PANEL ON MEDIATION AND OTHER MIRACLES OF THE 21ST CENTURY

Judea Pearl
UCLA

With Kosuke Imai and many others

THE QUESTION OF MEDIATION (direct vs. indirect effects)

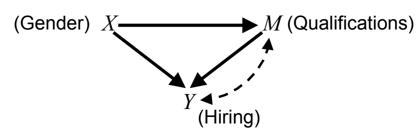
1. Why decompose effects?
2. What is the definition of direct and indirect effects?
3. What are the policy implications of direct and indirect effects?
4. When can direct and indirect effect be estimated consistently from experimental and nonexperimental data?

WHY DECOMPOSE EFFECTS?

1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions: Signal re-routing and mechanism deactivating, rather than variable fixing

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y ?

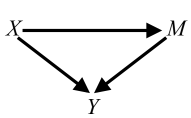
$$CDE = E(Y|do(x_1), do(m)) - E(Y|do(x_0), do(m))$$

(m-dependent) Adjust for M? No! No!

CDE identification is completely solved (Tian et al. 2002)

NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – Pearl (2001)



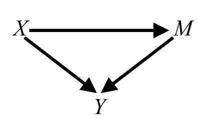
$m = f(x, u)$
$y = g(x, m, u)$

Natural Direct Effect of X on Y: $DE(x_0, x_1; Y)$
The expected change in Y, when we change X from x_0 to x_1 and, for each u, we keep M constant at whatever value it attained before the change.

$$E[Y_{x_1 M_{x_0}} - Y_{x_0}]$$

Note the 3-way symbiosis

DEFINITION OF INDIRECT EFFECTS



$m = f(x, u)$
$y = g(x, m, u)$

No controlled indirect effect

Indirect Effect of X on Y: $IE(x_0, x_1; Y)$
The expected change in Y when we keep X constant, say at x_0, and let M change to whatever value it would have attained had X changed to x_1.

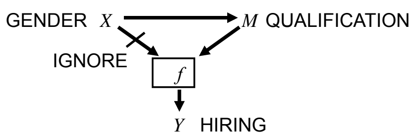
$$E[Y_{x_0 M_{x_1}} - Y_{x_0}]$$

In linear models, *IE = TE - DE*

POLICY IMPLICATIONS OF INDIRECT EFFECTS

What is the indirect effect of X on Y?

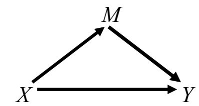
The effect of Gender on Hiring if sex discrimination is eliminated.



Deactivating a link – a new type of intervention

7

THE MEDIATION FORMULAS IN UNCONFOUNDED MODELS



$m = f(x, u_1)$
$y = g(x, m, u_2)$
u_1 independent of u_2

$$DE = \sum_m [E(Y \mid x_1, m) - E(Y \mid x_0, m)] P(m \mid x_0)$$

$$IE = \sum_m [E(Y \mid x_0, m)[P(m \mid x_1) - P(m \mid x_0)]$$

$$TE = E(Y \mid x_1) - E(Y \mid x_0) \qquad TE \neq DE + IE$$

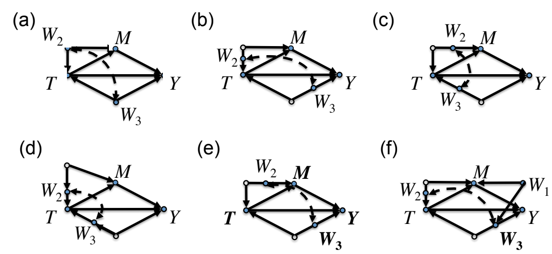
$IE =$ Fraction of responses explained by mediation (sufficient)

$TE - DE =$ Fraction of responses owed to mediation (necessary)

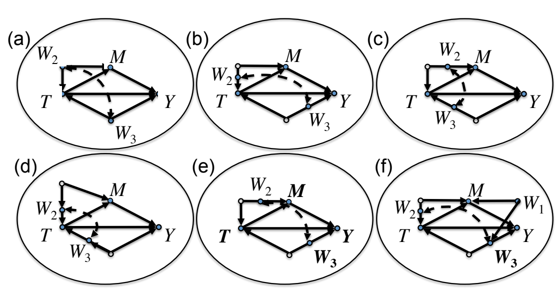
NONPARAMETRIC IDENTIFICATION (Natural mediation is a solved problem)

- The nonparametric estimability of natural (and controlled) direct and indirect effects can be determined in polynomial time given any causal graph G with both measured and unmeasured variables.
- If NDE (or NIE) is estimable, then its estimand can be derived in polynomial time.
- The algorithm is complete and was extended to any path-specific effects (Shpitser, 2013).

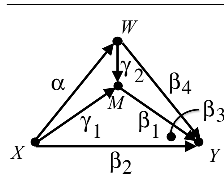
WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



WHAT CAN MEDIATION FORMULA DO FOR PARAMETRIC ANALYSTS?



Multi-mediators non-linear model

$$y = \beta_1 m + \beta_2 x + \beta_3 xm + \beta_4 w + u_1$$
$$m = \gamma_1 x + \gamma_2 w + u_2$$
$$w = \alpha x + u_3$$

What combination of parameters gives the effect mediated by M?

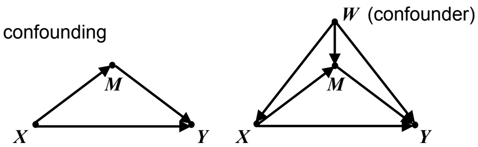
$$IE(M) = \beta_1(\gamma_1 + \alpha\gamma_2)$$

What combination of parameters gives the effect owed to M?

$$TE - DE(M) = (\beta_1 + \beta_3)(\gamma_1 + \alpha\gamma_2)$$

IGNORABILITY CONDITIONS FOR NDE IDENTIFICATION (Sequential Ignorability) (Imai et al (2010))

No confounding



(a) (b)

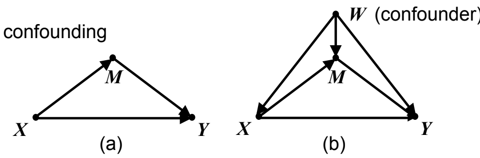
There exists a set W of measured covariates such that:

SI-1 $M_x \perp\!\!\!\perp Y_{xm} \mid (X, W)$

SI-2 $X \perp\!\!\!\perp (Y_{xm}, M_x) \mid W$

WEAKER AND TRANSPARENT CONDITIONS FOR NDE IDENTIFICATION

No confounding



(a) (b)

There exists a set W such that:

A-1 No member of W is a descendant of X.

A-2 W blocks all back-door paths from M to Y, disregarding the one through X.

A-3 The W-specific effect of X on M is identifiable.

$P(m \mid do(x), w)$

A-4 The W–specific effect of $\{X, M\}$ on Y is identifiable.

$P(y \mid do(x,m), w)$

FRIENDLY EXCHANGE CONCERNING IGNORABILITY VS. GRAPHICAL ASSUMPTIONS

- Psychological Methods (2014) Imai et al. proved that graphical and ignorability assumptions are identical for randomized treatments.
- Concensus achieved regarding transparency of graphical assumptions.
- Semi-concensus regarding other aspects of the graphical vs. ignorability languages.

DAGS VS. POTENTIAL COUTCOMES AN UNBIASED PERSPECTIVE

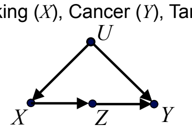
1. Semantic Equivalence
2. Both are abstractions of Structural Equation Models (SEM).

$$\boxed{Y_x(u) = Y_{M_x}(u)} \qquad \boxed{\begin{array}{c} X \rightarrow Y \\ y = f(x,z,u) \end{array}}$$

$Y_x(u)$ = All factors that affect Y when X is held constant at $X{=}x$.

FORMULATING A PROBLEM IN THREE LANGUAGES

1. English: Smoking (X), Cancer (Y), Tar (Z), Genotypes (U)



2. Counterfactuals: $Z_x(u) = Z_{yx}(u)$,

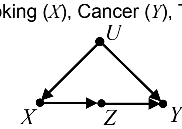
$X_y(u) = X_{zy}(u) = X_z(u) = X(u)$,

$Y_z(u) = Y_{zx}(u), \quad Z_x \perp\!\!\!\perp \{Y_z, X\}$

Not too friendly:
Consistent?, complete?, redundant?, plausible?, testable?

FORMULATING A PROBLEM IN THREE LANGUAGES

1. English: Smoking (X), Cancer (Y), Tar (Z), Genotypes (U)

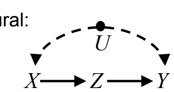


2. Counterfactuals: $Z_x(u) = Z_{yx}(u)$,

$X_y(u) = X_{zy}(u) = X_z(u) = X(u)$,

$Y_z(u) = Y_{zx}(u), \quad Z_x \perp\!\!\!\perp \{Y_z, X\}$

3. Structural:



$x = f_1(u, \varepsilon_1) \quad y = f_3(z, u, \varepsilon_3)$

$z = f_2(x, \varepsilon_2) \quad \varepsilon_1 \perp\!\!\!\perp \varepsilon_2 \perp\!\!\!\perp \varepsilon_3$

THE STRUCTURAL-COUNTERFACTUAL SYMBIOSIS

1. Express theoretical assumptions in structural language.
2. Express queries in counterfactual language.
3. Translate (1) into (2) for algebraic analysis,
   Or (2) into (1) for graphical analysis.
4. Use either graphical or algebraic machinery
   to answer the query in (2).

Thank you

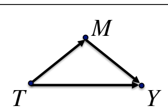
JON'S QUESTIONS TO PANEL

1. Do you think an experiment has any value without mediational analysis?
2. Is a separate study directly manipulating the mediator useful?
   How is the second study any different from the first one?
3. Imai's correlated residuals test seems valuable for distinguishing fake from genuine mediation . Is that so?
4. Why isn't it easy to test whether participants who show the largest increases in the posited mediator show the largest changes in the outcome?
   $Y_1 - Y_0 = f_m(Z_0 - Z_1) \qquad f_m$ monotonic
5. Why is mediational analysis any "worse" than any other method of investigation?

CHRISTIAN'S QUESTIONS TO PANEL

1. Can we go beyond "if assumptions, then conclusions"?
   - Yes, testable implications, experimental evidence, other studies.
2. How your framework would use the results of one mediation analysis to inform the setup of a second, new mediation analysis?
   - It is not a question of "framework" but of "information"

THE MEDIATION FALLACY OR HOW TRADITIONALISTS CONFUSED AN ENTIRE CENTURY

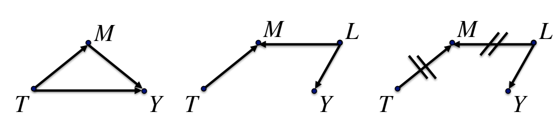


A

A – Standard model
   DE = Whatever changes we see in Y when we vary X and "hold M constant"
- Holding M constant ≠ controlling for M
- Statistics has no operator for "holding M constant"
- "Controlling for M" leads to fallacies in the presence of unobserved confounders

---

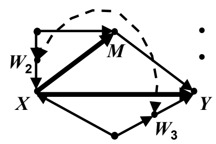
THE MEDIATION FALACY OR HOW TRADITIONALISTS CONFUSED AN ENTIRE CENTURY



A B C

B – Conditioning on M creates dependency between T and Y though DE = 0
C – "Fixing M" correctly shows DE = 0 ⇒ No dependence

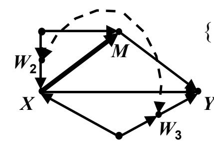
DIVIDE AND CONQUER MAKES A DIFFERENCE



- W_2 deconfounds $X \rightarrow M$
- W_3 deconfounds $X \rightarrow Y$

Two separate sets can accomplish what their union can't.

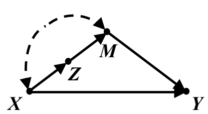
DIVIDE AND CONQUER MAKES A DIFFERENCE



$\{W_2, W_3\}$ confounds $X \rightarrow M$

Two separate sets can accomplish what their union can't.
To identify NDE, we must condition on W_2 and W_3 separately.

SEQUENTIAL IGNORABILITY IS NOT NEEDED



$W = \emptyset$

No set can deconfound $X \rightarrow M$.
Measuring Z permits the identification of $P\,(M = m \mid do(x))$ through the front-door formula.
NDE is identified.