Comparing Recent Advances in Causal Mediation:
How Medical Researchers and Practitioners Can Better Understand Causal Mediation and use it for Personalized Medicine

Emil Coman¹, Judith Fifield¹ & L. Suzanne Suggs ²
1. UConn TRIPP Center for Translating Research into Policy and Practice 2. University of Lugano
Uconn Health coman@uchc.edu

M³, PSTAT, CIRA, CHIP & blog
Motivating ‘day-to-day’ examples

4. Comparative Effectiveness (CE) of coronary artery bypass surgery (Tx=1) to medical therapy (Tx=2) in the treatment of coronary artery disease. [1]

5.

S1: “If Oswald did not kill Kennedy, someone else did”

S2: “If Oswald had not killed Kennedy, someone else would have.”

[2,3]

What is a ‘causal effect’?

Effect of Tx on Y

\[ E[Y_{Tx=1}] - E[Y_{Tx=0}] \text{ or } E[Y_1] - E[Y_0] \]

Many times this means comparing 2 sample means:

Treatment Tx \( \rightarrow \) effect \( \rightarrow \) Outcome Y

Outcome \( Y_0 \) \( \bar{Y}_0 \)
Outcome \( Y_1 \) \( \bar{Y}_1 \)

Or in Healthcare: ‘No harm’ or ‘near misses’: why something didn’t happen (‘root cause analysis’)
The effect of interest is biased when a confounder of it exists.

Causality is a complex topic; some references:


Mediation as inherent confounding

But this is exactly what any RCT/observation with an indirect effect/mechanism of change!
Why is mediation important?

In Translational Research [1] in particular: Comparative Effectiveness [2,3] questions: HOW/WHY Tx1 vs. Tx2. + [4,5]:

1. Anticipate policy impact
   Mediators may split a non-existing effect into 2 (or>) opposite effects
3. Understand the mechanisms of change

---

Isn’t mediation as simple as the a,b,c’s?

There are some problems in estimating the right effects; Judd & Kenny (2009) detail them visually, see also Smith (1982).

Figure 4.2 Mediation Model with Specification Error: Reverse Causal Effects, Omitted Variables, and Measurement Error in X and M


Structural Causal Modeling (SCM) IV-DV model

When no specific form (e.g. linear) is assumed for the \( f \) functions, one ‘falls back’ on the more general SCM or causal diagrams.

The \( f \) functions can become sets of parameters when the functional forms are translated into linear/nonlinear equations.

Structural Causal Modeling (SCM) multiple ‘regression’

Other M Causes

$M = f_M(U_M)$

Other Y Causes

$Y = f_Y(Tx,M,U_Y)$

Other X Causes

$X = f_X(U_X)$

Note: the ‘other causes’ are likely correlated.

Simplifying the notation:

✓ X is a random variable, x a fixed value of it (x₀, x₁, x₂, etc). Fixing X to x if at the core of Pearl’s counterfactual theory of causation: who fixes it under what conditions.

✓ Subscript – X, superscript - M

\[ Y(x, m) \]

represents the value of Y for \( X=x \) and \( M=m \)

\[ *Y^m_{x_0} \]

represents the value of Y for \( X=x_0 \) and \( M=M^x_{x_1} \), or M for when \( X=x_1 \); this is a contrary-to-fact quantity, hence the * in front of it, as X is set to x, and M is set to a value it would take, if X had taken a different value \( x^* \).

Establishing causation using random assignment

Random Assignment cuts off the links leading to/determining treatment reception. Can make the leap (for a ‘later treated’ p):

\[ k* = j* \text{ matching} \]

\[ Y_{p1} - \frac{1}{2} Y_{p0} \equiv Y_{k1} - Y_{j[Tx]0} \]

Notes: Half of the p cases become j’s, not-treated, half become k, treated.

j[Tx] means the p case who became k was matched (across Tx, in controls) to a j case, similar on all Z’z that could determine Tx assignment (rings of propensity scores).
Regular SEM

\[ M_i = \alpha_1^S + \delta_1^S \cdot Z_i + a \cdot X_i + e_i(1) \]  \hspace{1cm} (12.11)  

\[ Y_i = \alpha_2^S + \delta_2^S \cdot Z_i + c' \cdot X_i + b^S \cdot M_i + e_i(2) \]  \hspace{1cm} (12.12)  

Potential outcomes SEM-analogue for observed outcomes only:

\[ M_i(x) = \alpha_1^C + \delta_1^C \cdot Z_i + a \cdot x + e_i(x)^A \]  \hspace{1cm} (12.22)  

\[ Y_i(x,m) = \alpha_2^C + \delta_2^C \cdot Z_i + c' \cdot x + b^C \cdot m + e_i(x,m)^A \]  \hspace{1cm} (12.23)  

In randomized studies:

\[ \tau^S = \tau^C \quad \text{and} \quad a^S = a^C \]

Yet:

\[ c'^S \neq c'^C \quad \text{and} \quad b^S \neq b^C \text{ (hence } a^S \cdot b^S \neq a^C \cdot b^C \text{) } \]

Total and Tx->M only effects are estimated as causal.

A: these are ‘potential errors’.
How does one establish causality

Old problem: how to move from correlation to causation [1-5].

Warning: “Causal modeling is open to abuse.” [2:9]

“Three Basic Strategies to Estimate Causal Effects”:

1. “Condition on variables (with procedures such as stratification, matching, weighting, or regression) that block all back-door paths from the causal variable to the outcome variable.

2. Use exogenous variation in an appropriate instrumental variable to isolate covariation in the causal and outcome variables.

3. Establish an isolated and exhaustive mechanism that relates the causal variable to the outcome variable and then calculate the causal effect as it propagates through the mechanism” (5:26).

A: = treatment assignment is conditionally ignorable: Y (x) _⊥⊥ X | Z.
SisterTalk project

* A community-based intervention conducted in African American churches (SisterTalk) aimed at weight loss among women through lifestyle behavioral changes. Some unique features:

* 1. The protocol and program content were developed with the church community.

* 2. The principles of weight loss and weight control were translated into faith-based messages. [URL](#)

* 3. Faith-based messages were written and delivered by church leaders and health messages delivered by African American and Black professionals.
* 91 participants were randomized to the control condition and 110 to the intervention (valid 6 month data retained for these analyses).

* Measures were gathered at baseline and 6 months.

### Measures

<table>
<thead>
<tr>
<th>Physical Measures</th>
<th>Food/eating Measures</th>
<th>Activity Measures</th>
<th>Other measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Food Preparation</td>
<td>Physical activity</td>
<td>Social Support</td>
</tr>
<tr>
<td>Weight</td>
<td>Eating out</td>
<td>Stages of Change</td>
<td>Self-Efficacy</td>
</tr>
<tr>
<td>Waist circ.</td>
<td>(bad) Food Habits</td>
<td></td>
<td>SF-12</td>
</tr>
<tr>
<td>BMI</td>
<td>Eating behavior</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M_B=Habits</td>
<td>1.01*</td>
<td>0.88</td>
<td>36.23</td>
<td>35.17</td>
<td>0.89</td>
<td>0.65</td>
<td>36.52</td>
<td>A</td>
<td>34.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_B=BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>2</td>
<td>2</td>
<td>58</td>
<td>59</td>
<td>2</td>
<td>1</td>
<td>58</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.36</td>
<td>0.34</td>
<td>8.67</td>
<td>7.02</td>
<td>0.31</td>
<td>0.27</td>
<td>8.35</td>
<td>6.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>91</td>
<td>110</td>
<td>91</td>
<td>110</td>
<td>91</td>
<td>110</td>
<td>91</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A: p = .099 from an ANOVA ; B = Baseline; F = Follow-up
SisterTalk outcomes

6 month (bad) Food Habits

M=food habits is intermediate outcome

Causal Mediation

MMM 2015
SisterTalk outcomes

Y=BMI is a target outcome

Y_{F}=BMI

36.52^A

controls

interv.

Y_{F}=BMI

34.75

6 month Body Mass Index by condition

Y=BMI is a target outcome
Brief statistical modeling review

Nonparametric models

\[ y = f(x, u_y) \]

Classic linear models

\[ E[y] = c' \cdot x \quad \text{or} \quad y = c' \cdot x + u_y \]

Simultaneous linear models (Structural Equation Models SEM or ‘Seemingly Unrelated Regressions’ SUR)

“developed by econometricians in the 1940-50s [2-5]” [1:978]

\[ E[y] = c' \cdot x + b \cdot m \quad \text{or} \quad y = c' \cdot x + b \cdot m + u_y \]

\[ E[m] = a \cdot x \quad \text{or} \quad m = a \cdot x + u_m \]

with the possibility now of correlating the errors \( u_y \) & \( u_m \)

‘Possible world’ can be matched unto a Model (sub)space.

A direct geometrical view was advanced by Saville & Wood, who view a ‘sample of N observations’ as a N-dimensional vector $y$ which can then become the focus of analysis. E.g. the cosine of the angle between 2 such vectors is the correlation.

Modeling statistically, i.e. estimating model based values, means projecting $y$ unto a model subspace $M_m$ and unto an orthogonal error subspace $E_{n-m}[1]:61$, i.e. decomposing the vector into the nearest vector to $y$ in $M$ and the nearest vector to $y$ in $E$: the smaller the error projection, the better the model.

Figure 4.5: The first process: fitting the model by projecting the observation vector $y$ onto the model space $M$.

SEM imposes a causal structure on a covariance matrix (and means), i.e. makes some of them x’s (exogenous) others y’s (endogenous), based on theory and provides a measure of how well the structure replicates the observed covariances (and the means). SEM can be tested in 1 or more observed or latent groups. There are in fact 8 matrices involved in fitting SEM, (p. 107). Hayduk, 1987 [1:116]: B, Γ, Λₓ, Λᵧ, Φ, Ψ, Qₑ and Qδ ⇒ model implied covariance matrix

\[
Σ = \begin{bmatrix}
\text{Covariances among the y’s} & \text{Covariances between y’s & x’s} \\
\text{Covariances between x’s} & \text{Covariances among the x’s}
\end{bmatrix}
\]

\[
Σ = \begin{bmatrix}
Λᵧ [(I-B)^{-1} (ΓΦΓ’+Ψ)(I-B)^{-1}] + Qₑ & Λᵧ [(I-B)^{-1} ΓΦΛₓ] \\
Λₓ Φ Γ’ [(I-B)^{-1} ΓΦΛₓ]’ & ΓΦΛₓ’
\end{bmatrix}
\]

The key problem for causal mediation comes from the fact that the b effect, M->$Y$ i.e. $E[Y|M=m_2] - E[Y|M=m_1]$ is in fact comprised of 2 such effects at the level of observables M and Y, but models estimate 1 such b effect: the nature of how $Y$ changes with an incremental ‘change’ in M has been altered from the $Tx=0$ ‘natural’ state to a ‘desired’ $Tx=1$ state.

It is then likely that b varies by $Tx$.

---

<table>
<thead>
<tr>
<th>#</th>
<th>X</th>
<th>$M_F$</th>
<th>$Y_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1.56</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.37</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1.44</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.75</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.46</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.71</td>
<td>43</td>
</tr>
</tbody>
</table>

$i = \text{indirect effect and } c = \text{total (τ) effect: not representable in this setup, because they are derived effects of the 3 ‘ingredient effects’: } c = i + c'$

3 variables are observed/observable and off the 3 effects estimable 2 more are also estimated.

$M=\text{Food Habits; } Y = \text{BMI.}$

Cases with $M$ and $Y$ values around BMI mean and +/− 1SD chosen.
The PO is not as counterintuitive at it seems: Dave Kenny says: “potential outcomes are not really new to SEMers. They simply equal the predicted value for endogenous variable, once we fix the values of its causal variables.”

Note that among its ‘causal variables’ one needs to count the ‘all-other-unknown-causes’ captured by U disturbances [2]

\[
\begin{align*}
t &= 1 \cdot u_T \\
m &= \alpha \cdot t + 1 \cdot u_M \\
y &= \beta \cdot t + \gamma \cdot m + 1 \cdot u_Y
\end{align*}
\]

Note: some PO are CF = contrary-to-fact (CF) outcomes [1] (or unrealizable).

**Mediation: potential outcomes and counterfactuals**

- $p$ indexes all $N$ people
- $j$ indexes the $N/2$ controls
- $k$ indexes the $N/2$ treated

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\frac{1}{2} M^0_p$</th>
<th>$\frac{1}{2} M^1_p$</th>
<th>$\frac{1}{2} Y^0_p M^0$</th>
<th>$Y^0_0 M^1$</th>
<th>$Y^1_0 M^0$</th>
<th>$\frac{1}{2} Y^1_p M^1$</th>
<th>$T_x^{jk}$</th>
<th>$M^{jk}$</th>
<th>$Y^{jk}$</th>
</tr>
</thead>
</table>

2 directs:
- all $p$’s = 0: $*d^p_0 = *Y^1_p M^0 - \frac{1}{2} *Y^0_p M^0$
- all $p$’s = 1: $*d^p_1 = \frac{1}{2} *Y^1_p M^1 - *Y^0_p M^1$

2 indirects:
- all $p$’s = 0: $*i^p_0 = *Y^0_p M^1 - \frac{1}{2} *Y^0_p M^0$
- all $p$’s = 1: $*i^p_1 = \frac{1}{2} *Y^1_p M^1 - *Y^1_p M^0$

1 total:
- all $p$’s = $c^p = \frac{1}{2} *Y^1_p M^1 - \frac{1}{2} *Y^0_p M^0$

Generally:
- $c^p = \frac{1}{2} (*d^p_0 + *d^p_1 + *i^p_0 + *i^p_1)$
- If no interaction, i.e. $d = *d_0 = *d_1$ and $i = *i_0 = *i_1$,
- then $c = d + i$
### Sequential Ignorability

**Sequential Ignorability a** - treatment assignment is independent of potential $M$ and $Y$ given pretreatment covariates $Z$.

\[
\rho(Tx, M_{0/1}) \mid Z = 0
\]
\[
\rho(Tx, Y_{0/1}) \mid Z = 0
\]

**Sequential Ignorability b** - mediator is independent of potential $Y$, given $Tx$ and $Z$.

\[
\rho(M, Y_{0/1}) \mid Z = 0
\]

Withing subgroups of $Z$ $Tx$ is independent on $M_{po}$.  

---

X is ignorable = statistically independent of $M_p$ and $Y_p$

**Sequential Ignorability a** - Tx occurs at random (re: Y) conditional on pretreatment covariates

**Sequential Ignorability b** - M occurs at random (re: Y) conditional on pretreatment covariates and Tx

=> There are no unmeasured pre- or post-treatment confounders of M-\(\rightarrow\)Y
Specific indirect

\text{INTERVENTION-FOOD HABITS}_{6m} \rightarrow \text{BMI}_{6m} \ -0.184 \ 0.067 \ -2.733 \ 0.006

Baseline same-measure used as covariates, along with age and education.
Chen & Pearl example:

(i) Identify three testable implications of this model

(ii) Identify a testable implication assuming that only $X$, $Y$, $W_3$, and $Z_3$ are observed

(iii) Suppose $X$, $Y$, and $W_3$ are the only variables observed. Which parameters can be identified from the data?

(iv) If we regress $Z_1$ on all other variables in the model, which regression coefficient will be zero?

(v) The model implies that certain regression coefficients will remain invariant when an additional variable is added as a regressor. Identify five such coefficients with their added regressors.

The Q. of ‘Causal Mediation’ CM is central when there is XbyM interaction possible, which can be tested. Or at least incorporate it and get better estimates of the indirect effect (CM), accounting for interaction; we can then compare it to the classic=’naïve’ $a \cdot b$ one.

INTERACTION though is simply the variation of $b$ along levels of $X$, which is made clear by the random b slope model (Muthen, CM:13, eq. 18&19):

$$y_i = \alpha_Y + b_i \cdot m_i + c' \cdot x_i + \epsilon_{yi} \quad (18)$$

and

$$b_i = b_0 + b_x \cdot x_i + \epsilon_{bi} \quad (19)$$

which can be better described by the reduced form with interaction

$$y_i = \alpha_Y + b \cdot m_i + c' \cdot x_i + b_x \cdot x_i \cdot m_i + \epsilon_{yi} \quad (17)$$

Total effect of $X$ on $Y$ is the change in $Y$ when one changes $X$ from a $x_0$ to a $x_1$:

$$TE = E[Y_{x_1} - Y_{x_0}] = E[Y_{x_1}^{Mx_1} - Y_{x_0}^{Mx_0}]$$

One can add & subtract $*Y_{x_1}^{Mx_0}$ [1]:

$$E[Y_{x_1}^{Mx_1} - Y_{x_0}^{Mx_0} + *Y_{x_1}^{Mx_0} - *Y_{x_1}^{Mx_0}] =$$

$$= E[Y_{x_1}^{Mx_1} - *Y_{x_1}^{Mx_0} + *Y_{x_1}^{Mx_0} - Y_{x_0}^{Mx_0}] =$$

$$= E[Y_{x_1}^{Mx_1} - *Y_{x_1}^{Mx_0}] + E[*Y_{x_1}^{Mx_0} - Y_{x_0}^{Mx_0}]$$

$$= TNIE + PNDE = TE$$

$$= \text{Total (Natural) Indirect } E + \text{Pure (Natural) Direct } E$$

Change ‘in treated’ $Y$ due to $M_{x_1} - M_{x_0}$ Change in $Y$ due to $Tx$ at same $M_{x_0}$

Causal Mediation formulas [2]

Or one can add & subtract instead [1] \( *Y_{x0}^{Mx1} \):

\[
E[Y_{x1}^{Mx1} - Y_{x0}^{Mx0} + *Y_{x0}^{Mx1} - *Y_{x0}^{Mx1}] = \\
= E[Y_{x1}^{Mx1} - *Y_{x0}^{Mx1} + *Y_{x0}^{Mx1} - Y_{x0}^{Mx0}] = \\
= E[Y_{x1}^{Mx1} - *Y_{x0}^{Mx1}] + E[*Y_{x0}^{Mx1} - Y_{x0}^{Mx0}] = \\
= TNDE + PNIE = TE
\]

Change in \( Y \) due to \( Tx \) at same \( M_{x1} \) Change in controls \( Y \) due to \( M_{x1}-M_{x0} \)

So

\[ TE = TNIIE + PNDE = TNDE + PNIE \]

None of these effects are that ‘natural’: they operate in the PO space, in which some variables like \( M_0 \) and \( M_1 \) are \( \frac{1}{2} \) unobserved, others like \( *Y_{x0}^{Mx1} \) are 100% unobservable by design.

From now on, we’ll drop the N, since it appears in all (as in [1]).

So

\[ TE = TIE + PDE = TDE + PIE \]

Causal Mediation definitions

<table>
<thead>
<tr>
<th>Effects</th>
<th>Interpretations of Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Directs</strong></td>
<td></td>
</tr>
<tr>
<td>PDE</td>
<td>Change in Y due to Tx at same $M_0$</td>
</tr>
<tr>
<td>TDE</td>
<td>Change in Y due to Tx at same $M_1$</td>
</tr>
<tr>
<td><strong>Indirects</strong></td>
<td></td>
</tr>
<tr>
<td>TIE</td>
<td>Change due to $M_1 - M_0$ at same $Tx=1$</td>
</tr>
<tr>
<td>PIE</td>
<td>Change due to $M_1 - M_0$ at same $Tx=0$</td>
</tr>
</tbody>
</table>

Several of these effects *do not appear at all natural*, in terms of realistic or observational quantities.

That’s because they operate however at the PO (Potential Outcomes) level.
There is also a controlled direct effect, when one holds/forces $M$ to a value $m$ (central to policy evaluation):

$$CDE^m = E[Y_{x1}^m - *Y_{x0}^m]$$

Note: $CDE^{M0} = PDE$

When the unit controlled direct effects $CDE$ is constant across levels of $m$, i.e. the additivity assumption [1], or the no-interaction assumption [2], then

$$TDE = PDE \text{ and } TIE = PIE \quad [3]$$

## Names of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>We</th>
<th>Imai</th>
<th>Pearl/V. Weele</th>
<th>Robins</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{2}Y_{1M1}^0 - *Y_{1M0}^1$</td>
<td>$i_1 = \text{TIE}_1$</td>
<td>ACME(treated)</td>
<td>TIE</td>
<td>TIE</td>
</tr>
<tr>
<td>$Y_{0M1}^1 - \frac{1}{2}Y_{0M0}^0$</td>
<td>$i_0 = \text{PIE}_0$</td>
<td>ACME(control)</td>
<td>PIE</td>
<td>PIE</td>
</tr>
<tr>
<td>$Y_{1M0}^0 - \frac{1}{2}Y_{0M0}^0$</td>
<td>$d_0 = \text{PDE}_0$</td>
<td>ADE(control)</td>
<td>PDE</td>
<td>DE</td>
</tr>
<tr>
<td>$\frac{1}{2}Y_{1M1}^0 - *Y_{0M1}^0$</td>
<td>$d_1 = \text{TDE}_1$</td>
<td>ADE(treated)</td>
<td>TDE</td>
<td>-</td>
</tr>
<tr>
<td>$\frac{1}{2}Y_{1M1}^0 - \frac{1}{2}Y_{0M0}^0$</td>
<td>$C$</td>
<td>Total</td>
<td>Total</td>
<td>Total</td>
</tr>
</tbody>
</table>

- **TE** - Total Effect
- **TIE** - Total indirect effect
- **PIE** - Pure indirect effect
- **ACME** - Average causal mediated effect
- **ADE** - Average direct effect
- **PNIE** - Pure natural indirect effect
- **PNDE** - Pure natural direct effect
- **DE** - direct effect
- **CDE** - Controlled Direct Effect
- **PDE** - Pure direct effect
- **INT_{ref}** - Reference Interaction
- **INT_{med}** - Mediated Interaction
- **PE** - Proportion Eliminated
- **TNIE** - Total natural indirect effect
- **TNDE** - Total natural direct effect

---

Tyler VanderWeele’s unification of mediation and moderation

\[
\begin{align*}
TE &= CDE + INT_{\text{ref}} + INT_{\text{med}} + IE_{BK} \\
&= d_0 + i_1 + d_1 + i_0 + PE
\end{align*}
\]

- **TE** - Total Effect
- **CDE** - Controlled Direct Effect
- **INT_{\text{ref}}** - Reference Interaction
- **INT_{\text{med}}** - Mediated Interaction
- **IE_{BK}** - Baron-Kenny indirect effect
- **d_0** = Pure direct effect
- **i_1** = Total indirect effect
- **d_1** = Total direct effect
- **i_0** = Pure indirect effect
- **PE** = Proportion Eliminated
- **ACME** - Average causal mediated effect
- **PNIE** - Pure natural indirect effect
- **DE** - direct effect
- **ADE** - Average direct effect
- **PNDE** - Pure natural direct effect
- **IE_{BK}** - Baron-Kenny indirect effect
- **TNIE** - Total natural indirect effect
- **TNDE** - Total natural direct effect


**MMM 2015 Causal Mediation** 34
\[ y_i = \alpha_Y + b_i \cdot m_i + c' \cdot x_i + \varepsilon_{yi} \quad (0Y) \& \]
\[ m_i = \alpha_m + a \cdot x_i + \varepsilon_{mi} \quad (0M) \& \]
\[ b_i = b_0 + b_{\text{INT}} \cdot x_i + \varepsilon_{bi} \quad (0b) = \text{random } b \text{ effect} \]

One can substitute \( m \) and \( b_i \) into the \( y \) equation to get the reduced form (overall \( Y \) prediction)
Muthen Causal Mediation (CM) SEM & Mplus 7.2

\[ y_i = \alpha_Y + b \cdot m_i + c' \cdot x_i + b_{INT} \cdot x_i \cdot m_i + \epsilon_{yi} \quad \text{and} \quad m_i = \alpha_m + a \cdot x_i + \epsilon_{mi} \]

can better be described by the reduced form with interaction

\[ \mathbb{E}(y|x) = \alpha_Y + b \cdot \alpha_m + a \cdot b \cdot x + c' \cdot x + b_{INT} \cdot \alpha_m \cdot x + b_{INT} \cdot a \cdot x^2 \]

Total Natural Indirect Effect : \quad TIE = a \cdot b + a \cdot b_{INT} \quad \text{(1)}

Pure Natural Direct effect : \quad PNDE = c' + \alpha_M \cdot b_{INT} \quad \text{(2)}

Pure Natural Indirect Effect : \quad PIE = a \cdot b \quad \text{should replicate the classic} \quad \text{(3)}

Total Natural Direct Effect : \quad TDE = c' + \alpha_M \cdot b_{INT} + a \cdot b_{INT} \quad \text{(4)}

Total effect \quad TE = (1)+(2) = (3)+(4)

When = 0, causal mediation effects fall back on the classic effects.

Causal Mediation current implementation overview

$main lines:

1. Bengt Muthen\textsuperscript{1} = Mplus tradition – constantly expanding
   - Mplus syntax – classic IND statement and the new MOD statement
2. Imai, Keele, Tingley & Yamamoto\textsuperscript{2,3}.
   - R (library"mediation“\textsuperscript{4}) and Stata syntaxes paramed\textsuperscript{5}
   - medeff & medsens
   - Stata 13 now has a SEM module incorporated
3. Valeri & VanderWeele\textsuperscript{6}
   - SPSS, SAS (mediation.sas) and Stata (paramed) syntaxes
   + Andrew Hayes’ SPSS PROCESS\textsuperscript{7}
4. VanderWeele ‘unification’\textsuperscript{8} SAS syntaxes
   + Pearl –not implemented in software to our knowledge

---

# Causal Mediation software comparison

<table>
<thead>
<tr>
<th>Macro</th>
<th>Different covariates Y vs. M</th>
<th>Measurem. error</th>
<th>Correlate errors</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mplus</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>SPSS</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>Stata</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>SAS</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>R</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
</tbody>
</table>


M: With both X and M latents a bit clunky.


1. The intervention caused a 3.6 lbs drop (.63 in BMI, average woman height 63.8), reflecting somewhat the sample BMI mean changes in the treated group: 35.2 baseline to 34.7 follow-up.

2. The DE and IE look differently in their Pure and Natural forms: one needs a PO translation here.

Note: p values for parameters in superscripts

Findings:
- $c = -0.614^{.023}$
- $\text{PNDE} = -0.449^{.061}$
- $\text{TNIE} = -0.430^{.160}$
- $\text{PNIE} = -0.184^{.006}$
- $\text{TNDE} = -0.165^{.015}$
Imai’s (et al.) interpretation (2) of correlated errors

\[
\text{Cov}(\epsilon_{Mi}, \epsilon_{Yi}) = V(X_i) \text{Cov} (\epsilon_{ai}, \epsilon_{ci'}) + a V(X_i) \text{Cov} (\epsilon_{ai}, \epsilon_{bi})
\]

Heterogeneity of effects - each effect varies by individuals

When the M and Y confounder is considered, by allowing their errors to correlate, the picture gets muddier.

Findings:

- $c = -0.630^{.018}$
- $PNDE = -0.632^{.011}$
- $TNIE = -0.597^{.053}$
- $PNIE = -0.033^{.662}$

Mplus

MODEL:
Y2 on Y1 M2 INTERV age  xbym;  !Y
M2 on M1 INTERV age   ;!M
MODEL INDIRECT:
Y2 MOD M2 xbym INTERV;

SAS 1

%mend;

%mediation(data=data,yvar=y,avar=x,mvar=m,cvar=m0 y0 age
edu,a0=0,a1=1,m=0,nc=4,yreg=linear,mreg=linear, interaction=true); run;
Stata__________________________________________________
medeff (regress M2 inter M1 age) (regress Y2 inter M2 Y1 age),
treat(interv) mediate(M2) sims(1000)
medsens (regress M2 inter M1 age) (regress t2bmi inter M2 Y1
age), treat(interv) mediate(M2) sims(1000)
paramed Y2, avar(interv) mvar(M2 ) cvars(age M1 Y1) a0(0) a1(1)
m(1) yreg(linear) mreg(linear) nointer
R _____________________________________________________
med.fit <- glm(M2 ~ inter + M1 + AGE, data =sistlk)
> summary(med.fit)
out.fit <- glm(t2bmi ~ M2 + inter + M2 + Y1 + AGE, data = sistlk)
> summary(out.fit)
> med.out <- mediate(med.fit, out.fit, treat = "interv", mediator = 
"M2",robustSE = TRUE, sims = 100)
> summary(med.out)
proc nlmixed data=sistk;
parms im=0 cprime=0 b=0 gamint=0 yage=0 yfh0=0 ybmi0=0 iy=0 tx=0 mage=0 mfh0=0 mbmi0=0 ss_m=1 ss_y=1;
tx1=1; tx0=0; mstar=0; cc1=50.8646; cc2=.9405; cc3=35.6517;
mu_y=im + cprime*TX + b*M + gamint*A*M + yage*C1 + yM0*C2 + yY0*C3;
mu_m =iy + a*TX + mage*C1 + mM0*C2 + mY0*C3;
[...]
estimate "PDE Pure Direct Effect" (cde + intref);
estimate "TIE Total Indirect Effect" (intmed + pie);
estimate "TDE Total Direct Effect" (cde + intref + intmed);
Causal Mediation (CM) Mplus Illustration

<table>
<thead>
<tr>
<th>Diagram Markings</th>
<th>Food Habits&lt;sub&gt;6m&lt;/sub&gt;</th>
<th>BMI&lt;sub&gt;6m&lt;/sub&gt;</th>
<th>Tx</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε&lt;sub&gt;M&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α&lt;sub&gt;M&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ε&lt;sub&gt;Y&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Estimate** | **S.E.** | **Est./S.E.** | **P-Value**
--- | --- | --- | ---
DIRECT effect c’ (FOOD HABITS<sub>6m</sub>→BMI<sub>6m</sub>) | -0.401 | 0.425 | -0.943 ↑ | 0.346 |
a effect (INTERVENTION→FOOD HABITS<sub>6m</sub>) | -0.161 | 0.029 | -5.526 | 0.000 |
b effect (FOOD HABITS<sub>6m</sub>→BMI<sub>6m</sub>) | 1.141 | 0.363 | 3.145 | 0.002 |
INTERVENTION*by*FOOD HABITS<sub>6m</sub> | -0.115 | 0.515 | -0.224 | 0.823 |
TOTAL, INDIRECT, AND DIRECT EFFECTS BASED ON COUNTERFACTUALS (CAUSALLY-DEFINED EFFECTS)

- Pure natural DIRECT effect | -0.449NS | 0.240 | -1.871 | 0.061 |
- Tot natural INDIRECT effect | -0.165* ↓ | 0.068 | -2.440 | 0.015 |
- Total effect | -0.614* | 0.269 | -2.281 | 0.023 |
- Other effects
  - Tot natural DIRECT effect | -0.430NS | 0.307 | -1.404 | 0.160 |
  - Pure natural INDIRECT effect | -0.184* ↓ | 0.067 | -2.733 | 0.006 |
  - Total effect | -0.614* | 0.269 | -2.281 | 0.023 |
The correlation between M’s residual and Y’s residual was significant and positive, indicating something outside the model makes people’s bad habits stronger and at the same time contributes to making them heavier; shall we say… *advertising*?

<table>
<thead>
<tr>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M WITH Y</td>
<td>0.076</td>
<td>0.024</td>
<td>3.220</td>
</tr>
</tbody>
</table>

More importantly, such a potential confounder becomes even MORE problematic IF the Tx itself impacts it! [1]

---

Stata results

medeff (regress M2 interv M1 age d Educ) (regress t2bmi interv M2 Y1 age d Educ), treat(interv) mediate(M2) sims(1000> )

<table>
<thead>
<tr>
<th>Effect</th>
<th>Mean Paramed</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACME</td>
<td>-.176*↓ nie = -.319*</td>
<td>-.294   -.080</td>
</tr>
<tr>
<td>Direct Effect</td>
<td>-.491*↑ cde = -.406*</td>
<td>-.813   -.172</td>
</tr>
<tr>
<td>Total Effect</td>
<td>-.667*↓ te = -.725</td>
<td>-.975   -.371</td>
</tr>
<tr>
<td>% of Tot Eff mediated</td>
<td>.263*</td>
<td>.181     .475</td>
</tr>
</tbody>
</table>

----- ACME = Average Causal Mediation Effect------

medsens (regress M2 interv M1 age d Educ) (regress t2bmi interv M2 Y1 age d Educ), treat(interv) mediate(M2) sims(100> )

Sensitivity results

Rho at which ACME = 0
| .3567 |

R^2_M*R^2_Y* at which ACME = 0:
| .1272 |

R^2_M~R^2_Y~ at which ACME = 0:
| .001 |

--- MMM 2015 ---

Causal Mediation
Stata results

```stata
t2bmi, avar(interv) mvar(M2) cvars(age d_educ M1 Y1) a0(0) a1(1) m(1) yreg(linear) mreg(linear)
|   Est.   Std Err.   P>|z|  [95% Conf.Interval]
-------------+----------------------------------
  CDE |  -.456   .201  0.023  -.850  -.063
  NDE |  -.423   .167  0.011  -.750  -.097
  NIE |  -.302   .090  0.001  -.479  -.124
  MTE |  -.725   .160  0.000  -1.040  -.410
```

CDE: Controlled Direct Effect, NDE: Natural Direct Effect, NIE: Natural Indirect Effect, MTE: Marginal Total Effect

Findings: significant INDIRECT effect and TOTAL effect, but also a significant DIRECT effect.

Note: Stata can estimate (see next slide) CDE and NIE without interaction!
paramed t2bmi, avar(interv) mvar(M2) cvars(age d educ M1 Y1) a0(0) a1(1) m(1) yreg(linear) mreg(linear) nointer

|       | Estimate | Std Err | P>|z| | [95% Conf Interval] |
|-------|----------|---------|-------|------------------|
| CDE   | -.406    | .161    | 0.012 | -.722            | -.091 |
| NIE   | -.320    | .083    | 0.000 | -.481            | -.157 |
| TE    | -.725    | .160    | 0.000 | -1.039           | -.411 |

Cde: Controlled Direct Effect, NIE: Natural Indirect Effect, TE: Total Effect

Stata replicates the finding of significant INDIRECT effect and TOTAL effect (computed with NO INTERACTION!), but finds a significant DIRECT effect too.
glm(formula = M2 ~ interv + M1 + AGE + d_educ, data = sistlk)

glm(formula = t2bmi ~ M2 + interv + M2 + Y1 + AGE + d_educ, data = sistlk)

med.out <- mediate(med.fit, out.fit, treat = "interv", mediator = "M2", robustSE = TRUE, sims = 100)

summary(med.out)

### Causal Mediation Analysis

### Quasi-Bayesian Confidence Intervals

<table>
<thead>
<tr>
<th>Estimate</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACME</td>
<td>-0.175</td>
<td>-0.303</td>
<td>-0.089</td>
</tr>
<tr>
<td>ADE</td>
<td>-0.507</td>
<td>-0.760</td>
<td>-0.226</td>
</tr>
<tr>
<td>Total Effect</td>
<td>-0.683</td>
<td>-0.943</td>
<td>-0.397</td>
</tr>
<tr>
<td>Prop. Mediated</td>
<td>0.262</td>
<td>0.143</td>
<td>0.448</td>
</tr>
</tbody>
</table>

R replicates closer the medeff Stata results: significant INDIRECT effect (26% of total) and TOTAL effect (computed with NO INTERACTION!), and a significant DIRECT effect.
%mediation (data=data, yvar=y, avar=x, mvar=m, cvar=m0 y0 age edu, a0=0, a1=1, m=0, nc=4, yreg=linear, mreg=linear, interaction=true);
run;

There are some major differences in estimates: direct effect is SIG. indirect is NS:

<table>
<thead>
<tr>
<th>#</th>
<th>Effect</th>
<th>Est.</th>
<th>SE</th>
<th>p</th>
<th>L_95_CI</th>
<th>U_95_CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cde</td>
<td>-0.572</td>
<td>0.364</td>
<td>0.116</td>
<td>-1.285</td>
<td>0.141</td>
</tr>
<tr>
<td>2</td>
<td>nde</td>
<td>-0.550</td>
<td>0.177</td>
<td>0.002</td>
<td>-0.896</td>
<td>-0.203</td>
</tr>
<tr>
<td>3</td>
<td>nie</td>
<td>-0.129</td>
<td>0.071</td>
<td>0.070</td>
<td>-0.268</td>
<td>0.010</td>
</tr>
<tr>
<td>4</td>
<td>mar.tot. ef.</td>
<td>-0.679</td>
<td>0.167</td>
<td>0.000</td>
<td>-1.006</td>
<td>-0.352</td>
</tr>
</tbody>
</table>

How does one find out what would have happened?

How to obtain an individual $E[Y_{X=x}|Ev.=e]$

_Update knowledge based on evidence, then predict._

1. Abduction – Update $P[\varepsilon]$ by evidence $Ev.$ to $P[\varepsilon | Ev. = e]$

2. Action – Modify the model $M$ to $M_X$ by replacing $X$ equations to an $X$ equality.

3. Prediction – Use $M_X$ and the updated probabilities over the variable $\varepsilon$ to compute expectation of $Y$, the consequence of the counterfactual.


MMM 2015

Causal Mediation 53
What would have happened to average ‘control’ Jill, had she been exposed to treatment?

A 2-step process. The key is capturing $\varepsilon_{pM}$ and $\varepsilon_{pM}$ for ‘average’ Jill first, which anchors her position in the ‘patient state space’.

**Note:** significance of parameters here is moot.

$\hat{M}_0: \hat{M}^{\text{Jill}} = -.19$, so $\varepsilon_{M}^{\text{Jill}} = .56$

$\hat{Y}_{0M^0}: \hat{Y}^{\text{Jill}} = 27.4$, so $\varepsilon_{Y}^{\text{Jill}} = 6.48$

**Tx 0->1:** $\hat{Y}_{1M1} = 34.9$, so Jill would lose 7.4 Lbs.

**Note:** we use the ‘causal mediation model’ estimates, along with age and prior same-variable.

How much weight would Jill lose, had he magically been exposed to the intervention?

We make $\text{Tx}=1$ now, so Jill’s equation has a 1 instead of a 0 in the Tx factor (-0.40·1), which triggers 2 types of effects, through $M$ and directly; if we add her $M$ uniqueness to get $\overline{M_{\text{Jill}}}$ and her BMI uniqueness to get $\overline{Y_{\text{Jill}}}$, we get:

* $\text{BMI}_{1, \text{Jill}} = 34.94$

So Jill’s Potential Outcome would be 6.1 lbs. lower than her observed weight of 208 lbs, had she been ‘treated’.
What would have happened to Jill, had her Food Habits changed for the better (magically)?

We cut Jill’s M observed value to a .20, so Y now changes only due to M and directly due to Tx, but Jill’s condition does not ‘change (still not treated). She ‘drops’ 5.9 lbs.

Conclusions about Jill’s potential outcomes

* Jill, a average (BMI=36.2) woman, happened to improve her food habits (.44 to .37), better than her average controls buddies vs. .89.

* Had her food habits improved still (to a .20, what her mediator level would have been if treated), she would have lost 5.9 Lbs.

* Had she participated in the SisterTalk intervention, she would have lost 7.4 Lbs. A good chunk of this effect (42%) is due improving food habits, while a 58% is NOT through improving food habits.

Other Mediation concerns

* Measurement error is more troublesome in terms of predictors, more so the mediator [1-4]

“measurement error in the mediator causes an underestimated indirect effect and an overestimated direct effect in line with regression on one predictor measured with error and the other without error” [1:6]

* Powers (1-β) to detect effects differ: indirect > direct, & indirect >> total [5].

* The effect of measurement error and mediator-outcome confounding seem to be opposite [6].

The Causal mediation estimation yielded a significant INDIRECT Effect, both in the form of PIE and TIE, and non-significant DIRECT Effect, both in the form of PDE and TDE; total effect of intervention was significant.

1. The intervention does reduce BMI in the treated participants at follow-up, compared to controls.

2. The reduction in BMI can be attributed largely to improvements in food habits (eating healthier), rather than through other mechanisms.
Findings about indirect effects may have to be stated, unfortunately, more qualitatively [1].

Sensitivity analyses can tell us whether our conclusion is likely or not, given reasonable assumptions and missed factors/processes: specifically confounders.

M&Y errors were correlated positively => investigate

Different software may attack the same model slightly different, most flexible appears nowadays to be Mplus.

Comparing outputs side-by-side can aid in arriving at a common language and approach to properly test mediation.

Causal Mediation can be intuitive, but a common language is still negotiated.

Visual models of impact/effects can fully replace regression-based equations.

Judea Pearl’s Structural Causal Modeling graphical ‘language’ can be incorporated into simple/complex statistical models.

* Good translators of his work are still badly needed.

Estimating causal direct and indirect effects implies making statements about partial and total unknowns, and updating knowledge using possible scenarios.

coman@uchc.edu
http://evaluatehelp.blogspot.com/

Thank you!