Applying Mediation Analysis to Understand How Interventions Work

David P. MacKinnon

Arizona State University

Office of Disease Prevention: Mind the Gap Seminar Series

April 8, 2016

*This research was supported in part by the National Institute on Drug Abuse (R25 DA09757).

Outline

1. Mediating Variable Examples and Applications
2. Statistical Mediation Analysis
3. Advanced Mediation Models
4. Future Directions

Website:

http://www.public.asu.edu/~davidpm/

Book:

MacKinnon, D. P. (2008) Introduction to Statistical Mediation Analysis. Mahwah, NJ: Erlbaum.

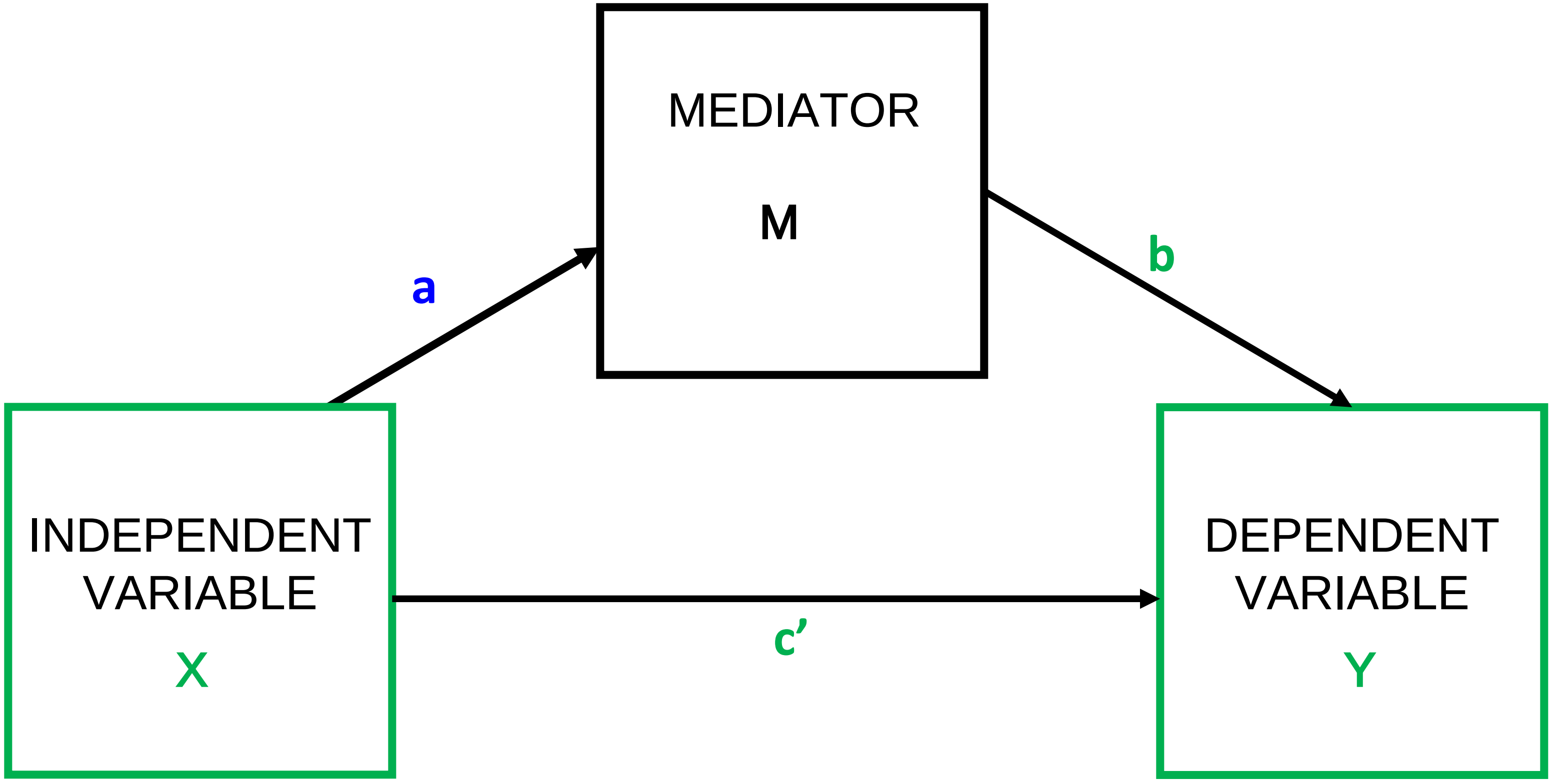
Mediator

A variable that is intermediate in the causal process relating an independent to a dependent variable.

Some Examples:

1) Intervention has beneficial effects on **exercise** which leads to reduced depression.
2) Tobacco prevention program promotes **anti-tobacco norms** which reduce tobacco use (MacKinnon et al., 1991)
3) Screening program increases **identification of early stage cancer** which reduces cancer deaths.
4) Wellbutrin (Bupropion) increases participant’s **willingness to quit and self-efficacy** which are associated with one month abstinence from tobacco (McCarthy et al., 2008)
5) Your Examples?

Single Mediator Model



Mediator Definitions

- A mediator is a variable in a chain whereby an independent variable causes the mediator which in turn causes the outcome variable (Sobel, 1990)
- The generative mechanism through which the focal independent variable is able to influence the dependent variable (Baron & Kenny, 1986)
- A variable that occurs in a causal pathway from an independent variable to a dependent variable. It causes variation in the dependent variable and itself is caused to vary by the independent variable (Last, 1988)

Two, three, four variable effects

- Two variables: $X \rightarrow Y$, $Y \rightarrow X$, $X \leftrightarrow Y$ are reciprocally related. Measures of effect include the correlation, covariance, regression coefficient, odds ratio, mean difference.
- Three variables: $X \rightarrow M \rightarrow Y$, $X \rightarrow Y \rightarrow M$, $Y \rightarrow X \rightarrow M$, and all combinations of reciprocal relations. Special names for third-variable effects: confounder, mediator, moderator/interaction.
- Four variables: many possible relations among variables, e.g., $X \rightarrow Z \rightarrow M \rightarrow Y$

Confounder and Moderator

- Confounder is a variable related to two variables of interest that falsely obscures or accentuates the relation between them (Meinert & Tonascia, 1986; Greenland & Morgenstern, 2001). It is not in a causal sequence like a mediator.
- Moderator is a variable that affects the strength of the relation between two variables. The variable is not intermediate in the causal sequence so it is not a mediator.

Mediation is important because …

- Central questions in many fields are about mediating processes.
- Important for basic research on mechanisms of effects.
- Critical for applied research, especially prevention and treatment to identify critical ingredients leading to more efficient interventions.
- Many interesting statistical and mathematical issues.

New Focus on Mediating Mechanisms

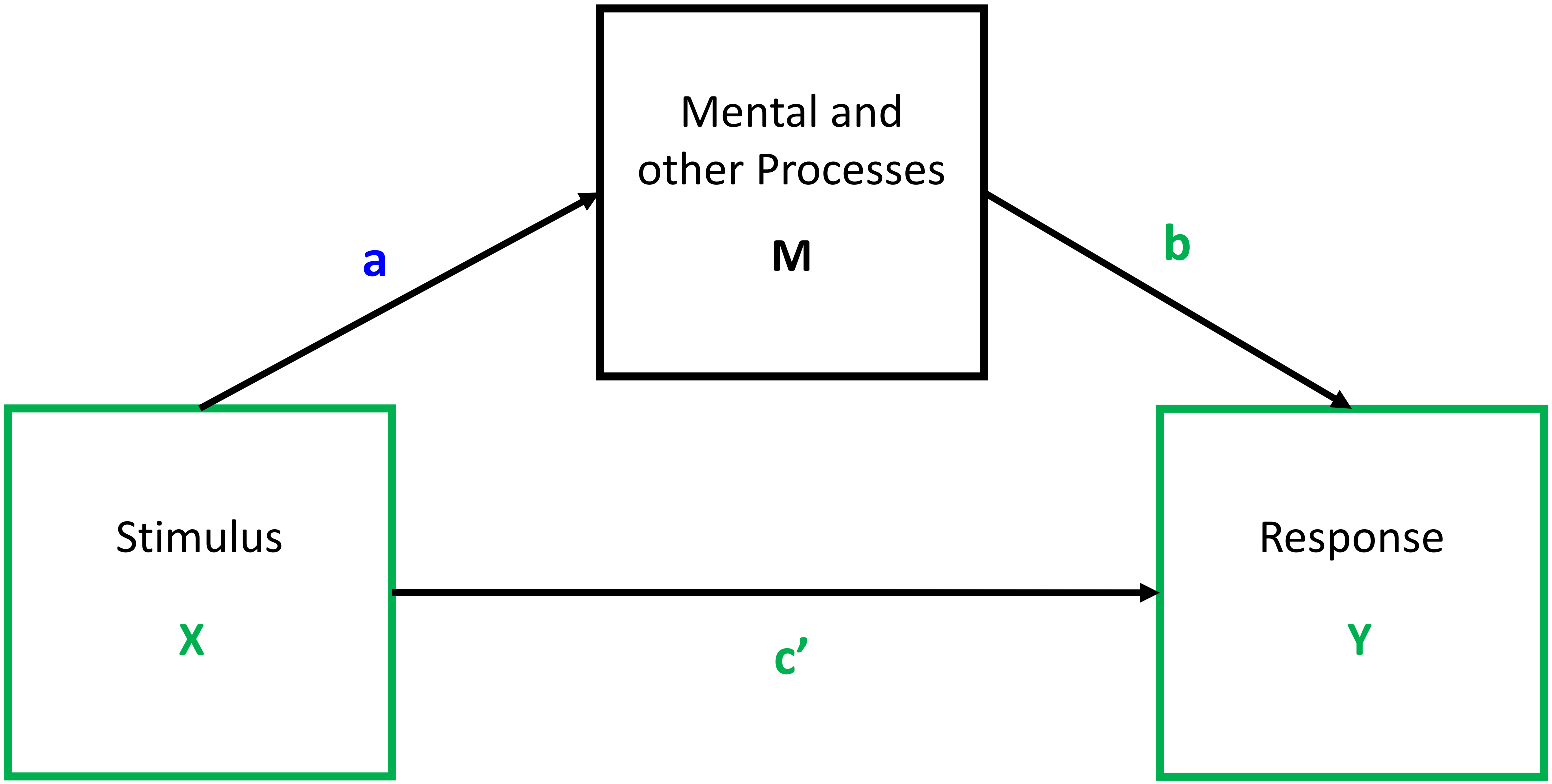
"First, future trials will follow an experimental medicine approach in which interventions serve not only as potential treatments, but as probes to generate information about the mechanisms underlying a disorder. …. It offers us a way to understand the mechanisms by which these treatments are leading to clinical change. "

Thomas Insel, M.D. NIMH Director: http://www.nimh.nih.gov/about/director/2014/a-new-approach-to-clinical-trials.shtml

S→O→R Theory I

- Stimulus→ Organism → Response (SOR) theory whereby the effect of a Stimulus on a Response depends on mechanisms in the organism (Woodworth, 1928). These mediating mechanisms translate the Stimulus to the Response. SOR theory is ubiquitous in psychology.
- Stimulus: Multiply 24 and 16
- Organism: You
- Response: Your Answer
- Organism as a Black Box

S-O-R Mediator Model



Applications

Two overlapping applications of mediation analysis:

1) **Mediation for Explanation**

2) **Mediation by Design**

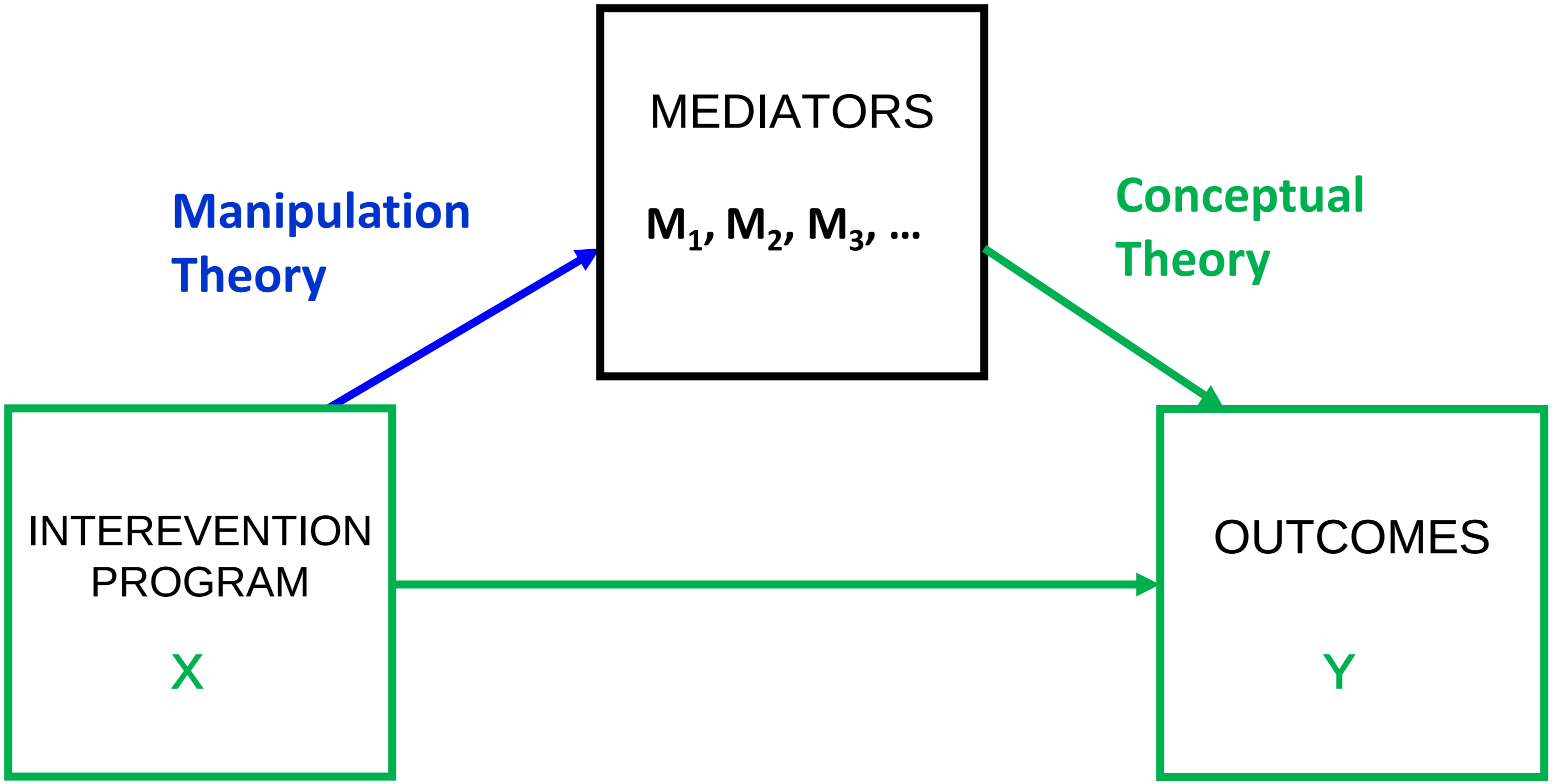
Mediation for Explanation

- Observe relation and then try to explain it.
- Elaboration method described by Lazarsfeld and colleagues (1955; Hyman, 1955) where third variables are included in an analysis to see if/how the observed relation changes.
- Replication (Covariate)
- Explanation (Confounder)
- Intervening (Mediator)
- Specification (Moderator)

Mediation by Design

- Select mediating variables that are causally related to an outcome variable.
- Intervention is designed to change these mediators.
- If mediators are causally related to the outcome, then an intervention that changes the mediator will change the outcome.
- Common in applied research like prevention and treatment.

Intervention Mediation Model



If the mediator changed is causally related to Y, then changing the mediator will change Y.

.

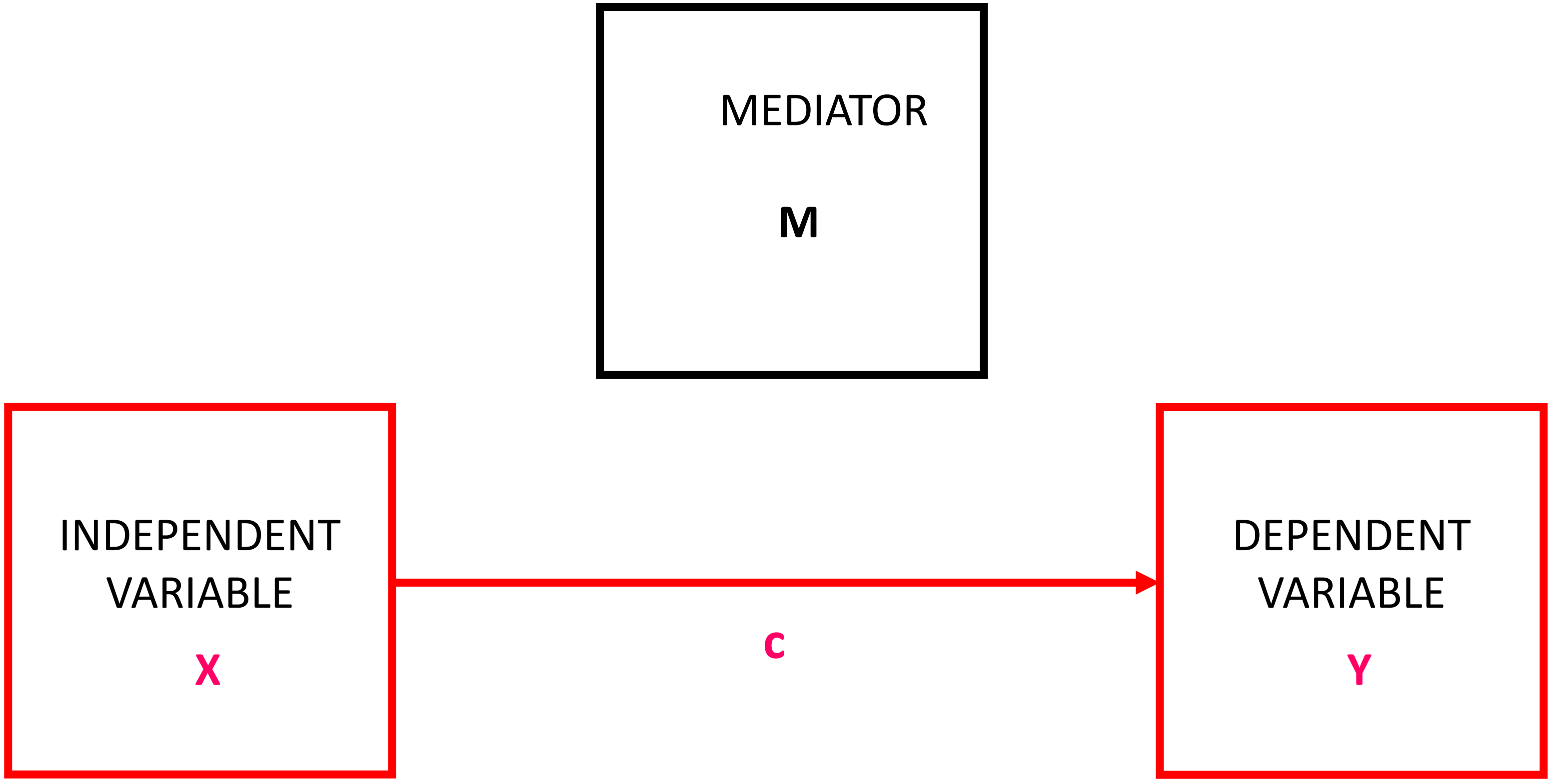
Mediation in Intervention Research

- A theory based approach focuses on the processes underlying programs. Mediators play a primary role. **Manipulation Theory** corresponds to how the manipulation will affect mediators. **Conceptual Theory** focuses on how the mediators are related to the dependent variables (Chen, 1990, Lipsey, 1993; MacKinnon, 2008).
- Identifying mediators is important for basic and applied science. Practical implications include reduced cost and more effective interventions if true mediators are identified.

Mediation Regression Equations

- Tests of mediation for a single mediator use information from some or all of three equations.
- The coefficients in the equations may be obtained using methods such as ordinary least squares regression, covariance structure analysis, or logistic regression.
- The product of coefficients test is the method of choice. It extends to more complicated models such as the multiple mediator model.

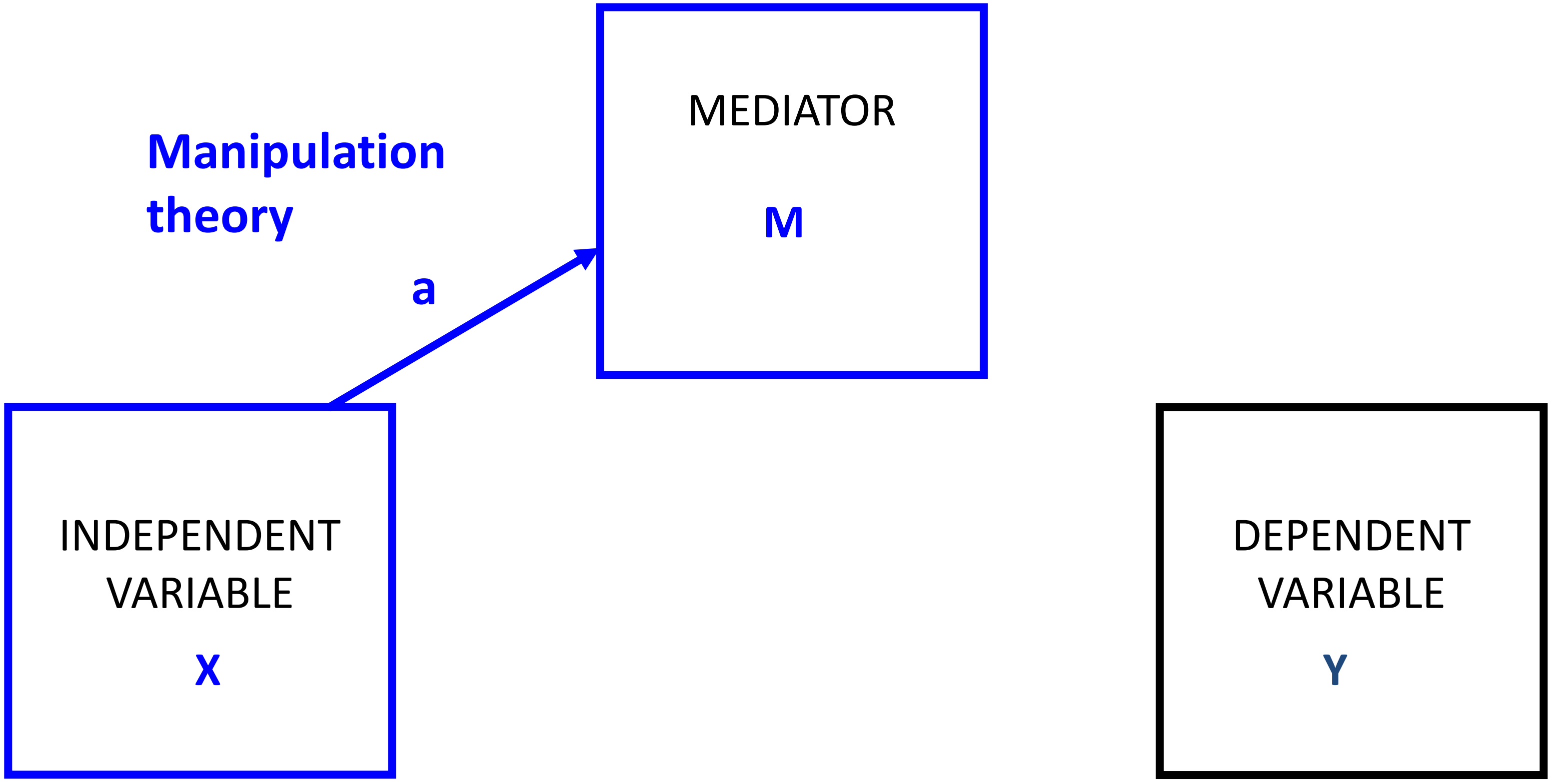
Regression Equation 1



1. The independent variable is related to the dependent variable:

$$Y = i_1 + \hat{c}X + e_1$$

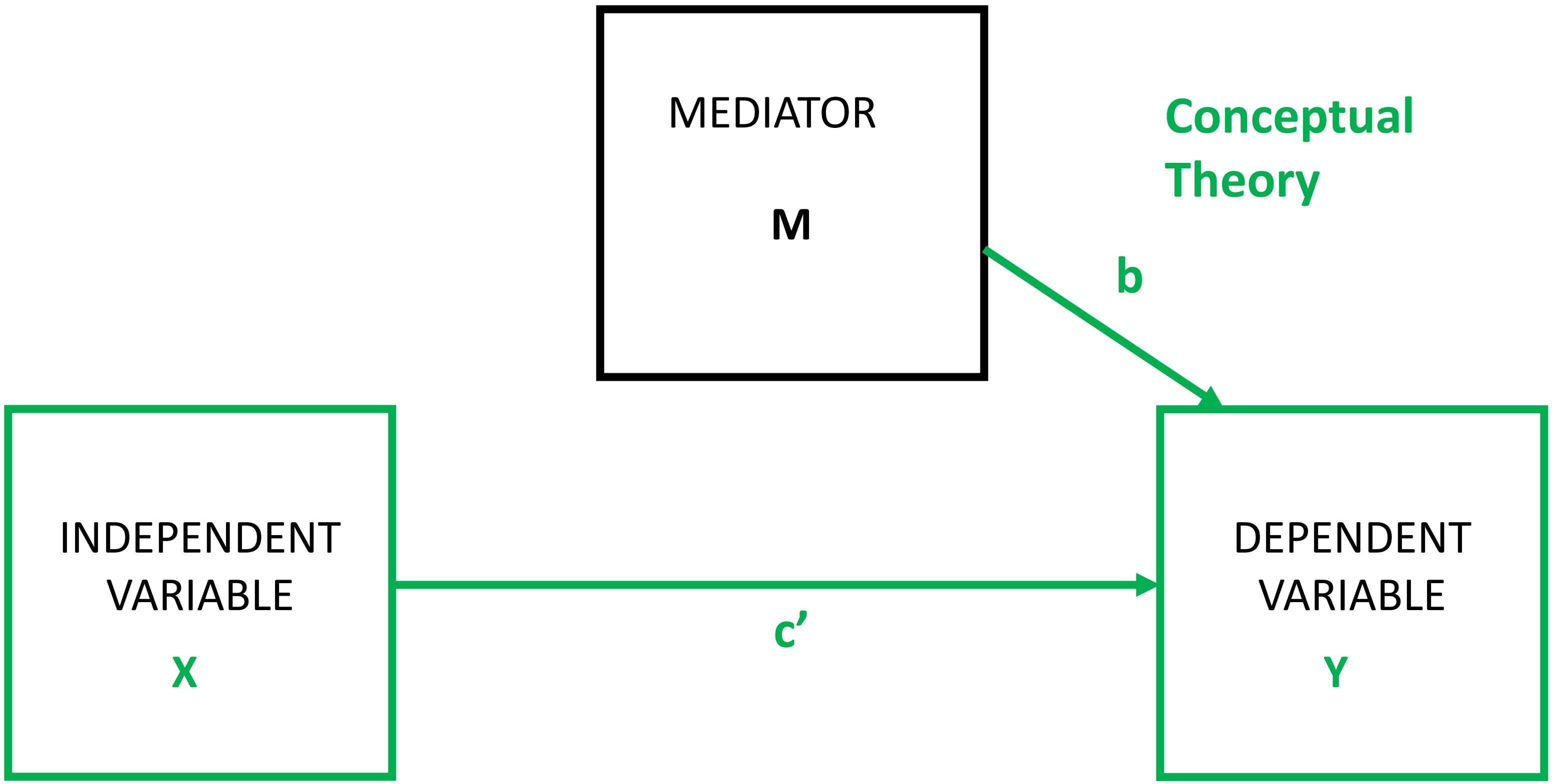
Regression Equation 2



2. The independent variable is related to the potential mediator:

$$M = i_2 + \hat{a}X + e_2$$

Regression Equation 3



3. The mediator is related to the dependent variable controlling for exposure to the independent variable:

$$Y = i_3 + \hat{c}' X + \hat{b} M + e_3$$

Mediated Effect Measures

Mediated effect = ab Product of Coefficients

Mediated effect = c-c' Difference in Coefficients

Mediated effect = ab = c-c'
(see MacKinnon et al., 1995 for a proof)

Direct effect = c'   &   Total effect = ab+c' = c

Mediated Effect, $\hat{a}\hat{b}$, Standard Error

Mediated effect = $\hat{a}\hat{b}$, Standard error = $\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$

Multivariate delta method standard error

(Sobel 1982)

Test for significant mediation:

$$z' = \frac{\hat{a}\hat{b}}{\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}}$$

Compare to empirical distribution of the mediated effect

Assumptions

- For each method of estimating the mediated effect based on Equations 1 and 3 (c - c') or Equations 2 and 3 (ab):
- Reliable and valid measures
- Coefficients, a, b, c' reflect true causal relations and the correct functional form. No omitted influences.
- Mediation chain is correct: Temporal ordering is correct X before M before Y.
- Homogeneous effects across subgroups: It assumed that the relation from X to M and from M to Y are homogeneous across subgroups or other characteristics of participants in the study. No moderators.

Significance Testing and Confidence Limit Estimation

- Product of coefficients estimation of the mediated effect, *ab*, and standard error is the most general approach with best statistical properties.
- Best tests are the **Joint Significance**, **Distribution of the Product**, and **Bootstrap** for confidence limit estimation and significance testing (MacKinnon et al., 2004; 2007).

Empirical Sample size estimates for .8 power to detect the mediated effect

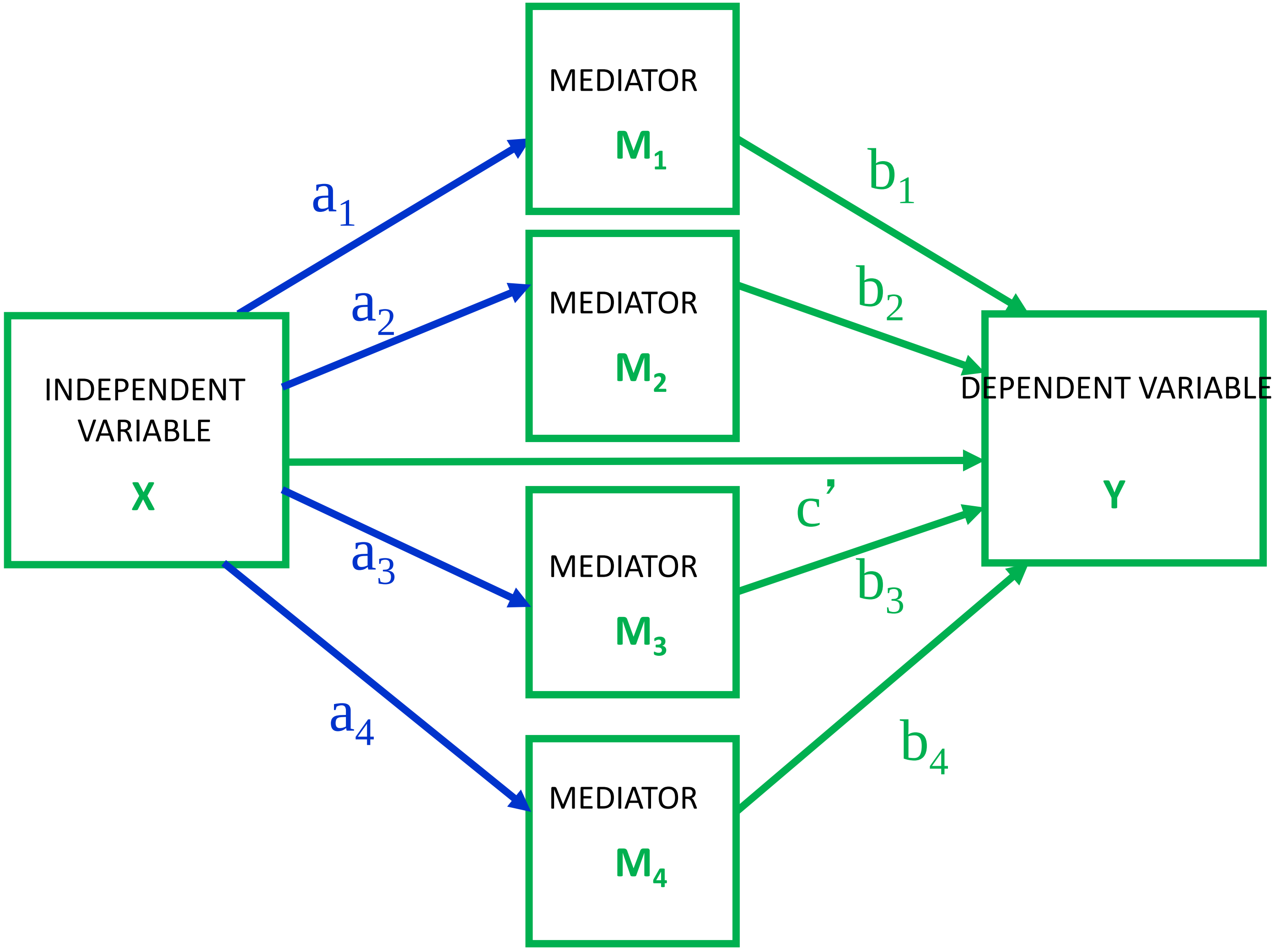
Test	S-S	S-M	M-M	L-L
Causal Steps (c' = 0)	20886	3039	397	92
Normal	667	422	90	42
Dist. Product	539	401	74	35

Note: N required for a complete mediation model, *c'* = 0;. Table entries are based on empirical simulation so they are not exact (Fritz & MacKinnon, 2007). S=small, M= medium, and L=large approximate effect size. For example, S-S means small effect size for the *a* path and small effect size for the *b* path.

Mediation and Nonsignificant X on Y Effect

- It is important to conduct mediation analysis whether an overall effect of X on Y is statistically significant or not.
- It is possible to obtain a nonsignificant overall effect of X on Y but statistically significant mediation (O'Rourke & MacKinnon, 2015).
- Mediation analysis also provides information about **Manipulation theory** (X on M) and **Conceptual Theory** (M on Y). Failure of one or both theories could lead to a nonsignificant effect of X on Y.

Parallel Four Mediator Model



Mediation Effects

Mediated effects = a_1b_1, a_2b_2, a_3b_3, a_4b_4

Standard error = $\sqrt{a_i^2 s_{b_i}^2 + b_i^2 s_{a_i}^2}$

Total mediated effect= $a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 = c - c'$

Direct effect= c' Total effect= $a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 + c' = c$

Test for significant mediation:

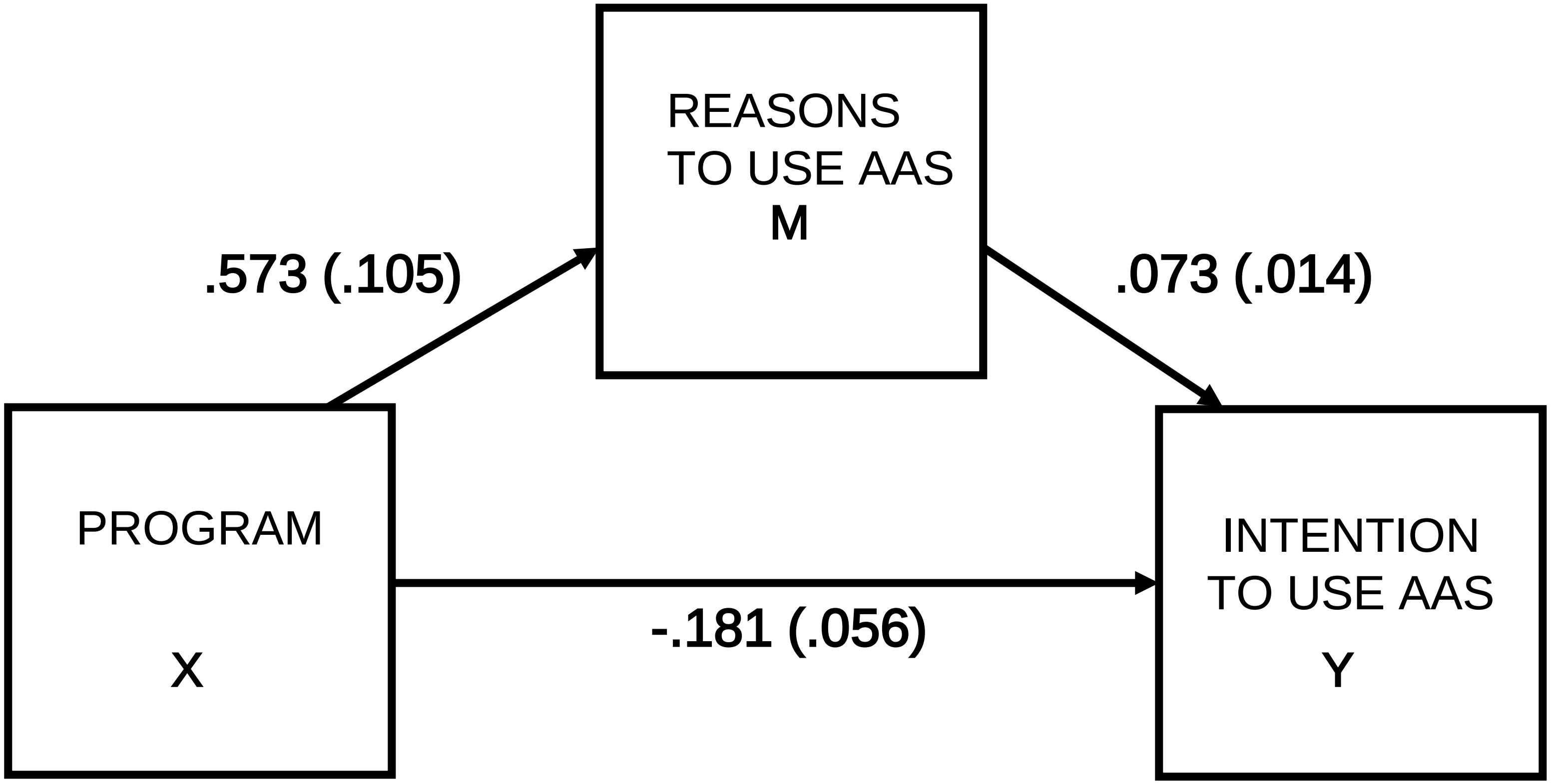
$$z' = \frac{a_1b_1}{\sqrt{a_i^2 s_{b_i}^2 + b_i^2 s_{a_i}^2}}$$

Compare to empirical distribution of the mediated effect

Inconsistent Mediation Models

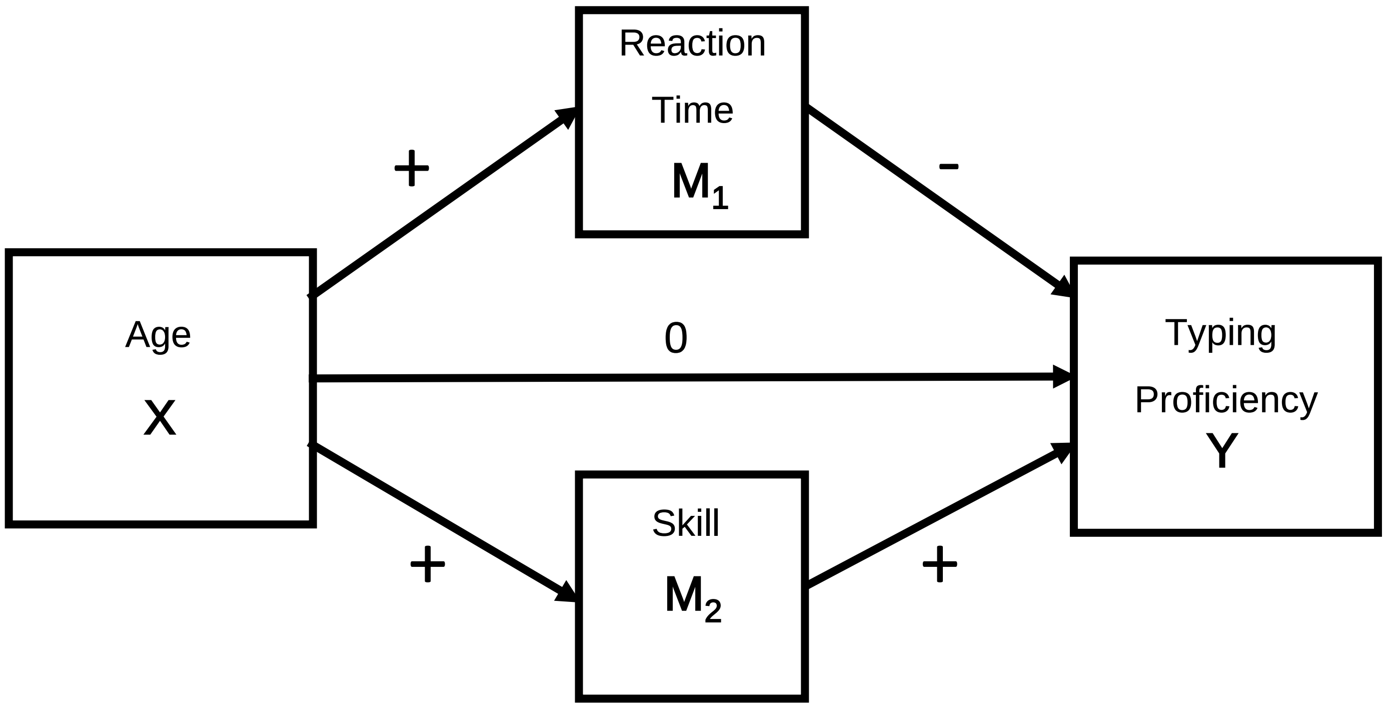
- An inconsistent mediation model has at least one mediated effect with a different sign than the direct effect or other mediated effects (MacKinnon et al., 2000)
- There is mediation because the mediator transmits the effect of the independent variable to the dependent variable. Inconsistent mediation can occur whether or not $\hat{c}$ is statistically significant.
- Intervention studies may have a mediator that is counterproductive. The best way to find these variables is to use mediation analysis.

Inconsistent Mediation in Steroid Prevention Study



Mediated effect = .042
Standard error = .011

Mediators of the null effect of age on typing (Salthouse, 1984)



Compensation - compensate for loss of capacity with other methods. Compensation implies opposing mediational processes for the effect of aging (Baltes, 1997).

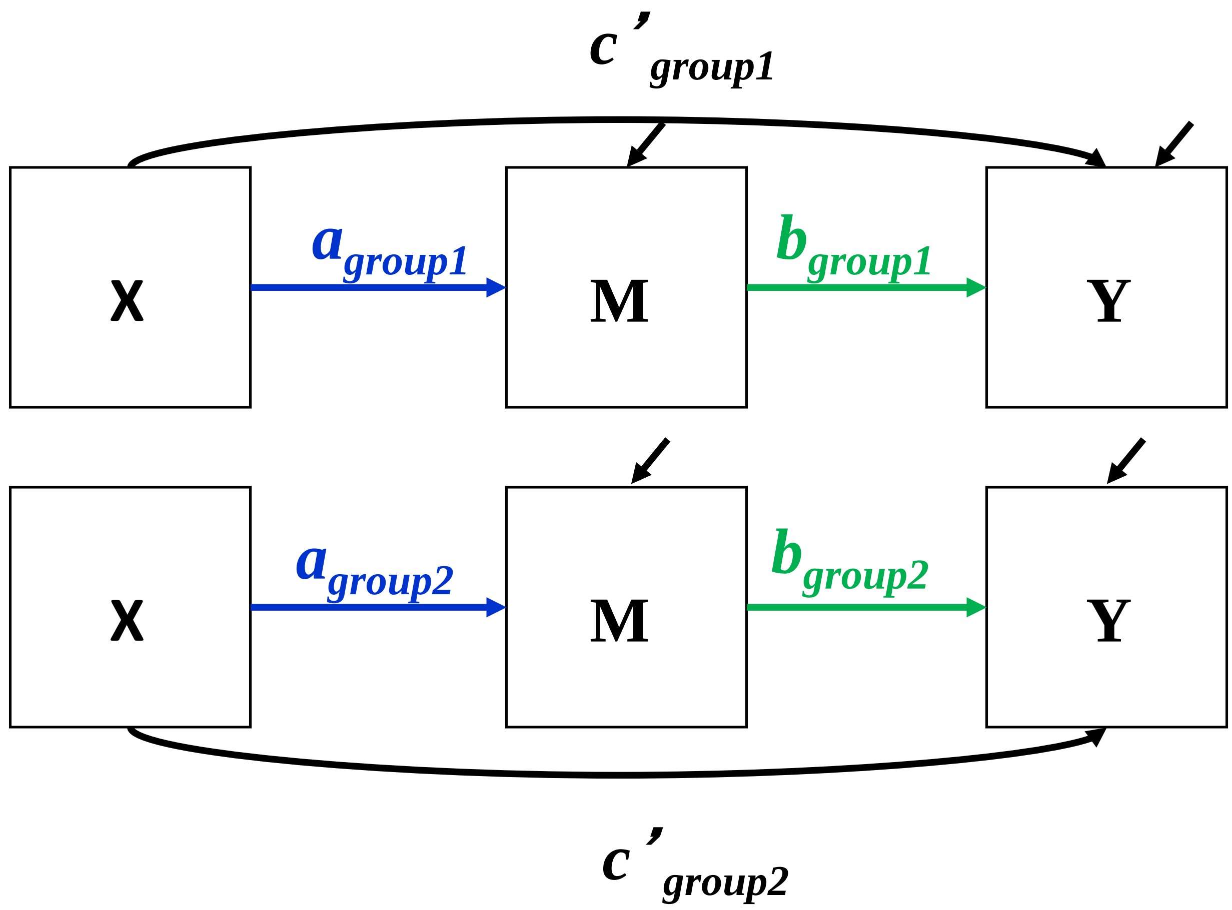
Path Model for Testing Homogeneity of Effects across Groups

GROUP 1

Mediated effect: $a_{grp1}b_{grp1}$

GROUP 2

Mediated effect: $a_{grp2}b_{grp2}$



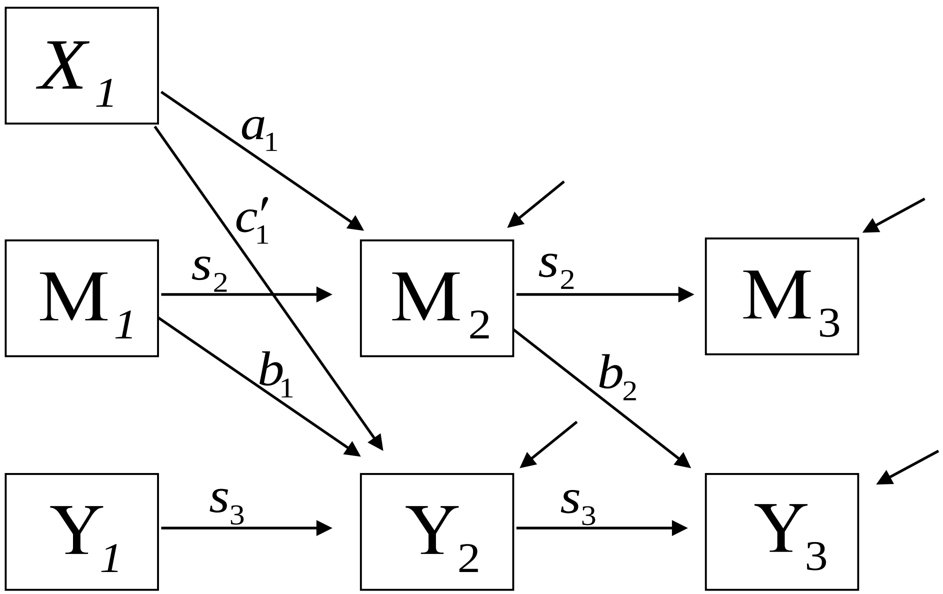
Longitudinal Mediation Analysis

- Mediation is a longitudinal model. Assume correct temporal ordering: X before M before Y.
- Relations among X, M, and Y are at some equilibrium so the observed relations are not solely due to when they are measured, i.e., if measured 1 hour later a different model would apply.
- Correct timing and spacing of measures to detect effects.
  - When does X affect M and M affect Y
  - Triggering, cascading, and other timing processes may be at work (Tang & DeRubeis, 1999; Howe et al., 2002)
  - Timing is crucial for deciding when to collect longitudinal measures (Collins & Graham, 2002)

What if Repeated Measures of X, M, and Y are Available?

- Measures of X, M, and Y at two time points: difference score, ANCOVA, residualized change score.
- Measures of X, M, and Y at three or more time points: Autoregressive, Latent Growth, Latent Change Score Models, Survival Models, and methods to reduce to a few measures, e.g. Area Under the Curve.
- For intervention research, X is usually measured once and represents random assignment of participants to one of two groups.

Autoregressive Model with Time-Ordered Mediation



Note: All residuals are correlated

(Cole & Maxwell, 2003; MacKinnon, 1994; 2008)

Causal Inference in Mediation

- Methods above assume true causal relations and no omitted variables for mediation analysis.
- Blalock's (1979) presidential address--about 50 variables are involved in sociological phenomenon. Comprehensive health psychology models. How many variables are relevant for your research?
- Problem with mediation analysis because M is not randomly assigned but is self-selected.

Counterfactual/ Potential Outcome Models

- Most modern causal inference approaches are based on a counterfactual or potential outcome model.
- These models consider a treatment participant if instead they were in the control group and a control participant if instead they were in the treatment group.
- All the possible counterfactual and actual conditions of an experiment are considered and the statistical model is based on all these potential conditions.

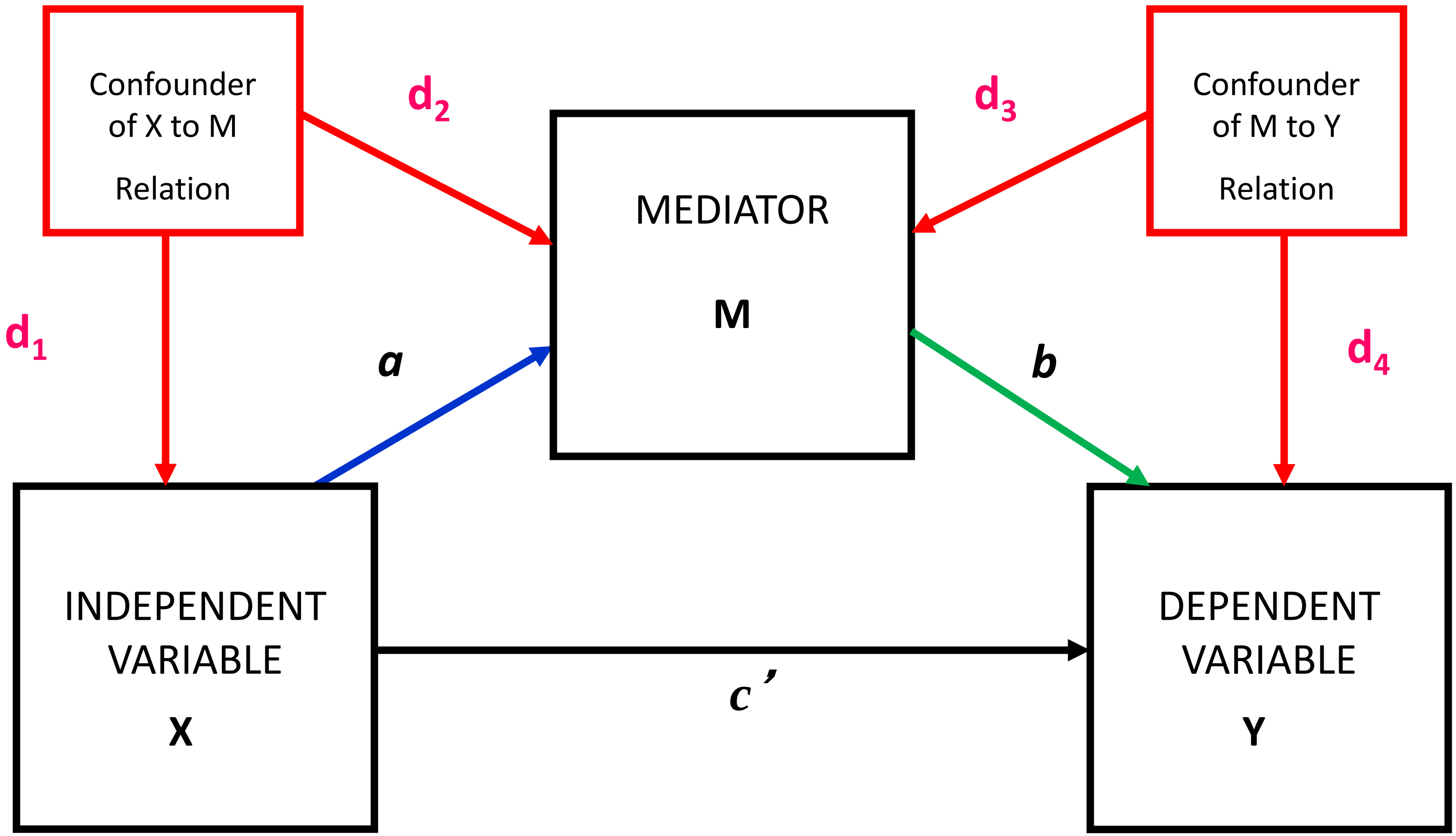
Randomized Two Group Design

- Ideally we need the same individual in both the treatment and control conditions at the same time. Units (individual level) usually have observed data for one of two conditions but not the other—the fundamental problem of causal inference (Holland, 1986).
- Randomization of a large number of persons resolves the fundamental problem of causal inference. The average in each group can be compared and is an estimator of a causal effect. It is called an average causal effect (ACE).

Why *b* and *c*′ Do Not Reflect a Causal Relation

- Because M is not under experimental control, *b* and *c*′ do not necessarily represent causal effects.
- Need: The relation between M and Y for participants in the treatment group if they were in the control group; the relation between M and Y for control participants if they instead were in the treatment group. Coefficients *b* and *c*′ are not Average Causal Effects, because M is not randomized making the counterfactuals for these relations complicated.

Confounders of Mediation Relations



True model needs d_1,d_2, d_3, d_4, otherwise coefficients are confounded.

40

Sensitivity Analysis for Confounding

- How will results change with confounding of the M to Y relation, e.g. when X is randomized?
- Adaptation of Left Out Variables Error (LOVE; Mauro, 1990) based on the correlation of a confounder with Y and the correlation of a confounder with M.
- VanderWeele (2010), confounder effect on Y and difference in proportions of the confounder between groups at level of M.
- Imai et al. (2010), confounder effect as the correlation between error terms.
- See Cox et al., 2014, *Evaluation Review*.

Statistical Methods for Confounding

- Statistical approaches to improve causal inference from a mediation study. A way to deal with omitted variable bias.

  1) Instrumental Variable Methods
  2) Principal Stratification
  3) Inverse Probability Weighting
  4) G-estimation

- Active area of research (MacKinnon & Pirlott, 2015, *Personality and Social Psychology Review*)…

Inverse Probability Weighting

- Method to adjust results for confounders.
- Assumes no unmeasured confounding.
- Weights observations as a way to deal with confounding, missing data etc.
- With X randomized, weights are used to adjust for confounding of the M to Y relation.
- Robins, Hernan, & Brumbeck (2000) and Coffman (2011).

Design Approaches to Improving Causal Inference

- Statistical mediation analysis answers, "How does a researcher use measures of the hypothetical intervening process to increase the amount of information from a research study?"
- Another question is, "What is the best next study or studies to conduct after a statistical mediation analysis to test mediation theory."

1. Designs to address **Consistency** of the mediation relation.
2. Designs to address **Specificity** of the mediation relation.

MacKinnon, 2008; MacKinnon, Cheong, & Pirlott, 2013 related to Hill's (1971) considerations. Also SMART designs (Almirall et al., 2014)

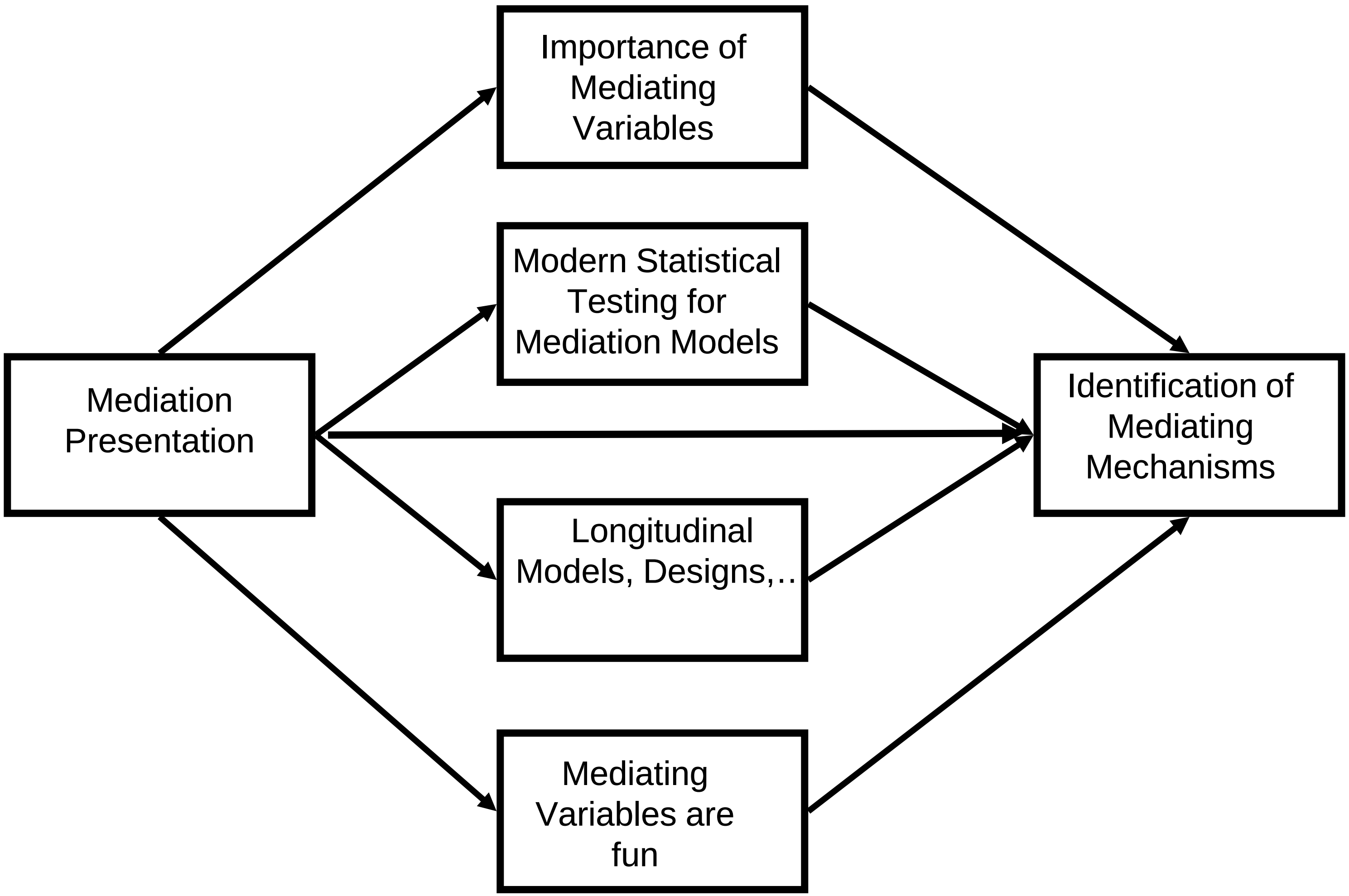
Future Directions

- Mediation Meta-Analysis and Data Synthesis to combine information across studies.
- Person-oriented Methods.
- Qualitative Mediation Methods.
- Mediation for nonlinear models, logistic regression, survival analysis. Potential outcomes models important for this work.
- Bayesian Mediation Analysis to include prior information in a mediation analysis.

Summary

- Mediation analysis is important because it provides information on how variables are related, e.g., how an intervention achieved its effects, how effects unfold over time… It tests theory.
- Tests of mediation based on product *ab with* distribution of product/bootstrap are the most accurate.
- Multiple Mediator Models, Models with Moderation and Mediation, Longitudinal Mediation Models and others are available.
- Causal inference is an active research area generating new methods to investigate confounder bias and experimental designs are available.

Hypothesized Effects of Mind the Gap Presentation



Thank You

References available by contacting David.MacKinnon@asu.edu

Questions?

Submit questions to

prevention@mail.nih.gov

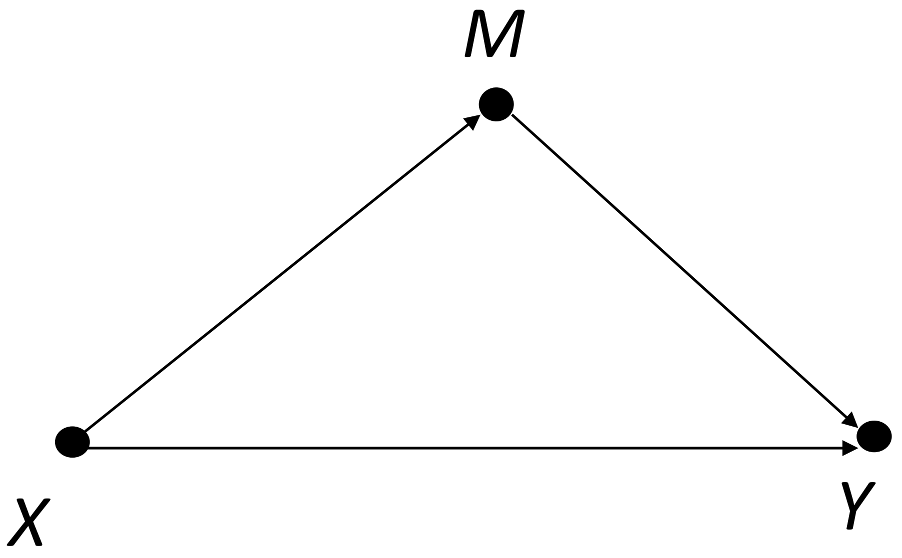
OR

Use @NIHprevents & #NIHMtG on Twitter

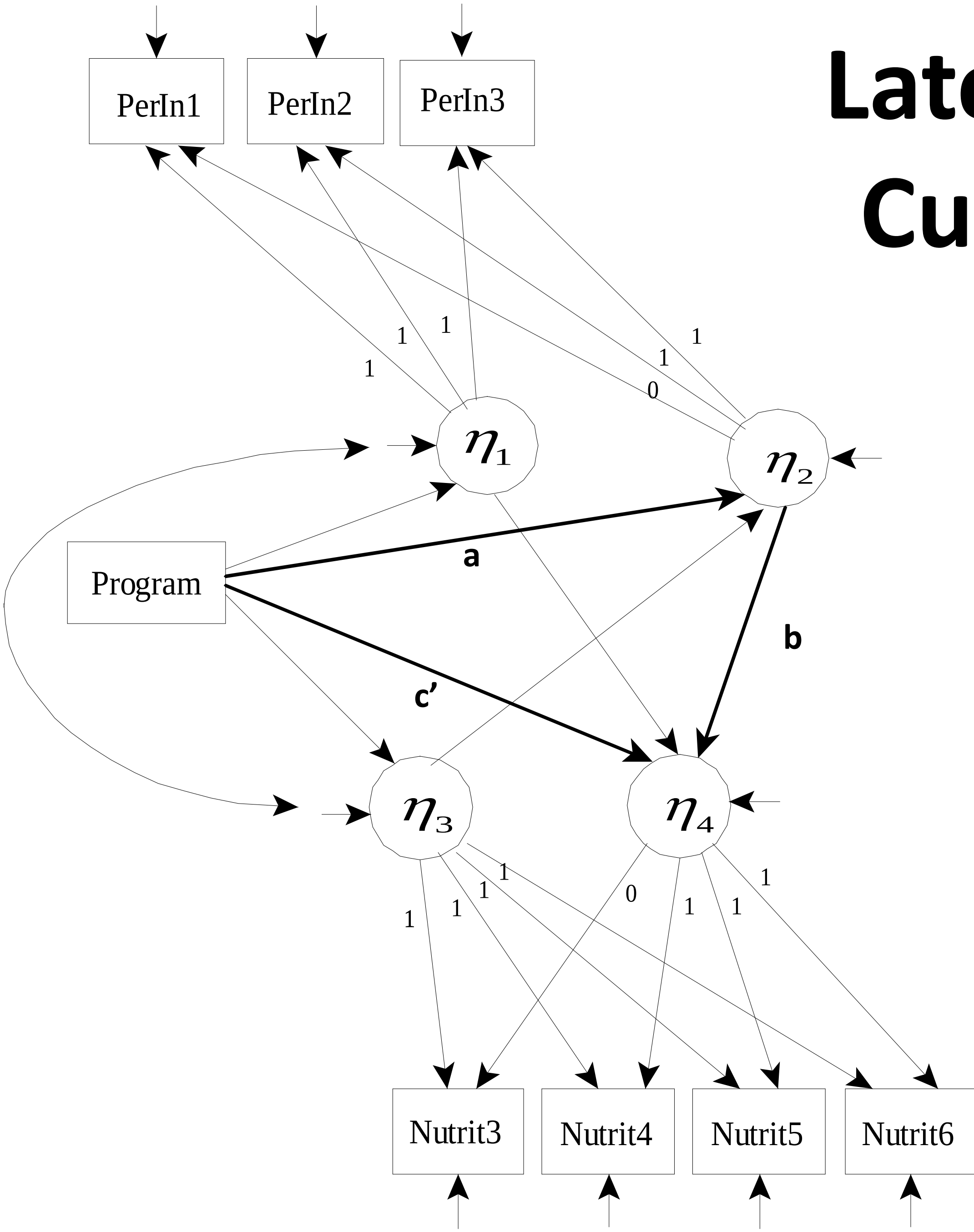
Mediation and Theory

- Mediating mechanisms are central to theory.
- How does an intervention lead to changes?
- What mediators are changed by an intervention that affect the outcome?
- Why does a manipulation change one mediator and not another?
- It is important to specify mediating theory when planning research. Just the task of describing this theory is beneficial.

Directed Acyclic Graph



Latent Growth Curve Model



Cheong et al. (2003)