Mediation for the 21st Century

Ross Boylan `ross@biostat.ucsf.edu`

Center for Aids Prevention Studies and

Division of Biostatistics

University of California, San Francisco

Fixed formula and page 27 since the presentation.

Motivating Problem

Ethnic Group	HIV+	Jail
AA	51%	41%
API	14%	13%
Latino	43%	35%

- Logistic regression indicates that Jail raises the odds ratio of being HIV+ by 50%.
- How much of the group differences in HIV+ rates are accounted for by differences in incarceration?
- Binary outcome, 3-category predictor, binary mediator, categorical and continuous covariates.

Overview

- “Classical” Mediation and Moderation
- Some Updates
- Limitations
- Structural Causal Models
- Causal Modeling Approach to Mediation

Scope

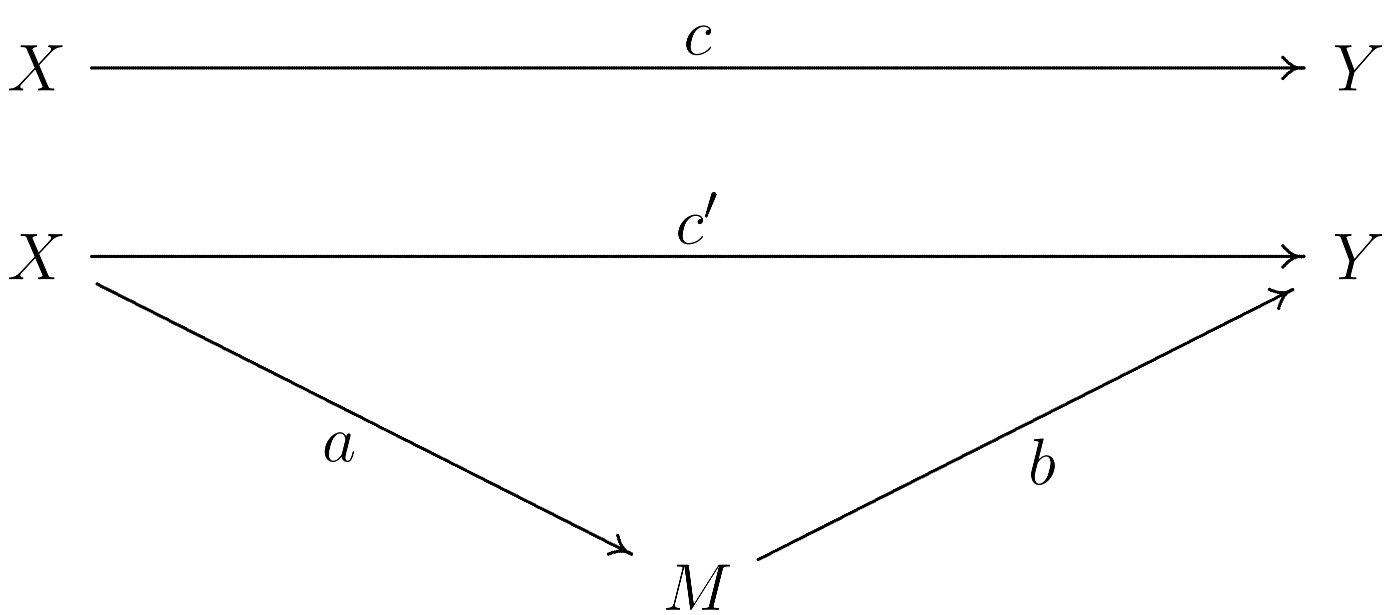
This *not* a talk about determining true causal relationships or estimating them from data.
It is about how to *interpret* models once they are in hand.

“Classical” Mediation

Started with Baron & Kenny 1986:

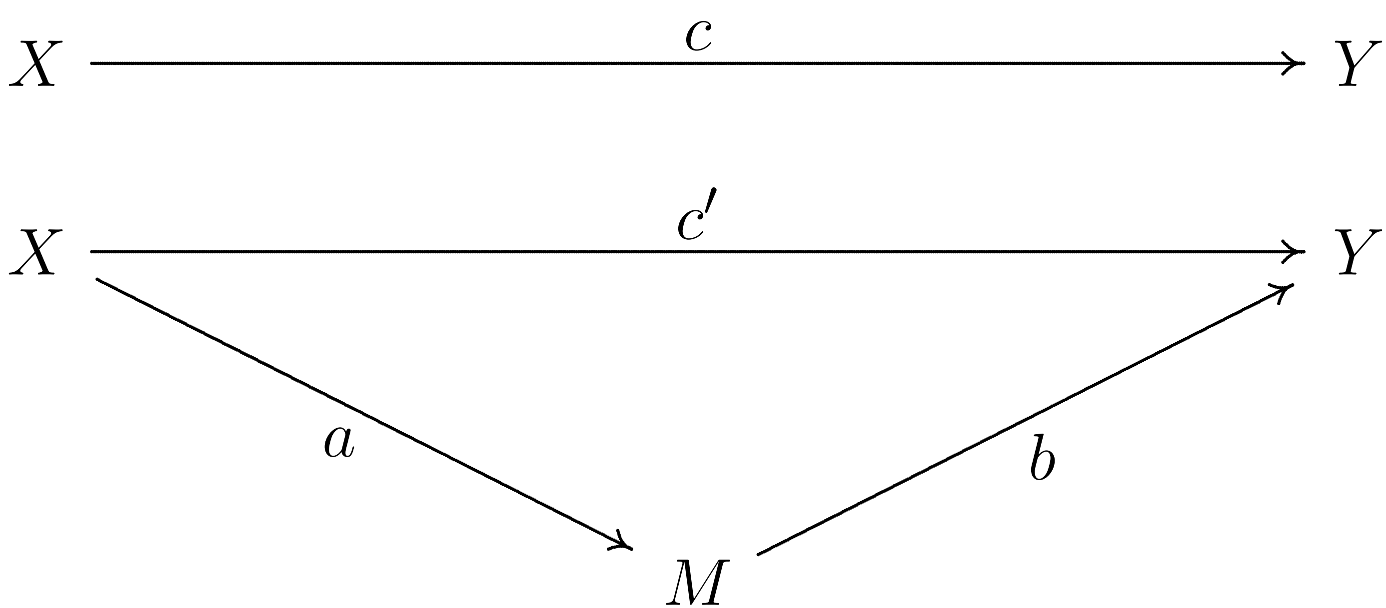
> In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the criterion.

4 Step Test For Mediation



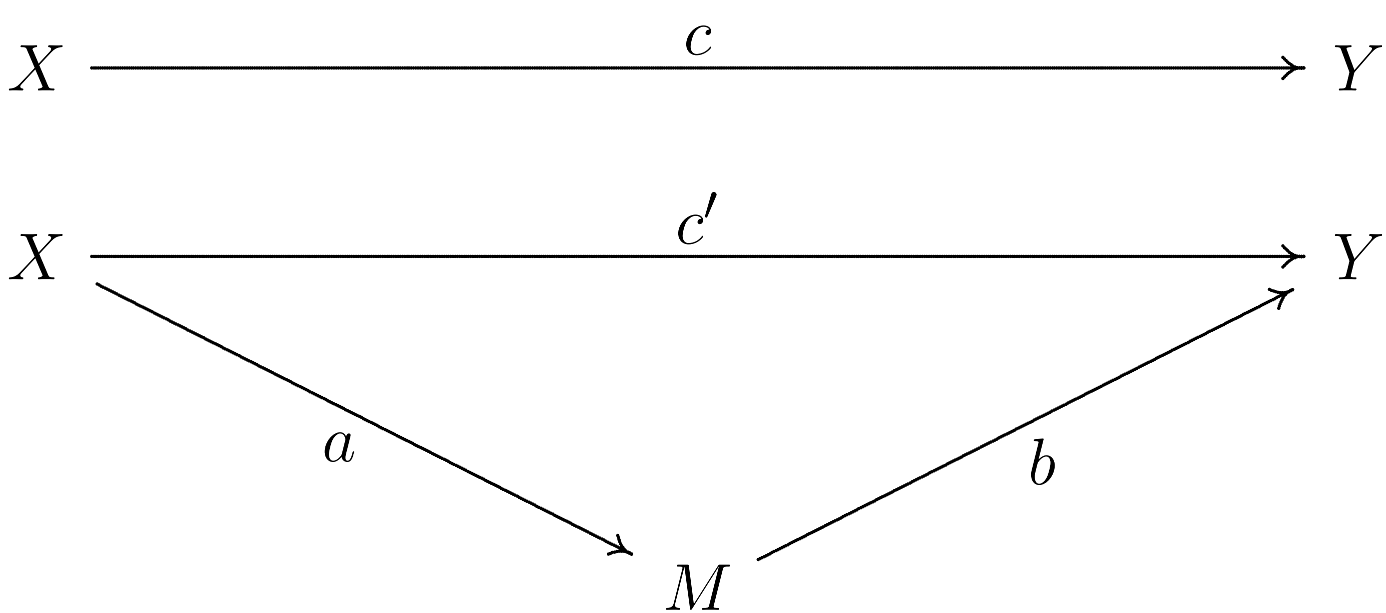
1. c is significant.

4 Step Test For Mediation



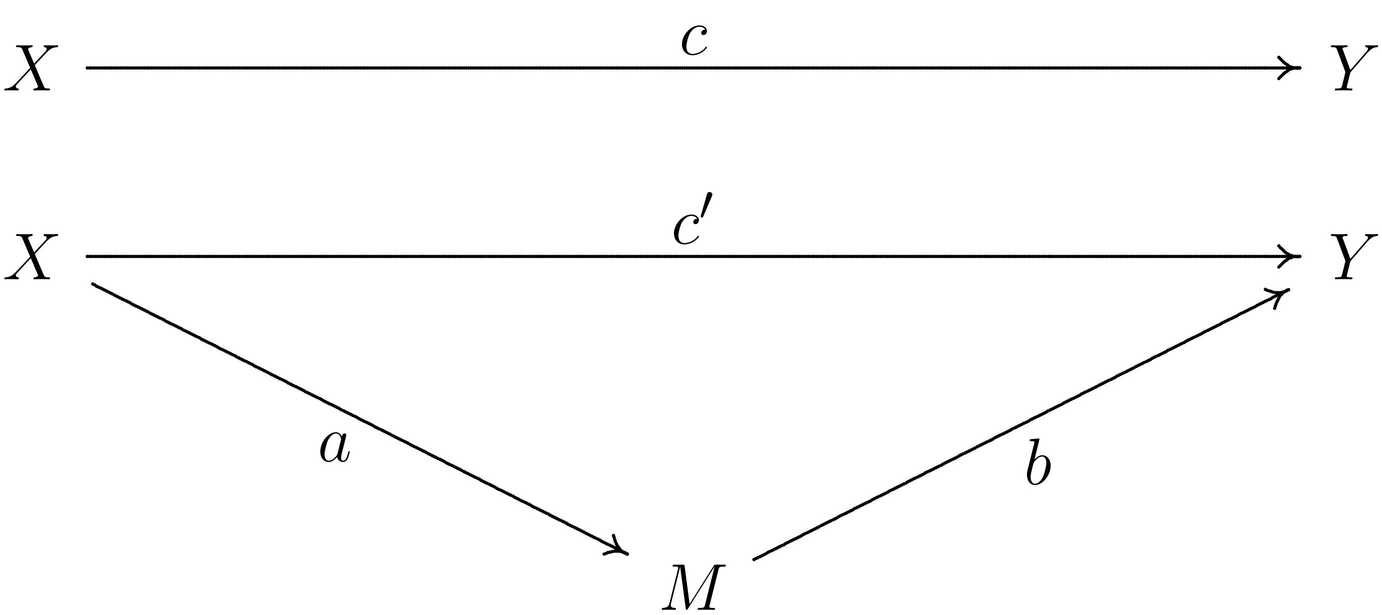
1. c is significant.
2. a is significant.

4 Step Test For Mediation



1. c is significant.
2. a is significant.
3. b is significant.

4 Step Test For Mediation



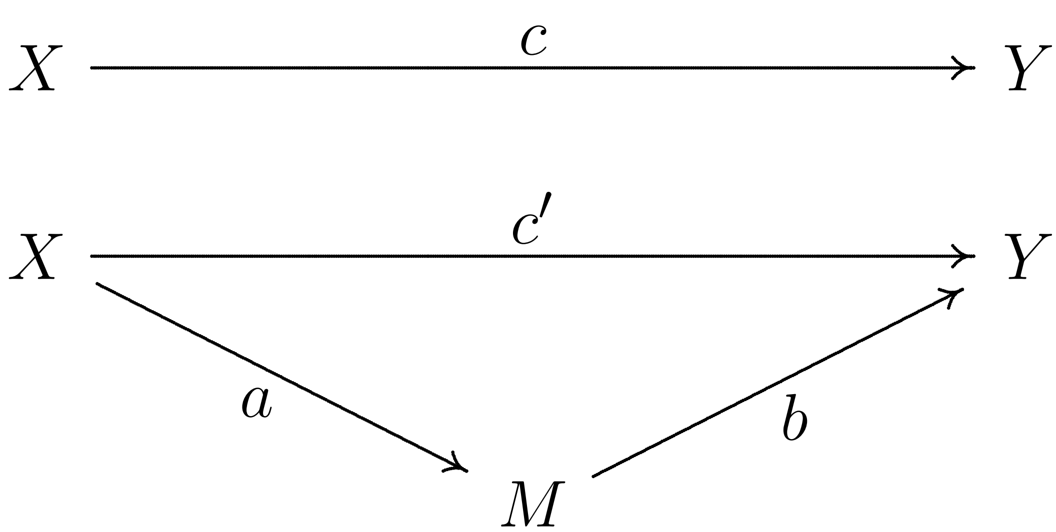
1. c is significant.
2. a is significant.
3. b is significant.
4. $c' < c$. If c' is not significant, total mediation.

Modern Reconsiderations

$$X \xrightarrow{\quad c \quad} Y$$

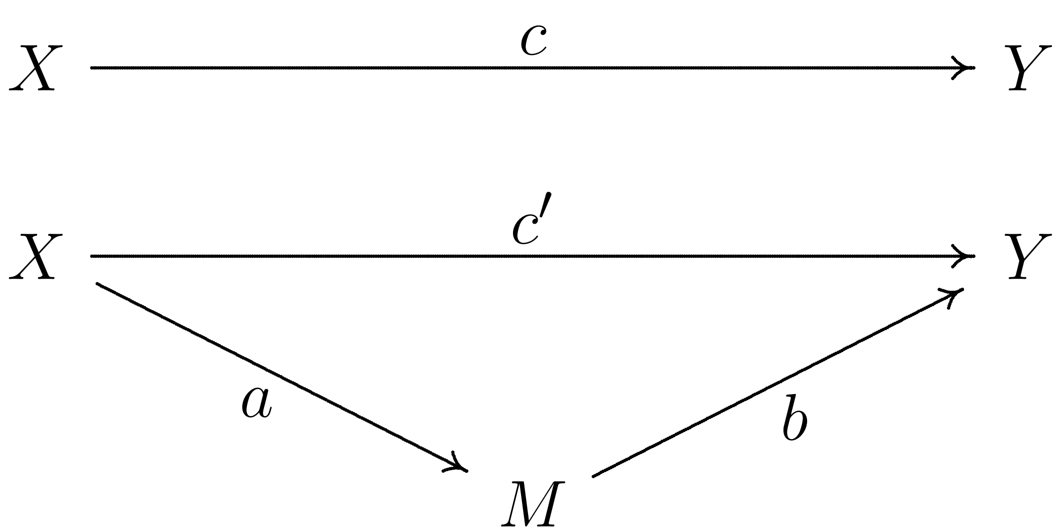
$$\begin{tikzcd}
X \arrow[rr, "c'"] \arrow[dr, "a"'] & & Y \\
& M \arrow[ur, "b"'] &
\end{tikzcd}$$

Modern Reconsiderations



1. ab is significant.

Modern Reconsiderations



1. ab is significant.

That's it!

Modern Reconsiderations

$$X \xrightarrow{c} Y$$

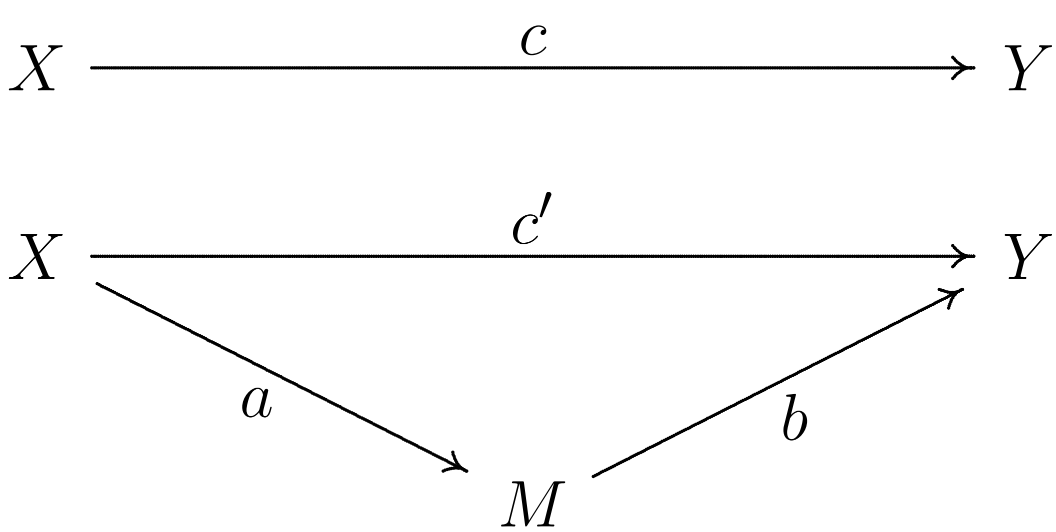
$$\begin{tikzcd} X \arrow[rr, "c'"] \arrow[dr, "a"'] & & Y \\ & M \arrow[ur, "b"'] & \end{tikzcd}$$

1. ab is significant.

That's it!

- c doesn't matter because of *inconsistent mediation*, e.g. c and c' have opposite signs.

Modern Reconsiderations



1. ab is significant.

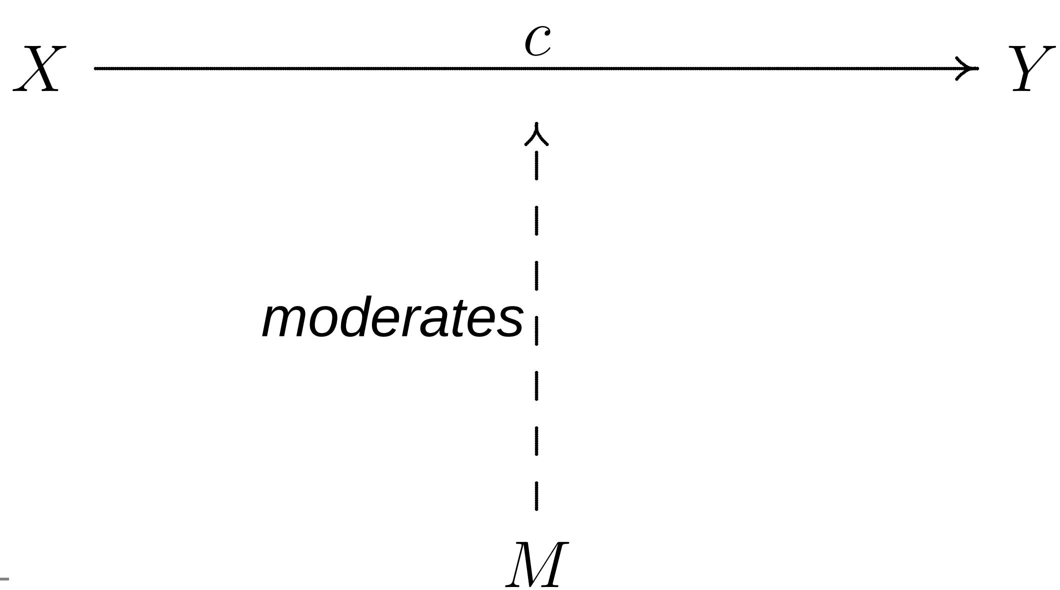
That's it!

- c doesn't matter because of *inconsistent mediation*, e.g. c and c' have opposite signs.
- Focus on ab rather than separate tests on each.

Moderation

In general terms, a moderator is a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable.

Baron and Kenny (1986)



Limitations

- Big mess if M is a mediator *and* a moderator
- Little guidance in handling categorical variables and non-linear models.
- Abuse of the null hypothesis.

The Null in Practice

- If any of a, b or c is not statistically significant, conclude there is no mediation.
- In more modern forms, if ab is not statistically significant, conclude there is no mediation.
- If the interaction term d in

$$Y = c'X + bM + dMX + \epsilon$$

is not statistically significant, conclude there is no moderation.

Some Smaller Problems

- The relation might not be linear, and you probably didn't check.
- Even the tests of ab require huge sample sizes to have a reasonable chance of rejecting the null when the effects are "small."

The Bigger Problem

Absence of evidence is not evidence of absence.
Don Rumsfeld.

The Bigger Problem

Absence of evidence is not evidence of absence.
Don Rumsfeld.

- The fact that the data are consistent with null does not mean the null is true.

The Bigger Problem

> Absence of evidence is not evidence of absence.
> Don Rumsfeld.

- The fact that the data are consistent with null does not mean the null is true.
- Fighting this is an uphill battle.

Recommendations

- Say what you know, not what you don’t know.
- Give the range of values covered by the confidence intervals and their substantive interpretations.
- In this context, that means focusing on how much mediation there is, rather than whether there is mediation.
- If there really is no mediation, no amount of data will ever suffice to tell us the effect is 0. Data can tell us the effect is almost certainly trivial.

An Example

Conference Abstract, first draft:

Background The deleterious effects of racism on a wide range of health outcomes including HIV risk is well documented among racial and ethnic minority groups in the United States. However, little is known about how MSM of color cope with stress from racism and *whether coping with racism moderates the association between stress from racism and HIV risk among these men*.

Methods . . .

Example, cont'd

Results . . . *None of the interactions* of stress with race/ethnicity, the four coping measures with race/ethnicity , and stress with the four coping measures *was statistically significant.*

Conclusions Stress from racism within the gay community increased the likelihood of engaging in unprotected anal intercourse among U.S. MSM of color, but this *association was not moderated by coping responses to racism*. . . .

Example, cont'd

Results . . . *None of the interactions* of stress with race/ethnicity, the four coping measures with race/ethnicity , and stress with the four coping measures *was statistically significant.*

Conclusions Stress from racism within the gay community increased the likelihood of engaging in unprotected anal intercourse among U.S. MSM of color, but this *association was not moderated by coping responses to racism*. . . .

1. Strike all references to moderation.

Example, cont'd

Results . . . *None of the interactions* of stress with race/ethnicity, the four coping measures with race/ethnicity , and stress with the four coping measures *was statistically significant.*

Conclusions Stress from racism within the gay community increased the likelihood of engaging in unprotected anal intercourse among U.S. MSM of color, but this *association was not moderated by coping responses to racism*. . . .

1. Strike all references to moderation.
2. "We can not draw any firm conclusions about moderation."

Example concluded

The wording as submitted:

Background . . . However, little is known about how MSM of color cope with stress from racism and whether coping with racism *buffers* the impact of stress from racism on HIV risk among these men.

Conclusions . . . However, we found *little evidence* that coping responses to racism buffered stress from racism. . . .

Discussion

How do you think statistical significance should be handled?

Moderation: A Slippery Concept

When the model is non-linear, it's not clear moderation is a meaningful concept. Suppose the true model for a binary outcome is the logistic:

$$E\left[\log\left(\frac{p}{1-p}\right)\right] = -1 + 3X + 4Y + 0XY.$$

Since the interaction term is 0, there appears to be no moderation.

Moderation: A Slippery Concept

		X		Effect of ΔX
		0	1	
Y	0	27%	50	23
	1	50	88	38

Moderation: A Slippery Concept

		X		Effect of ΔX
		0	1	
Y	0	27%	50	23
	1	50	88	38

The effect of X on the outcome varies depending on the level of Y, and we have moderation by the usual definition, even though the XY coefficient is 0.

Moderation and Nonlinear Models

This line of argument implies that any conceivable coefficients on X and Y, except ones that are exactly 0, produce moderation.
This is likely to be true for most nonlinear models. So it does not seem useful to be worrying about whether there is or is not moderation.

Structural Causal Models

Recent work in Causal Inference (e.g., Pearl, 2010) has sought to clarify and generalize previous work on causation.

Stochastic Effects operate by changing the probability distributions of outcomes.

Non-Parametric The critical assumptions concern the *absence* of certain causal relationships. No functional forms are assumed.

General Works with non-linear relations and discrete and continuous variables, both as effects and causes.

Causal Inference Concerned with conditions under which we can extract causal relations from the world.

Vocabulary Either graphs or functions.

News You Can Use

Good News We can think about and calculate mediation for any kind of model.

News You Can Use

Good News We can think about and calculate mediation for any kind of model.

Bad News We have to be much more precise about exactly what questions we are asking.

News You Can Use

Good News We can think about and calculate mediation for any kind of model.

Bad News We have to be much more precise about exactly what questions we are asking.

News No magic: identification of true causal effects remains challenging to impossible.

SCM Vocabulary and Concepts

$$m = f_M(x, \mu_M)$$

is a claim that X and a random term determine the value of M. Crucially, it claims that Y does *not* determine the value of M.

$$f_M\left(\text{do}(x_0), \mu_M\right)$$

is a statement about the distribution of M if we intervened and set X to x_0.
An alternate notation is the *potential outcome* notation:

$$M_{x_0}(u)$$

refers to the value individual u would have had if X had been x_0.

Controlled Direct Effects

$$\text{CDE}(m) = E(Y|X = 1, M = m) - E(Y|X = 0, M = m).$$

This is the effect of a 1 unit change in X on Y when the mediator is fixed at m.
The expression assumes independence of the error terms for X and M. Without that assumption the more general expression is

$$\text{CDE}(m) = E[Y|\text{do}(X = 1, M = m)] - E(Y|\text{do}(X = 0, M = m)).$$

This is the effect of X if we fix the mediator at m.

Natural Direct Effects

$$\text{NDE}_{x,x'}(Y) = \sum_m \left\{E(Y|X = x', M = m) - E(Y|X = x, M = m)\right\} \Pr(M = m|X = x),$$

or, more compactly,

$$\text{NDE}_{x,x'} = \sum_m \left\{E(Y|x', m) - E(Y|x, m)\right\} \Pr(m|x).$$

The NDE is the weighted average of the CDE's using the baseline $X = x$ distribution of M as the weights.

Indirect Effects

$$\mathsf{IE}_{x,x'} = \sum_m E(Y|x,m)\left[\Pr(m|x') - \Pr(m|x)\right]$$

- X operates here only through its effects on M. Note that we evaluate those effects at the baseline level of X.

Indirect Effects

$$\text{IE}_{x,x'} = \sum_m E(Y|x,m)\left[\Pr(m|x') - \Pr(m|x)\right]$$

- X operates here only through its effects on M. Note that we evaluate those effects at the baseline level of X.
- $\sum_m E(Y|x,m)\Pr(m|x')$ is the expected value of Y for group x when the mediator has been changed to the levels for group x'.

Indirect Effects

$$\mathsf{IE}_{x,x'} = \sum_m E(Y|x,m)\left[\Pr(m|x') - \Pr(m|x)\right]$$

- X operates here only through its effects on M. Note that we evaluate those effects at the baseline level of X.
- $\sum_m E(Y|x,m)\Pr(m|x')$ is the expected value of Y for group x when the mediator has been changed to the levels for group x'.
- $\sum_m E(Y|x,m)\Pr(m|x)$ is the expected value of Y for group x when the mediator has been left at the levels for group x.

Total Effects

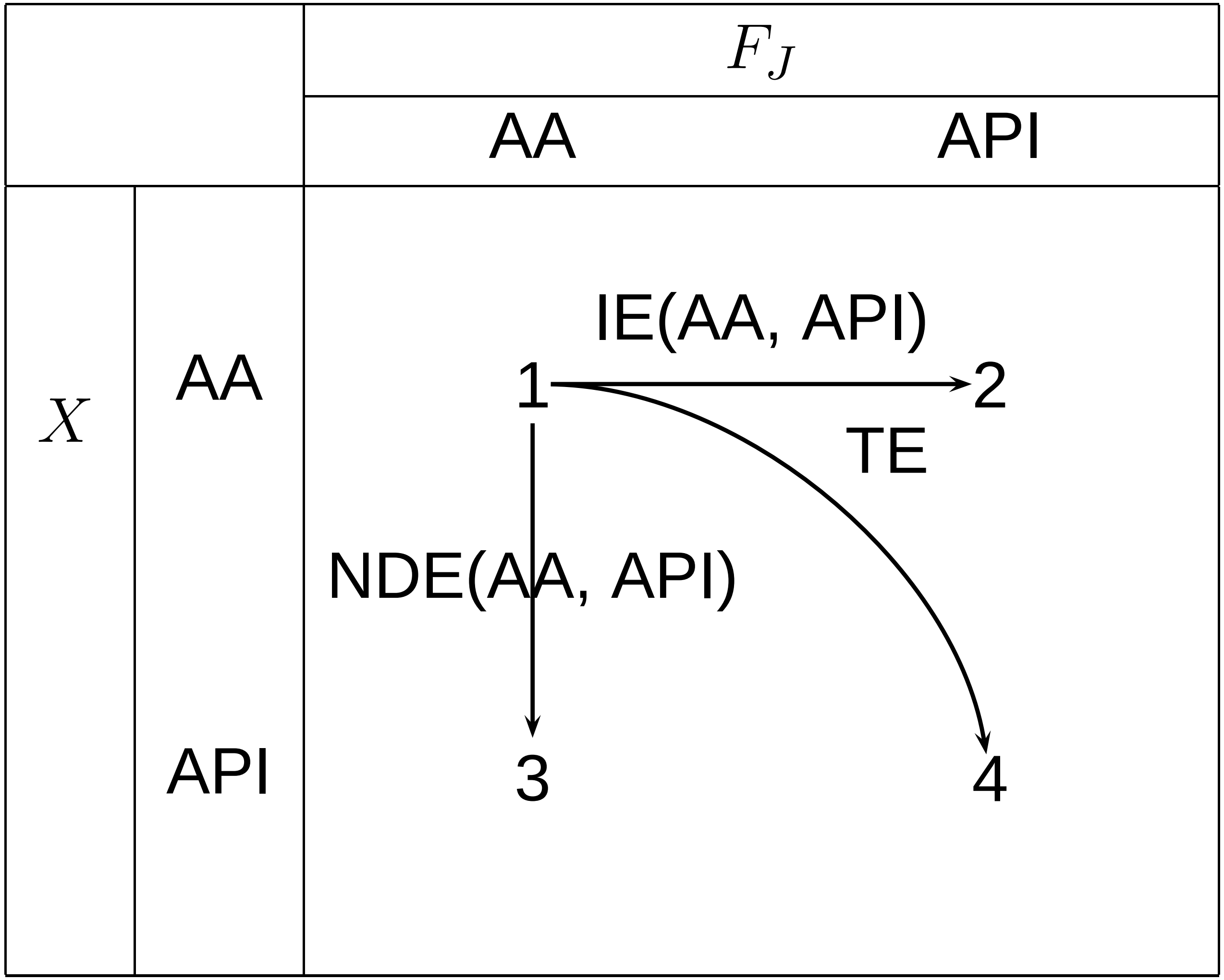
Both the NDE and IE evaluate from a baseline of $X = x$, but the total effect has 3 sources:

1. The direct effect of raising X, holding M constant (NDE)
2. The indirect of effect of raising X on M, holding baseline X constant (IE).
3. The interaction of the changed M values with the changed X values.

So we conclude that in general

$$\text{TE} \neq \text{NDE} + \text{IE}.$$

Relation Between Effects



Total and Reverse Effect

Surprisingly, the total effect appears in an expression that includes travel backward along the paths. By going backward we get a term that starts from the interaction and steps down from it.

$$\mathsf{TE}_{x,x'}(Y) = \mathsf{NDE}_{x,x'}(Y) - \mathsf{IE}_{x',x}(Y).$$

Caution: Multivariate Dangers

		F_J	
		AA	API
X	AA	1	2
	API	3	4

In a multivariate model, going from 1 to 2 to 4 does *NOT* turn AA's into API's. It turns AA's into API's who still have the AA distribution of other variables.
The total effect is not the same as observed or modeled group differences, since those depend on the other covariates as well.

No Estimation of Damaged Models

None of the previous analysis relied on a statistical estimate of the gross effect of race on HIV status.
A lot of the traditional procedures rely on estimating models with and without mediators. The SCM approach does not; it draws out the implications of the "true" model by manipulating it.

Not Quite Home

- The SCM tools let us compare the direct and indirect effects of, e.g., being AA on an API.
- But we wanted to know what share of the differences between the three groups owed to indirect effects of jail.

An Extension

- Define a target measure $G(F) = \sum_{g_i,g_j} |\bar{Y}_{g_i} - \bar{Y}_{g_j}|$. This is the sum of the the absolute differences in means between all pairs of groups in a population with distribution F.

An Extension

- Define a target measure $G(F) = \sum_{g_i, g_j} |\bar{Y}_{g_i} - \bar{Y}_{g_j}|$. This is the sum of the the absolute differences in means between all pairs of groups in a population with distribution F.
- T is the distribution of the population under the model $m = F_M(x)$, $y = F_Y(x, m)$. This is the Theoretical distribution under the model.

An Extension

- Define a target measure $G(F) = \sum_{g_i, g_j} |\bar{Y}_{g_i} - \bar{Y}_{g_j}|$. This is the sum of the the absolute differences in means between all pairs of groups in a population with distribution F.
- T is the distribution of the population under the model $m = F_M(x)$, $y = F_Y(x, m)$. This is the Theoretical distribution under the model.
- A is an alternate distribution of the population under the model $m = F_m(\text{do}(*))$, $y = F_y(x, m)$. In this model all groups have the same distribution of Jail. do$(*)$ means we use the overall population distribution, disregarding race.

The Answer

$$\frac{G(T) - G(A)}{G(T)}$$

is the fraction of group differences mediated by Jail.

Conclusion

This is a useful, general purpose approach. Try it!