Tutorial In Modern Mediation Analysis

David P. MacKinnon
Arizona State University
American Psychological Association
San Francisco, CA
August 11, 2018

Introductions
Tutorial Goals
Definitions
Examples of Mediating Variables
History

Introductions

- Undergraduate Social Psychology Class from Charles Judd around 1978 at Harvard University
- Graduate School at the University of California, Los Angeles Quantitative Psychology
- Drug Prevention Research at University of Southern California
- Support from the **National Institute on Drug Abuse** including MERIT award

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

- MacKinnon, D. P. (2008) Introduction to Statistical Mediation Analysis, Mahwah, NJ: Erlbaum. 2nd Edition in process.
- Introductions in small groups

Introduction Questions

What is your name?

Where are you from?

Why are you taking this tutorial?

What is your area of interest?

Tutorial Activities

- Lecture
- Small Group Activities
- Examples
- Questions and Feedback

Tutorial Goals

- Understand Theoretical Motivation for Mediating Variables.
- Understand Practical Motivation for Mediating Variables.
- Understand Single Mediator Analysis Model.
- Exposure to Multiple Mediator Models, Inconsistent Mediation Models, Models with Mediation and Moderation, Longitudinal Mediation Models, and Causal Mediation Analysis.
- Realize Mediation is Fun and Useful.

Quotes

Nursing ".. Should consider hypotheses about mediators …. that could provide additional information about why an observed phenomenon occurs" (Bennett, 2000).

Children's programs ".. Including even one mediator ….. in a program theory and testing it with the evaluation .. will yield more fruit…." (Petrosino, 2000)

Child mental health "rapid progress … depends on efforts to identify … mediators of treatment outcome. We recommend randomized clinical trials routinely include and report such analyses" (Kraemer et al., 2002).

"Everyone talks about the weather but nobody does anything about it." (Mark Twain)

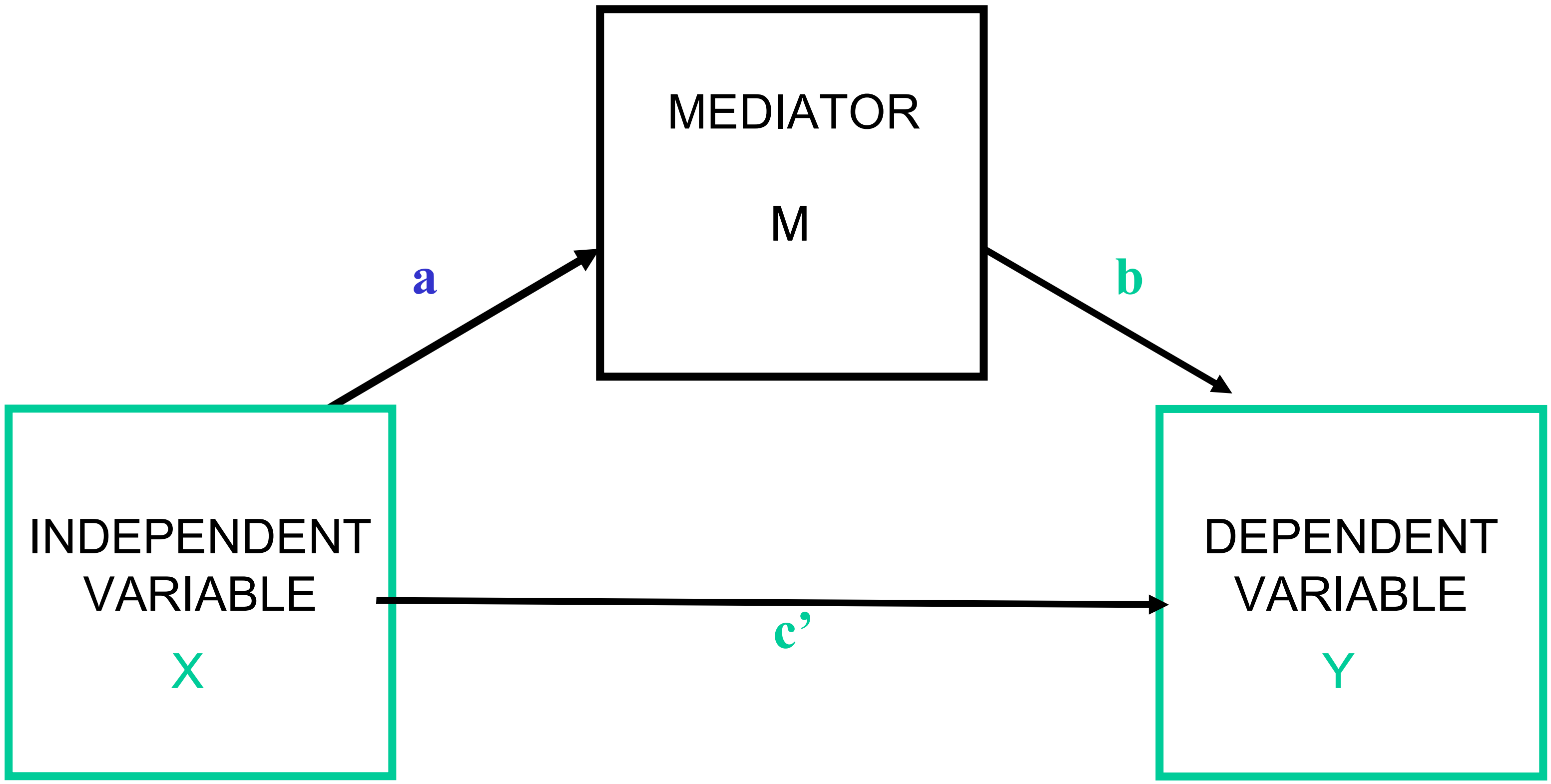
Part I: Introduction

- Overview
- Examples
- Definitions
- History

Three Ways to Specify a Model

- Verbal description: A variable M is intermediate in the causal sequence relating X to Y.
- Diagram
- Equations

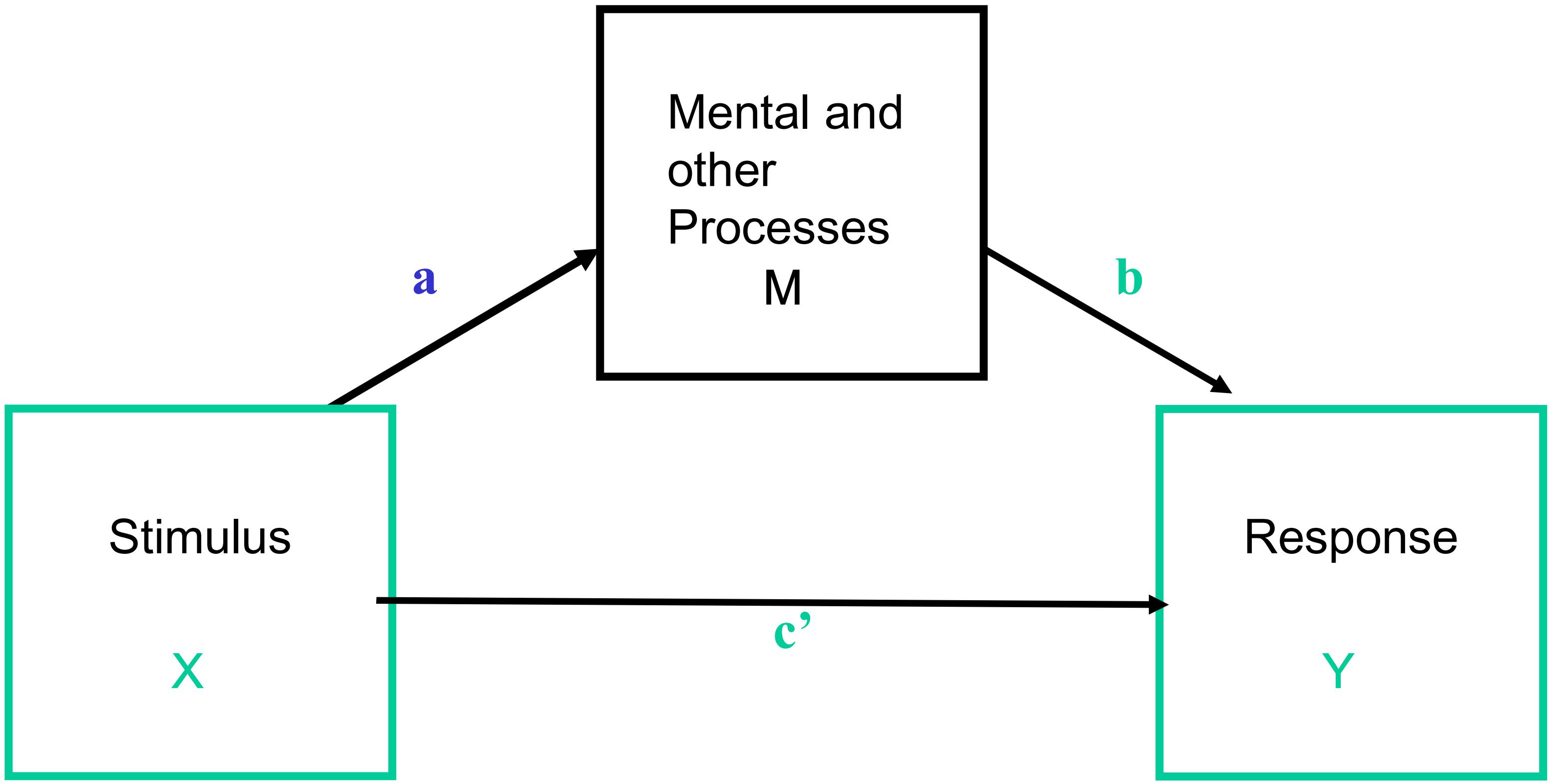
Single Mediator Model



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



S→O→R Theory II

- Note that the mediation process is usually unobservable.
- Process may operate at different levels, individuals, neurons, cells, atoms, teams, schools, states etc.
- Mediating processes may happen simultaneously.
- Mediating process may be part of a longer chain. The researcher needs to decide what part of a long mediation chain to study, the micromediatonal chain.
- Mediation as a measurement problem.

Mediation Statements

- If norms become less tolerant about smoking then smoking will decrease.
- If you increase positive parental communication then there will be reduced symptoms among children of divorce.
- If children are successful at school they will be less anti-social.
- If unemployed persons can maintain their self-esteem they will be more likely to be reemployed.
- If pregnant women know the risk of alcohol use for the fetus then they will not drink alcohol during pregnancy.

Mediating Variable

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

Attitudes cause intentions which then cause behavior (Azjen & Fishbein, 1980)

Prevention programs change norms which promote healthy behavior (Judd & Kenny, 1981)

Increasing exercise skills increases self-efficacy which increases physical activity (Bandura, 1977)

Exposure to an argument affects agreement with the argument which affects behavior (McGuire, 1968)

Clinical Psychology Examples

Psychotherapy induces catharsis, insight, and other mediators which lead to a better outcome (Freedheim & Russ, 1992)

Psychotherapy changes attributional style which reduces depression (Hollon, Evans, & DeRubies, 1990)

Parenting programs reduce parents' negative discipline which reduces symptoms among children with ADHD (Hinshaw, 2002).

Developmental Psychology Examples

- Influence of childhood experiences on later behavior.
- Neglect/Abuse in childhood (X) to impaired threat appraisal (M) to aggressive behavior in adolescence (Y).
- Positive Parenting (X) of an infant predicts self-esteem (M) which predicts positive parenting as an adult (Y).
- Equifinality (different start same end) and Multifinality (same start different end) (Cicchetti & Rogosch, 1996)

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

Many interesting statistical and mathematical issues

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, collider, moderator/interaction.
- Four variables: many possible relations among variables, e.g., $X \rightarrow Z \rightarrow M \rightarrow Y$

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)

Other names for Mediators and the Mediated Effect

- Intervening variable is a variable that comes in between two others.
- Process variable because it represents the process by which X affects Y.
- Intermediate or surrogate endpoint is a variable that can be used in place of an ultimate endpoint.
- Indirect Effect for Mediated Effect to indicate that there is a direct effect of X on Y and there is an indirect effect of X on Y through M.

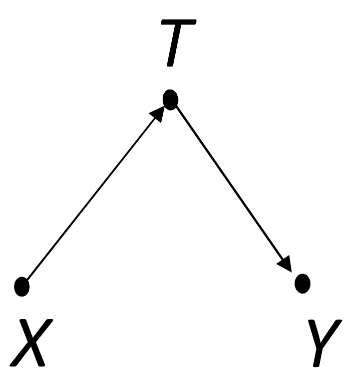
Other names for Variables in the Mediation Model

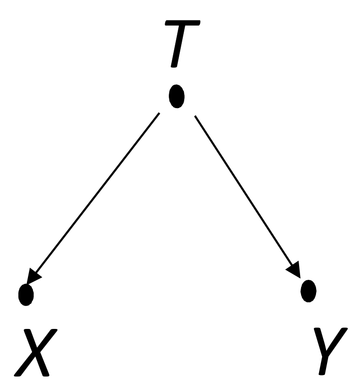
- Initial to Mediator to Outcome (Kenny, Kashy & Bolger, 1998)
- Antecedent to Mediating to Consequent (James & Brett, 1984)
- Program to surrogate (intermediate) endpoint to ultimate endpoint
- Independent to Mediating to Dependent used here.

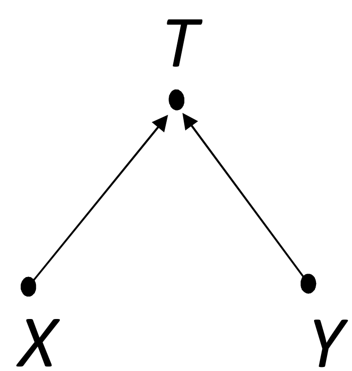
Third-Variable (T) Effects

Mediation is one of three possible causal relations for three variables. There are three fundamental causal relations for three variables, (1) **Mediator**, (2) **Confounder**, and (3) **Collider**.

1. T is Mediator (Chain) 2. T is a Confounder (Fork) 3. T is a Collider (Inverted Fork)

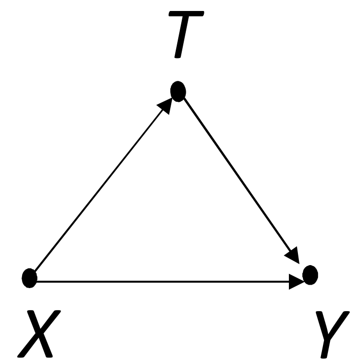




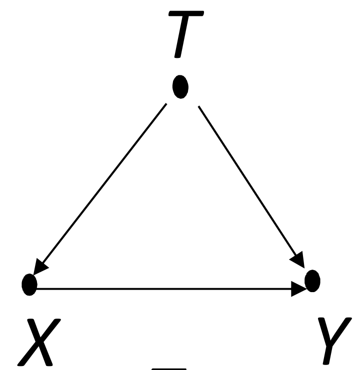


Third-Variable (T) Effects

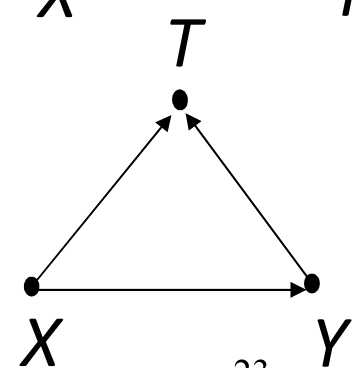
Mediator: This is the focus of the tutorial. A variable that is intermediate in a causal process between X and Y.



Confounder: A variable that causes X and Y such that if it is not included in the analysis an incorrect estimate of the relation between X and Y will be obtained.



Collider: A variable that is caused by X and Y so that it should not be adjusted in the analysis of X and Y because it will incorrectly change the relation between X and Y.



Mediator versus Confounder

- Confounder is a variable related to two variables of interest that falsely obscures or accentuates the relation between them (Meinert & Tonascia, 1986)
- The definition below is also true of a confounder because a confounder also accounts for the relation but it is not intermediate in a causal sequence.
- In general, a mediator is a variable that accounts for all or part of the relation between a predictor and an outcome (Baron & Kenny, 1986, p.1176)

Mediator versus Collider

- Collider is caused by X and Y (Elwert & Winship, 2015). Inaccurate estimates of X to Y will be obtained if adjusted for the collider because the collider causes X and Y. It is not in a causal sequence X to T to Y, as for a mediator.
- Don't adjust for a collider.
- X and Y collide at the collider.
- Talent and Beauty cause Fame example.
- Other examples..

Summary: Mediator, Confounder, Collider

- Mediator-a variable that is intermediate in a causal sequence such that X causes the mediator and the mediator causes Y. The relation between X and Y changes when adjusted for the mediator.
- Confounder-a variable that is related to both X and Y but is not in a causal mediation sequence. The relation between X and Y changes when adjusted for the confounder. Should adjust for confounder.
- Collider-a variable that is caused by X and Y. The relation between X and Y changes when adjusted by the collider but it should not be adjusted for the collider because the collider is caused by X and Y.

Mediator, Collider, or Confounder?

- The effect of **age** is removed from the relation between stress and health symptoms.
- Physical fitness affects **feelings of athletic competence** which then affects body image.
- The relation between risk-taking and alcohol use are evaluated in **drivers who were in a car crash**.
- Intervention changes **norms** which reduces tobacco use.

Mediator versus Moderator

- 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 but it could be in a causal sequence.
- Moderator is usually an interaction, the relation between X and Y depends on a third variable. There are other more detailed definitions of a moderator.

Mediator versus Covariate

- Covariate is a variable that is related to X or Y, or both X and Y, but is not in a causal sequence between X and Y, and does not change the relation between X and Y. Because it is related to the dependent variable it reduces unexplained variability in the dependent variable.
- A covariate is similar to a confounder but does not appreciably change the relation between X and Y so it is related to X and Y in a way that does not affect their relation with each other.

Mediator versus Redundant Measure

- A Redundant Measure (MacKinnon, 2018) is actually another measure of X, another measure of Y, or another measure of both X and Y. If X is randomized, the third variable could be a redundant measure of Y. The measure is most accurately used as an additional measure of Y, rather than as a variable in a causal sequence.
- A redundant measure would change the relation between X and Y because it is related to Y and evidence for mediation may be mistakenly found when the redundant measure is really just another measure of Y. (more in MacKinnon, 2nd Edition).

Summary: Covariate, Moderator, and Redundant Measure

- Covariate- a variable that is related to X or Y or both. The relation between X and Y does not appreciably change when adjusted for the covariate. Not a mediator, confounder, or collider.
- Moderator-a variable where the relation of X to Y is different at different values of the moderator. Moderation can be present for mediators, confounders, and colliders.
- Redundant Measure-a variable that is actually another measure of X, Y, or X and Y. Redundant measures can be present for mediators, confounders, and colliders

Mediator, Moderator, Covariate Redundant Measure, or Confounder?

- The relation of stress to cortisol differs in the **morning compared to the evening**.
- Marriage changes **expectations** regarding alcohol and alcohol expectations affect alcohol use.
- Exposure to violent themes in a music video increases aggressiveness but only among **males**.
- Intervention changes both **Hamilton** and Beck inventory measures of depression.
- Intervention changes **parenting consistency** which reduces externalizing behaviors.

History: Wright's Path Analysis

- Sewall Wright (1923) developed path analysis to investigate hereditary and environmental influences on the color patterns of piebald guinea pigs. Path analysis was based on correlations among measures. Equations and path diagrams were used to represent the path models. Mediation was described as products of coefficients, "**the correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients**." p. 330.

History: Modern Mediation Analysis

- Sociologist O. D. Duncan rediscovers Path Analysis as a way to investigate systems of relations.
- Jöreskog and others combine psychometrics with path analysis models.
- Alwin & Hauser (1975) describe methods of effect decomposition. Sobel (1982) derives standard error of the mediated effect.
- Kenny and colleagues (e.g., Baron & Kenny, 1986) describe mediation analysis in psychology and MacKinnon & Dwyer (1993) describe mediation for interventions.
- Holland (1988) causal mediation model.

Now

- Best methods for testing for mediation.
- Causal inference for mediation models, including evaluation of assumptions.
- Development and evaluation of advanced models: including longitudinal mediation models, path models, models for binary and count data.
- Best program of research to investigate mediation relations…

Most Psychological Experiments do not Measure Mediators

- Historically, psychological experiments are designed to change a mediator but do not measure the mediator.
- Theory is that feeling good leads to helping behavior.
- Gave some participants cookies, that got them in a good mood which increased helping behavior (Isen & Levin, 1972).
- Set up a situation where persons found a dime (It was a long time ago) in a telephone coin return and they were then in a situation where they could help a person. If they found the dime they were more likely to help. (Levin & Isen, 1975).

Manipulations to change mediators

Manipulation designed to change the mediator of feeling good. Feeling good was not measured so there was not a measure of the mediator.

Many experimental studies manipulate the mediator but do not measure it.

Mediation analysis is a method that incorporates measures of the mediator in a statistical analysis.

Applications

Two overlapping reasons for mediation analysis: (1) Mediation for Explanation and (2) Mediation for Design

Mediation for Explanation

- Observed relation and 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 variable (Mediator)
- Specification (Moderator)

Mediation by Design

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

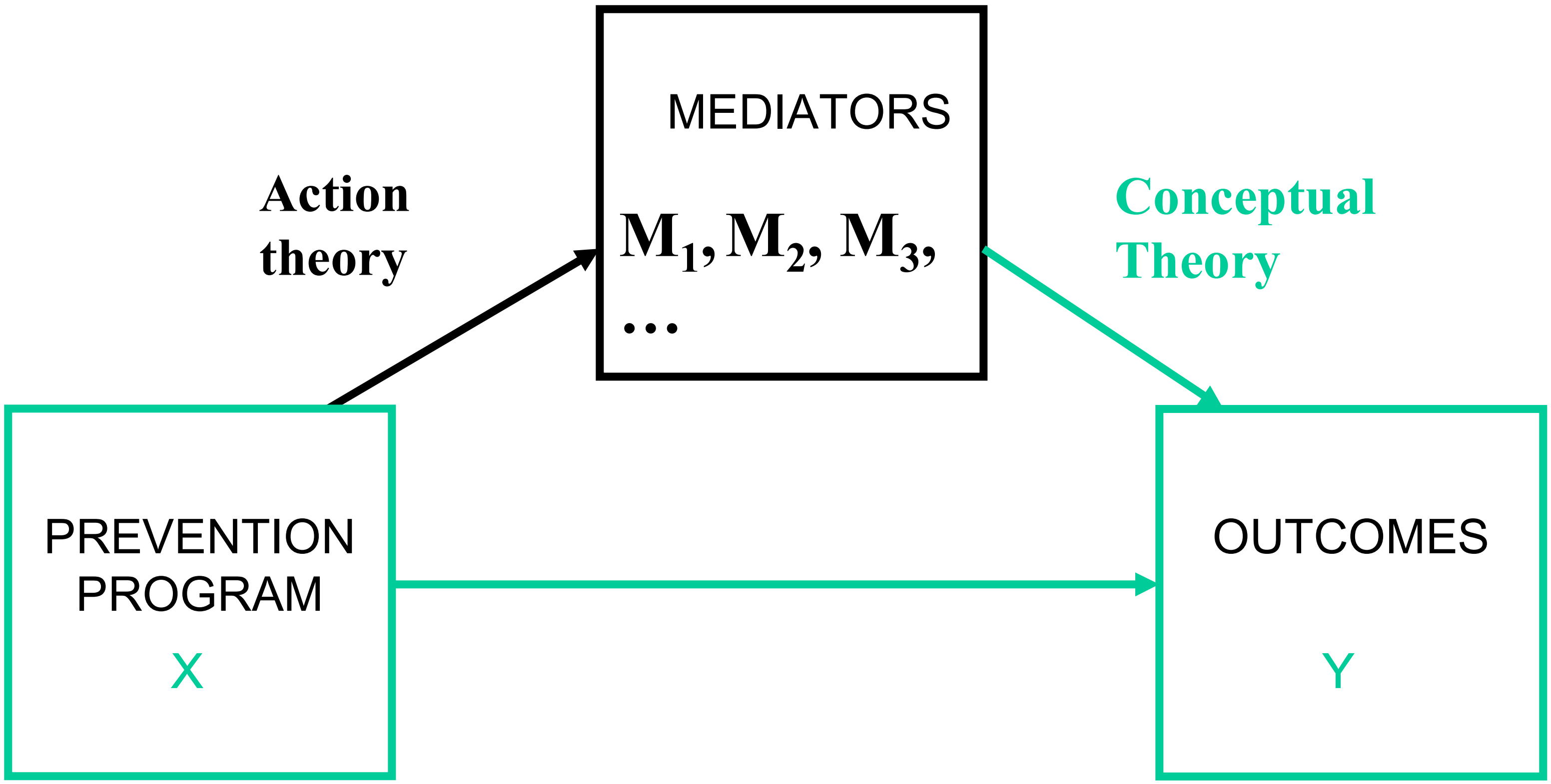
Treatment and Prevention

- Mediators selected for change because they are thought to be causally related to the dependent variable. Often the relation that prevention researchers are most confident about is the M to Y relation.
- Many large scale prevention efforts, alcohol, tobacco, drug use, AIDS/HIV prevention, obesity, poverty….
- Mediation model is the basis of all of them.

Mediation in Intervention Research Theory

- Mediation is important for intervention science. Practical implications include reduced cost and more effective interventions if the mediators of programs are identified. Mediation analysis is an ideal way to test theory.
- A theory based approach focuses on the processes underlying interventions. Mediators play a primary role. **Action Theory** corresponds to how the program will affect mediators. **Conceptual Theory** focuses on how the mediators are related to the dependent variables (Chen, 1990, Lipsey, 1993; MacKinnon, 2008).

Intervention Mediation Model



If the mediators selected are causally related to Y, then changing the mediators will change Y.

Mediators in your research.

Small group activity:

Describe a single mediator model in your research.
X is ?
M is ?
Y is ?

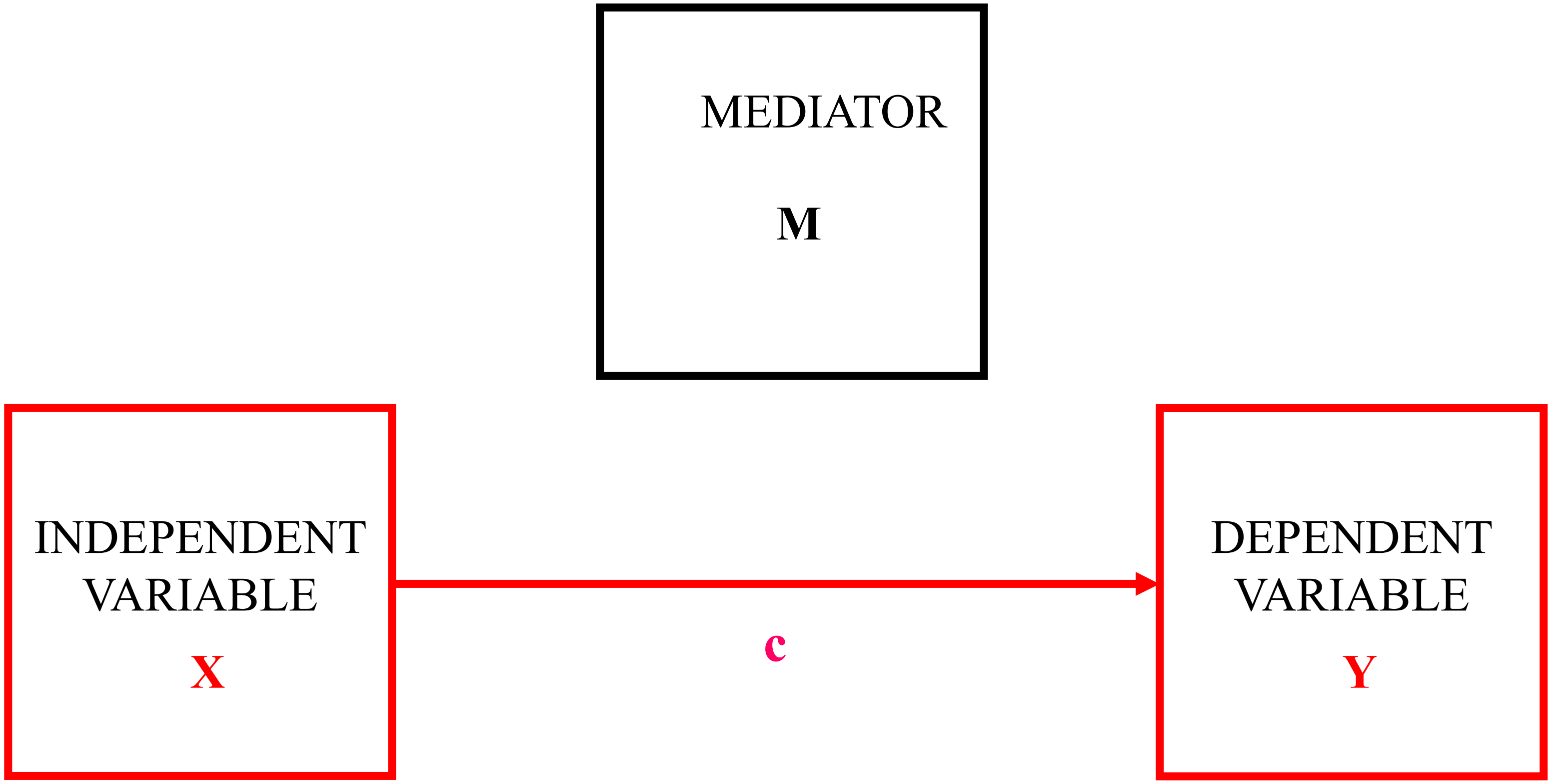
Part II: Statistical Mediation Analysis

- Single mediator model in some detail
- Exposure to Advanced Mediation Models

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.

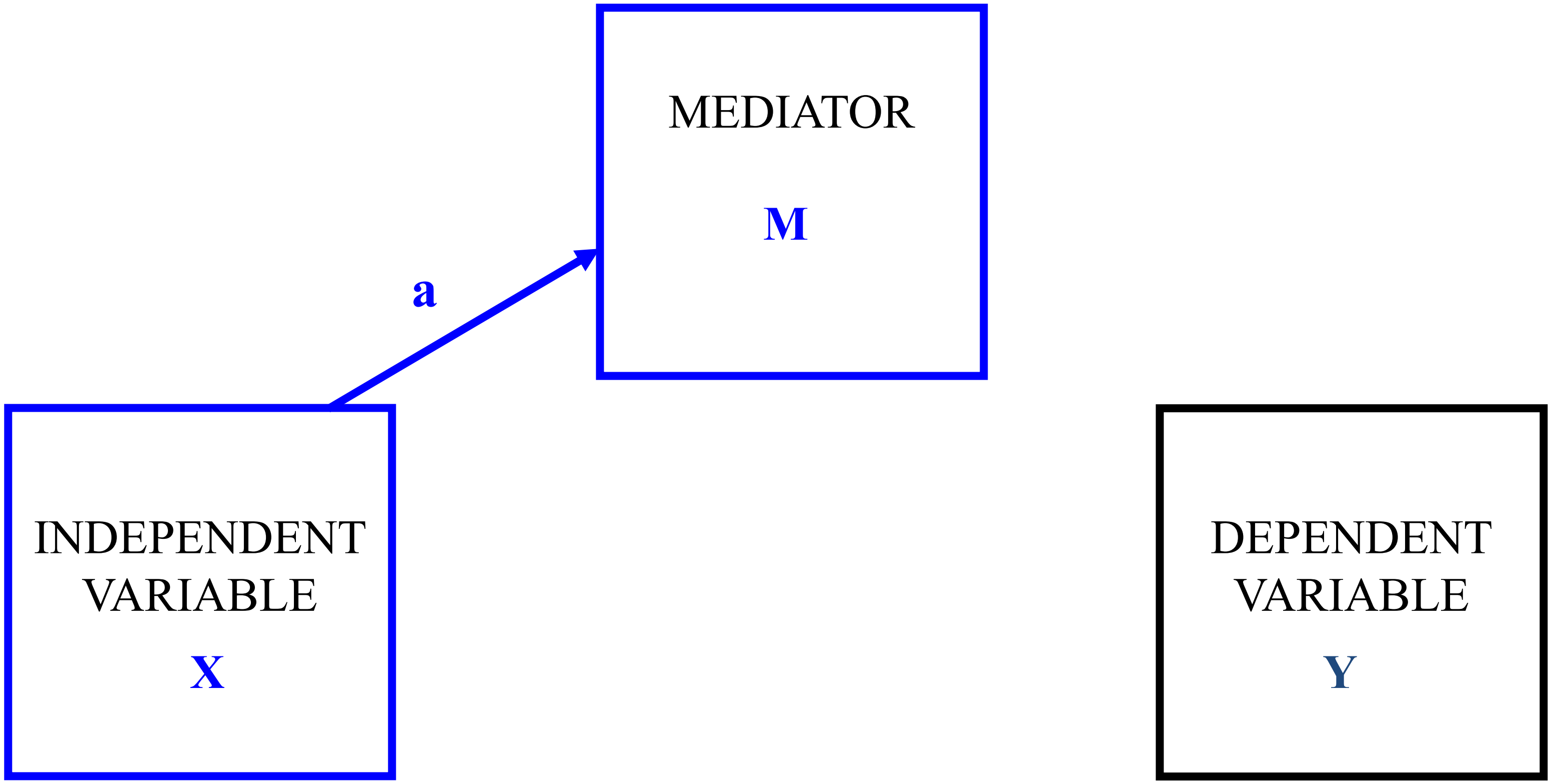
Regression Equation 1



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

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

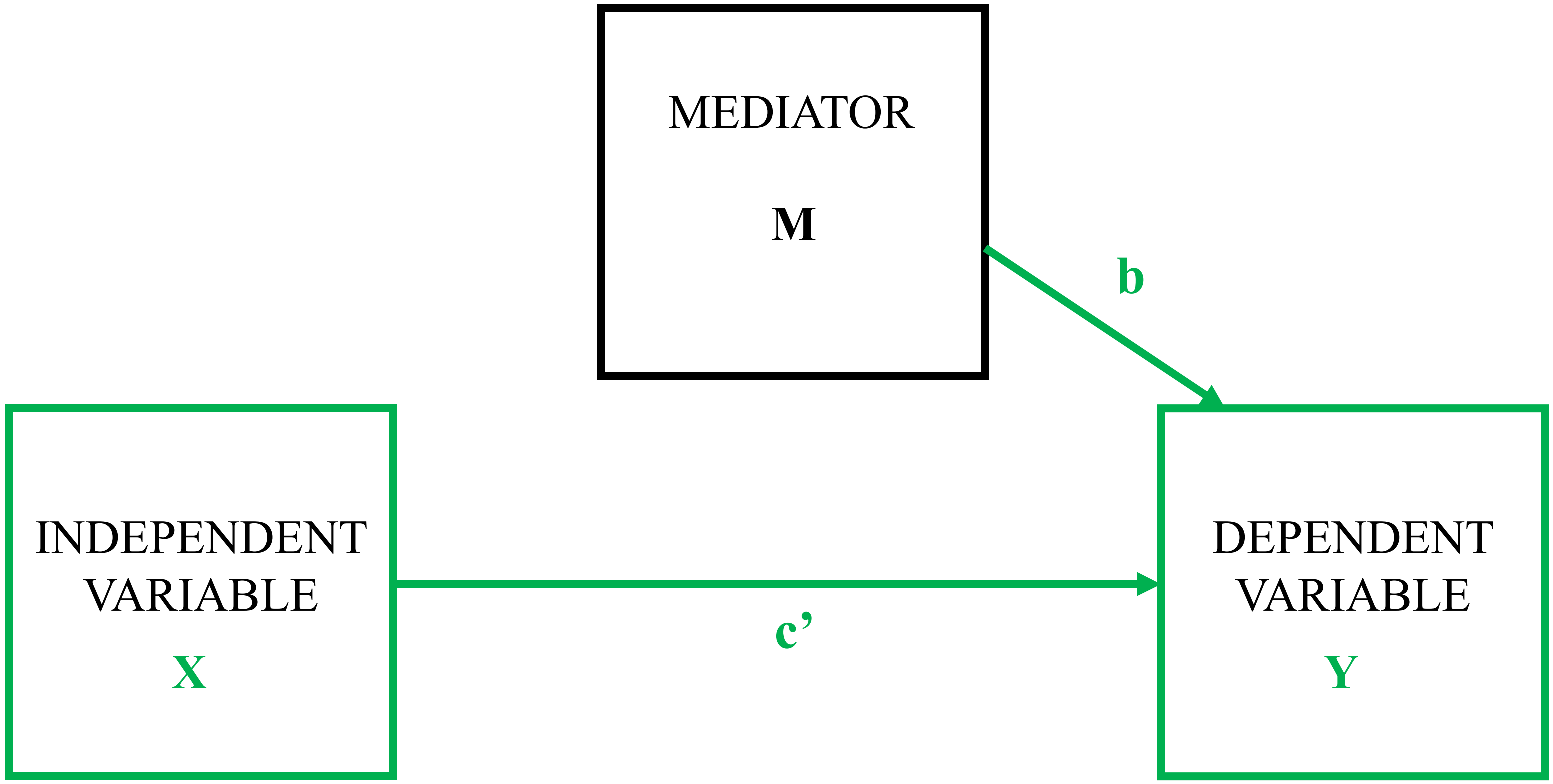
Regression Equation 2



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 = $c\text{-}c'$ **Difference in Coefficients**

Mediated effect = $ab = c\text{-}c'$
(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; Folmer 1981)

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 I

- 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 is 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.

Identification Assumptions

1) No unmeasured X to Y confounders given covariates.
2) No unmeasured M to Y confounders given covariates.
3) No unmeasured X to M confounders given covariates.
4) There is no effect of X that confounds the M to Y relation.

(VanderWeele & Vansteelandt, 2009).

Randomized X satisfies Assumptions 1 and 3 but not 2 and 4.

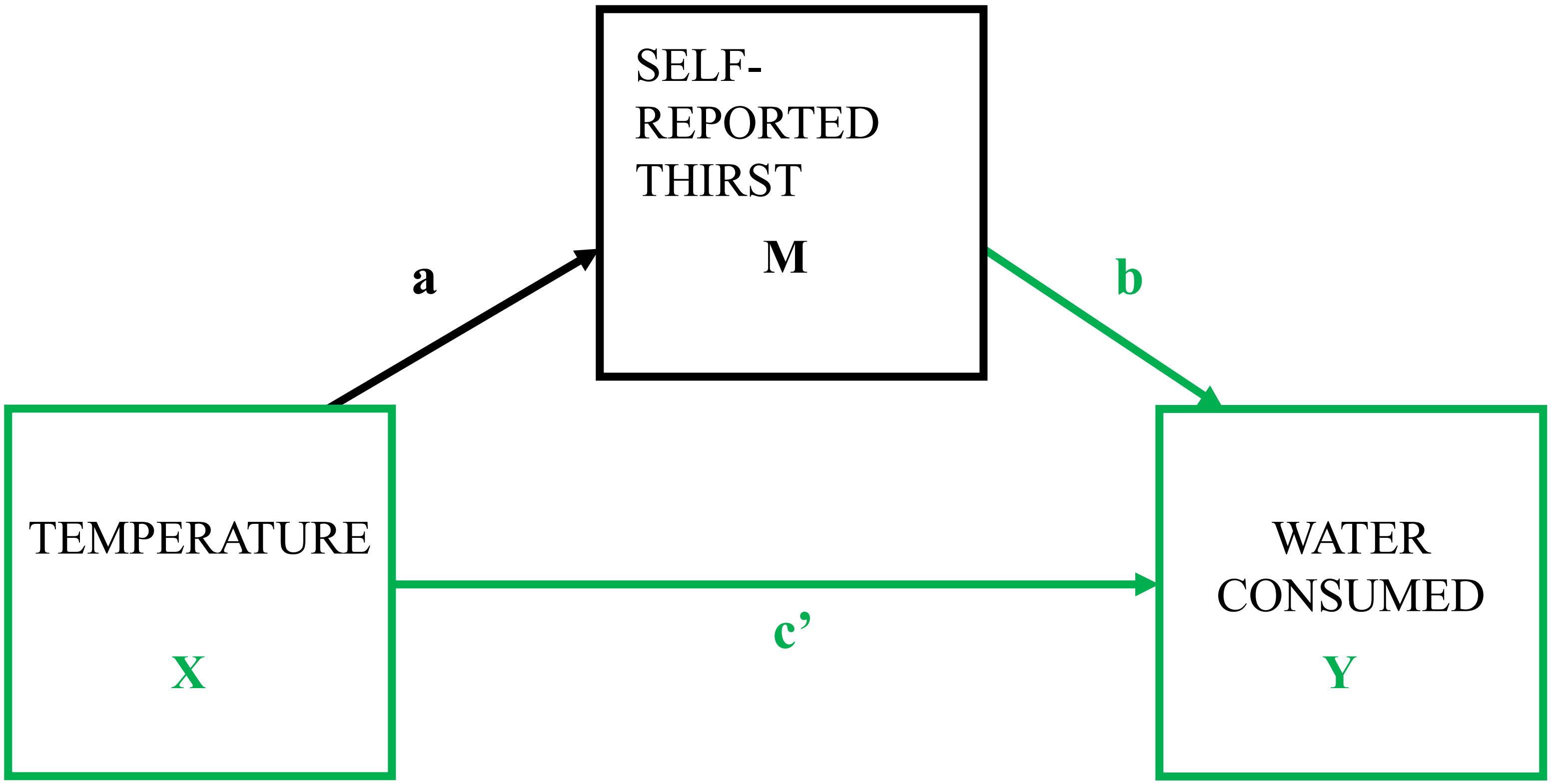
Water Consumption Study Variables

- Stimulus→Organism→Response study
- X is the temperature in degrees Fahrenheit
- M is self-report of thirst at the end of the first two hours of the study
- Y is the number of deciliters of water consumed during the last two hours of the study
- 50 participants were in a room for four hours doing a variety of tasks including sorting objects, tracking objects on a computer screen, and communicating via an intercom system

Water Consumption Study Purpose

- The purpose of the study was to investigate whether persons can judge their water needs. Temperature should affect self-reported thirst which then should affect water consumption.
- The accuracy of self-reported thirst is important because persons in self-contained environments need to monitor their own hydration.
- The mediated effect of temperature on water consumption through self-reported thirst estimates the extent to which persons were capable of gauging their own need for water.

Water Consumption Study



Temperature (X) to self-reported thirst (M) to water consumption (Y).

SAS Program

```
proc reg;
model y=x;
model y=x m;
model m=x;
```

model y=x; ← Estimate *c*.

model y=x m; ← Estimate *c'* and *b*.

model m=x; ← Estimate *a*.

See handout for output.

SPSS Program

```
regression
    /variables x y m
    /dependent=y
    /enter=x.
regression
    /variables x y m
    /dependent=y
    /enter=x m.
regression
    /variables x y m
    /dependent=m
    /enter x.
```

Estimate c.

Estimate c' and b.

Estimate a.

See handout for output.

Estimates of *a*, *b*, *c*, and *c'*

(1) Temperature (X) was significantly related to water consumption (Y) ($\hat{c}$=.3604, $s_{\hat{c}}$=.1343, $t\hat{c}$ = 2.683).

(2) Temperature was significantly related to self-reported thirst (M) ($\hat{a}$=.3386, $s_{\hat{a}}$=.1224, $t\hat{a}$=2.767).

(3) Self-reported thirst was significantly related to water consumption controlling for temperature ($\hat{b}$=.4510, $s_{\hat{b}}$=.1460, $t\hat{b}$=3.090).

-The adjusted effect of temperature was not statistically significant ($\hat{c}'$=.2076, $s_{\hat{c}'}$=.1333, $t\hat{c}'$=1.558) and there was a drop to $\hat{c}'$ = .2076 from $\hat{c}$=.3604.

Mediation Models for Water Consumption Data

$$Y = \hat{i}_1 + \hat{c}\,X$$

$$Y = -22.0505 + \underset{(.1343)}{.3604}\,X$$

$$Y = \hat{i}_2 + \hat{c}'\,X + \hat{b}\,M$$

$$Y = -12.7129 + \underset{(.1333)}{.2076}\,X + \underset{(.1460)}{.4510}\,M$$

$$M = \hat{i}_3 + \hat{a}\,X$$

$$M = -20.7024 + \underset{(.1224)}{.3386}\,X$$

Mediated Effect Measures

Mediated effect

$\hat{a}\hat{b}$= (.3386) (.4510) = $\hat{c} - \hat{c}'$ =.3604-.2076 =.1527

Standard error = $s_{First} = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$

Standard error = $\sqrt{.3386^2(.1460)^2 + .4510^2(.1224)^2} = .0741$

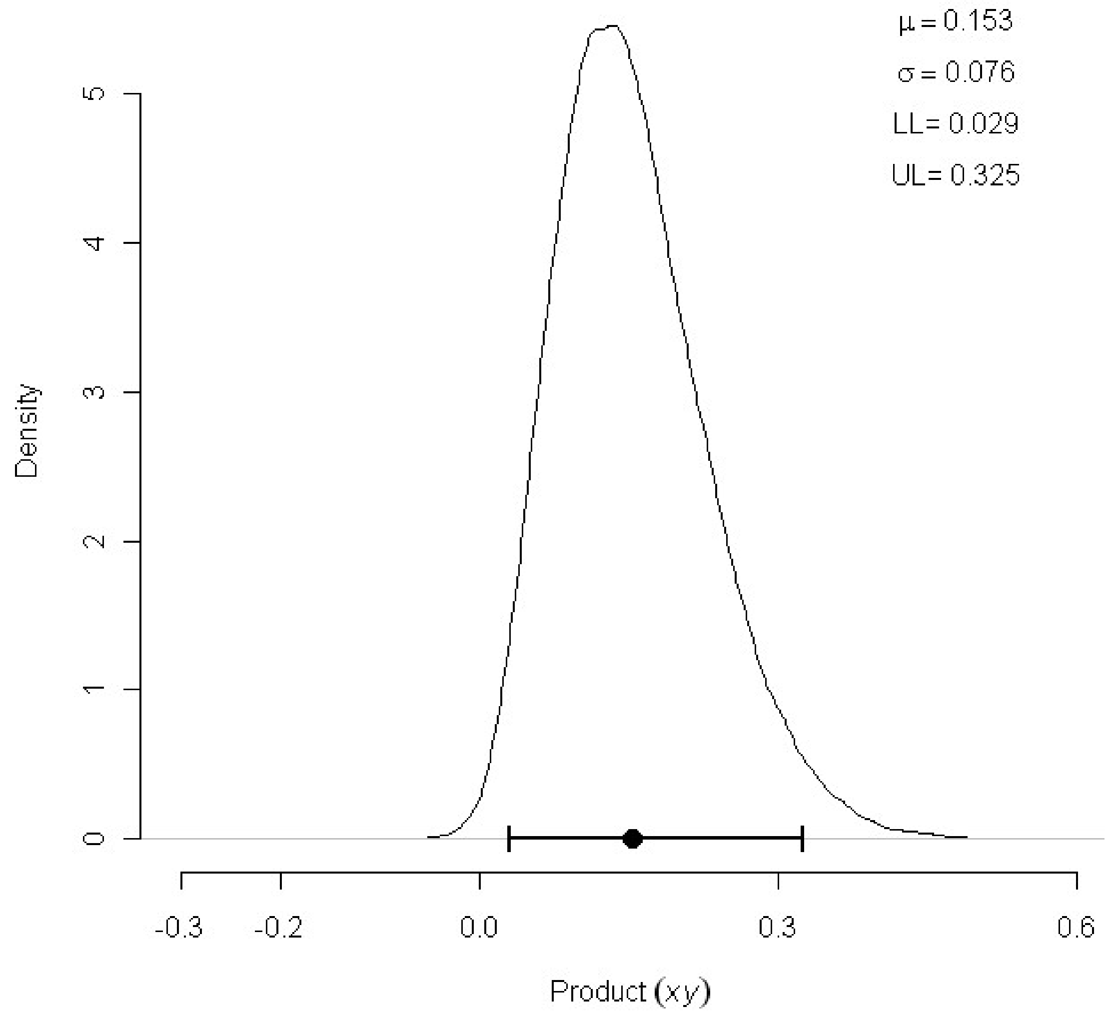
Confidence Intervals for the Mediated Effect Using the Normal Distribution

- Confidence intervals are advocated by researchers for several reasons: effect size, range of possible values, not just null hypothesis binary significance testing. For 95% confidence intervals:
- Upper Confidence Interval (UCL) = $\hat{a}\hat{b} + z_{.975}\, s_{\hat{a}\hat{b}}$
- Lower Confidence Interval (LCL) = $\hat{a}\hat{b} + z_{.025}\, s_{\hat{a}\hat{b}}$
- For water consumption data.
- UCL = .1527 + (1.96)(.0741) = .2979
- LCL = .1527 + (-1.96) (.0741) =.0075
- 95% Confidence Interval from .0075 to .2979. The effect is statistically significant because 0 is not in the interval.

Example Calculations Using the Distribution of the Product

- For example, $\hat{a} = .3386$, $s_{\hat{a}} = .1224$, $\hat{b} = .4510$, $s_{\hat{b}} = .1460$. Enter these values in the PRODCLIN program.
- PRODCLIN uses the critical values for the 2.5% percentile, $M_{lower} = -1.6175$ and $M_{upper} = 2.2540$ the critical value for the 97.5% percentile.
- Use the critical values to calculate upper and lower confidence limits.
- $LCL = \hat{a}\hat{b} + M_{upper}\, s_{\hat{a}\hat{b}} = .1527 + (-1.6175)(.0741)$
  $UCL = \hat{a}\hat{b} + M_{lower}\, s_{\hat{a}\hat{b}} = .1527 + (2.2540)(.0741)$
- Asymmetric Confidence Limits are (.0329, .3197) and (.0294, .3245) from new PRODCLIN.

Plot and Confidence Limits from RMediation (Chapter 3 Data)



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).
- The mediated effect *ab*, does not follow a normal distribution so methods that allow the distribution to be nonnormal are more accurate—the bootstrap and a method based on the distribution of the product (MacKinnon et al., 2004).

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

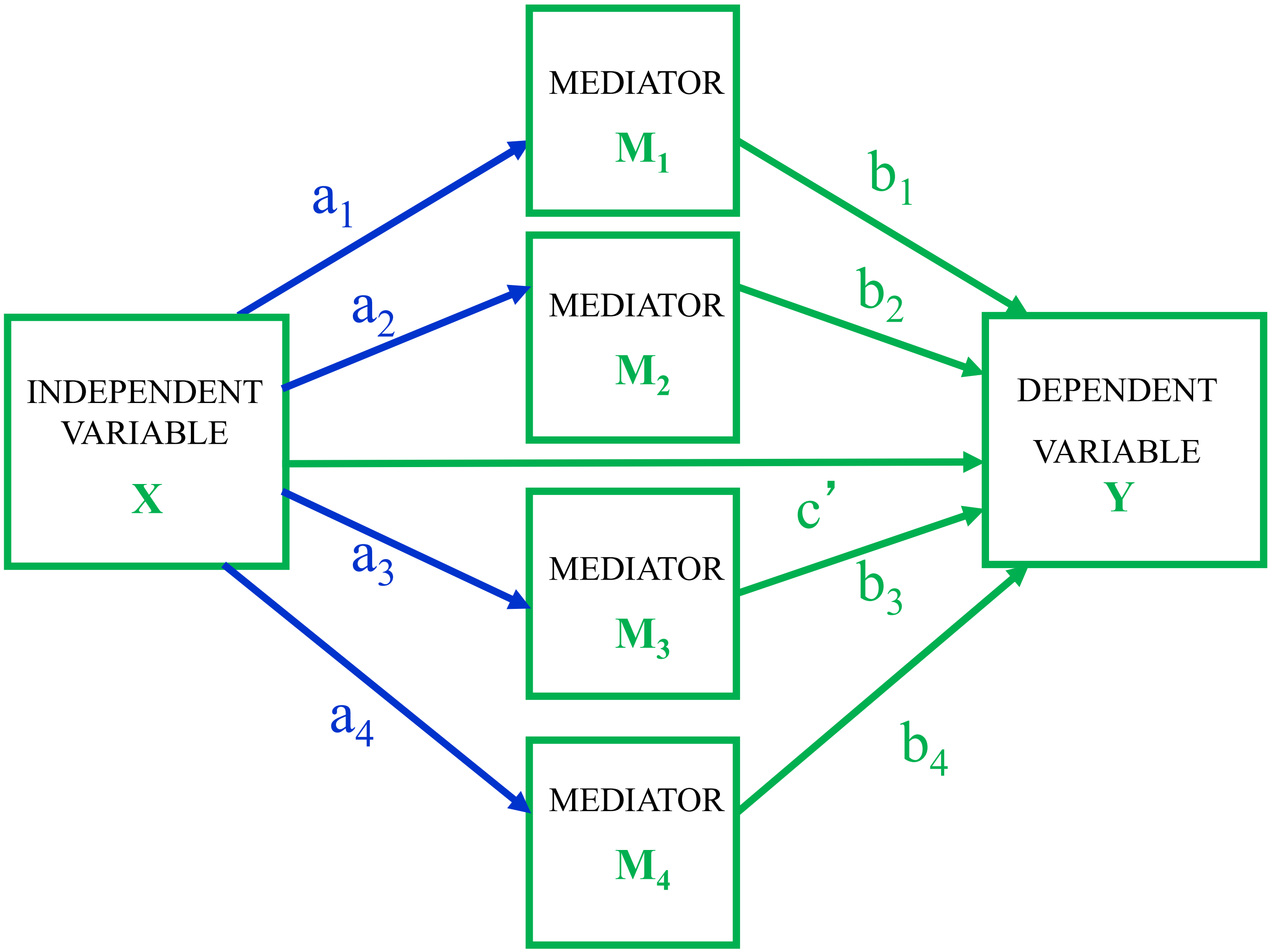
Test	S-S	S-M	M-M	L-L
Causal Steps (c' = 0)	20886	3039	397	92
Normal	667	422	90	42
Dist. Product	539	401	74	35

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. S-S means small effect size for the *a* path and small effect size for the *b* path.

Testing Mediation When the Total Effect is Not Statistically Significant

- Test of $\hat{a}\hat{b}$ can be more powerful than test of $\hat{c}$, i.e., mediation more precisely explains how X affects Y (O'Rourke & MacKinnon, 2014).
- Lack of statistically significant $\hat{c}$ is very important for mediation analysis because failure of action, conceptual theory, or both theories is critical for future studies.
- Note the test of $\hat{c}$ is important in its own right but is a different test than the test for mediation.

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:

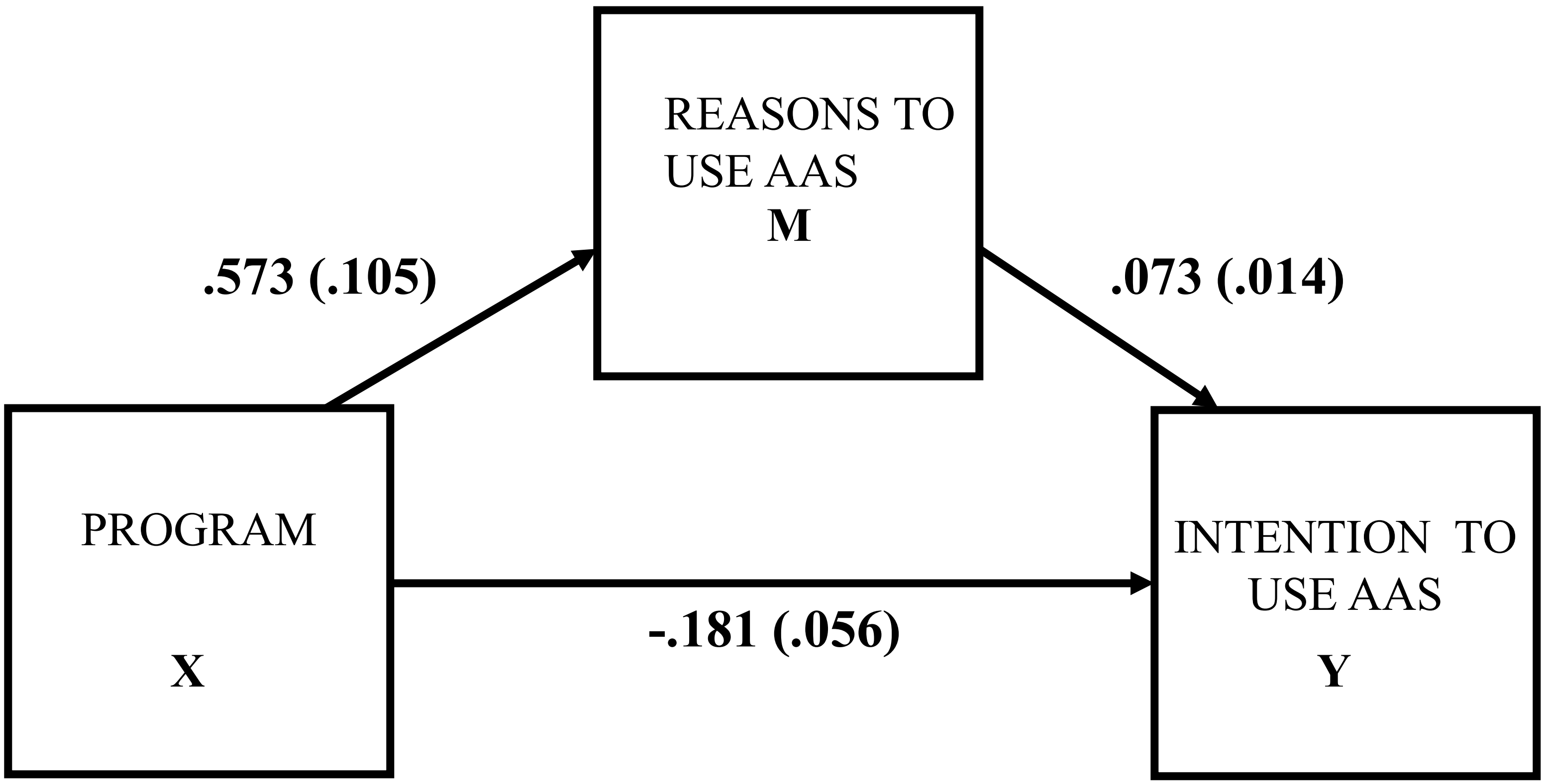
$$z' = \frac{a_1b_1}{\sqrt{a_i^{\ 2} s_{b_i}^{\ 2} + b_i^{\ 2} s_{a_i}^{\ 2}}}$$

Compare to empirical distribution of the mediated effect

Inconsistent Mediation Models

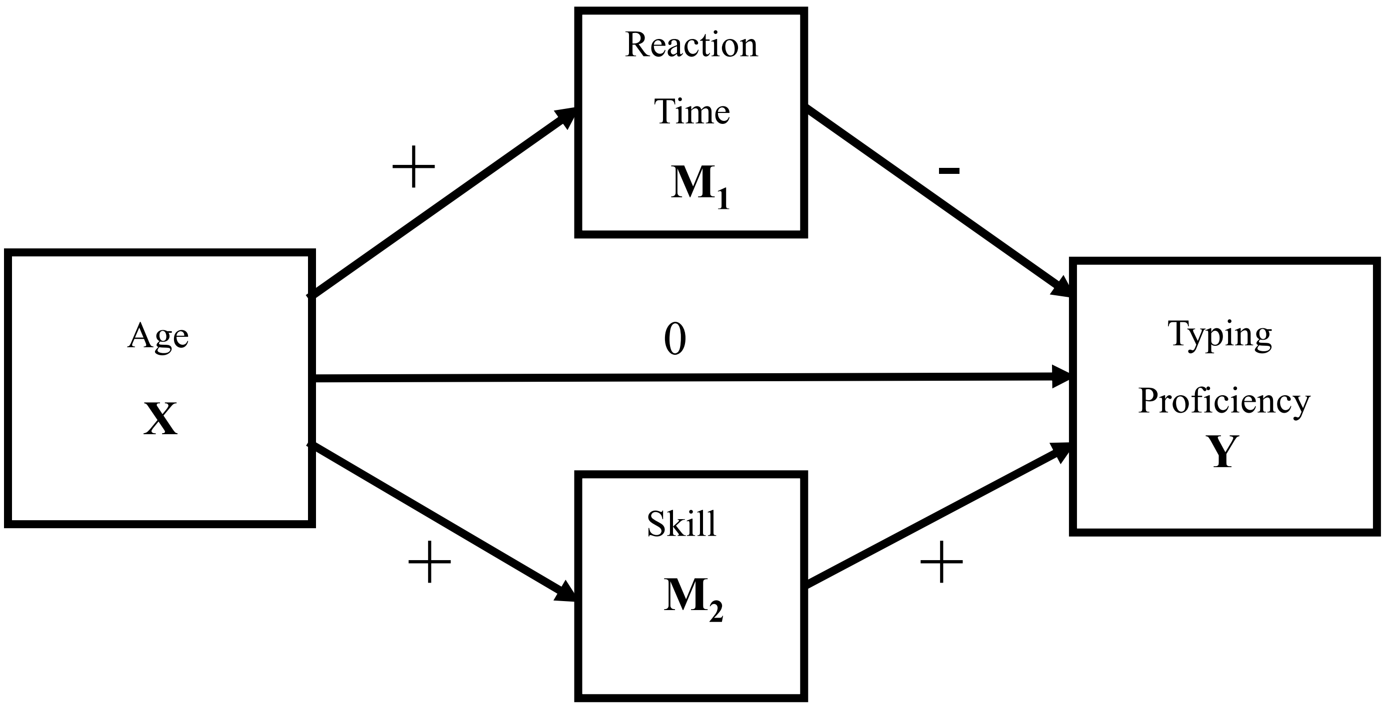
- An inconsistent mediation model has at least one mediated effect with a different sign than the direct effect or other mediated effects (MacKinnon et al., 2000)
- There is mediation because the mediator transmits the effect of the independent variable to the dependent variable. Inconsistent mediation 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 a Steroid Prevention Study



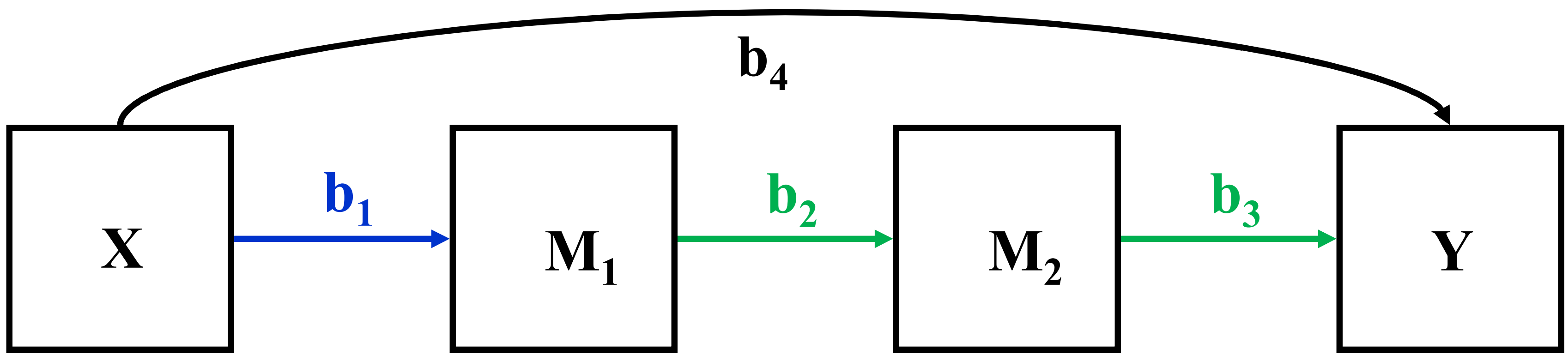
Mediated effect = .042
Standard error = .011

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).

Three-Path Sequential Mediation Model



Mediated effect = $b_1 b_2 b_3$

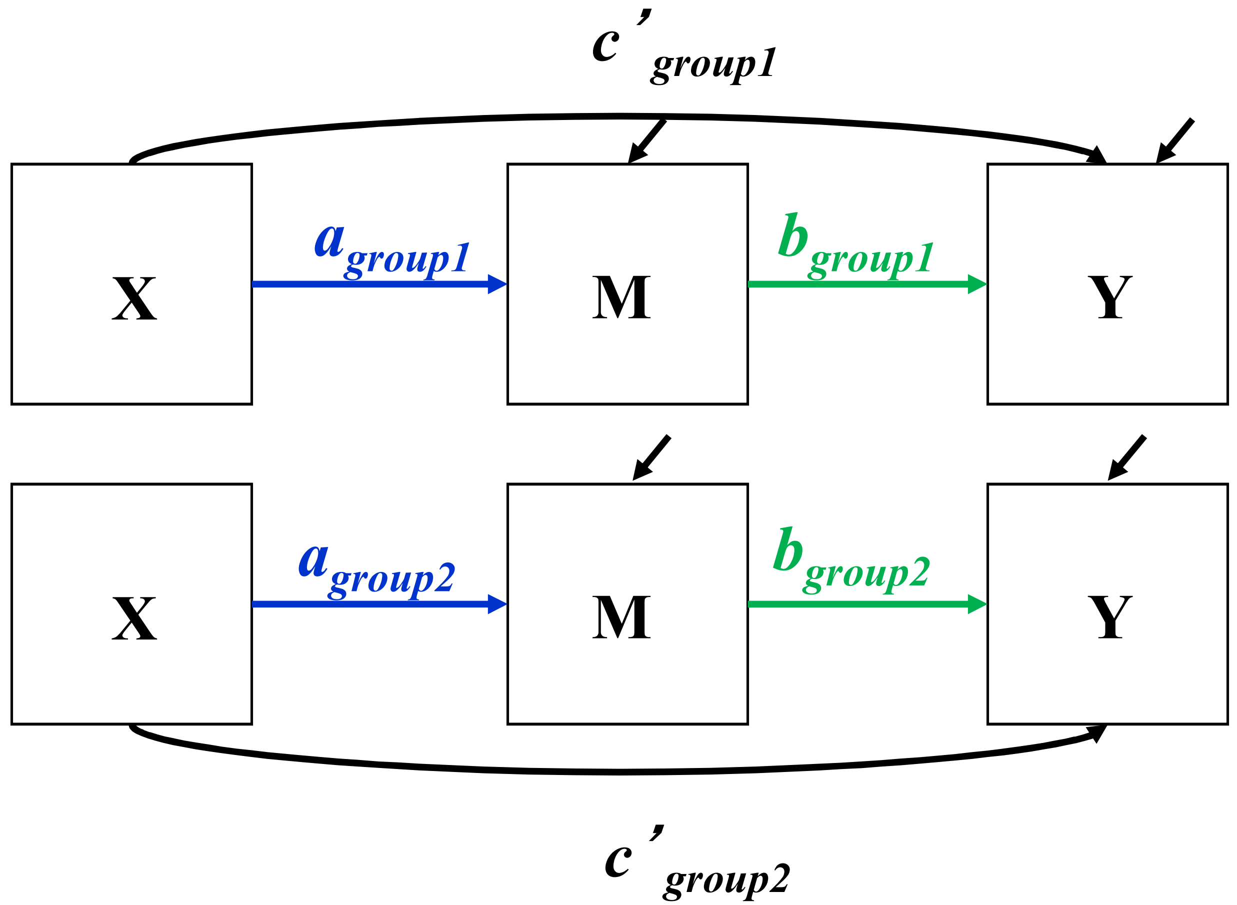
Path Model for Testing Homogeneity of Effects across Groups

GROUP 1

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

GROUP 2

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



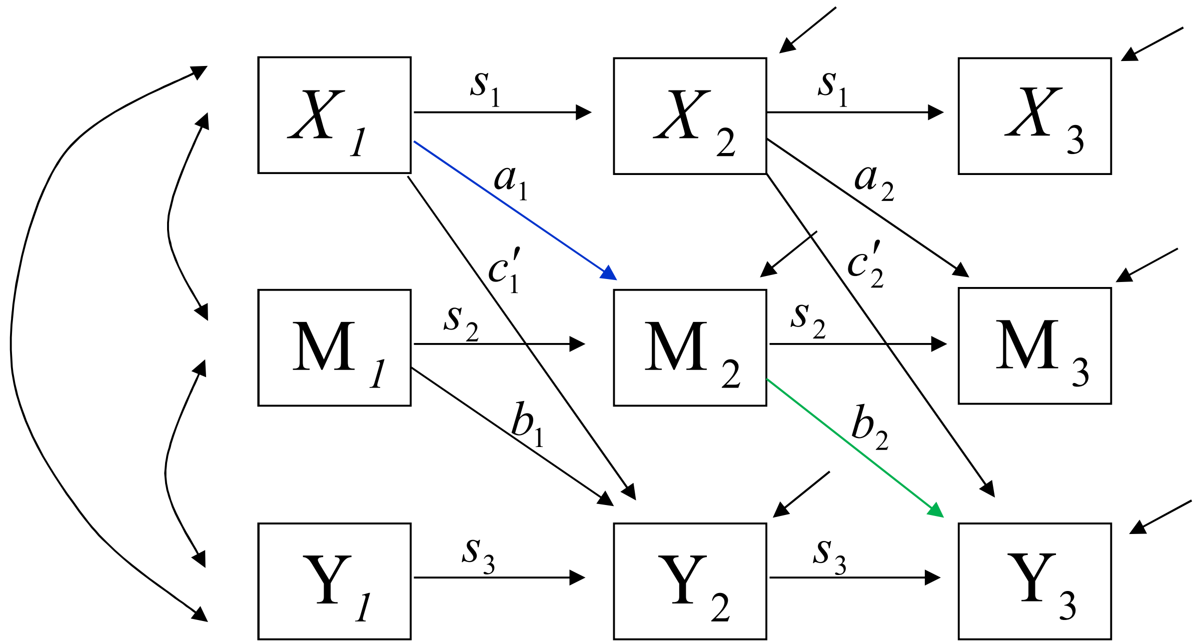
Longitudinal Mediation Analysis

- Assume correct temporal ordering: X before M before Y. Mediation is a longitudinal model.
- 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. Stability and stationarity assumptions also.
- Correct timing and spacing of measures to detect effects.
  - Important to consider when X affects M and when M affects 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 allow for several options: difference score, ANCOVA, residualized change score (Valente & MacKinnon, 2017).
- Measures of X, M, and Y at three or more time points allow for many alternative longitudinal models: 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.

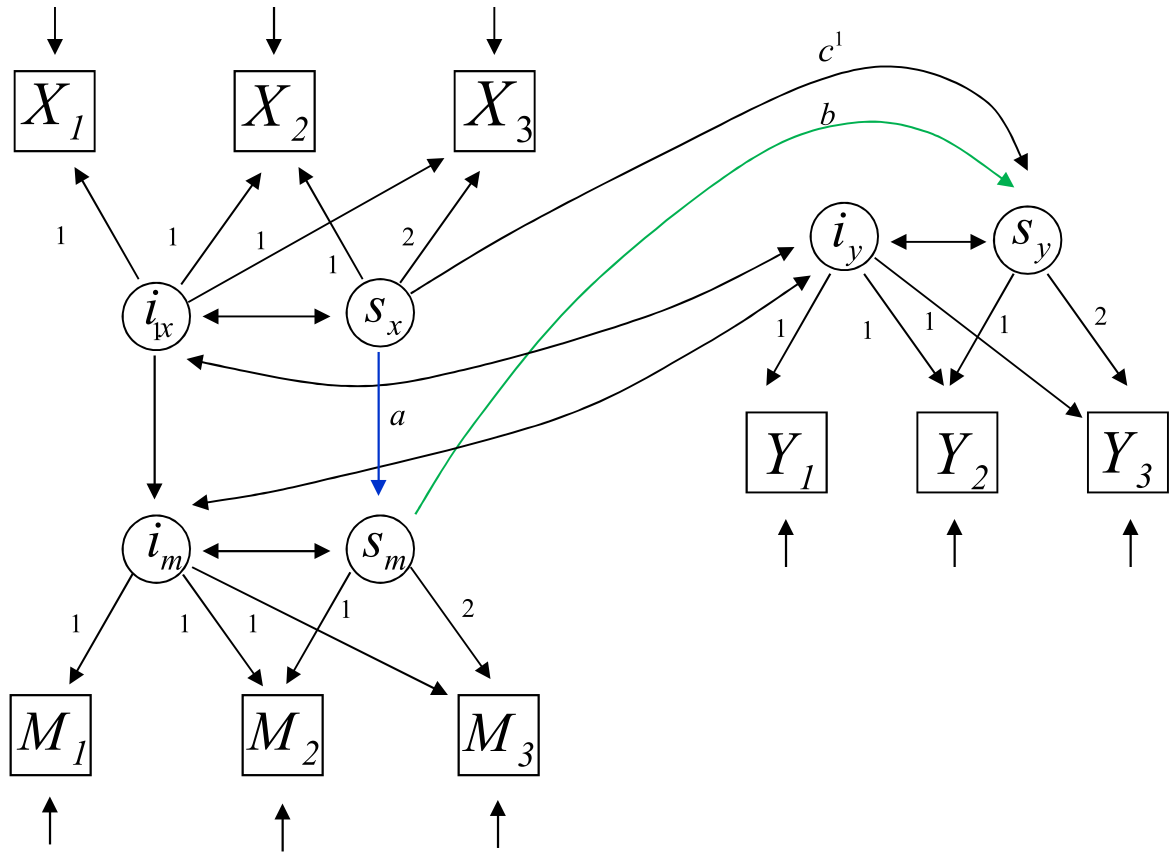
Autoregressive Model with Time-Ordered Mediation



Note: All residuals are correlated

Cole & Maxwell (2003)

Latent Growth Curve Mediation Model



See Cheong et al., 2003 for more on latent growth curve (LGC) mediation models. A related model, the Latent Change Score (LCS) models fixes loadings to conduct analysis of the change between adjacent waves (see McArdle, 2001). See MacKinnon (2008; Chapter 8) for more on longitudinal mediation models.

Causal Inference in Mediation

- Assumptions of true causal relations and self-contained/comprehensive model for regression analysis for mediation.
- Blalock (1979; Presidential address) stated that about 50 variables are involved in sociological phenomenon. How many variables are relevant in your research area?
- Problem with mediation analysis because M is not randomly assigned but is self-selected.
- Causal inference for mediation is an active research area (Frangakis & Rubin, 2002; Pearl, 2001; Pearl, 2009).

Counterfactual/ Potential Outcome Models

- Most modern causal inference approaches are based on a counterfactual or potential outcome model.
- In these models, all the possible counterfactual and actual conditions of an experiment are considered and the statistical model is based on all these possible or potential conditions.
- Requires consideration of conditions that did not occur.
- Counterfactual thinking is common, e.g., If I had three egg McMuffins instead of one this morning, I would not be as tired.

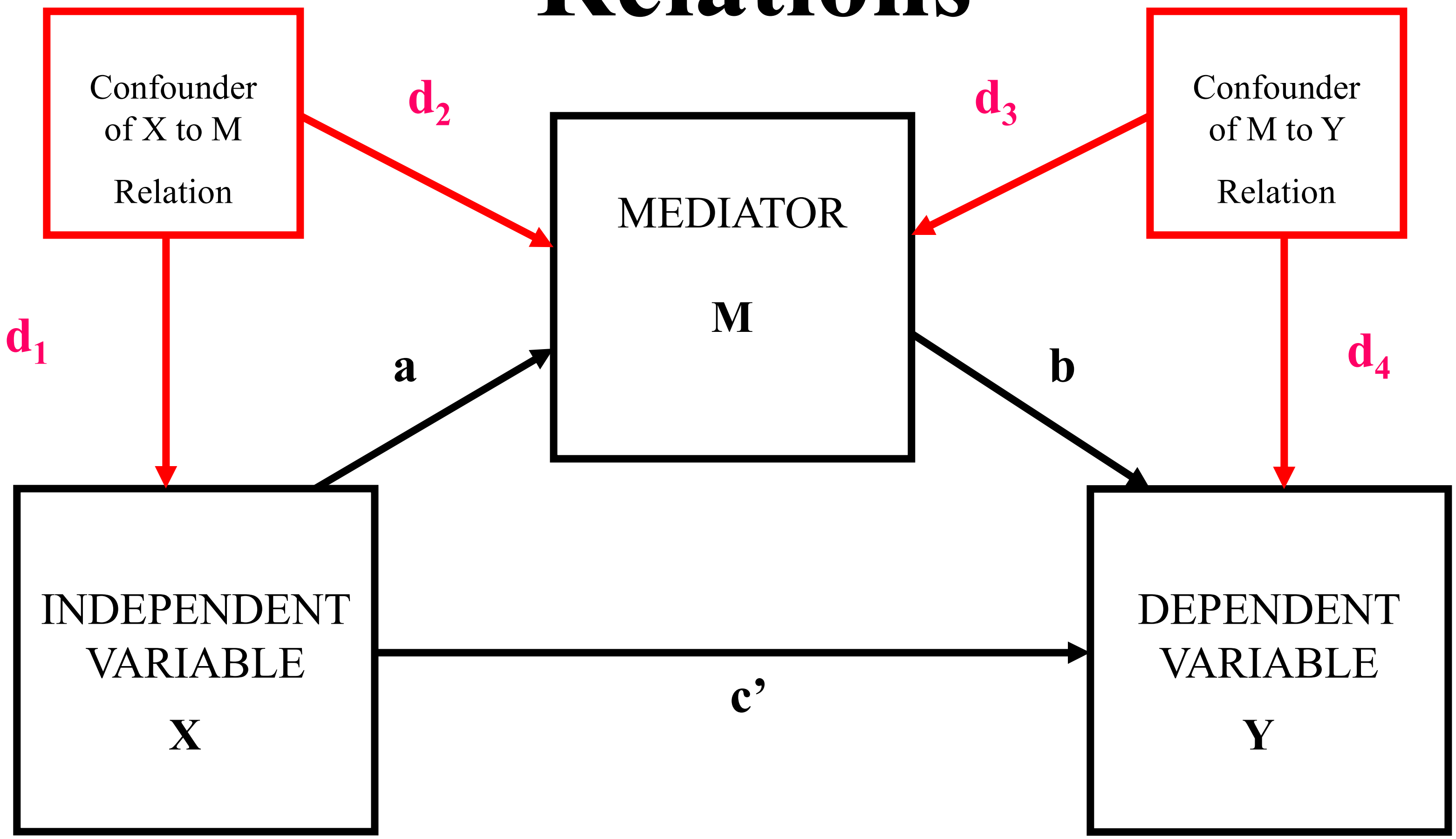
Randomized Two Group Design

- Ideally we need the same individual in both the treatment and control conditions at the same time. 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 solves 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. M is both a dependent and independent variable.
- 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 the counterfactuals for these relations are complicated because M is not randomly assigned.

Confounders of Mediation Relations



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

Sensitivity Analysis for Confounding

- How will results change with confounding of the M to Y relation, e.g. when X is randomized.
- 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.
- 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.
- 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;* Valente et al., 2017, *Journal of Counseling Psychology*)…

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 the following question, "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."

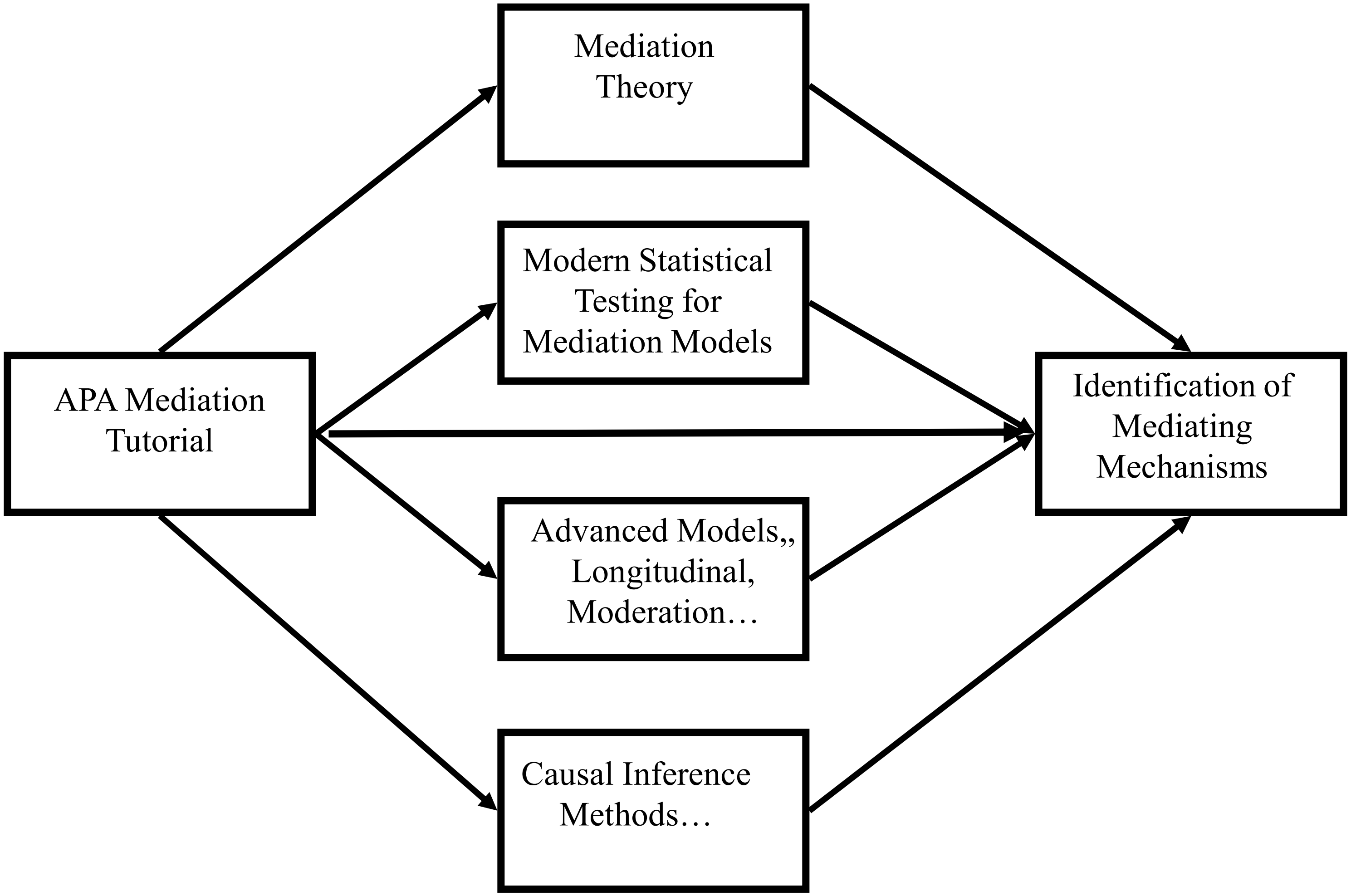
1. Designs to address **Consistency** of the mediation relation.
2. Designs to address **Specificity** of the mediation relation.

MacKinnon, 2008; MacKinnon & Pirlott, 2012 related to Hill's (1971) considerations. Also SMART designs (Almirall et al., 2014)

Summary

- Mediation analysis is important because it provides information about how variables are related, e.g., how and why an effect occurs, how an intervention achieved its effects, how effects unfold over time…
- Tests of mediation based on product *ab*; distribution of product/bootstrap are the most accurate.
- Multiple Mediator Models, Models with Moderation and Mediation, Experimental Mediation Designs, Longitudinal Mediation Models, and other models are available.
- Longitudinal data are ideal for testing mediation.
- Causal inference is an active research area with new methods to investigate confounder bias and experimental designs are available.

Hypothesized Effects of APA Mediation Analysis Tutorial



Thank You

Reference List available by contacting David.MacKinnon@asu.edu