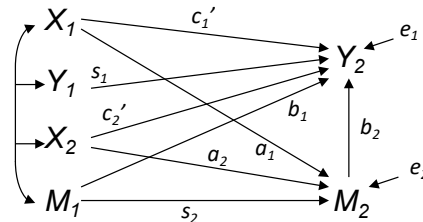


## Mediation Analysis with Two Waves

*Mediation* involves an intervening mechanism in a causal chain of events,  $X \rightarrow M \rightarrow Y$ . The variable  $M$  is the mediator.<sup>1</sup> A mediation analysis implies that there is an *indirect effect* of variable  $X$  on variable  $Y$ , which is defined as the product of the two regression coefficients ( $Y$  regressed on  $M$  and  $M$  regressed on  $X$ ). Mediation can be tested with cross-sectional data (although causal conclusions may not be appropriate), but we are concerned with longitudinal data here. In particular, in this section, we considered longitudinal data for two time points. Ideally, we would prefer to have a temporal lag for the  $X$ - $M$  relation and the  $M$ - $Y$  relation. Although a three-wave design might be the best suited for a three variable causal chain, we can examine each link longitudinally using only two waves of data (Cole & Maxwell, 2003). This analysis assumes the researcher has available the predictor, mediator, and outcome measured at both time points.



In this figure  $X_1$ ,  $M_1$ , and  $Y_1$  are all measured at Wave 1, and  $X_2$ ,  $M_2$ , and  $Y_2$  are all measured at Wave 2. The analysis can be represented by a system of two regression equations, where  $a$ ,  $b$ ,  $c'$ , and  $s$  are regression coefficients, and  $X$ ,  $M$ , and  $Y$  are variables with subscripts indicating time of measurement (MacKinnon, 2008).

$$Y_2 = i_1 + c'_1 X_1 + c'_2 X_2 + b_1 M_1 + b_2 M_2 + s_1 Y_1 + e_1$$

$$M_2 = i_2 + a_1 X_1 + a_2 X_2 + s_2 M_1 + e_2$$

The most appropriate estimate of the indirect effect then uses the product of the two longitudinal paths between the predictor and the mediator,  $a_1$ , and between the mediator and the outcome,  $b_1$ . The use of these paths for the computation of the indirect effect takes advantage of the temporal lag and longitudinal regression for both links of the causal chain. One can estimate the indirect effect from this model using two regression models or a single structural equation model.

This approach to mediation with two waves may not always be the most appropriate. In some cases, the  $X \rightarrow M$  relationship or the  $M \rightarrow Y$  relationship may be hypothesized to be more immediate. In which case, the lag time of the study may not be best match for one of these relationships.

An important limitation of this analysis is that measurement error may impact the regression estimates of any of the paths. A latent variable model would address this limitation as well as allow for a more accurate estimate of autoregressive effects by including correlated errors over time. An addition concern is that the measurement properties of each measure are not consistent over time, a condition known as *longitudinal invariance*, creating to a methodological artifact. Longitudinal invariance is important, because, if the reliability of the measure changes over time, it is possible that the change observed is not a result of a theoretical processes but a result of measurement properties instead. The above described analysis assumes measurement invariance, whereas the use of latent variables with multiple indicators would allow the researcher to test for measurement invariance.

## References and Further Reading

- Cole, D.A., & Maxwell, S.E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558-577.
- MacKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- Roth, D.L., & MacKinnon, D.P. (2012). Mediation analysis with longitudinal data. In J.T. Newsom, R.N. Jones, & S.M. Hofer, *Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Sciences* (pp. 181-216). New York: Routledge.

<sup>1</sup> It is important to distinguish *mediation* from *moderation* (Baron & Kenny, 1986). Moderation involves a statistical interaction, in which the effect of one independent variable on the dependent variable is different for different levels of another independent variable (i.e., the moderator variable). With moderation, there is no implied association or causal relationship between the independent variable and the moderator as there is with mediation.