Causally-Defined Direct and Indirect Effects in Mediation Modeling

Bengt Muthén

Mplus
www.statmodel.com

Presentation at Utrecht University
August 2012

New Thinking About Mediation Drawing on the Causal Effect Literature

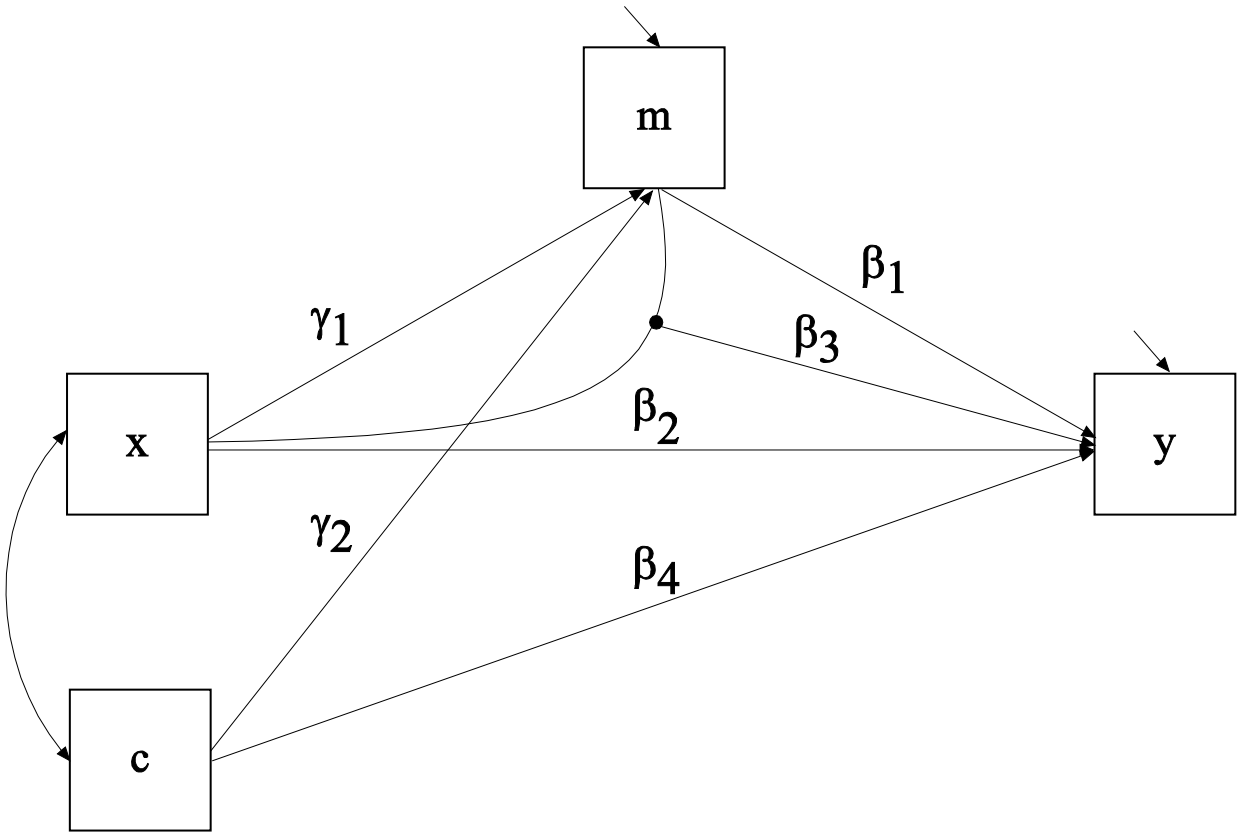
Muthén (2011). Applications of Causally Defined Direct and Indirect Effects in Mediation Analysis using SEM in Mplus.

New ways to estimate mediation effects with categorical and other non-normal mediators and distal outcomes

The paper, an appendix with formulas, and Mplus scripts are available at www.statmodel.com under Papers, Mediational Modeling

A Mediation Model with Interactions

The filled circle represents an interaction term consisting of the variables connected to it without arrow heads, in this case x and m



Causal Effect Definitions

- $Y_i(x)$: Potential outcome that would have been observed for that subject had the treatment variable X been set at the value x, where x is 0 or 1 in the example considered here
- The $Y_i(x)$ outcome may not be the outcome that is observed for the subject and is therefore possibly counterfactual
- The causal effect of treatment for a subject can be seen as $Y_i(1) - Y_i(0)$, but is clearly not identified given that a subject only experiences one of the two treatments
- The average effect $E[Y(1) - Y(0)]$ is, however, identifiable if X is assigned randomly as is the case in a randomized controlled trial.
- Similarly, let Y(x, m) denote the potential outcome that would have been observed if the treatment for the subject was x and the value of the mediator M was m

The Direct Causal Effect (DE)

The direct effect (often called the pure or natural direct effect) does not hold the mediator constant, but instead allows the mediator to vary over subjects in the way it would vary if the subjects were given the control condition. The direct effect is expressed as

$$DE = E[Y(1,M(0)) - Y(0,M(0)) \mid C = c] = \tag{1}$$

$$= \int_{-\infty}^{\infty} \{E[Y \mid C = c, X = 1, M = m] - E[Y \mid C = c, X = 0, M = m]\}$$

$$\times f(M \mid C = c, X = 0)\, \partial M, \tag{2}$$

where f is the density of M. A simple way to view this is to note that in Y's first argument, that is x, changes values, but the second does not, implying that Y is influenced by X only directly. The right-hand side of (2) is part of what is referred to as the Mediation Formula in Pearl (2009, 2011c).

The Total Indirect Effect (TIE)

The total indirect effect is defined as (Robins, 2003)

$$TIE = E[Y(1,M(1)) - Y(1,M(0)) \mid C = c] = \tag{3}$$

$$= \int_{-\infty}^{\infty} E[Y \mid C = c, X = 1, M = m] \times f(M \mid C = c, X = 1)\, \partial M$$

$$- \int_{-\infty}^{\infty} E[Y \mid C = c, X = 1, M = m] \times f(M \mid C = c, X = 0)\, \partial M. \tag{4}$$

A simple way to view this is to note that the first argument of Y does not change, but the second does, implying that Y is influenced by X due to its influence on M.

The Total Effect (TE)

The total effect is (Robins, 2003)

$$TE = E[Y(1) - Y(0) \mid C = c] \tag{5}$$
$$= E[Y(1, M(1)) - Y(0, M(0)) \mid C = c]. \tag{6}$$

A simple way to view this is to note that both indices are 1 in the first term and 0 in the second term. In other words, the treatment effect on Y comes both directly and indirectly due to M. The total effect is the sum of the direct effect and the total indirect effect (Robins, 2003),

$$TE = DE + TIE. \tag{7}$$

The Pure Indirect Effect (PIE)

The pure indirect effect (Robins, 2003) is defined as

$$PIE = E[Y(0, M(1)) - Y(0, M(0)) \mid C = c] \tag{8}$$

Here, the effect of X on Y is only indirect via M. This is called the natural indirect effect in Pearl (2001) and VanderWeele and Vansteelandt (2009).

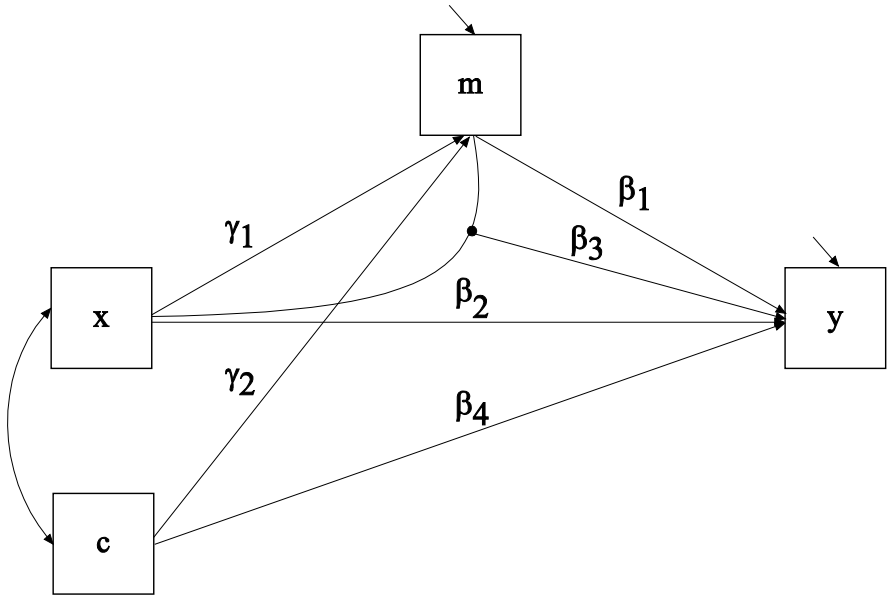
A General Approach

The DE, TIE, and PIE effects are expressed in a general way and can be applied to many different setttings

- Continuous mediator, continuous distal outcome
- Categorical mediator, continuous distal outcome
- Continuous mediator, categorical distal outcome
- Categorical mediator, categorical distal outcome

The direct and indirect effects can be estimated in Mplus using maximum-likelihood. Standard errors of the direct and indirect causal effects are obtained by the delta method using the Mplus MODEL CONSTRAINT command. Bootstrapped standard errors and confidence intervals are also available, taking into account possible non-normality of the effect distributions. Furthermore, Bayesian analysis is available in order to describe the posterior distributions of the effects.

Applied to Continuous Variables in the Model in the Figure



$$DE = \beta_2 + \beta_3\ \gamma_0 + \beta_3\ \gamma_2\ c. \tag{9}$$

$$TIE = \beta_1\ \gamma_1 + \beta_3\ \gamma_1. \tag{10}$$

The pure indirect effect excludes the interaction part,

$$PIE = \beta_1\ \gamma_1. \tag{11}$$

Using the general definition, the causal total indirect effect is expressed as the probability difference

$$TIE = \Phi[probit(1,1)] - \Phi[probit(1,0)], \tag{12}$$

using the standard normal distribution function Φ, and where for x, x' = 0, 1 corresponding to the control and treatment group,

$$probit(x,x') = [\beta_0 + \beta_2\, x + \beta_4\, c + (\beta_1 + \beta_3\, x)(\gamma_0 + \gamma_1\, x' + \gamma_2\, c)]/\sqrt{v(x)}, \tag{13}$$

where the variance $v(x)$ for x = 0, 1 is

$$v(x) = (\beta_1 + \beta_3\, x)^2\, \sigma_2^2 + 1. \tag{14}$$

where σ_2^2 is the residual variance for the continuous mediator m. Although not expressed in simple functions of model parameters, the quantity of (12) can be computed and corresponds to the change in the y=1 probability due to the indirect effect of the treatment (conditionally on c when that covariate is present).

Categorical Distal Outcome, Continued

Using the general definition, the pure indirect effect is expressed as the probability difference

$$PIE = \Phi[probit(0,1)] - \Phi[probit(0,0)]. \quad (15)$$

and the direct effect expressed as the probability difference

$$DE = \Phi[probit(1,0)] - \Phi[probit(0,0)]. \quad (16)$$

Conventional versus Causal Mediation Effects with a Categorical Distal Outcome

With a categorical distal outcome, conventional product formulas for indirect effects are only valid for an underlying continuous latent response variable behind the categorical observed outcome (2 linear regressions), not for the categorical outcome itself (linear plus non-linear regression).

Similarly, with a categorical mediator, conventional product formulas for indirect effects are only relevant/valid for a continuous latent response variable behind the mediator.

Example: Aggressive Behavior and Juvenile Court Record

- Randomized field experiment in Baltimore public schools
- Classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students
- Mediator is the aggression score in Grade 5 after the intervention ended
- Distal outcome is a binary variable indicating whether or not the student obtained a juvenile court record by age 18 or an adult criminal record
- $n = 250$ boys in treatment and control classrooms

Aggressive Behavior and Juvenile Court Record: Mplus Input for Causal Effects

```
Analysis:
        estimator = mlr;
        link = probit;
        integration = montecarlo;

model:
        [juvcrt$1] (mbeta0);
        juvcrt on tx (beta2)
        agg5 (beta1)
        xm (beta3)
        agg1 (beta4);
        [agg5] (gamma0);
        agg5 on tx (gamma1)
        agg1 (gamma2);
        agg5 (sig2);
```

Aggressive Behavior and Juvenile Court Record: Mplus Input for Causal Effects, Continued

```
model constraint:

new(ind dir arg11 arg10 arg00 v1 v0
probit11 probit10 probit00 indirect direct
total iete dete compdete orind ordir);
dir=beta3*gamma0+beta2;
ind=beta1*gamma1+beta3*gamma1;
arg11=-mbeta0+beta2+beta4*0+(beta1+beta3)*(gamma0+gamma1+gamma2*0);
arg10=-mbeta0+beta2+(beta1+beta3)*gamma0;
arg00=-mbeta0+beta1*gamma0;
v1=(beta1+beta3)^2*sig2+1;
v0=beta1^2*sig2+1;
probit11=arg11/sqrt(v1);
probit10=arg10/sqrt(v1);
probit00=arg00/sqrt(v0);
! Version 6.12 Phi function needed below:
indirect=phi(probit11)-phi(probit10);
direct=phi(probit10)-phi(probit00);
total=phi(probit11)-phi(probit00);
orind=(phi(probit11)/(1-phi(probit11)))/(phi(probit10)/(1-phi(probit10)));
ordir=(phi(probit10)/(1-phi(probit10)))/(phi(probit00)/(1-phi(probit00)));
```

Aggressive Behavior and Juvenile Court Record: Estimates

The causal direct effect is not significant. The causal indirect effect is estimated as -0.064 and is significant. This is the drop in the probability of a juvenile court record due to the indirect effect of treatment.

The odds ratio for the indirect effect is estimated as 0.773 which is significantly different from one ($z = (0.773 - 1)/0.092 = -2.467$).

The conventional direct effect is not significant and the conventional product indirect effect is -0.191 (z$= -1.98$).

Binary Mediator and Binary Distal Outcome

Recalling that the general formulas for the direct, total indirect, and pure indirect effects are defined as

$$DE = E[Y(1,M(0)) - Y(0,M(0)) \mid C], \tag{17}$$

$$TIE = E[Y(1,M(1)) - Y(1,M(0)) \mid C], \tag{18}$$

$$PIE = E[Y(0,M(1)) - Y(0,M(0)) \mid C], \tag{19}$$

it can be shown that with a binary mediator and a binary outcome these formulas lead to the expressions

$$DE = [F_Y(1,0) - F_Y(0,0)]\,[1 - F_M(0)] + [F_Y(1,1) - F_Y(0,1)]\,F_M(0), \tag{20}$$

$$TIE = [F_Y(1,1) - F_y(1,0)]\,[F_M(1) - F_m(0)], \tag{21}$$

$$PIE = [F_Y(0,1) - F_y(0,0)]\,[F_M(1) - F_m(0)]. \tag{22}$$

where $F_Y(x,m)$ denotes $P(Y = 1 \mid X = x, M = m)$ and $F_M(x)$ denotes $P(M = 1 \mid X = x)$, where F denotes either the standard normal or the logistic distribution function corresponding to using probit or logistic regression. These formulas agree with those of Pearl (2010, 2011a).

Pearl's Hypothetical Binary-Binary Case

Pearl (2010, 2011a) provided a hypothetical example with a binary treatment X, a binary mediator M corresponding to the enzyme level in the subject's blood stream, and a binary outcome Y corresponding to being cured or not. This example was also hotly debated on SEMNET in September 2011.

Pearl's Hypothetical Binary-Binary Case, Continued

Treatment X	Enzyme M	Percentage Cured Y = 1
1	1	$F_Y(1,1) = 80\%$
1	0	$F_Y(1,0) = 40\%$
0	1	$F_Y(0,1) = 30\%$
0	0	$F_Y(0,0) = 20\%$

Treatment	Percentage M=1
0	$F_M(0) = 40\%$
1	$F_M(1) = 75\%$

The top part of the table suggests that the percentage cured is higher in the treatment group for both enzyme levels and that the effect of treatment is higher at enzyme level 1 than enzyme level 0: Treatment-mediator interaction.

Generalizations in Muthén (2011)

- Nominal mediator
- Count distal outcome
- General latent variable framework (e.g. latent class variable as a nominal mediator)

Effects are Causal Only Under Strong Assumptions

To claim that effects are causal, it is not sufficient to simply use the causally-derived effects

The underlying assumptions need to be fulfilled, such as no mediator-outcome confounding

- Sensitivity analysis

Mediator-Outcome Confounding

Violation of the no mediator-outcome confounding can be seen as an unmeasured (latent) variable Z influencing both the mediator M and the outcome Y. When Z is not included in the model, a covariance is created between the residuals in the two equations of the regular mediation model. Including the residual covariance, however, makes the model not identified.

Imai et al. (2010a, b) proposed a sensitivity analysis where causal effects are computed given different fixed values of the residual covariance. This is useful both in real-data analyses as well as in planning studies. As for the latter, the approach can answer questions such as how large does your sample and effects have to be for the lower confidence band on the indirect effect to not include zero when allowing for a certain degree of mediator-outcome confounding?

Sensitivity plots can be made in Mplus.

Indirect Effect Based on Imai Sensitivity Analysis with ρ Varying from -0.9 to +0.9 and True Residual Correlation 0.25