption that X does not cause Y (represented by the absence of an X→Y arrow in SEM) the PO analyst imagines a hypothetical variable Y_x (standing for the value that Y would attain had treatment X=x been administered) and writes Y_x = Y, meaning that, regardless of the value of x, the potential outcome Y_x will remain unaltered, and will equal the observed value Y. Likewise, the SEM assumption of independent disturbances is expressed in the PO framework as an independence relationship between counterfactual variables such asY_x_1, Y_x_2, X_y_1 , Z_x_2 etc. A systematic analysis of the syntax and semantics of the two notational systems reveals that they are logically equivalent (Galles and Pearl, 1998;

[4] A researcher could use the specific effects techniques proposed in Bollen (1987) to eliminate indirect effects originating with or going through any "attributes" when performing effect decomposition.

Halpern 1998); a theorem in one is a theorem in the other, and an assumption in one has a parallel interpretation in the other. Although counterfactuals variables do not appear explicitly in the SEM equations, they can be derived from the SEM using simple rules developed in Pearl (2009, pp. 101) and illustrated in Pearl (2012).

Remarkably, despite this equivalence, potential outcome advocates have continued to view SEM as a danger to scientific thinking, labeling it an "unprincipled" "confused theoretical perspective," "bad practical advice," "theoretical infatuation," and "nonscientific ad hockery".(Rubin SIM (2009). The ruling strategy in this criticism has been to lump SEM, graphs and regression analysis under one category, called "observed outcome notation," and blame the category for the blemishes of regression. "The reduction to the observed outcome notation is exactly what regression approaches, path analyses, directed acyclic graphs, and so forth essentially compels one to do." (Rubin, 2010, p. 39)." "These fundamental ideas with the science fixed cannot be clearly represented by graphs or paths, which are effectively wedded to the observed outcome notation, and these other approaches do not create a clear distinction between the science and the assignment mechanism;" (ibid, pp. 44)

The scientific merits of this assault surface in the fact that none of the critics has thus far acknowledged the 1998 proofs of the logical equivalence of SEM and PO, and none has agreed to compare the cognitive transparency of the two notational systems (which favors SEM, since PO becomes utterly incomprehensible when the number of variables exceeds three (Pearl, 2011).

Instead, the critics continue to discredit SEM from a safe distance of hands-off dismissal: "[we] are unconvinced that directed graphical models (DGMs) are generally useful for "finding causal relations" or estimating causal effects." (Lindquist & Sobel, 2011).

Notwithstanding these critics, a productive symbiosis has emerged that combines the best features of the two approaches (Pearl, 2010). It is based on encoding causal assumptions in the transparent language of (nonparametric) SEM, translating these assumptions into counterfactual notation, and then giving the analyst an option of either pursuing the analysis algebraically in the calculus of counterfactuals or use the inferential machinery of graphical models to derive conclusions concerning identification, estimation and testable implications. This symbiosis has become routine in epidemiology and the health sciences and is making its way slowly into the social sciences (Morgan and Winship, 2007; Muthén, 2011).

Myth #5 SEMs are not equipped to handle nonlinear causal relationships.

The SEM presented so far is indeed linear in variables and in the parameters. We can generalize this in several ways. First, there is a fair amount of work on including interactions and quadratics of the latent variables into the model (e.g., Schumacker & Marcoulides, 1996). These models stay linear in the parameters, though they are

nonlinear in the variables. Another nonlinear model arises when the endogenous observed variables are not continuous. Here dichotomous, ordinal, counts, censored, and multinomial observed variables might be present. Fortunately, such variables are easy to include in SEMs, often by formulating an auxiliary model that links the noncontinuous observed variables to an underlying continuous variable via a series of thresholds or through formulations that deal directly with the assumed probability distribution functions without threshold models (e.g., Muthen, 1984; Skrondal & Rabe-Hesketh, 2005).

We could further develop the classic SEM in a more general nonlinear or nonparametric form by writing the latent variable model as

$$\boldsymbol{\eta}_i = f_\eta(\eta_i, \boldsymbol{\xi}_i, \boldsymbol{\zeta}_i)$$

and the two equation measurement model as

$$\mathbf{y}_i = f_y(\boldsymbol{\eta}_i, \boldsymbol{\varepsilon}_i)$$
$$\mathbf{x}_i = f_x(\boldsymbol{\xi}_i, \boldsymbol{\delta}_i)$$

The symbols in these equations are the same as defined earlier. The new representations are the functions which provide a general way to represent the connections between the variables within the parentheses to those on the left hand side of the equation.

Graphical models are natural for representing nonparametric equations (See chapter xx of this volume) for they highlight the assumptions and abstract away unnecessary algebraic details. In contrast to the usual path diagrams, no commitment is made to the functional form of the equations.

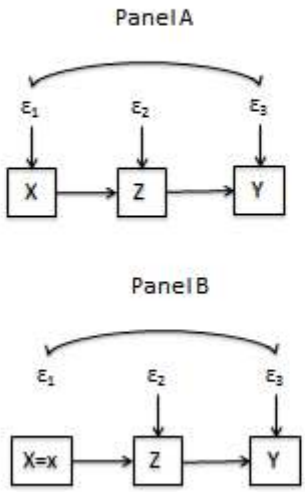
To illustrate, consider the following model:

$x = f(\varepsilon_1)$ $\quad z=g(x, \varepsilon_2)$ $\quad y=h(z,\varepsilon_3)$

with ε_2 is independent of $\{ \varepsilon_1, \varepsilon_3\}$ (See Pearl 2000, figure 3.5). Figure 5 is a graph of the model where the single headed arrows stand for nonlinear functions and the curved two headed arrow connecting $\{ \varepsilon_1, \varepsilon_3\}$ represents statistical dependence between the two error terms, coming from an unspecified source.

[FIGURE 5 ABOUT HERE]

Figure 5



Assume that we face the task of estimating the causal effect of X on Y from sample data drawn from the joint distribution Pr(x,y,z) of the three observed variables, X, Y and Z. Since the functions f, g, and h are unknown, we cannot define the effect of X on Y, written Pr(Y=y|do(X=x)), in terms of a coefficient or a combination of coefficients, as is usually done in parametric analysis. Instead, we need to give the causal effect a definition that transcends beyond parameters and captures the essence of intervening on X and setting it to X=x, while discarding the equation $x = f(\varepsilon_1)$ that previously governed X

This we do by defining Pr(Y=y|do(X=x)) as the probability of Y=y in a modified model, in which the arrow from ε_1 to X is removed, X is set to the value x and all the other functions and covariances remain intact. See Figure 5B.

Remarkably, despite the fact that no information is available on the functions f, g, and h, we can often identify causal effects and express them in terms of estimable quantities. In the example above (Pearl, 2000, p.81) the resulting expression is (assuming discrete variables)

$$\Pr(Y = y \mid do(X = x)) = \sum_{z} \Pr(Z = z \mid X = x) \sum_{x'} \Pr(Y = y \mid X = x', Z = z) \Pr(X = x')$$

All terms in the right hand side of the equation are estimable from data on the observed variables X, Y, and Z. Moreover, a logical machinery (called do-calculus) can derive such expressions automatically from any given graph, whenever a reduction to estimable quantities is possible. Finally, a complete graphical criterion has been derived that enables a researcher to inspect the graph and write down the estimable

expression, whenever such expressions exist.(Shpither and Pearl, 2008a).

This example also demonstrates a notion of "identification" that differs from its traditional SEM aim of finding a unique solution to a parameter, in terms of the means and covariances of the observed variables. The new aim is to find a unique expression for a policy or counterfactual question in terms of the joint distribution of observed variables. This method is applicable to both continuous and discontinuous variables and has been applied to a variety of questions, from unveiling the structure of mediation to finding causes of effects, to analyzing regrets for actions withheld (Shpitser and Pearl, 2009) Concrete examples are illustrated in (Pearl, 2009, 2012)

Myth # 6 SEMs are less applicable to experiments with randomized treatments.

This misunderstanding is not as widespread as the previous ones. However, the heavy application of SEMs to observational (nonexperimental) data and its relative infrequent use in randomized experiments have led to the impression that there is little to gain from using SEMs with experimental data. This is surprising when we consider that in the 1960s through 1980s during the early spread of SEMs, there were several papers and books that pointed to the value of SEMs in the analysis of data from experiments (e.g., Blalock, 1985; Costner, 1971; Miller, 1971; Kenny, 1979: chapter 10).

Drawing on these sources we summarize valuable aspects of applying SEMs to experiments. In brief, SEMs provide a useful tool to help to determine (1) if the randomized stimulus actually affects the intended variable ("manipulation check"), (2) if the output measure is good enough to detect an effect, (3) if the hypothesized mediating variables serve as the mechanism between the stimulus and effect, and (4) if other mechanism, possibly confounding ones, link the stimulus and effect. These tasks require assumptions, of course, and SEM's power lies in making these assumptions formal and transparent.

Figure 6a illustrates issues (1) and (2). Suppose X is the randomized stimulus intended to manipulate the latent variable η_1 and η_2 is the latent outcome variable measured by Y. A social psychologist, for instance, might want to test the hypothesis that frustration (η_1) is a cause of aggression (η_2). The stimulus (X) for frustrating the experiment subjects is to ask them to do a task at which they fail whereas an easier task is given to the control group. The measure of frustration is Y.

Even if frustration affects aggression (i.e., $\eta_1 \rightarrow \eta_2$), it is possible that the ANOVA or regression results for Y and X are not statistically or substantively significant. One reason for this null result could be that the stimulus (X) has a very weak effect on frustration (η_1), that is, the $X \rightarrow \eta_1$ effect is near zero. Another reason could be that Y is a poor measure of aggression and the path of $\eta_2 \rightarrow Y$ is near zero. The usual ANOVA/regresson approach would not reveal this.

Figure 6

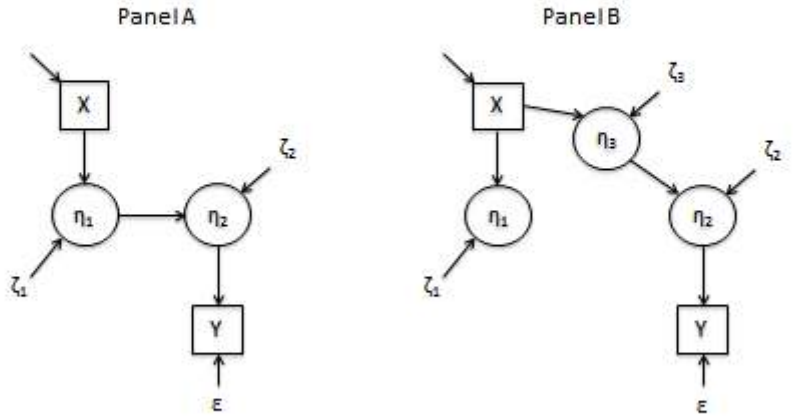
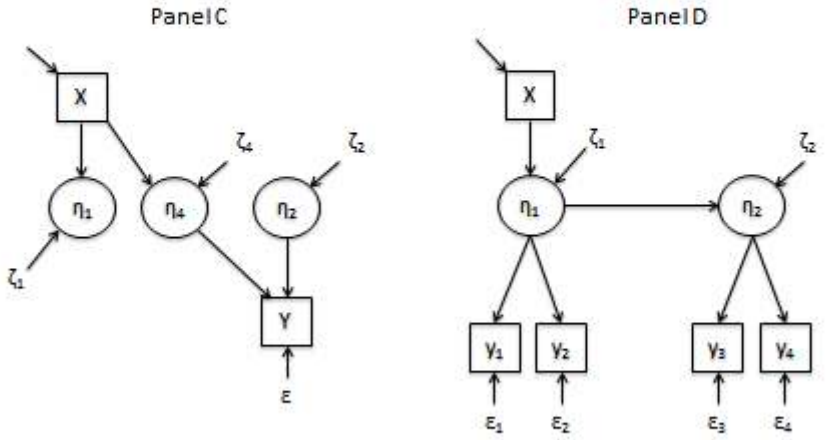


Figure 6



[FIGURE 6 ABOUT HERE]

Points (3) and (4) are illustrated with Figures 6b and 6c. In Figure 6b the stimulus causes another latent variable (η_3) besides frustration which in turn causes aggression (η_2). Here frustration is not the true cause of aggression and is not the proper mechanism for explaining an association of *Y* and *X*. Rather it is due to the causal paths of $X \rightarrow \eta_3 \rightarrow \eta_2 \rightarrow Y$. The η_3 variable might be demand characteristics where the subject shapes her response to please the experimenter or it could represent experimenter biases. Figure 6c is another case with a significant *Y* and *X* association, yet the path $\eta_1 \rightarrow \eta_2$ is zero. Here the stimulus causes a different latent variable (η_4) which does not cause η_2 but instead causes *Y.*

A SEM approach that explicitly recognizes the latent variables hypothesized to come between the experimental stimulus and the outcome measure provides a means to detect such problems. Costner (1971), for instance, suggests that a researcher who collects two effect indicators of η_1 (say, Y_1 and Y_2) and two effect indicators of η_2 (say, Y_3 and Y_4) can construct a model as in Figure 6d.

This model is overidentified and has testable implications that must hold if it is true. We talk more about testing SEMs below, but for now suffice it to say that under typical conditions this model would have a poor fit if Figures 6b or 6c were true. For instance, a stimulus with a weak effect on frustration (η_1) would result in a low to zero R-squared for η_1. A weak measure of aggression would be reflected in a weak R-squared for the measure of aggression.

Our discussion just scratches the surface of the ways in which SEM can improve the analysis of experiments. But this example illustrates how SEM can help aid manipulation checks, assess the quality of outcome measures, and test the hypothesized intervening mechanisms while controlling for measurement error.

Myth # 7 -- SEM is not appropriate for mediation analysis

Mediation analysis aims to uncover causal pathways along which changes are transmitted from causes to effects. For example, an investigator may be interested in assessing the extent to which gender disparity in hiring can be reduced by making hiring decisions gender-blind, compared with say eliminating gender disparity in education or job qualifications. The former concerns the "direct effect" (of gender on hiring) and the latter the "indirect effect" or the "effect mediated via qualification".

The myth that SEM is not appropriate for mediation analysis is somewhat ironic in that much of the development of mediation analysis occurred in the SEM literature. Wright (1923; 1934) used path analysis and tracing rules to understand the various ways in which one variable's effect on another might be mediated through other variables in the model. The spread of path analysis through the social sciences from the 1960s to 1980s also furthered research on decomposition of effects and the study of mediation. Much research concentrated on simultaneous equations without latent variables (e.g.,

Duncan, 1975; Fox, 1980, Baron & Kenny, 1983). More general treatments that include latent variables also were developed (e.g., Jöreskog & Sörbom, 1981) which included asymptotic standard error estimates of indirect effects (Folmer, 1981; Sobel, 1986; Bollen & Stine, 1990) and the ability to estimate the effects transmitted over any path or combination of paths in the model (Bollen, 1987).

Although these methods were general in their extension to latent as well as observed variable models, they were developed for linear models. There was some limited work on models with interaction terms or quadratic terms (Stolzenberg, 1979) and other work on limited dependent models (Winship & Mare, 1983). But these discussions were not generalized to all nonparameteric models where the meaning of "mediation" or its policy implications are not clear cut. Pearl (2001) has extended SEM meditational analysis to nonparametric models in a symbiotic framework based on graphs and counterfactual logic. It generalizes earlier SEM work on mediation by treating general nonparametric models, but it concentrates on models that focus on observed variables.

This symbiotic mediation theory has led to three major advances:

1. Formal definitions of direct and indirect effects that are applicable to models with both continuous and categorical variables. These definitions set objective standards for mediation analysis, for they are universally applicable across domains, and retain their validity regardless of the underlying data-generating models.

2. The establishment of conceptually meaningful conditions (or assumptions) under which direct and indirect effects can be estimated from either experimental or observational studies, while making no commitment to distributional or parametric assumptions.

3. The derivation of a simple estimand, called *Mediation Formula*, that measures (subject to the conditions in (2)) the extent to which the effect of one variable (X) on another (Y) is mediated by a set (W) of other variables in the model. This estimand assesses both the extent to which mediation is necessary and the extent to which it is insufficient for explaining a given effect (the two are not necessarily the same). This development allowed researchers to cross the linear-nonlinear barrier and has spawned a rich literature in nonparametric mediation analysis (Imai et al., 2010; Muthen, 2011; Pearl, forthcoming, Intervention Science, VanderWeele & Vansteelandt, 2009). These were shunned however by PO researchers who, constrained by the "no causation without manipulation" paradigm, could not accommodate a mediator that is not manipulable. Instead, a new framework was proposed under the rubric "Principal Strata Framework" which defines direct effect with no attention to structure or mechanisms.

Whereas the structural interpretation of "direct effect" measures the effects that would be transmitted in the population with all mediating paths (hypothetically) deactivated, the Principal Strata Direct Effect (PSDE) was defined as the effects transmitted in those subjects only for whom mediating paths *happened to be inactive* in the study. This

seemingly mild difference in definition leads to paradoxical results that stand in glaring contradiction to common usage of direct effects, and excludes from the analysis all individuals who are both directly and indirectly affected by the causal variable *X* (Pearl, 2009 (R-350); 2011 (R-382)). For example, a grandfather would be deemed to have no direct effect on his grandson's behavior in families where he has had some effect on the father. This precludes from the analysis all typical families, in which a father and a grandfather have simultaneous, complementary influences on children's upbringing. In linear systems, to take an even sharper example, the PSDE direct effect would be undefined whenever indirect paths exist from the cause to its effect. The emergence of such paradoxical conclusions underscores the wisdom, if not necessity of a symbiotic analysis, in which counterfactual relationships are governed by their structural interpretation.

Indeed, in a recent discussion concerning the utility of the Principal Strata framework, the majority of discussants have concluded that "There is nothing within the principal stratification framework that corresponds to a measure of an "indirect" or "mediated" effect," (Vander Weele 2011),that "it is not the appropriate tool for assessing "mediation" (ibid), that it contains "good ideas taken too far" (Joffe, 2011, p. 1) that "when we focus on PSDEs we effectively throw the baby out with the bath-water [and] ... although PSDE is a proper causal effect, it cannot be interpreted as a *direct* effect" (Sjolander 2011, p.1-2). Even discussants, who found the principal stratification framework to be useful for some purposes, were quick to discount its usefulness in mediation analysis.

As we remarked earlier, the major deficiency of the PO paradigm is its rejection of structural equations as a means of encoding causal assumptions and insisting instead on expressing all assumptions in the opaque notation of “ignorability” conditions. Such conditions are extremely difficult to interpret (unaided by graphical tools); and "are usually made casually, largely because they justify the use of available statistical methods and not because they are truly believed." (Joffe, et al., 2010).

Even the most devout advocates of the ”ignorability" language use ”omitted factors" when the need arises to defend assumptions in any real setting (e.g., Sobel, 2008). SEM's terminology of "omitted factors", "confounders," "common causes," and "path models" has become the standard communication channel among mediation researchers

In short, SEM largely originated and it remains at the core of mediation analysis.

Myth #8 SEMs do not test any major part of the theory against the data.

In a frequently cited critique of path analysis, Freedman (1987, page 112) argues that “path analysis does not derive the causal theory from the data, or test any major part of it against the data.”[5] This statement is both vacuous and complimentary. It is vacuous

[5] The first part of the statement represents an earlier misunderstanding under point (1) above where critics have made the false claim that SEM researchers believe that they can derive causal theory from associations in the data.

in that no analysis in the world can derive the causal theory from nonexperimental data and it is complimentary because SEM tests ALL the testable implications of the theory, no analysis can do better.

While it is true that no causal assumption can be tested in isolation and that certain combinations of assumptions do not have testable implications (e.g., a saturated model), SEM researchers are assured that those combinations that do have such implications will not go untested and those that do not, will be recognized as such. More importantly, researchers can verify whether the assumptions necessary for the final conclusion have survived the scrutiny of data and how severe that scrutiny was (Pearl, 2004).

The SEM literature has developed a variety of global and local tests of the implications of these causal assumptions that can lead to their rejection. In the classic SEM, the best known test is a likelihood ratio test that compares the model implied covariance matrix that is a function of the model parameters to a saturated covariance matrix. Formally, the null hypothesis is

$$H_o: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

where $\boldsymbol{\Sigma}$ is the population covariance matrix of the observed variables and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the model implied covariance matrix that is a function of $\boldsymbol{\theta}$, the parameters of the model (e.g., Bollen, 1989).[6] The maximum likelihood estimator or other estimators such as Weighted Least Squares are available to estimate **θ**. Once estimated, the researcher can form test statistics to test $H_o: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$.[7] The model implied covariance matrix is based on the causal assumptions that are embedded in the path diagram or equations of the model. If the model is true, then H_o should hold. Alternatively, if this equality does not hold, then it casts doubt on one or more of the causal assumptions that led to the SEM.

See our above discussion under Myth #1 that refutes this view. The second part, that SEM does not test any major part of the causal theory (assumptions) is ambiguous in that we do not know what qualifies as a "major" part of the theory.

[6] If the means and intercepts of the model are included, then the null hypothesis includes a test of whether the population means of the observed variables equals the model implied means that are a function of the model parameters.

[7] If some parameters are not identifiable, then the estimator might fail to converge or the run might be interrupted by SEM software that detects the identification problem. It is sometimes possible to estimate values for those parameters and functions of parameters that are identified and to test the fit of the overidentified parts of the model (see Shapiro, 1986). But for most researchers, it would be prudent to abandon the test unless they have sufficient expertise on the problem. An alternative is to use a tetrad or partial correlation test statistics for models that are under identified as long as vanishing tetrad or vanishing partial correlation are implied by the structure (see Bollen & Ting, 1993; Pearl 2000, page 144-154).

This issue is complicated in that like all likelihood ratio tests it assumes a large sample and that distributional assumptions are satisfied. Fortunately, there are distributionally robust corrections (e.g., Bollen & Stine, 1993; Satorra & Bentler, 1994) and some small sample corrections (e.g., Bentler & Yuan, 1999). There also is discussion about how to take account of the approximate nature of most models when the null hypothesis is one of exact fit where fit indexes are often used to supplement the chi square test.

One advantage of the chi square likelihood ratio test is that it is an omnibus simultaneous test of all of the restrictions on the implied covariance matrix. As such there is one test rather than a series of tests. However, this is a two-edged sword. If the chi square test is significant, the source of the lack of fit is unclear. The causal relationships of primary interest might hold, even though other parts of the model of less interest do not and these lead to the significant chi square test statistic. Additionally, the statistical power of the chi square test to detect a particular misspecification is lower than a local test aimed directly at that misspecification. For instance, a specific causal relation of interest might fail, but go undetected if the model includes many other variables with essentially correct specifications.

A less frequently used test statistic for SEM is the simultaneous tetrad test (Bollen, 1990). It is scalable in that it can provide a test of the full model or components of the model. This is based on the confirmatory tetrad analysis (CTA) proposed in Bollen & Ting (1993).[8] A tetrad is the difference in the product of pairs of covariances (e.g., $\sigma_{12}\sigma_{34}$-$\sigma_{13}\sigma_{24}$). The structure of a SEM typical implies that some of the tetrads equal zero whereas others do not. Rejection of the model implied tetrads that are supposed to be zero is a rejection of the specified SEM structure and hence a rejection of at least some of its causal assumptions.

The tetrad test applied to parts of the model permits local tests. This can be helpful in better isolating the sector of the model suffering from structural misspecifications. It is limited however to linear models.

Another local test that derives from the model is based on partial correlations or, more generally conditional independence conditions that are implied by the model structure. Simon (1954) and Blalock (1961) discussed the use of partial correlations to check the empirical predictions of some linear models. However, recent advances in graphical models have resulted in a complete systematization of conditional independence tests, to the point where they can be used to test nonparametric models which include latent variables (see Verma and Pearl 1990, Spirtes, et al. 1993 , Ali et al., 2007.see chapter on graphical models). Nonparametric models with no latent variables further enjoy the fact that ALL testable implications are of the conditional independence variety and the

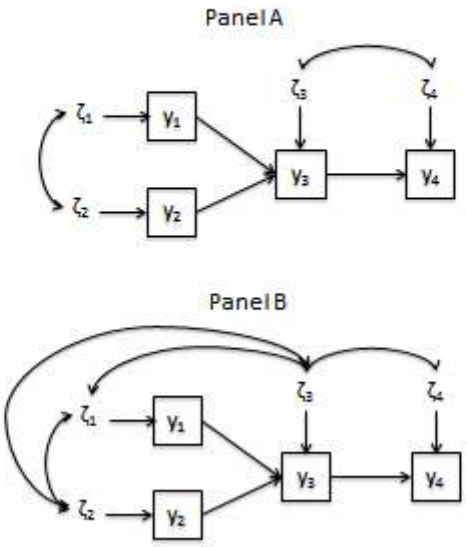
[8] Exploratory tetrad analysis which is designed to look for the different models that are consistent with the data is more oriented to creating models rather than testing models. Generally, ETA using tests single tetrads rather than simultaneous tests of multiple tetrads. See, e.g., Glymour, et al. (1987).

number of necessary tests is equal to the number of missing edges in the graph. This means that every conceivable test must be a derivative of one or more conditional independence tests that are advertised by the graph.

Another way to test the model using one equation at a time comes from the Model Implied Instrumental Variable (MIIV) approach proposed in Bollen (1996; 2001). Instrumental variables (IVs) are a method to estimate equations when one or more of the covariates of an equation correlate with the equation disturbance. IVs should correlate with the covariates and be uncorrelated with the equation disturbance. In addition there should be a sufficient number of IVs to permit estimation of effects.[9] The MIIV approach of Bollen begins with a SEM. All equations are transformed into observed variable equations by replacing all latent variables with their scaling indicators minus their errors. Then a set of rules is applied to determine those observed variables that are uncorrelated with the disturbance for each equation. (Bollen & Bauer, 2004). Those equations that have more than the bare minimum of MIIVs permit an overidentification test. The overidentification test reflects the presence of two or more separate equations that could estimate a coefficient of interest and, if the model is correct the solutions should result in the same coefficient values in the population. The overidentification test is a test of that equality. This is a logical consequence of the fact that all instrumental variables for the equation are uncorrelated with the equation disturbance. Sargan's (1958) test performs well in this context (Kirby & Bollen, 2009). Rejection of the null hypothesis is a rejection of the causal assumptions that led to the MIIV for that equation, and means that at least one of the IV's tested is misspecified.

[9] For instance, suppose that we have a model with two correlated covariates, one of which correlates with the equation disturbance and another that does not. This covariate that is uncorrelated with the disturbance and correlated with the other covariate meets the two preceding conditions, but will not permit IV estimation. We need other variables excluded from the equation that are uncorrelated with the disturbance and correlated with the covariate. In addition, the IV should have sufficient (partial) correlation with the problematic covariate to avoid the weak IV problem.

Figure 7



[FIGURE 7 ABOUT HERE]

As a simple illustration, consider Figure 7. In Figure 7a y_1 and y_2 correlate and both have direct effects on y_3 which in turn directly affects y_4. Also, the equation errors of y_3 and y_4 correlate. This model is overidentified so that we could test it using the likelihood ratio chi square test. A significant chi square implies that one or more of the causal assumptions of this model are false (assuming that the other distributional assumptions are satisfied). A more localized test is possible for the y_4 equation using the MIIV approach. Both y_1 and y_2 are MIIV for y_3 in the y_4 equation. Using these MIIVs separately we can write

$$\beta_{43} = \frac{COV(y_4, y_1)}{COV(y_3, y_1)} = \frac{COV(y_4, y_2)}{COV(y_3, y_2)}$$

as two solutions for β_{43} that should be identical under the hypothesized model, which constitutes an indirect test for the assumption that both MIIVs are uncorrelated with the error (ζ_4) for y_4. Thus, we have global and local tests of the model readily available.

Now consider the model in Figure 7b. Here we have a similar structure to Figure 7a except we have added all possible correlations among ζ_1, ζ_2, and ζ_3. ζ_1 and ζ_2 are still uncorrelated with ζ_4, and ζ_3 and ζ_4 are still correlated as in Figure 7a. However, a simple inspection of the graph(using the d-separation criterion see Chapter by Felix) or using the MIIV method of Bollen (1996; Bollen & Bauer, 2004) reveals that y_1 and y_2 are still MIIV for y_3 in the y_4 equation and we can apply an the identity above as a local test.

Graphical methods for enumerating all instrumental variables for a given equation are

illustrated in Kyono (2010), including variables that turn into instruments by conditioning on auxiliary variables (Britto and Pearl, 2002a). Additional tests, falling outside the partial correlation and instrumental variable variety are known as “dormant independencies” and are applicable to both linear and nonparametric models (Shpitser and Pearl, 2008b)

In sum, critics imply that the causal assumptions that lead to a SEM are not seriously tested. In fact, there are global tests such as the likelihood ratio chi square test for the match of the model implied covariance matrix to the covariance matrix of the observed variables and there are local tests such as partial correlations, conditional independence, vanishing tetrads, and MIIV overidentification tests. When these tests fail, then one or more of the causal assumptions of the model fail. True, the inability to distinguish between equivalent models may place some conclusions at the mercy of scientific judgment. Likewise, issues of statistical power, the treatment of approximate models, and the use of fit indexes are all complicating factors, but to say that SEM provides little empirical information to reject the causal assumptions of a model is false in all those models that allow such rejection. Not less important, and unique to SEM methods, models that do not allow such rejection can be identified as such, in which case the credibility of their conclusions would be recognized as supported entirely by the plausibility of the input assumptions.

CONCLUSIONS

SEM has had a long and turbulent encounter with causality. It was conceived and motivated by needs to solve causal inference problems, it was attacked and misunderstood on account of these needs and, today, it is emerging as a universal formalism that unifies all approaches to causation around simple and transparent principles.

There are many myths and misunderstandings about SEMs. In this chapter we have addressed just eight of them. Repetition of these myths among critics has led some to believe they are true. Our goal was to point out the inaccuracies of these beliefs and what the actual limitations of SEM are.

Indeed, we have our own critiques of common practices in the application of SEMs beyond those mentioned in the preceding discussion. Several come to mind. For one, practitioners give insufficient attention to the strong causal assumptions that are part of their model, and even seasoned analysts are often not clear on what those assumptions are. The rationale for them is not always provided and plausible alternative structures are not considered. A related point is that SEM researchers tend to focus too heavily on global tests of model fit such as the likelihood ratio chi square test and often ignore local tests such as partial correlations, MIIV tests, and tetrad tests which are indispensable for model diagnosis and respecification. These can be improved significantly through the advent of graphical models. We also feel that insufficient attention is given to the replication of models in different settings and among different populations. This

handicaps the cumulative development of knowledge. Finally, measurement issues are often given short shrift where the dimensionality of measures and appropriateness of indicators are not always discussed. This list of our criticisms is not complete, but it is enough to illustrate that we see much room for improvement in SEM applications.

Part of the way to improve SEM practice is to better separate the true from the false limitations of SEM and, not less important, to make the unique advantages of SEM explicit and vivid to practitioners. We hope that the current paper goes some ways towards accomplishing this goal.

REFERENCES

Ali, R., Richardson, T. and Spirtes, P. 2009. “Markov equivalence for ancestral graphs.” *The Annals of Statistics* 37:2808–2837.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Baron, Reuben M. and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51:1173-1182.

Baumrind, Diana. 1983. "Specious Causal Attributions in the Social Sciences: The Reformulated Stepping-Stone Theory of Heroin Use as Exemplar." *Journal Of Personality and Social Psychology* 45:1289-1298.

Bentler, Peter M. and Ke-Hai Yuan. 1999. "Structural Equation Modeling with Small Samples: Test Statistics." *Multivariate Behavioral Research* 34:181-197.

Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique*. Newbury Park, CA: Sage.

Bernert, Christopher. 1983. "The Career of Causal Analysis in American Sociology." *British Journal of Sociology* 34:230-254.

Bhrolcháin, Máire Ní and Tim Dyson. 2007. "On Causation in Demography: Issues and Illustrations." *Population and Development Review* 33:1-36.

Blalock Jr., Hubert M. 1960. "Correlation Analysis and Causal Inferences." *American Anthropologist* 62:624-631.

—. 1961. *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: University of North Carolina Press.

—. 1962. "Four-Variable Causal Models and Partial Correlations." *American Journal of Sociology* 68:182-194.

—. 1963. "Making Causal Inferences for Unmeasured Variables from Correlations Among Indicators." *American Journal of Sociology* 69:53-56.

—. 1969. "Multiple Indicators and the Causal Approach to Measurement Error." *American Journal of Sociology* 75:264-272.

—. 1985. *Causal Models in the Social Sciences*. Hawthorne, NY: Aldine de Gruyter.

Bollen, Kenneth A. 1987. "Total, Direct, and Indirect Effects in Structural Equation Models." *Sociological Methodology* 17:37-69.

—. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

—. 1990. "Outlier Screening and a Distribution-Free Test for Vanishing Tetrads." *Sociological Methods & Research* 19:80-92.

—. 1996. "An Alternative 2SLS Estimator for Latent Variable Models." *Psychometrika* 61:109-121.

—. 2001. "Two-stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications." Pp. 119-138 in *Structural Equation Modeling: Present and Future, a Festschrift in Honor of Karl Jöreskog*, edited by R. Cudeck, S. D. Toit, and D. Sörbom. Lincoln, IL: Scientific Software.

—. 2002. "Latent Variables in Psychology and the Social Sciences." *Annual Review of Psychology* 53:605-634.

Bollen, Kenneth A. and Daniel J. Bauer. 2004. "Automating the Selection of Model-Implied Instrumental Variables." *Sociological Methods & Research* 32:425-452.
Bollen, Kenneth A. and Robert Stine. 1990. "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability." *Sociological Methodology* 20:115-140.
Bollen, Kenneth A. and Kwok-fai Ting. 1993. "Confirmatory Tetrad Analysis." *Sociological Methodology* 23:147-175.
Cliff, Norman. 1983. "Some Cautions Concerning the Application of Causal Modeling Methods." *Multivariate Behavioral Research* 18:115-126.
Costner, Herbert L. 1971. "Utilizing Causal Models to Discover Flaws in Experiments." *Sociometry* 34:398-410.
de Leeuw, Jan. 1985. "Reviews." *Psychometrika* 50:371-375.
Duncan, Otis Dudley. 1966. "Path Analysis: Sociological Examples." *American Journal of Sociology* 72:1-16.
—. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.
Everitt, B. S. 1984. *An Introduction to Latent Variable Models*. New York: Chapman and Hall.
Fisher, Ronald A. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.
Folmer, H. 1981. "Measurement of the effects of Regional Policy Instruments by Means of Linear Structural Equation Models and Panel Data." *Environment and Planning A* 13:1435-1448.
Fox, John. 1980. "Effect Analysis in Structural Equation Models: Extensions and Simplified Methods of Computation." *Sociological Methods & Research* 9:3-28.
Freedman, David A. 1981. "Pitfalls in Large Econometric Models: A Casey Study." *Journal of Business* 54:479-500.
—. 1987. "As Others See Us: A Case Study in Path Analysis." *Journal of Educational and Behavioral Statistics* 12:101-128.
—. 2004. "Graphical Models for Causation, and the Identification Problem." *Evaluation Review* 28:267-293.
Glymour, Clark. 1986. "Statistics and Causal Inference: Comment: Statistics and Metaphysics." *Journal of the American Statistical Association* 81:964-966.
Glymour, Clark, Richard Scheines, Peter Spirtes, and Kevin Kelly. 1987. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling*. Orlando, FL: Academic Press.
Goldberger, Arthur S. 1972. "Structural Equation Methods in the Social Sciences." *Econometrica* 40:979-1001.
—. 1973. "Structural Equation Models: An Overview." Pp. 1-18 in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar Press.
Goldberger, Arthur S. and Otis Dudley Duncan. 1973. "Structural Equation Models in the Social Sciences." New York: Seminar Press.
Goldthorpe, John H. 2001. "Causation, Statistics, Sociology." *European Sociological Review* 17:1-20.
Granger, Clive W. 1969. "Investigating Causal Relationships by Econometric Models and Cross-Spectral Methods." *Econometrica* 37:424-438.

Guttman, Louis. 1977. "What is Not What in Statistics." *Journal of the Royal Statistical Society: Series D (The Statistician)* 26:81-107.
Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945-970.
—. 1995. "Some Reflections on Freedman's Critiques." *Foundations of Science* 1:50-57.
Imai, Kosuke, Booll Jo, and Elizabeth A. Stuart. 2011. "Commentary: Using Potential Outcomes to Understand Causal Mediation Analysis." *Multivariate Behavioral Research* 46:861-873.
Imai, K., Keele, L. and Yamamoto, T. 2010. "Identification, inference, and sensitivity analysis for causal mediation effects." *Statistical Science* 25:51–71.
James, C. R., Stanley A. Mulaik, and Jeanne M. Brett. 1982. *Causal Analysis: Assumptions, Models, and Data*. Beverly Hills, CA: Sage.
Joffe, Marshall. 2011. "Principle Stratification and Attribution Prohibition: Good Ideas Taken Too Far." *International Journal of Biostatistics* 7.
Joffe, Marshall, Wei Peter Yang, and Harold I. Feldman. 2010. "Selective Ignorability Assumptions in Causal Inference." *International Journal of Biostatistics* 6.
Jöreskog, Karl G. 1969. "A General Approach to Confirmatory Maximum Likelihood Factor Analysis." *Psychometrika* 34:183-202.
—. 1970. "A General Method for the Analysis of Covariance Structures." *Biometrika* 57:239-251.
—. 1973. "A General Model for Estimating a Linear Structural Equation System." in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar Press.
Jöreskog, Karl G. and Dag Sörbom. 1978. *LISREL IV [Computer software]*. Chicago, IL: Scientific Software International, Inc.
—. 1981. *LISREL V [Computer software]*. Chicago, IL: Scientific Software International, Inc.
Kenny, David A. 1979. *Correlation and Causality*. New York: Wiley.
Kirby, James B. and Kenneth A. Bollen. 2009. "Using Instrumental Variable Tests to Evaluate Model Specification in Latent Variable Structural Equation Models." *Sociological Methodology* 39:327-355.
Lindquist, Martin A. and Michael E. Sobel. 2011. "Graphical Models, Potential Outcomes and Causal Inference: Comment on Ramsey, Spirtes and Glymour." *NeuroImage* 57:334-336.
Long, J. Scott. 1983. *Confirmatory Factor Analysis*. Newbury Park, CA: Sage.
—. 1983. *Covariance Structure Models: An Introduction to LISREL*. Newbury Park, CA: Sage.
Miller, Alden Dykstra. 1971. "Logic of Causal Analysis: From Experimental to Nonexperimental Designs." Pp. 273-294 in *Causal Models in the Social Sciences*, edited by H. M. Blalock Jr. Chicago: Aldine Atherton.
Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
Mulaik, Stanley A. 2009. *Linear Causal Modeling with Structural Equations*. Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. 2011. “Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus.” Tech. rep., Graduate School of Education and Information Studies, University of California, Los Angeles, CA.
Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press. 2nd edition 2009
—. 2009. "Remarks on the Method of Propensity Score." *Statistics in Medicine* 28:1415-1416.
—. 2009. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3:96-146.
—. 2010. "The Foundations of Causal Inference." *Sociological Methodology* 40:75-149.
—. 2011a "Principal Stratification - A Goal or a Tool?" *International Journal of Biostatistics* 7.
---. 2011b. “The causal mediation formula – a guide to the assessment of pathways and mechanisms.” Tech. Rep. R-379, <http://ftp.cs.ucla.edu/pub/stat ser/r379.pdf>, Department of Computer Science, University of California, Los Angeles, CA., forthcoming, Intervention Science
----2011c “Trigveh Haavelmo and the Emergence of Causal Calculus” Technical Report R-391 Presented in Haavelmo Centennial Symposium Oslo, Dec. 14, 2011.
---- 2012. “The causal foundation of structural equation modeling” R-370, Rick Hoyle (ed) *Handbook of Structural Equation Modeling*. Newbury Park, CA: Sage.
Provine, William B. 1986. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.
Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688-701.
—. 2004. "Direct and Indirect Causal Effects Via Potential Outcomes." *Scandinavian Journal of Statistics* 31:162-170.
—. 2009. "Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions Across Treatment Groups." *Statistics in Medicine* 28:1420-1423.
—. 2010. "Reflections Stimulated by the Comments of Shadish (2010) and West and Thoemmes (2010)." *Psychological Methods* 15:38-46.
Sargan, J. D. 1958. "The Estimation of Economic Relationships using Instrumental Variables." *Econometrica* 26:393-415.
Saris, W. and H. Stronkhorst. 1984. *Causal Modeling in Nonexperimental Research*. Amsterdam: Sociometric Research Foundation.
Satorra, Albert and Peter M. Bentler. 1994. "Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis." in *Latent Variable Analysis: Applications for Developmental Research*, edited by A. v. Eye and C. C. Clogg. Thousand Oaks, CA: Sage.
Shapiro, Alexander. 1986. “Asymptotic Theory of Overparameterized Structural Models.” *Journal of the American Statistical Association* 81:142-49.
Shpitser, I. and Judea Pearl. 2008a. “Complete Identification Methods for the Causal Hierarchy” *Journal of Machine Learning* 9:1941-1979.
------- 2008b**,** "Dormant Independence" In *Proceedings of the Twenty-Third Conference on Artificial Intelligence*, 1081-1087, 2008.

___. 2009. “Effects of Treatment on the Treated: Identification and Generalization.” In J. Bilmes and A. Ng (eds) *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence.* AUAI Press, Montreal, Quebec.

Simon, Herbert A. 1954. "Spurious Correlation: A Causal Interpretation." *Journal of the American Statistical Association* 49:467-479.

Sjolander, Arvid. 2011. "Reaction to Pearl's Critique of Principal Stratification." *International Journal of Biostatistics* 7

Skrondal, Anders and Sophia Rabe-Hesketh. 2005. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Sobel, Michael E. 1986. "Some New Results on Indirect Effects and their Standard Errors in Covariance Structure Models." *Sociological Methodology* 16:159-186.

—. 2008. "Identification of Causal Parameters in Randomized Studies with Mediating Variables." *Journal of Educational and Behavioral Statistics* 33:230-251.

Spirtes, P., Glymour, C. and Scheines, R. 2000. *Causation, Prediction, and Search*. 2nd ed. MIT Press, Cambridge, MA.

Stolzenberg, Ross M. 1980. "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models." *Sociological Methodology* 11:459-488.

VanderWeele, Tyler J. 2011. "Principal Stratification -- Uses and Limitations." *International Journal of Biostatistics* 7.

VanderWeele, T. & Vansteelandt, S. 2009. “Conceptual issues concerning mediation, interventions and composition.” *Statistics and Its Interface* 2:457–468.

Verma, T. and Pearl, J. 1990. “Equivalence and synthesis of causal models.” In *Uncertainty in Artificial Intelligence*, Proceedings of the Sixth Conference. Cambridge, MA.

West, Stephen G. 2011. "Editorial: Introduction to the Special Section on Causal Inference in Cross Sectional and Longitudinal Mediational Models." *Multivariate Behavioral Research* 46:812-815.

Winship, Christopher and Robert D. Mare. 1983. "Structural Equations and Path Analysis with Discrete Data." *American Journal of Sociology* 89:54-110.

Wright, Sewall S. 1918. "On the Nature of Size Factors." *Genetics* 3:367-374.

—. 1921. "Correlation and Causation." *Journal of Agricultural Research* 20:557-585.

—. 1934. "The Method of Path Coefficients." *Annals of Mathematical Statistics* 5:161-215.

Yu, Chong Ho. 2007. "Causation in Quantitative Research Methodologies from Path Modeling, SEM to TETRAD." *Theory and Science* 9.