

# Mediation analysis first step: Defining effects based on what we want to learn

Trang Quynh Nguyen (with Ian Schmid, Elizabeth A. Stuart)

Johns Hopkins Bloomberg School of Public Health

[arxiv.org/abs/1904.08515](https://arxiv.org/abs/1904.08515) (Psych Methods) | [trang.nguyen@jhu.edu](mailto:trang.nguyen@jhu.edu)

SREE Webinar – 2020-07-15

# Synopsis

- ▶ Original desire: understand mechanisms of effect of  $A$  on  $Y$ 
  - ▶ effect through a causal pathway via an intermediate variable  $M$
  - ▶ total effect = direct + indirect components
- ▶ With this desire
  - ▶ Effect were traditionally model-centric, eg indirect effect =  $ab$ , where  $a, b$  are two regression coefs
  - ▶ Causal inference revised these effects using potential outcomes, freeing them from the models – *natural (in)direct effects*
- ▶ Causal inference brings in the idea of sequential intervention
  - ▶ Another genre of effects – *interventional effects*
  - ▶ Fit a different desire: effects of hypothetical conditions – in intervention research, disparity research
- ▶ Our proposal: carefully choose the target effect (*estimand*) **based on what we want to learn**

# The estimand should drive the analysis

- ▶ *define*: define the target estimand – what we want to learn
- ▶ *identify*: assess its identifiability – given study design, assumptions
- ▶ *estimate*: estimate or test it – using statistical methods

Clarity on the estimand leads to clarity in interpreting analysis results

# Effect definitions ← research questions

Many effects and effect types

Which one best matches my research question?

May require clarifying vague research questions

# If the research question is about explaining the causal effect of exposure on outcome

eg

- ▶ what are the mechanisms of this effect?
- ▶ what part of this effect is due to the exposure's influence on this intermediate variable and what part is not?
- ▶ is the effect partly due to the exposure's influence on this intermediate variable?

# If the research question is about explaining the causal effect of exposure on outcome

then the closest estimands are *natural (in)direct effects*

- ▶ they decompose the total effect
- ▶ a NIE can be interpreted as an effect on the outcome *of the exposure's effect on the mediator*

decompositions are not unique

# Notation and consistency

$A$  .....  $M$  .....  $Y$

Observed variables:  $A$  binary exposure (0/1)  
 $M$  mediator  
 $Y$  outcome

Potential variables:  $M_a$   $a = 0, 1$   
 $Y_a$   
 $Y_{am}$   $m$  is a mediator value  
 $Y_{aM_{a'}}$

Consistency assumptions: if  $A = a$   $M = M_a$   
(connecting potential and  $Y = Y_a = Y_{aM} = Y_{aM_a}$   
observed variables) if  $A = a, M = m$   $Y = Y_a = Y_{aM} = Y_{am}$   
if  $M_{a'} = m$   $Y_{aM_{a'}} = Y_{am}$

## Natural (in)direct effects

Defined at individual level, decompose individual total effect

$$\begin{aligned} TE &= Y_1 - Y_0 \\ &= Y_{1M_1} - Y_{0M_0} \end{aligned}$$



# Natural (in)direct effects

Defined at individual level, decompose individual total effect

$$\begin{aligned} TE &= Y_1 - Y_0 \\ &= Y_{1M_1} - Y_{0M_0} \end{aligned}$$

2 decompositions

▶ direct-indirect:  $TE = \underbrace{Y_{1M_1} - Y_{1M_0}}_{NIE_1} + \underbrace{Y_{1M_0} - Y_{0M_0}}_{NDE_0}$

▶ indirect-direct:  $TE = \underbrace{Y_{1M_1} - Y_{0M_1}}_{NDE_1} + \underbrace{Y_{0M_1} - Y_{0M_0}}_{NIE_0}$

NIE = an effect on the outcome *of the exposure's effect on the mediator*

NDE = an effect of the exposure when holding the mediator at a natural value

# Natural (in)direct effects

Target average effects (individual effects not identified and not of interest)

▶ direct-indirect: 
$$TE = \underbrace{E[Y_1] - E[Y_{1M_0}]}_{NIE_1} + \underbrace{E[Y_{1M_0}] - E[Y_0]}_{NDE_0}$$

▶ indirect-direct: 
$$TE = \underbrace{E[Y_1] - E[Y_{0M_1}]}_{NDE_1} + \underbrace{E[Y_{0M_1}] - E[Y_0]}_{NIE_0}$$

These definitions are model free

## Natural (in)direct effects

Target average effects (individual effects not identified and not of interest)

▶ direct-indirect: 
$$TE = \underbrace{E[Y_1] - E[Y_{1M_0}]}_{NIE_1} + \underbrace{E[Y_{1M_0}] - E[Y_0]}_{NDE_0}$$

▶ indirect-direct: 
$$TE = \underbrace{E[Y_1] - E[Y_{0M_1}]}_{NDE_1} + \underbrace{E[Y_{0M_1}] - E[Y_0]}_{NIE_0}$$

These definitions are model free

Which decomposition to use? – discussion in paper

# Natural (in)direct effects

Target average effects (individual effects not identified and not of interest)

▶ direct-indirect: 
$$TE = \underbrace{E[Y_1] - E[Y_{1M_0}]}_{NIE_1} + \underbrace{E[Y_{1M_0}] - E[Y_0]}_{NDE_0}$$

▶ indirect-direct: 
$$TE = \underbrace{E[Y_1] - E[Y_{0M_1}]}_{NDE_1} + \underbrace{E[Y_{0M_1}] - E[Y_0]}_{NIE_0}$$

These definitions are model free

Which decomposition to use? – discussion in paper

Not identified if exist mediator-outcome confounders influenced by exposure

now another effect type for another question type

# If the research question is a *what-if* question

eg

- ▶ in intervention development research: what if the program is modified
  - ▶ removing elements that affect the mediator
  - ▶ retaining only elements that affect the mediator
  - ▶ some other way
- ▶ in disparities research: what if could shift the distribution of a factor that contributes to disparity

then want to consider the class of *interventional effects*

# Interventional effects

Lage class, incl. total effect, controlled direct effect, generalized direct effects, interventional (in)direct effects, many other effects, NOT natural (in)direct effects

An effect in this class contrasts

- ▶ a (hypothetical) active intervention condition
- ▶ a comparison intervention (or no intervention) condition

An (hypothetical) intervention condition

- ▶ sets exposure and/or mediator each to a specific value or distribution that is known or is identified (based on data observed in current study)
- ▶ does not change anything else

# Selecting an interventional effect

2 key questions:

- ▶ Which condition best matches the *what-if* condition of scientific interest?
- ▶ What is the most appropriate comparison condition?

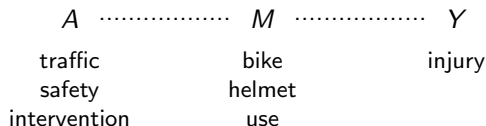
Note that an interventional effect

- ▶ generally does not tell us exactly about a *realistic* intervention  
BUT
- ▶ does tell us about an *ideal* intervention
- ▶ our job to judge how rough or fine the approximation is



## Some examples

## Controlled and generalized direct effects



In the context of new law requiring helmet use

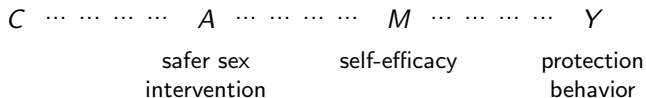
assuming 100% compliance, the effect of the intervention in the new context is a controlled direct effect:

$$\text{CDE}(100) = E[Y(\mathbf{1}, 100)] - E[Y(\mathbf{0}, 100)]$$

assuming compliance about  $75\% \pm 15\%$ , and representing this distribution by  $\mathcal{M}$ , the intervention's effect in the new context is a generalized direct effect:

$$\text{GDE}(\mathcal{M}) = E[Y(\mathbf{1}, \mathcal{M})] - E[Y(\mathbf{0}, \mathcal{M})]$$

## Effect of intervention if modified to remove indirect effect elements



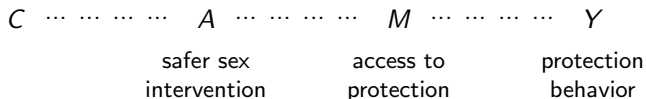
$$E[Y(\mathbf{1}, \mathcal{M}(0 | C))] - E[Y(0)]$$

The active intervention condition here sets the exposure to 1, but sets the mediator to the distribution of  $M(0)$  (conditional on pre-exposure covariates)

Note this is different from setting the mediator to  $M(0)$

The squiggly  $\mathcal{M}$  indicates the randomness of the mediator values assigned

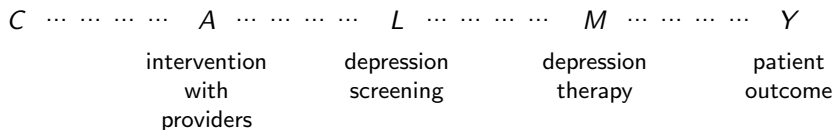
## Effect of intervention if modified to remove direct effect elements



$$E[Y(0, \mathcal{M}(1 | C))] - E[Y(0)]$$

The active intervention condition here sets the exposure to 0, but sets the mediator to the distribution of  $M(1)$  (conditional on pre-exposure covariates)

## Effect of alternative intervention that affects treatment but not screening for depression



$$E[Y(0, L(0), \mathcal{M}(1, L(0) | C))] - E[Y(0)]$$

Here the notation  $\mathcal{M}(1, L(0) | C)$  means the distribution of the mediator had  $A$  been set to 1 and  $L$  been set to the value of  $L(0)$

## Interventional (*in*)direct effects

Well-known cousins of natural effects. Also called stochastic (*in*)direct effects

Arguably not as relevant as some of the effects mentioned earlier

$$\text{IDE}(\cdot 0) = E[Y(\mathbf{1}, \mathcal{M}(0|C))] - E[Y(\mathbf{0}, \mathcal{M}(0|C))]$$

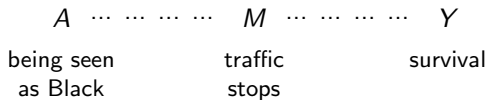
$$\text{IDE}(\cdot 1) = E[Y(\mathbf{1}, \mathcal{M}(1|C))] - E[Y(\mathbf{0}, \mathcal{M}(1|C))]$$

$$\text{IIE}(0 \cdot) = E[Y(\mathbf{0}, \mathcal{M}(\mathbf{1}|C))] - E[Y(\mathbf{0}, \mathcal{M}(\mathbf{0}|C))]$$

$$\text{IIE}(1 \cdot) = E[Y(\mathbf{1}, \mathcal{M}(\mathbf{1}|C))] - E[Y(\mathbf{1}, \mathcal{M}(\mathbf{0}|C))]$$

In special case with no intermediate confounders, equal to natural (*in*)direct effects

What if could reduce the frequency of traffic stops of Black folks down to half-way between their actual experience and that of non-Black folks



$$E[Y(\mathbf{1}, \mathcal{M}(0.5|C)) | A = 1] - E[Y(\mathbf{1}) | A = 1]$$

$\mathcal{M}(0.5|C)$  is a half-half mixture of two distributions

# To sum up

Wide range of effect definitions

- ▶ natural (in)direct effects
- ▶ very broad class of interventional effects

Flexibility in selecting/defining effects to match research questions