

Causal organic direct and indirect effects: closer to Baron and Kenny

Judith J. Lok

Associate Professor of Mathematics and Statistics

Boston University

jjlok@bu.edu

Adjunct Associate Professor of Biostatistics

Harvard T.H. Chan School of Public Health

Congratulations to Aad!

Leiden University, June 18, 2019

Joint work with Ronald J. Bosch

Center for Biostatistics in AIDS Research

Harvard T.H. Chan School of Public Health

Supported by: NSF DMS 1810837, NIH NIAID R01AI100762

(but not the view of NIH or NSF)

Motivation: direct and indirect effects

Direct and indirect effects decompose the effect of a treatment A on outcome Y into:

- part that is mediated through covariate M (*indirect effect*)
- part that is not (*direct effect*)

Example: A blood pressure lowering medication,
outcome Y : heart attack.

How much of the effect of the blood pressure lowering medication on heart attacks is mediated by its effect on blood pressure, and how much (if any) by other pathways?

Motivation: direct and indirect effects

Direct and indirect effects decompose the effect of a treatment A on outcome Y into:

- part that is mediated through covariate M (*indirect effect*)
- part that is not (*direct effect*)

Example: there are several potential/pre-clinical HIV curative treatments, typically targeting the HIV reservoir, which may prolong the time to viral rebound after ART interruption.

What is the effect of an HIV curative treatment on viral rebound mediated through the HIV reservoir?

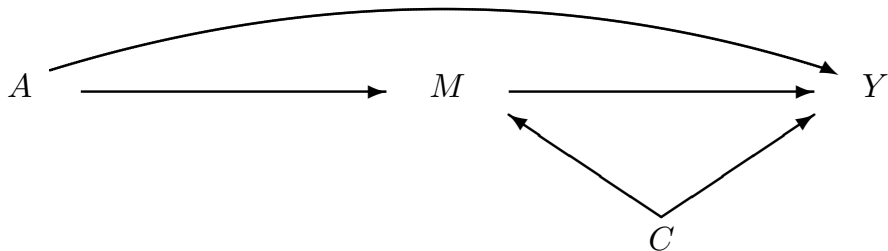
Motivation: direct and indirect effects

Direct and indirect effects decompose the effect of a treatment A on outcome Y into:

- part that is mediated through covariate M (*indirect effect*)
- part that is not (*direct effect*)

Example: How much of the effect of antidepressants on depressive symptoms is mediated through participants' expectations?

Figure 1: DAG summarizing the data



Motivation for mediation analysis

The seminal article on the topic of mediation analysis, Baron and Kenny (1986), has 82.945 citations in Google Scholar.

Mediation analysis is very important in the health sciences, like epidemiology and psychology.

Knowing the type of assumptions under which these analyses are valid is paramount.

From Robins and Richardson (2011):

“The nature of the relationship between the sentence expressing these causal conclusions and the statistical computer calculations performed on the strings of numbers has been obscure.”

Setting and notation

First: randomized treatment A .

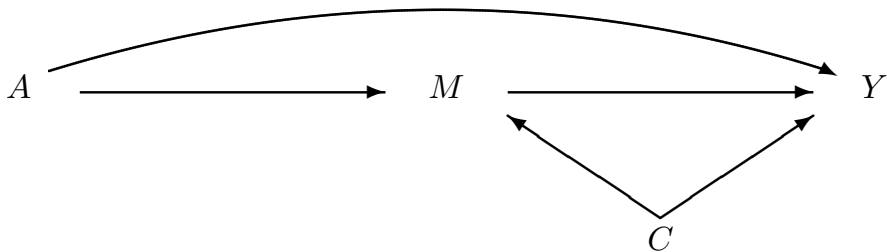
Pre-treatment common causes of mediator M and outcome Y : C .

No post-treatment common causes of mediator M and outcome Y . Can be relaxed; not this talk.

Subscript 0 indicates “without treatment”.

Subscript 1 indicates “under treatment”.

Figure 1: DAG summarizing the data



2 ways through which A affects Y :

- $A \rightarrow M \rightarrow Y$: indirect effect.
- $A \rightarrow Y$: direct effect.

Motivation: definition of causal direct and indirect effects

Causal direct and indirect effects are often defined in terms of the following counterfactuals:

The outcomes under treatment had the mediator been “set” to the mediator without treatment: Y_{1,M_0} .

Two issues with these counterfactual outcomes Y_{1,M_0} :

1. Under treatment, M_0 not observed. $\Rightarrow$ even if we could set the mediator, how to set it to M_0 under treatment ($A = 1$)?

Motivation: definition of causal direct and indirect effects

Causal direct and indirect effects are often defined in terms of the following counterfactuals:

The outcomes under treatment had the mediator been “set” to the mediator without treatment: Y_{1,M_0} .

Two issues with these counterfactual outcomes Y_{1,M_0} :

2. How to set the mediator is usually left unanswered $\Rightarrow$ outcomes undefined in many practical situations.

Cole and Frangakis (2009) provide an illustrative example: “There are many competing ways to assign (hypothetically) a body mass index of 25 kg/m² to an individual, and each of them may have a different causal effect on the outcome”.

Motivation: definition of causal direct and indirect effects

Currently, most causal focus on *natural* direct and indirect effects:

Natural direct effect: $E(Y_{1,M_0} - Y_0)$ (*not mediated*: $M \equiv M_0$).

Natural indirect effect: $E(Y_1 - Y_{1,M_0})$ (*mediated* through M).

See e.g. Pearl (2001), Imai et al. (2010), VanderWeele (2009), Robins and Richardson (2011) and Tchetgen-Tchetgen (2011).

Motivation: identification result for natural direct and indirect effects

Suppose randomized treatment A . Then estimation of EY_1 and EY_0 standard. Left to estimate: $E(Y_{1,M_0})$. Under (strong!) conditions, with C all pre-treatment common causes of mediator M and outcome Y : “**Mediation Formula**”:

$$E(Y_{1,M_0}) = \int_{(m,c)} E[Y|M=m, C=c, A=1] f_{M|C=c,A=0}(m) f_C(c) dm dc.$$

Appealing result. Lok (*Statistics In Medicine* 2016): Strong conditions not needed!

Motivation: identification result for natural direct and indirect effects

Under certain conditions (strong parametric assumptions, linear models, and no exposure-mediator interaction):

Resulting estimators same as Baron and Kenny (1986).

⇒ Causal literature adds a causal interpretation to those estimators.

But most methods need many counterfactual outcomes: not only Y_{1,M_0} , also all $Y_{a,m}$: sometimes not conceivable.

⇒ **Alternative: intervention-based approach.**

Lok (*Statistics In Medicine* 2016).

Here: adaptation.

Interventions on the mediator

I : intervention on the mediator that does not affect pre-treatment common causes C of mediator M and outcome Y .

$M_{0,I=1}$ and $Y_{0,I=1}$: mediator and outcome under “no treatment” *and* under intervention I on the mediator.

“Organic” interventions on the mediator (adapted from Lok (2016))

Definition: (Organic intervention). An intervention I is an organic intervention with respect to C if

$$M_{0,I=1} \mid C = c \sim M_1 \mid C = c \quad (1)$$

and

$$Y_{0,I=1} \mid M_{0,I=1} = m, C = c \sim Y_0 \mid M_0 = m, C = c, \quad (2)$$

where $\sim$ indicates having the same (conditional) distribution.

$$M_{0,I=1} \mid C = c \sim M_1 \mid C = c$$

says that I “puts the mediator at its distribution under treatment”: given C , there is no difference in the distribution of the mediator under intervention I and the distribution of the mediator under treatment.

$$Y_{0,I=1} \mid M_{0,I=1} = m, C = c \sim Y_0 \mid M_0 = m, C = c,$$

says that I “has no direct effect on the outcome”: whether mediator = m came about due to I or without I doesn't affect prognosis. Or, “how mediator came about is irrelevant”.

$$((P(Y_{0,I=1} \leq y \mid M_{0,I=1} = m, C = c) = P(Y_0 \leq y \mid M_0 = m, C = c)$$

for all y .)

$$Y_{0,I=1} \mid M_{0,I=1} = m, C = c \sim Y_0 \mid M_0 = m, C = c,$$

Need all pre-treatment common causes of the mediator M and the outcome Y : C .

Without C : “mediator under no treatment equals m ” likely implies different prognosis:

under intervention I ($M_{0,I=1}$) versus without intervention (M_0).

“Organic” intervention: an example

Example: $A = 1$ blood pressure lowering medicine, M blood pressure, Y occurrence of a heart attack. Does $A = 1$ have a direct effect on heart attacks?

E.g., $A = 1$ lowers blood pressure by 10, on average, without changing the shape of the blood pressure distribution. I intervention that decreases the blood pressure by 10, on average, without changing the shape of the blood pressure distribution. $\Rightarrow M_{0,I=1} \sim M_1|C$.

? I : Salt reduction in dosage dependent on $C = c$?

The effect of salt on heart attacks is believed to be through its effect on blood pressure (see for example the CDC website, <http://www.cdc.gov/vitalsigns/Sodium/index.html>).

⇒ Can hope

$$Y_{0,I=1} \mid M_{0,I=1} = m, C = c \sim Y_0 \mid M_0 = m, C = c$$

for this intervention I .

“Organic” intervention: an example

What does the salt reduction do?

- It doesn't have a direct effect on heart attacks.
- It affects the blood pressure the same way as $A = 1$ does.
- $\Rightarrow$ Its effect should be the effect of $A = 1$ mediated through blood pressure.

“Organic” direct and indirect effects (adapted from Lok (2016)):

$Y_{0,I=1}$: outcome under “no treatment” *and* under organic intervention I on the mediator.

Organic indirect effect of treatment A based on I :

$$EY_{0,I=1} - EY_0$$

((treatment= 0 for both $Y_{0,I=1}$ and Y_0 , direct effect of $I = 1$ is 0, so mediated)).

Organic direct effect of treatment A based on I :

$$EY_1 - EY_{0,I=1}$$

((mediator same distribution for Y_1 and $Y_{0,I=1}$, so not mediated)).

Consistency assumption

Usual Consistency Assumption relating the observed data to the counterfactual data:

Assumption: (*Consistency*). On $A = 0$, $M = M_0$ and $Y = Y_0$.
On $A = 1$, $M = M_1$ and $Y = Y_1$.

Identification result: Mediation Formula for organic direct and indirect effects

Under randomized treatment, estimation of EY_0 and EY_1 standard. For $E(Y_{0,I=1})$, I have proven:

Theorem (*Organic direct and indirect effects: the Mediation Formula for randomized data*). Under randomized treatment and consistency, for an intervention I that is organic with respect to C :

$$\begin{aligned} E(Y_{0,I=1}) \\ = \int_{(m,c)} E[Y|M=m, C=c, A=0] f_{M|C=c,A=1}(m) f_C(c) dm dc. \end{aligned}$$

Note: does not depend on choice of organic intervention I !

Mediation formula for organic direct and indirect effects

Similar identification result as previously found for the natural direct and indirect effects studied by previous authors, of note Pearl (2001), VanderWeele (2009), Imai et al. (2010), and Tchetgen-Tchetgen (2011).

Resulting expression in terms of observable quantities only. All three ingredients can be estimated using standard methods.

Contribution of Lok (2016): definition and therefore the interpretation of direct and indirect effects as well as the conditions under which estimators for these effects are valid can be considerably relaxed.

Organic direct and indirect effects: an intervention based approach

This intervention based approach answers questions about the effect of interventions, and what one might expect from interventions that satisfy certain conditions.

After Lok (2016): Combine organic interventions on the mediator with “no treatment”, that is, with $A = 0$.

Instead of with $A = 1$. <http://arxiv.org/abs/1903.04697>

Important advantages!

- ① Product method from Baron and Kenny (1986) does work in linear models, *regardless* of whether there is treatment-mediator interaction in the outcome model, $E[Y|M, A, C]$.
- ② Organic indirect effects can be estimated with:
 - Ⓐ Distribution of mediator under “treatment”, $A = 1$ and “no treatment”, $A = 0$, and
 - Ⓑ Expectation of the outcome Y given the mediator M and pre-treatment covariates C under “no treatment”, $A = 0$ (*only* under $A = 0$!).

This has important advantages in the pre-clinical stage, for selection of new treatments with promising indirect effects for clinical trials. Illustration: HIV cure.

One could select HIV curative drugs which are at the pre-clinical stage for clinical trials, based on their estimated indirect effect

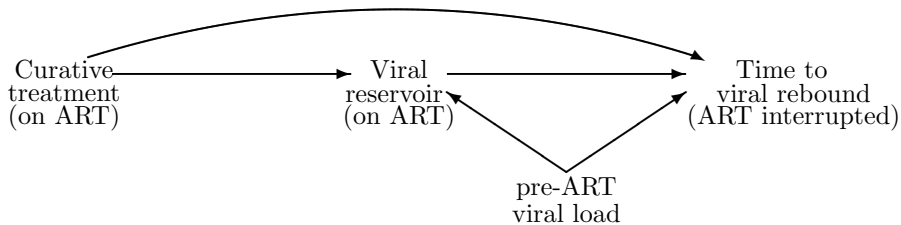
Most HIV-positive patients are on ART (recommended therapy).
HIV viral load often undetectable on ART.

Aim of HIV cure: delay time to HIV viral rebound, the time until the HIV virus is detectable in the blood. Delay indefinitely? Not yet. Delay for some weeks...

HIV viral rebound not observable while on ART.

But: ART treatment interruption carries risks for the patients.
Also expensive: close monitoring of patients off ART.

Figure 2: DAG summarizing the HIV cure setting



What will we obtain? Aim.

The organic *indirect* effect of a curative HIV drug that shifts the distribution of the viral reservoir M by $10\log$ given pre-treatment common causes C , estimated from *existing ART interruption data under no curative intervention* ($A = 0$).

That is, *the effect of a curative HIV drug that*

- shifts the distribution of the viral reservoir M by $10\log$ given C
- and that *has no direct effect* on viral rebound.

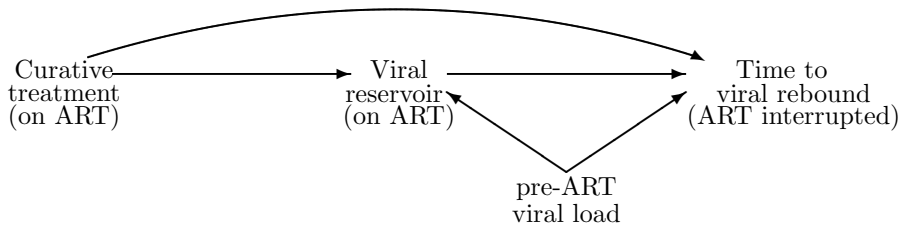
Selecting HIV cure treatments with promising indirect effects for clinical trials (!)

ART treatment interruption has been carried out in a substantial number of patients in the absence of additional curative treatments ($A = 0$).

To estimate the *indirect* effect of the HIV curative treatments: Can combine the effect of HIV curative treatments ($A = 1$) on the on-ART viral reservoir M with ART interruption data under “no curative treatment” ($A = 0$).

Then **select potential HIV curative treatments with most promising indirect effects** for ART interruption trials.

Figure 2: DAG summarizing the HIV cure setting



The organic *indirect* effect of a curative HIV drug which affects a mediator

Outcome Y : suppressed ($Y = 1$)/not suppressed ($Y = 0$) by week 8 of ART interruption.

Two potential on-ART mediators M , both measures of the viral reservoir: cell associated HIV RNA and single-copy plasma HIV RNA, both on the $^{10}\log$ scale.

Pre-ART viral load is a measure of the size of the viral reservoir before ART initiation, and predictive of the time to viral rebound. No information on the pre-ART viral load $\Rightarrow$ the pre-ART NADIR CD4 count is a surrogate, and our pre-treatment common cause of mediator and outcome C .

The organic *indirect* effect of a curative HIV drug which affects a mediator

We estimate the organic indirect effect of a new curative HIV drug A , that is, the effect of a new curative HIV drug without a direct effect, that shifts the distribution of the mediator M by one $^{10}\log$ given C :

$$M_1 \sim M_0 - 1^{10}\log \mid C. \quad (3)$$

Note: this is a shift in the *distribution* of the mediator rather than a shift of the actual values (no “rank preservation”).

Ron Bosch have requested the data from Li et al. (2016): data on > 100 HIV infected patients without curative interventions and with ART interruptions. Results to follow, but here is our plan:

The organic *indirect* effect of a curative HIV drug which affects a mediator

The organic *indirect* effect of a curative HIV drug that shifts the distribution of the mediator by $1^{10}\log$ given C equals

$$\begin{aligned} & \int_{(m,c)} P(Y = 1 | M = m, C = c, A = 0) f_{M|C=c,A=1}(m) f_C(c) dm dc \\ & \quad - P(Y_0 = 1) \\ & = \int_{(m,c)} P(Y = 1 | M = \tilde{m}(m), C = c, A = 0) f_{M|C=c,A=0}(m) f_C(c) dm \\ & \quad - P(Y_0 = 1), \end{aligned}$$

where $\tilde{m}(m) = m - 1^{10}\log$.

((if $m - 1^{10}\log$ is above the limit of detection; $\tilde{m}(m)$ is “below the limit of detection” otherwise.))

The organic *indirect* effect of a curative HIV drug which affects a mediator

We will fit the following logistic regression model for the probability of virologic suppression at week 8 of ART interruption, Y , given M and C , using the requested data (y_i, m_i, c_i) , $i = 1, \dots, n$, without HIV curative drug ($A = 0$):

$$\begin{aligned} \text{logit} p_{\beta; m, c} &= \text{logit} P(Y = 1 | M = m, C = c, A = 0) \\ &= \begin{cases} \beta_0 + \beta_1 m + \beta_2 c & \text{if } m \text{ above the limit of detection} \\ \beta_3 + \beta_2 c & \text{if } m \text{ below the limit of detection.} \end{cases} \end{aligned}$$

The organic *indirect* effect of a curative HIV drug which affects a mediator

⇒ The organic *indirect* effect of a curative HIV drug that shifts the distribution of the mediator by $1^{10}\log$ given C equals

$$\int_{(m,c)} P(Y = 1 | M = \tilde{m}(m), C = c, A = 0) f_{M|C=c,A=0}(m) f_C(c) dm dc$$

minus $P(Y_0 = 1)$.

⇒ Estimate the organic *indirect* effect of a curative HIV drug that shifts the distribution of the mediator by $1^{10}\log$ given C by:

$$\sum_{i=1}^n (p_{\hat{\beta}; \tilde{m}(m_i), c_i} - y_i) / n,$$

with $(y_i, m_i, c_i)_{i=1}^n$ the ART interruption data under no curative HIV drug.

What did we obtain?

The organic *indirect* effect of a curative HIV drug that shifts the distribution of the viral reservoir by $1^{10}\log$ given pre-treatment common causes C , estimated from *existing ART interruption data under no curative intervention* ($A = 0$).

That is, *the effect of a curative HIV drug that*

- shifts the distribution of the viral reservoir by $1^{10}\log$ given C
- and that *has no direct effect* on viral rebound.

Organic direct and indirect effects: benefits of alternative definition

Can be shown: under linear models and in the absence of treatment-mediator interaction in the outcome model, new definition leads to the same results as before.

Can be shown: with new definition, product method from Baron and Kenny (1986) does work in linear models regardless of whether there is treatment-mediator interaction in the outcome model.

Conclusions

Both for natural direct and indirect effects and for Didelez et al. (2006): need to be able to **set** mediator to any specific value.

“Organic” direct and indirect effects: need to be able to affect the **distribution** of the mediator.

Organic direct and indirect effects generalize natural direct and indirect effects.

Conclusions

To answer clinical questions, often useful to combine interventions on the mediator with “no treatment”, rather than with “treatment”.

These newly defined organic indirect effects can be estimated with information on

- 1 Distribution of mediator under treatment, $A = 1$, and “no treatment”, $A = 0$.
- 2 Expectation of the outcome given the mediator and pre-treatment covariates C under $A = 0$ (only under $A = 0$!).

Important advantages in the pre-clinical stage, for selection of new treatments with promising indirect effects for clinical trials.

Closer to Baron and Kenny (1986).

Thanks for listening!

Happy Birthday Aad!

Acknowledgements organic direct and indirect effects

This research was supported by the Division of AIDS of the National Institutes of Health R01AI100762 and the National Science Foundation DMS 1810837. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF or NIH.

We are extremely grateful to the HIV positive participants who volunteered for the ACTG ART interruption trials. The authors also thank Dr. Li (Brigham and Women Hospital, Boston) for his support, for leading the treatment interruption project and generating the mediator data to be used in our analysis.

- Baron, R. M. and D. A. Kenny (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51, 1173–1182.
- Cole, S. R. and C. E. Frangakis (2009). The consistency statement in causal inference: a definition or an assumption? *Epidemiology* 20(1), 3–5.
- DeGruttola, V. G., P. Clax, D. L. DeMets, G. J. Downing, S. S. Ellenberg, L. Friedman, M. H. Gail, R. Prentice, J. Wittes, and S. L. Zeger (2001). Considerations in the Evaluation of Surrogate Endpoints in Clinical Trials: Summary of a National Institutes of Health Workshop. *Controlled Clinical Trials* 22, 485–502.

- Didelez, V., A. P. Dawid, and S. Geneletti (2006). Direct and indirect effects of sequential treatments. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, pp. 138–146. Arlington, VA: AUAI Press: arxiv preprint arXiv 1206.6840.
- Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 25, 51–71.
- Li, J. Z., B. Etemad, H. Ahmed, E. Aga, R. J. Bosch, J. W. Mellors, D. R. Kuritzkes, M. M. Lederman, M. Para, and R. T. Gandhi (2016). The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. *AIDS* 30(3), 343–353.
- Lok, J. J. (2016). Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. *Statistics in Medicine* 35(22), 4008–4020.

- Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the 17th annual conference on uncertainty in artificial intelligence (UAI-01)*, pp. 411–442. San Francisco: Morgan Kaufmann.
- Robins, J. M. and T. S. Richardson (2011). Alternative Graphical Causal Models and the Identification of Direct Effects. In P. Shrout, K. Keyes, and K. Ornstein (Eds.), *Causality and psychopathology: finding the determinants of disorders and their cures*, Chapter 6, pp. 1–52. Oxford University Press.
- Sperling, R. S., D. E. Shapiro, R. W. Coombs, J. A. Todd, S. A. Herman, G. D. McSherry, M. J. O'Sullivan, R. B. VanDyke, E. Jiminez, C. Rouzioux, P. M. Flynn, and J. L. Sullivan (1996). Maternal viral load, Zidovudine treatment, and the risk of transmission of Human Immunodeficiency Virus type 1 from mother to infant. *New England Journal of Medicine* 335(22), 1621–1629.

- Tchetgen-Tchetgen, E. J. (2011). On causal mediation analysis with a survival outcome. *The International Journal of Biostatistics* 7(1), 1–38.
- Valeri, L. and T. J. VanderWeele (2013). Mediation analysis allowing for exposure-mediator interaction and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods* 18(2), 137–150.
- VanderWeele, T. J. (2009). Marginal Structural Models for the estimation of direct and indirect effects. *Epidemiology* 20(1), 18–26.

The Mediation Formula for organic direct and indirect effect: proof

Proof:

$$\begin{aligned} E(Y_{0,l=1}) &= E(E[Y_{0,l=1}|M_{0,l=1}, C]) \\ &= \int_{(m,c)} E[Y_{0,l=1}|M_{0,l=1} = m, C = c] f_{M_{0,l=1}|C=c}(m) dm f_C(c) dc \end{aligned}$$

because of the definition of conditional expectations.

$$= \int_{(m,c)} E[Y_0|M_0 = m, C = c] f_{M_1|C=c}(m) dm f_C(c) dc$$

because of the definition of organic intervention.

Product method for new organic indirect effect

Assumption:

$$E[Y|M = m, A = 0, C = c] = \beta_0 + \beta_1 c + \beta_2 m.$$

Note: $E[Y|M = m, A = a, C = c]$ can have an interaction term $\beta_3 am$.

Assumption:

$$M_1 \sim M_0 + \alpha_1 \mid C,$$

i.e.

$$f_{M_1|C=c}(m) = f_{M_0|C=c}(m - \alpha_1).$$

$$\begin{aligned}
& E(Y_{0,l=1}) - EY_0 \\
&= \int_{(m,c)} E[Y|M=m, A=0, C=c] \\
&\quad (f_{M|A=1,C=c}(m) - f_{M|A=0,C=c}(m)) f_C(c) dm dc \\
&= \int_{(m,c)} (\beta_0 + \beta_1 c + \beta_2 m) (f_{M|A=1,C=c}(m) - f_{M|A=0,C=c}(m)) \\
&\quad f_C(c) dm dc \\
&= \beta_2 \int_{(m,c)} m (f_{M|A=1,C=c}(m) - f_{M|A=0,C=c}(m)) f_C(c) dm dc \\
&= \beta_2 \int_{(m,c)} m f_{M|A=0,C=c}(m - \alpha_1) f_C(c) dm dc \\
&\quad - \beta_2 \int_{(m,c)} m f_{M|A=0,C=c}(m) f_C(c) dm dc \\
&= \beta_2 \left(\int_{(\tilde{m},c)} (\tilde{m} + \alpha_1) f_{M|C=c,A=0}(\tilde{m}) f_C(c) d\tilde{m} dc - \dots \right) \\
&= \beta_2 \alpha_1.
\end{aligned}$$

Uniqueness of organic direct and indirect effect

Next, I show: organic direct and indirect effect do not depend on C if C “has all common causes of mediator M and outcome Y ”.

Uniqueness of organic direct and indirect effect: common cause

Definition: (*common cause*). X is *not* a common cause of mediator and outcome given C if either equation (4) or equation (5) holds:

$$X \perp\!\!\!\perp M_0 \mid C \qquad \text{and} \qquad X \perp\!\!\!\perp M_1 \mid C \qquad (4)$$

or

$$X \perp\!\!\!\perp Y_0 \mid M_0, C. \qquad (5)$$

In graphical language: X is a common cause of mediator and outcome if in a DAG that has C , X , M , and Y , there is an arrow from X to M , and there is a direct arrow from X to Y . This definition is in line with, for example, Pearl (2000).

Uniqueness of organic direct and indirect effects

Let I^C be an intervention that is organic with respect to C and $I^{\tilde{C}}$ and intervention that is organic with respect to $\tilde{C}$. Assume that C is not a common cause of mediator and outcome given $\tilde{C}$, and $\tilde{C}$ is not a common cause of mediator and outcome given C ; hence there are 4 different cases, with either (4) or (5) holding for C and $\tilde{C}$, respectively. One can show that under any of the 4 different cases,

$$E(Y_{0,I^{\tilde{C}}=1}) = \int_{(\tilde{c},m,c)} E[Y_0 \mid M_0 = m, \tilde{C} = \tilde{c}, C = c] f_{M_1|\tilde{C}=\tilde{c},C=c}(m) f_{\tilde{C}}$$

Because of symmetry, it follows that also

$$E(Y_{0,I^C=1}) = \int_{(\tilde{c},m,c)} E[Y_0 \mid M_0 = m, \tilde{C} = \tilde{c}, C = c] f_{M_1|\tilde{C}=\tilde{c},C=c}(m) f_{\tilde{C}},$$

But then, $E(Y_{0,I^{\tilde{C}}=1}) = E(Y_{0,I^C=1})$.

From “natural” conditions to “organic” conditions

If can envision all $Y_{a,m}$: my interpretation of cross-worlds assumption of natural direct and indirect effects: nature did not have more information on the potential outcomes $Y_{a',m}$ to determine the value of the mediator M_a than recorded in C .

Needs that all common causes of mediator and outcome are recorded in the database. Seems strong condition, but usual. “Organic” conditions assume this as well.

Comparison with Lok (2016): how to define organic direct and indirect effects

In Lok (2016), I combined organic interventions on the mediator with “treatment”, that is, with $A = 1$ (instead of with $A = 0$).

Followed analogies with natural direct and indirect effect.

Product method from Baron and Kenny (1986) didn't work when there is treatment-mediator interaction in the outcome model.

Product method from Baron and Kenny (1986) did work in linear models when there is no treatment-mediator interaction in the outcome model.

Organic versus natural direct and indirect effects

Organic direct and indirect effects from Lok (2016)

generalize

Natural direct and indirect effects.

Provided that $M_{1,I=1} = M_0$ exists, the usual “cross-worlds” assumption implies that intervention I that sets the mediator to M_0 , $M_{1,I=1} = M_0$, is organic!

Organic direct and indirect effects generalize natural direct and indirect effects

Theorem: If $M_{1,l=1} = M_0$ exists, usual identifiability assumption for natural direct and indirect effects (cross-worlds assumption):

$$Y_{a',m} \perp\!\!\!\perp M_a | C = c, A = a, \quad (*)$$

implies that $M_{1,l=1} = M_0$ is organic intervention.

Proof : $?Y_{1,l=1} | M_{1,l=1} = m, C = c \sim Y_1 | M_1 = m, C = c?$

$$\Rightarrow ?Y_{1,m} | M_0 = m, C = c \sim Y_{1,m} | M_1 = m, C = c? \quad (6)$$

And yes: under randomization, $A = a$ can be left out of the conditioning event of (*), which hence implies (6): given C , $Y_{1,m}$ depends neither on M_0 nor on M_1 .

A weaker identifiability condition for natural direct and indirect effects

For natural direct and indirect effect, consider intervention I that results in $M_{1,I=1} = M_0$ (“set the mediator to its value without treatment”). If all $Y_{a,m}$ “exist”, this I is an organic intervention if:

$$Y_{1,m}|M_0 = m, C = c \quad \sim \quad Y_{1,m}|M_1 = m, C = c.$$

((Follows e.g. if (*) holds)). Then, natural and organic direct and indirect effects are the same.

Still cross-worlds assumption. No surprise: definition cross-worlds.

Relaxes (*) though.

Comparison with Didelez et al. (2006): randomize mediator and then set

Didelez et al. (2006) propose to let intervention I be: take a random draw from M_0 given C , then set the mediator to this random draw. Such intervention I is organic if

$$Y_{1,I=1} | M_{1,I=1} = m, C = c \quad \sim \quad Y_1 | M_1 = m, C = c.$$

Can be tested: not cross-worlds. E.g. (*) also implies that this $M_{1,I=1}$ is organic. Then, Didelez et al. (2006) direct and indirect effects same as organic direct and indirect effects, provided one can set the mediator to a random draw of M_0 and treat.

Observational data

Same definition, similar identification result provided “ C has all common causes of mediator and outcome”.

There may exist baseline covariates Z (beyond the common causes of mediator and outcome, C) that need to be included in the analysis in order to eliminate confounding:

Assumption: (*No Unmeasured Confounding*).

$$A \perp\!\!\!\perp (Y_0, M_0) \mid C, Z \quad \text{and} \quad A \perp\!\!\!\perp Y_1 \mid C, Z \quad \text{and} \quad A \perp\!\!\!\perp M_1 \mid C, Z.$$

Observational data: identification result

Theorem: (*Organic direct and indirect effects: the Mediation Formula for observational data*). Assume No Unmeasured Confounding, Consistency, intervention I is organic with respect to C , and given C , Z is not a common cause of mediator and outcome. Then

$$\begin{aligned} E(Y_{0,I=1}) &= \int_{(m,c,z)} E[Y|M=m, C=c, Z=z, A=0] \\ &\quad f_{M|C=c,Z=z,A=1}(m) f_{C,Z}(c, z) dm d(c, z). \end{aligned}$$

The organic *indirect* effect of a curative HIV drug which affects a mediator

$$\begin{aligned}
 & \int_{(m,c)} P(Y = 1 | M = m, C = c, A = 0) f_{M|C=c,A=1}(m) f_C(c) dm dc \\
 &= \int_{(m,c)} P(Y = 1 | M = m, C = c, A = 0) f_{M|C=c,A=0}(m + 1^{10}\log) \\
 &\quad f_C(c) dm dc \\
 &= \int_{(m,c)} P(Y = 1 | M = \tilde{m} - 1^{10}\log, C = c, A = 0) f_{M|C=c,A=0}(\tilde{m}) \\
 &\quad f_C(c) d\tilde{m} dc.
 \end{aligned}$$

In the last line we did a change of variables: $\tilde{m} = m + 1^{10}\log$.

The organic *indirect* effect of a curative HIV drug which affects a mediator

Caveat: we have assumed that it is irrelevant how far below the limit of detection the on-ART viral reservoir is. Measurement of the “true” on-ART viral reservoir would be better. Not possible now; some day?

Mother-to-child transmission of HIV/AIDS

HIV-positive mothers can transmit the HIV virus to their infants.

Effect of AZT treatment on mother-to-child transmission of HIV-1 is surprisingly large, given the limited effect of AZT mono-therapy on HIV-1 RNA (DeGruttola et al. (2001)).

Less than 20% of the effect of AZT on mother to child transmission can be explained through the effect of AZT on HIV-1 RNA (Sperling et al. (1996)).

Mother-to-child transmission of HIV/AIDS

???Likely effect on mother-to-child transmission of a potential new treatment that has the same effect on HIV-1 RNA as AZT but no “direct” effect on mother to child transmission???

Outcome Y : indicator “newborn baby is HIV-positive”.

Mediator M : HIV-1 RNA.

I' : intervention that, without AZT, causes the distribution of HIV-1 RNA, $M_{0,I'=1}$, to be the same as under AZT; the potential new treatment.

Quantity of interest: $EY_{0,I'=1} - EY_0$.

Mother-to-child transmission of HIV/AIDS

Quantity of interest: $EY_{0,I'=1} - EY_0$.

Note: quantity different from indirect effect defined in Lok (2016) and in most of the literature on direct and indirect effects:

$EY_1 - EY_{1,I=1}$, with I intervention on the mediator under treatment. $EY_{0,I'=1} - EY_0$, with I' intervention on the mediator without treatment, more relevant here. See

<http://arxiv.org/abs/1903.04697>

Mother-to-child transmission of HIV/AIDS

HIV viral load in the mother's blood cannot be “set”.

Once we will be able to set it, we will set it to 0.

⇒ For a treatment like AZT, organic direct and indirect effects are more natural than their natural counterparts.

Mother-to-child transmission of HIV/AIDS

Combining organic interventions with “no treatment” provides useful information on what to expect from a treatment that:

- ① Affects the HIV viral load in the mother's blood the same way as AZT does.
- ② Has no direct effect on mother-to-child transmission.