Causal mediation analysis: Effect definitions
(what is it we are trying to learn?)

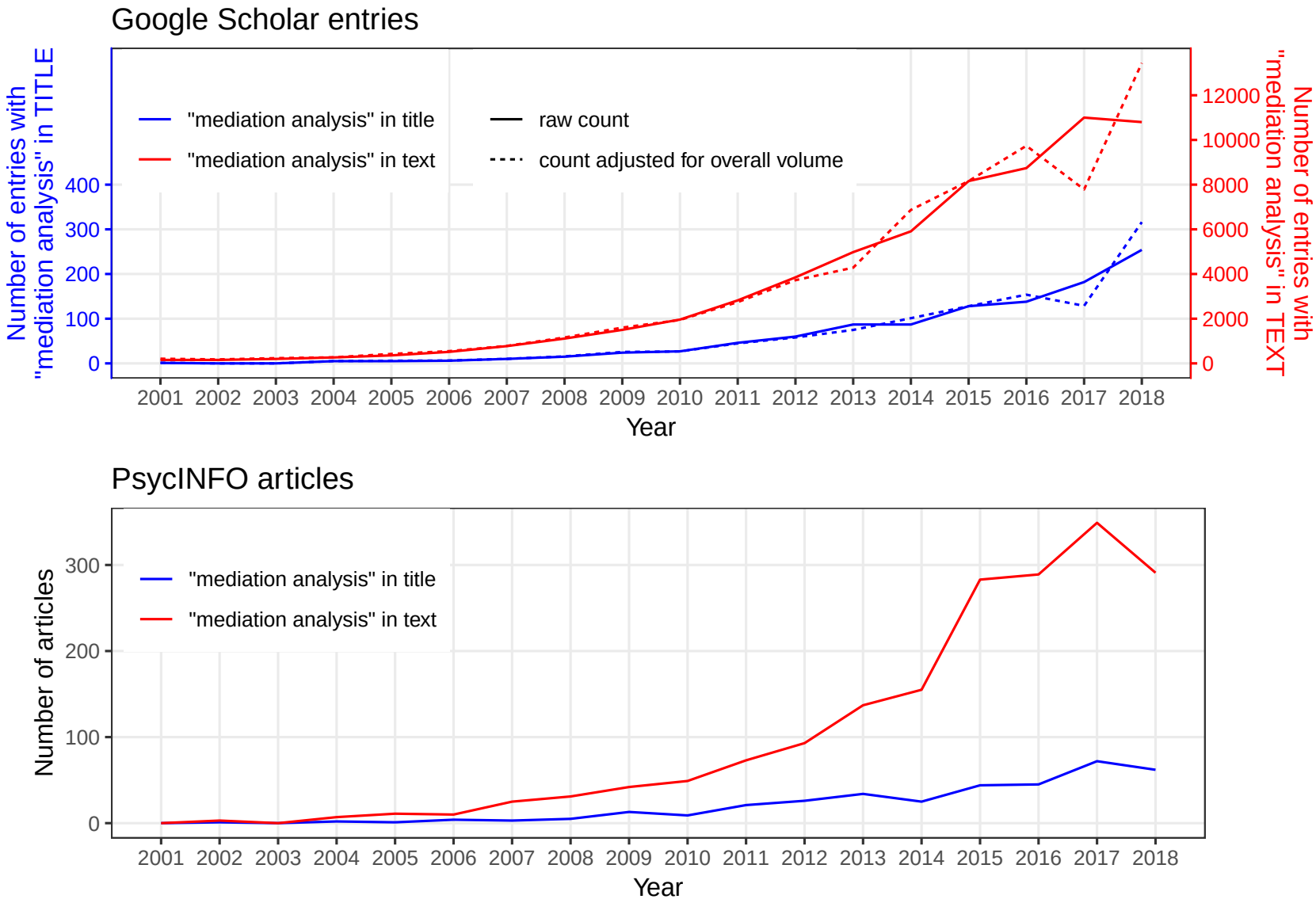
Trang Quynh Nguyen
tnguye28@jhu.edu | trang-q-nguyen.weebly.com

Johns Hopkins Bloomberg School of Public Health

with Training Interdisciplinary Educational Scientists Program
Penn State University, 2019/04/17

(Causal) Mediation Analysis

Trends


Google Scholar entries
Number of entries with "mediation analysis" in TITLE
Number of entries with "mediation analysis" in TEXT
"mediation analysis" in title
"mediation analysis" in text
raw count
count adjusted for overall volume
0
100
200
300
400
0
2000
4000
6000
8000
10000
12000
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
Year
PsycINFO articles
Number of articles
"mediation analysis" in title
"mediation analysis" in text
0
100
200
300
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
Year

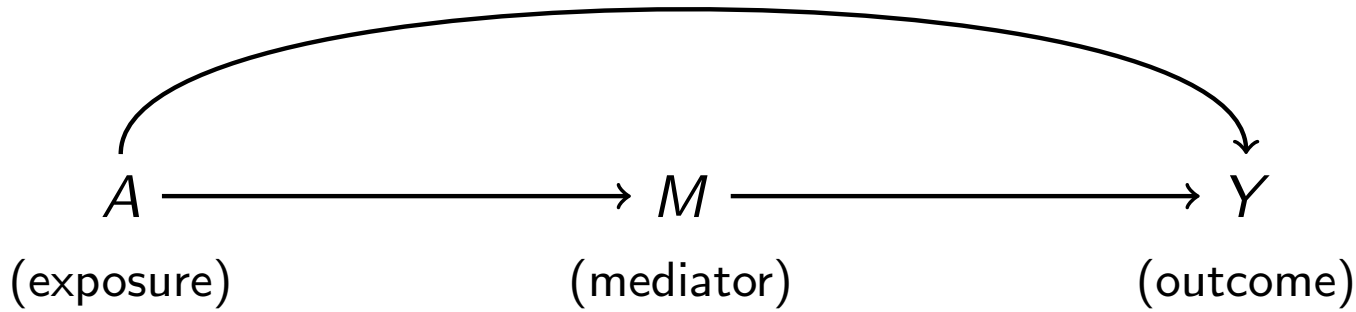
What is mediation analysis?

Have you done one? or seen one?

What method did you use? or did the authors use?

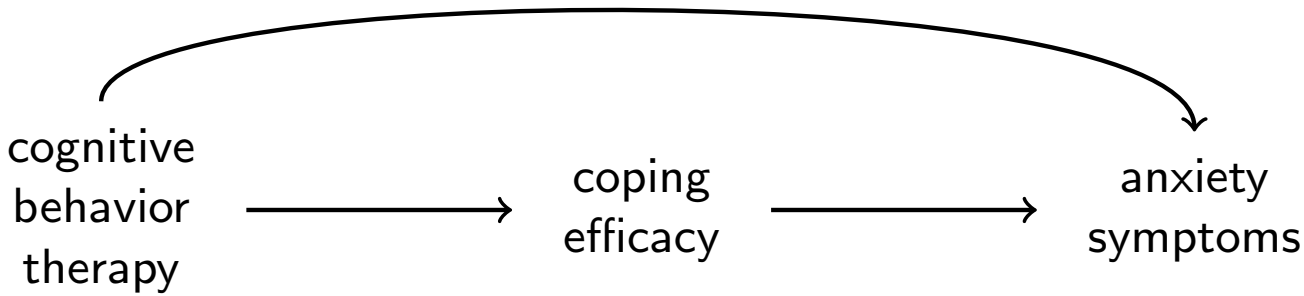
What type of effect did you estimate? or did they estimate?

Typical setting

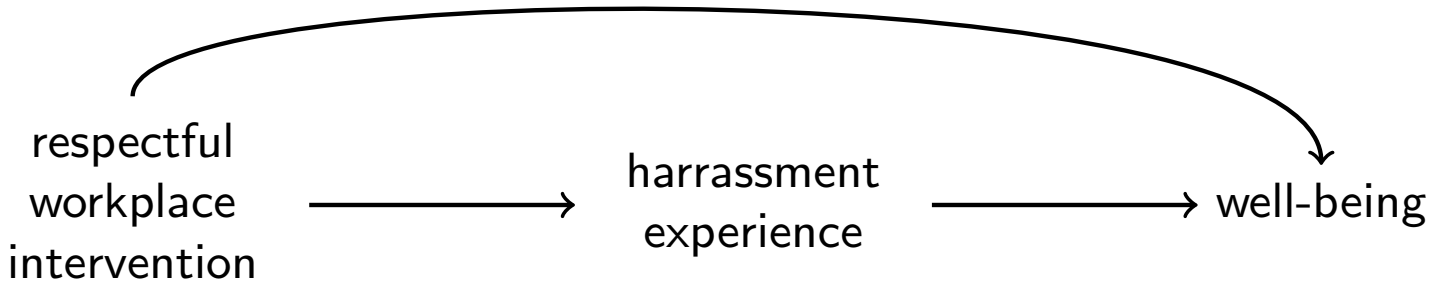


interested in effects of A on Y $\underbrace{\text{through } M}_{\text{(indirect effect)}}$ and $\underbrace{\text{not through } M}_{\text{(direct effect)}}$

Example


cognitive
behavior
therapy
coping
efficacy
anxiety
symptoms

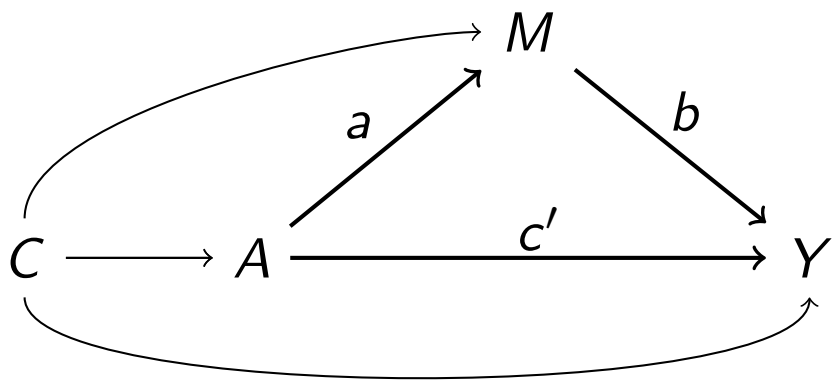
Example


respectful
workplace
intervention
harrassment
experience
well-being

History part 1: traditional mediation analysis

- Long history
  - roots in path analysis [**?**]
  - Baron & Kenny [1986] – linear models (>80K Google cites)
  - more complex situations – see [**?**, **?**]

- Key features
  - indirect effect = product of regression coefficients
  - no separation of effect definition and effect estimation



a: coef in model $M \sim A + C$
b, c': coefs in model $Y \sim A + M + C$

indirect effect $:= ab$
direct effect $:= c'$

History part 2: causal mediation analysis

- More recent
  - Robins & Greenland [1992] first definition
  - Pearl [2001] identification assumptions
  - explosion of theoretical and methodological work since then – 50 articles in 2015-2016 alone

- Key features
  - (in)direct effect = contrast of potential outcomes (or intervention regimes)
  - separates effect definition, identification, estimation
  - several types of effects: controlled, natural, interventional

Mediation analysis is, unavoidably, about causal effects

- The arrows we draw imply causal relationships
- Results are generally interpreted by research consumers in causal terms
  - indirect effect = influence of A on Y through M
  - direct effect = influence of A on Y not through M
- There is no convenient *association* (as opposed to *causal*) interpretation of an indirect effect

Current practice snapshot

A review of articles with mediation analysis published in 2013-2018 in ten top psychology and ten top psychiatry journals

- Use of causal mediation analysis methods?
  - less than 4%

Current practice snapshot

A review of articles with mediation analysis published in 2013-2018 in ten top psychology and ten top psychiatry journals

- Use of causal mediation analysis methods?
  - less than 4%
- Other elements of causal reasoning?
  - less than 20% have full A-M-Y temporal ordering
  - less than half adjust for covariates to control confounding

We need to do better!

Use explicit causal thinking

- Do we have A-M-Y temporal ordering?
  - in addition to: does theory suggest that M influences Y?
  - also ask: is it possible for M to influence Y in this study?
- What are the sources of bias?
  - 3 relationships that can be confounded: A-M, A-Y, M-Y
  - does study design take care of any of these?
  - what data did (can) we collect to help control confounding?

AND adopt the causal inference approach

We need to do better!

Use explicit causal thinking

AND adopt the causal inference approach

- Step 1 – effect definition (what do we want to learn?)
  - select effect that matches research question
- Step 2 – effect identification (can we learn it from data?)
  - what assumptions are required?
  - are they plausible in this study?
- Step 3 – effect estimation (how can we learn it?)
  - what strategies (eg regression, weighting, simulation)?

Now we will

Take simple setting: A (binary), M, Y

- focus on step 1 – effect definition

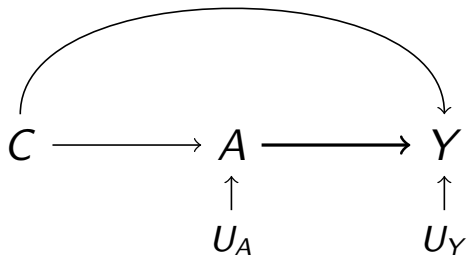
Different types

- total effect
- natural (in)direct effects
- interventional (in)direct effects
- interventional effects more generally
- controlled direct effects

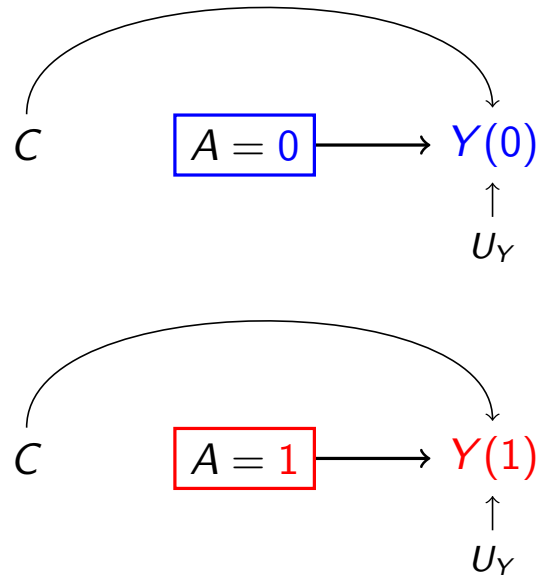
Total Effect

Total effect – a review

The regular world
without any manipulation



The two worlds contrasted
by the total effect

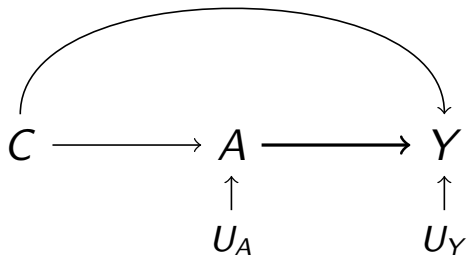


Potential outcome perspective: each individual has 2 potential outcomes

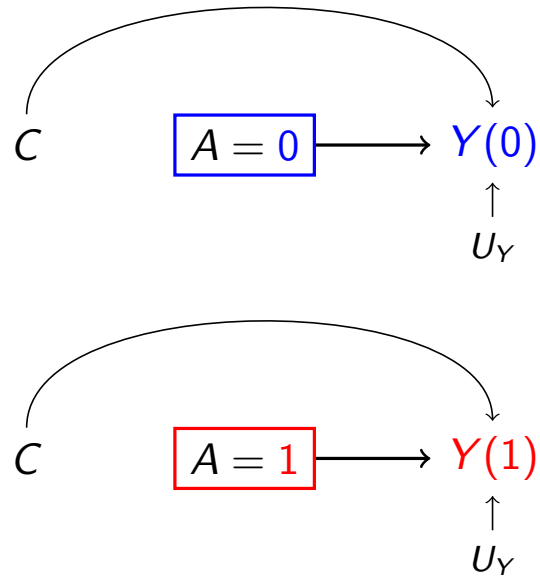
- Individual total effects: $TE_i = Y_i(1) - Y_i(0)$
- Average total effect: $\mathrm{TE} = \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)]$ (the familiar ATE)

Total effect – a review

The regular world
without any manipulation



The two worlds contrasted
by the total effect



Intervention regimes perspective: imagine 2 interventions, one setting A to 1, the other setting A to 0, for everyone

- $\mathrm{TE} = \mathrm{E}[Y \mid \mathrm{set}(A{=}1)] - \mathrm{E}[Y \mid \mathrm{set}(A{=}0)]$ (also the ATE)

Natural (In)Direct Effects

Analysis motivation: explain the total effect

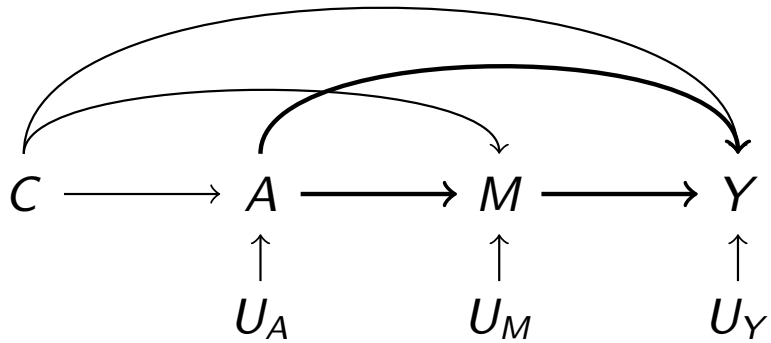
- what part (if any) of this effect went through M?
  (what part was due to the change in M in response to treatment?)
- what part went through other ways?

This implies splitting the total effect into a direct effect part and an indirect effect part.

The natural (in)direct effects best match this desire.

Prepping TE for decomposition

Suppose this is the regular world

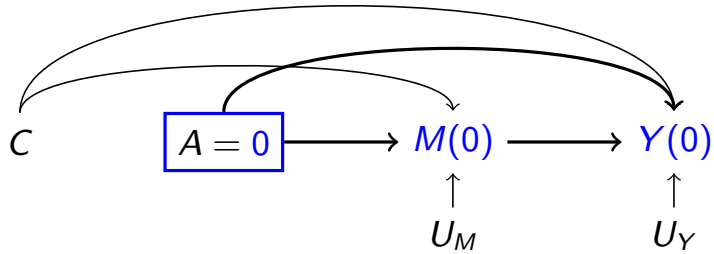


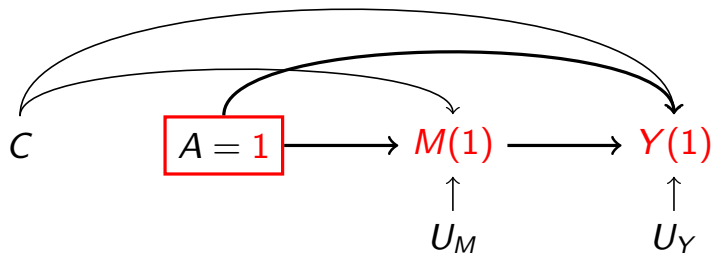
C stands for *covariates*, *confounders*, *common causes*

(This is a special case – for simple presentation. More generally, there may be common causes of M and Y that are influenced by A.)

Prepping TE for decomposition

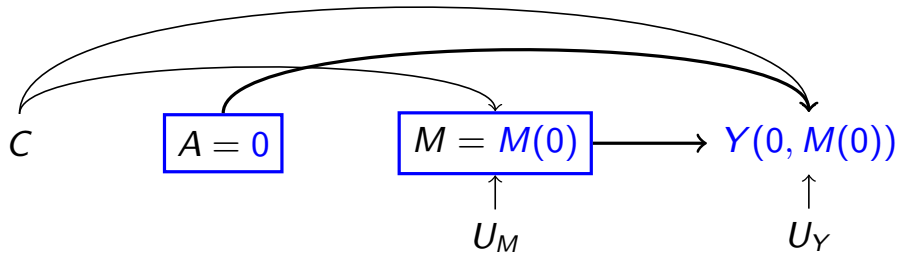
The two conditions contrasted by the total effect are

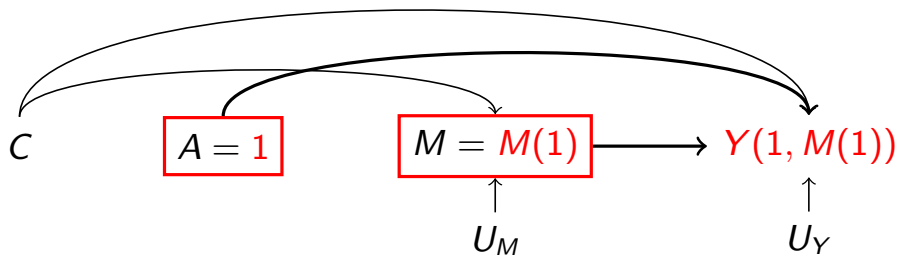




Prepping TE for decomposition

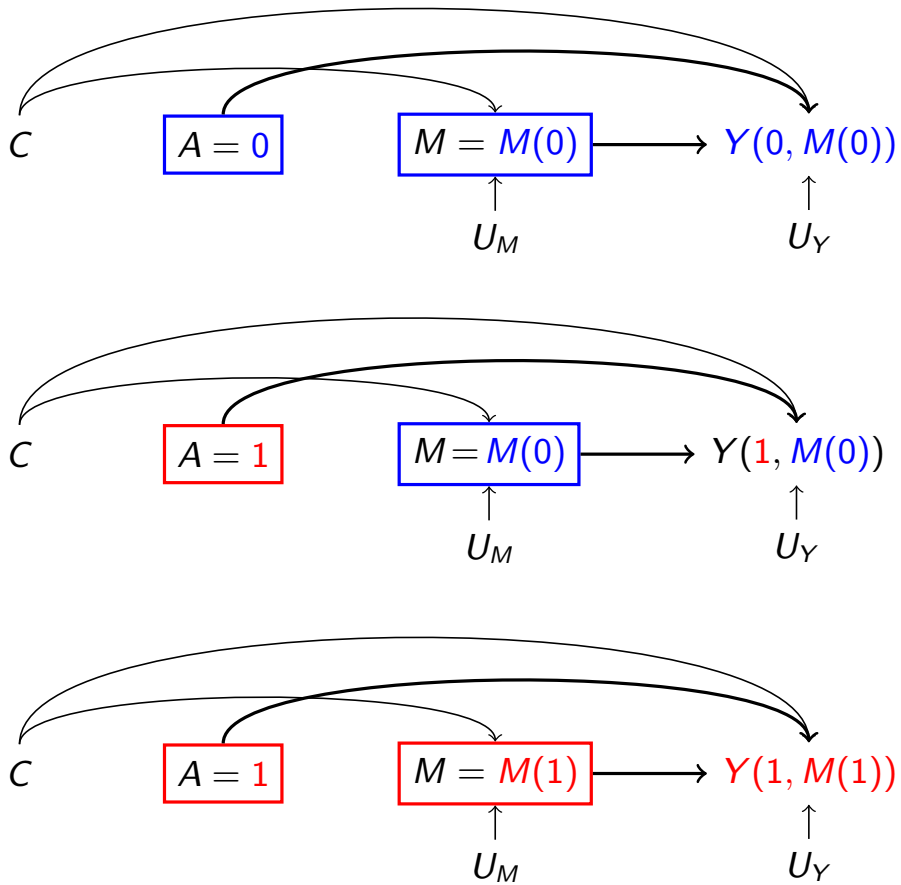
or





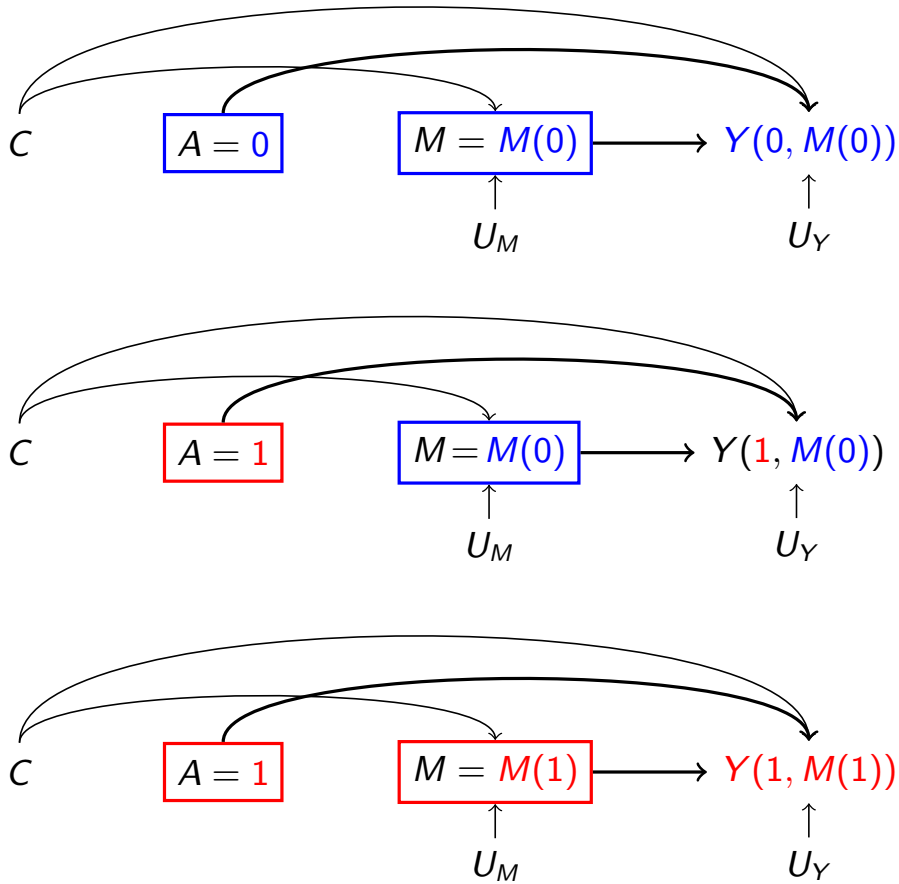
Decomposing TE: one pair of natural (in)direct effects

The direct-indirect decomposition (my label)



Decomposing TE: one pair of natural (in)direct effects

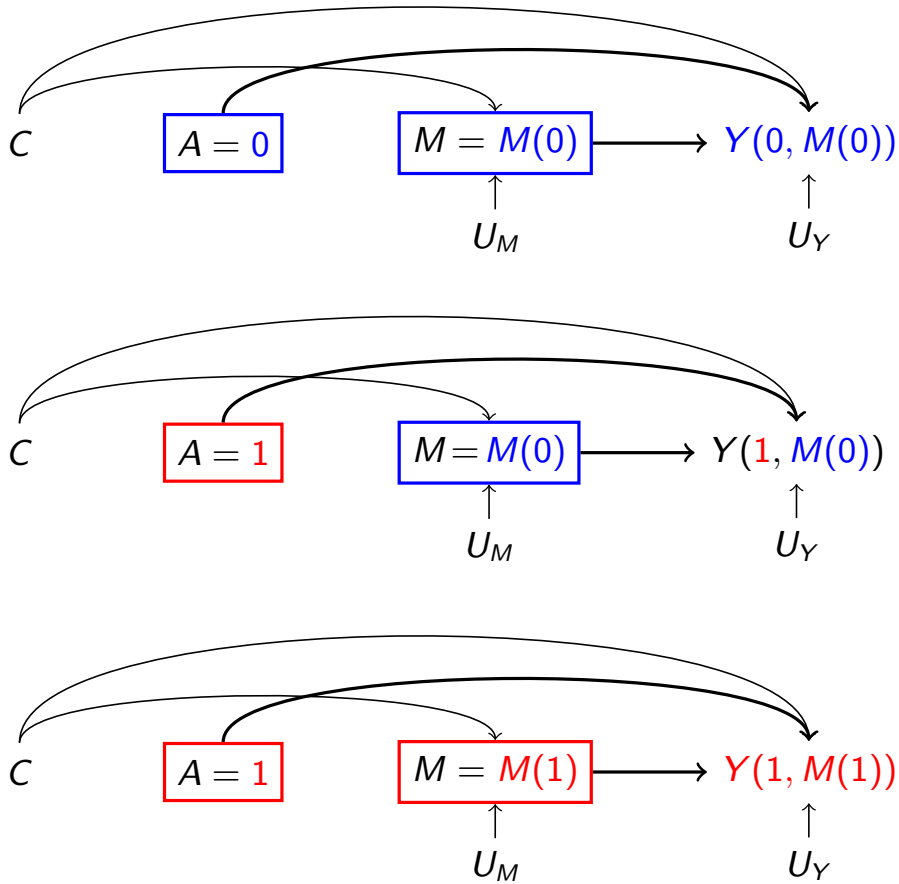
The direct-indirect decomposition (my label)



$$TE_i = \overbrace{Y_i(1, M_i(1)) - Y_i(1, M_i(0))}^{\text{a NIE}} + \overbrace{Y_i(1, M_i(0)) - Y_i(0, M_i(0))}^{\text{a NDE}}$$

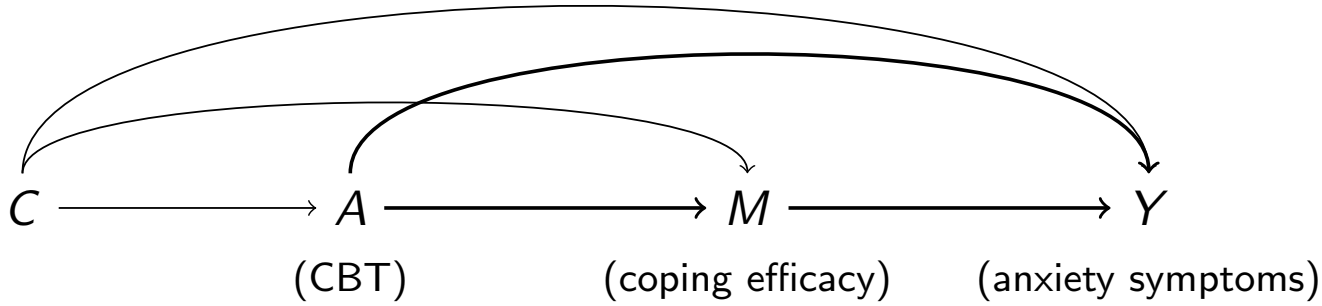
Decomposing TE: one pair of natural (in)direct effects

The direct-indirect decomposition (my label)



$$\text{TE} = \overbrace{\text{E}[Y(1, M(1))] - \text{E}[Y(1, M(0))]}^{\text{a NIE}} + \overbrace{\text{E}[Y_i(1, M(0))] - \text{E}[Y(0, M(0))]}^{\text{a NDE}}$$

CBT example

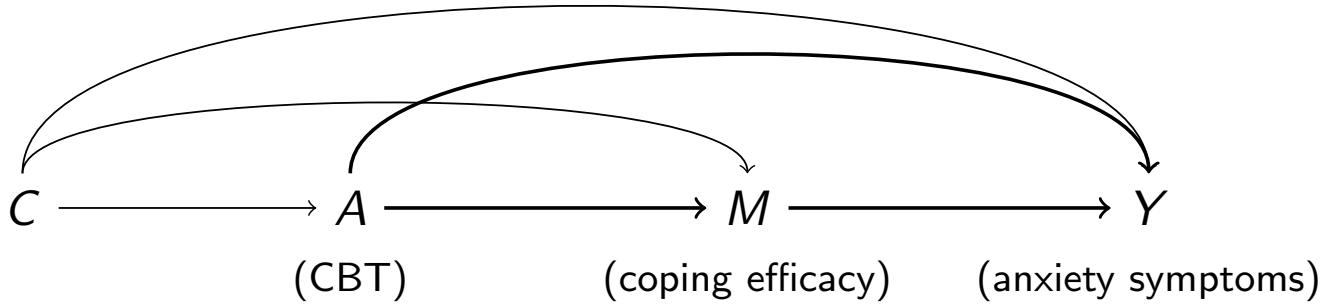


Consider Jamie, an anxious adolescent.

	Potential mediator	Potential outcome
non-CBT condition	$M_J(0) =$ low	$Y_J(0) =$ high
CBT condition	$M_J(1) =$ high	$Y_J(1) =$ low

TE (shift from non-CBT to CBT): reduction of symptoms from high to low

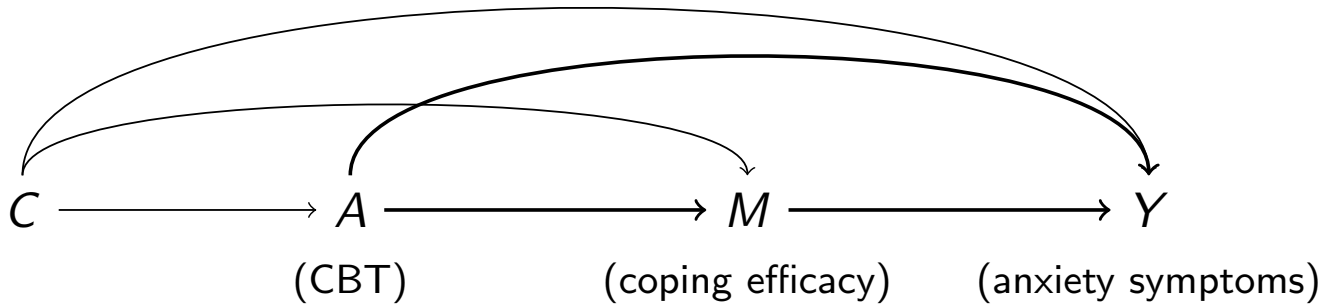
CBT example



Consider Jamie, an anxious adolescent.

	Potential mediator	Potential outcome
non-CBT condition	$M_J(0) = \text{low}$	$Y_J(0) = Y_J(0, M_J(0)) = Y_J(0, \text{low}) = \text{high}$
CBT condition	$M_J(1) = \text{high}$	$Y_J(1) = Y_J(1, M_J(1)) = Y_J(1, \text{high}) = \text{low}$

CBT example



Consider Jamie, an anxious adolescent.

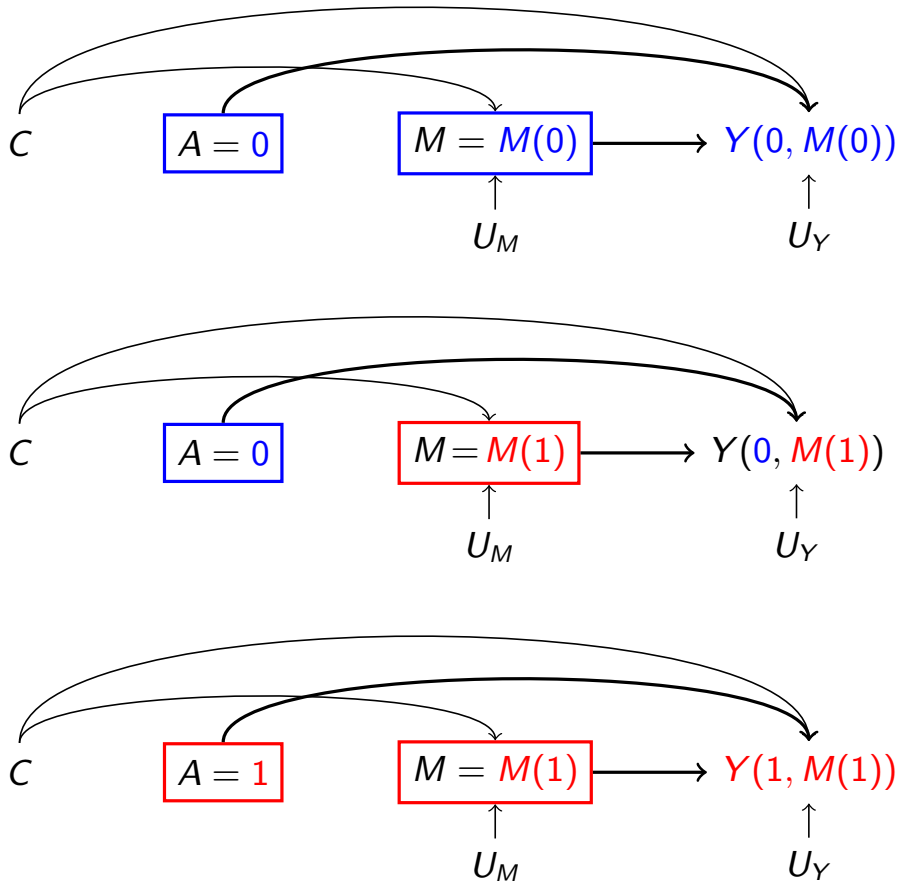
	Potential mediator	Potential outcome
non-CBT condition	$M_J(0) = \text{low}$	$Y_J(0) = Y_J(0, M_J(0)) = Y_J(0, \text{low}) = \text{high}$
in-between condition		$Y_J(1, M_J(0)) = Y_J(1, \text{low}) = \text{moderate}$
CBT condition	$M_J(1) = \text{high}$	$Y_J(1) = Y_J(1, M_J(1)) = Y_J(1, \text{high}) = \text{low}$

NDE (shift from non-CBT to in-between): symptoms reduction from high to moderate

NIE (shift from in-between to CBT): further reduction from moderate to low

Decomposing TE: another natural (in)direct effects pair

The indirect-direct decomposition



$$\text{TE} = \overbrace{\text{E}[Y(1, M(1))] - \text{E}[Y(0, M(1))]}^{\text{a NDE}} + \overbrace{\text{E}[Y(0, M(1))] - \text{E}[Y(0, M(0))]}^{\text{a NIE}}$$

Natural (in)direct effects: 2 TE decompositions

Direct-indirect decomposition:

$$\text{TE} = \underbrace{\text{E}[Y(1, M(1))] - \text{E}[Y(1, M(0))]}_{\text{NIE}(1\cdot)} + \underbrace{\text{E}[Y(1, M(0))] - \text{E}[Y(0, M(0))]}_{\text{NDE}(\cdot 0)}$$

Indirect-direct decomposition:

$$\text{TE} = \underbrace{\text{E}[Y(1, M(1))] - \text{E}[Y(0, M(1))]}_{\text{NDE}(\cdot 1)} + \underbrace{\text{E}[Y(0, M(1))] - \text{E}[Y(0, M(0))]}_{\text{NIE}(0\cdot)}$$

Natural (in)direct effects: why two pairs?

Ask Jamie!

$$Y_J(0, M_J(0)) = \text{high}$$

$$Y_J(1, M(0)) = \text{moderate}$$

$$Y_J(0, M(1)) = \text{moderate}$$

$$Y_J(1, M_J(1)) = \text{low}$$

$$\text{NDE}(\cdot 0) = \text{reduction from high to moderate}$$

$$\text{NDE}(\cdot 1) = \text{reduction from moderate to low}$$

are these the same?

Natural (in)direct effects: why two pairs?

Implicit in

What is the direct effect? What is the indirect effect?

is an assumption that these things are separate and don't interact,

which is arbitrary.

Natural (in)direct effects: which to use?

Case 1: Is there a mediated effect?
Is the causal effect (partly) mediated by this putative mediator?
What is the size of this effect?

- the direct-indirect decomposition
  - if no mediated effect, TE same as NDE($\cdot 0$)

Case 2: In addition to the mediated effect, is there a direct effect?
Does the exposure influence the outcome in ways not through M?
What is the size of this effect?

- the indirect-direct decomposition
  - if no direct effect, TE same as NIE($0\cdot$)

Case 3: What can we learn about the effect of this exposure, either through M or not through M? – no prior assumption or preferred question

- both decompositions

(case 1 probably most common)

Natural (in)direct effects: a couple of challenges

- In principle, not experimentally testable
  - barring feasibility and ethics, can't conceive of an (ideal) experimental study to test or estimate these effects

    (do not exist in the intervention regimes framework)
- Not identified if an M-Y confounder is influenced by A

  (will get to this later)

Interventional In(Direct) Effects

Analysis motivation: asking *what if* questions

- in intervention development research: what if the program is modified
  - removing elements that change the mediator
  - retaining only elements that change the mediator
  - some other way
- in health/social disparities research
  - what if could bring bullying experience at of LGB adolescents down to level experienced by heterosexual adolescents
  - what if a school intervention gets us half way there

Interventional (in)direct effects

- do not exactly tell us about realistic interventions

  BUT
- do tell us about *ideal* interventions that intervene on the mediator and change nothing else

- our job to judge how rough or fine the approximation
  (can tune the ideal interventions within some boundaries – later)

Interventional (in)direct effects

- are *interventional*: contrast intervention conditions
- are *(in)direct*: involve some type of fixing/swapping of the mediator under exposed and unexposed conditions
  - will see from the *individual* and *population* viewpoints

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

$$Y(0, M(0)), \quad Y(1, M(0)), \quad Y(0, M(1)), \quad Y(1, M(1))$$

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value
  - group the individual with others in a subpopulation that share pre-exposure covariates pattern $C = c$
  - randomly draw one value from the subpopulation's pool of $M(a)$ values
  - assign this value to the individual

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value
  - group the individual with others in a subpopulation that share pre-exposure covariates pattern $C = c$
  - randomly draw one value from the subpopulation's pool of $M(a)$ values
  - assign this value to the individual

  Labeling for clarity:

  $d_{M(a)|C}$: the distribution of $M(a)$ given C (those pools of values)
  $\mathcal{M}(a|C)$: a random value from $d_{M(a)|C}$ (value drawn from pool)

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value $\mathcal{M}(a|C)$

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

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value $\mathcal{M}(a|C)$

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

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value $\mathcal{M}(a|C)$

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

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value $\mathcal{M}(a|C)$

  Quick note: some authors call IDE/IIEs *stochastic* (in)direct effects (stochastic = randomly determined)

  - short for *stochastic interventional*
  - I prefer *interventional* which emphasizes distinction from *natural*

from the viewpoint of the individual

- NDE/NIEs are defined based on $Y(a', M(a))$

  i.e., mediator is set to value specific to the individual

- For IDE/IIEs, mediator is set to a random value $\mathcal{M}(a|C)$

- IDE/IIEs are not defined at the individual level
  - $Y_i(a', \mathcal{M}(a|C))$ is random

from the population viewpoint

IDE/IIEs contrast interventions that set the exposure to 0 or 1 and set the mediator distribution to $d_{M(0)|C}$ or $d_{M(1)|C}$

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!

$$\mathrm{IDE}(\cdot 0) + \mathrm{IIE}(1\cdot) = \mathrm{IIE}(0\cdot) + \mathrm{IDE}(\cdot 1) = \mathrm{OE}$$

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!

$$\text{IDE}(\cdot 0) + \text{IIE}(1\cdot) = \text{IIE}(0\cdot) + \text{IDE}(\cdot 1) = \text{OE}$$

OE contrasts 2 interventions

- one setting exposure to 1 and mediator dist. to $d_{M(1)|C}$
- the other setting exposure to 0 and mediator dist. to $d_{M(0)|C}$

while TE contrasts

- one intervention setting exposure to 1
- the other intervention setting exposure to 0

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!
  - analyses that aim to explain TE target a *pair* of natural effects
  - analyses that ask *what if* questions target one (or more) IDE/IIE(s) based on the research question – not pairs

Some comments on IDE/IIEs vs. NDE/NIEs

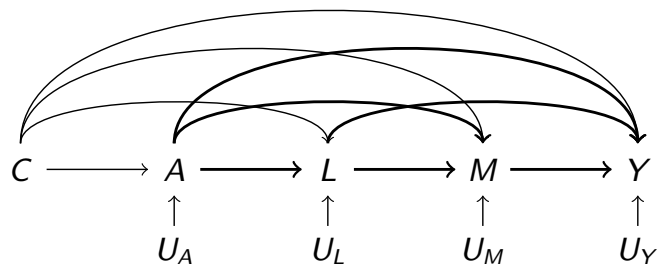
- IDE/IIEs do not decompose TE – not made for that purpose!
  - analyses that aim to explain TE target a *pair* of natural effects
  - analyses that ask *what if* questions target one (or more) IDE/IIE(s) based on the research question – not pairs
- IDE/IIEs are, in principle, experimentally testable

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!
  - analyses that aim to explain TE target a *pair* of natural effects
  - analyses that ask *what if* questions target one (or more) IDE/IIE(s) based on the research question – not pairs
- IDE/IIEs are, in principle, experimentally testable
- if no intermediate confounders & no unobserved confounding, IDE/IIEs $=$ NDE/NIEs; otherwise, IDE/IIEs $\neq$ NDE/NIEs

Some comments on IDE/IIEs vs. NDE/NIEs

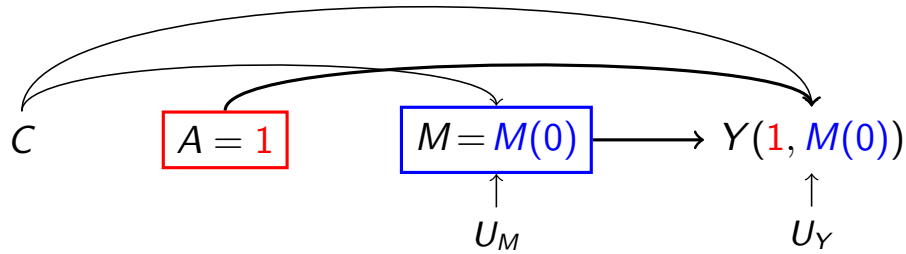
- IDE/IIEs do not decompose TE – not made for that purpose!
  - analyses that aim to explain TE target a *pair* of natural effects
  - analyses that ask *what if* questions target one (or more) IDE/IIE(s) based on the research question – not pairs
- IDE/IIEs are, in principle, experimentally testable
- if no intermediate confounders & no unobserved confounding, IDE/IIEs = NDE/NIEs; otherwise, IDE/IIEs $\neq$ NDE/NIEs
  - intermediate confounders: influenced by A, influencing M and Y



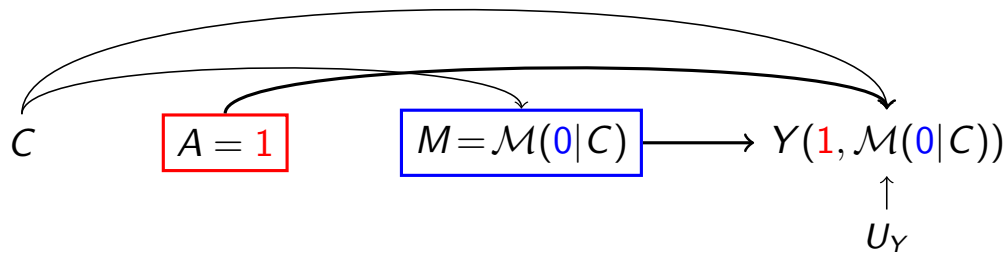
also called: post-exposure confounders, exposure-induced confounding

The special case with no intermediate confounders

outcome means in NDE/NIEs depend on distribution of $M(a)$ given C



outcome means in IDE/IIEs depend on distribution of $\mathcal{M}(a|C)$ given C

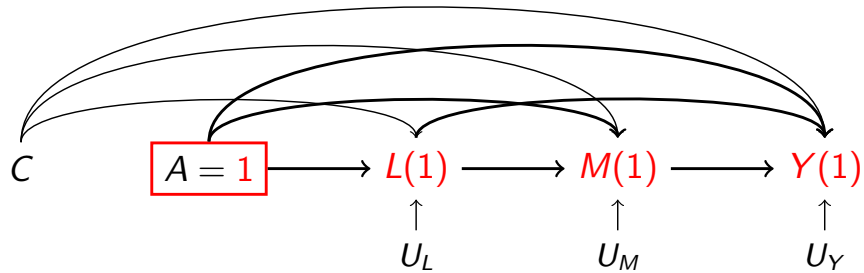


the mediator distributions are the same between the two types of effects

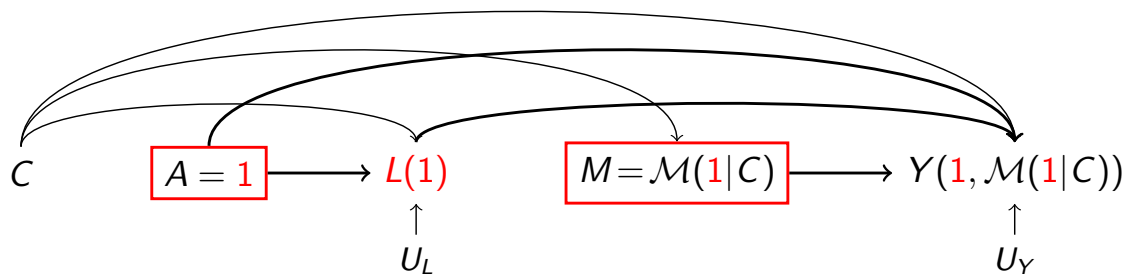
(assuming all confounders are observed in pre-exposure covariates C)

The general case with intermediate confounders L

outcome means in NDE/NIEs depend on distribution of $\{L(a'), M(a)\}$ given C



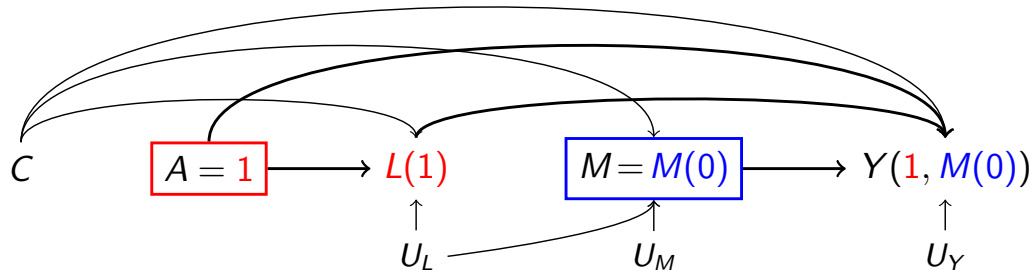
outcome means in IDE/IIEs depend on distribution of $\{L(a'), \mathcal{M}(a|C)\}$ given C



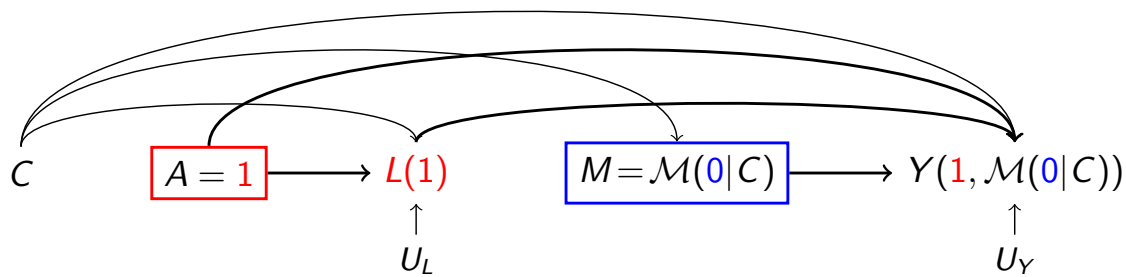
the mediator distributions are not the same between the two types of effects – conditional on C, $L(a'), \mathcal{M}(a|C)$ independent, but $L(a'), M(a)$ generally dependent

The general case with intermediate confounders L

outcome means in NDE/NIEs depend on distribution of $\{L(a'), M(a)\}$ given C



outcome means in IDE/IIEs depend on distribution of $\{L(a'), \mathcal{M}(a|C)\}$ given C



the mediator distributions are not the same between the two types of effects – conditional on C, $L(a'), \mathcal{M}(a|C)$ independent, but $L(a'), M(a)$ generally dependent

U_L to $M(a)$ path: $U_L \rightarrow L(a) \rightarrow M(a)$

Some comments on IDE/IIEs vs. NDE/NIEs

- IDE/IIEs do not decompose TE – not made for that purpose!
  - analyses that aim to explain TE target a *pair* of natural effects
  - analyses that ask *what if* questions target one (or more) IDE/IIE(s) based on the questions – not pairs
- IDE/IIEs are experimentally testable
- if no intermediate confounders & no unobserved confounding, IDE/IIEs $=$ NDE/NIEs; otherwise, IDE/IIEs $\neq$ NDE/NIEs
- if any intermediate confounder, NDE/NIEs are unidentified, but IDE/IIEs may be

Example

- *Real-world* questions: what if the intervention (college prep program for high school students) is modified,
  i. removing elements that change the mediator (self awareness)?
  ii. retaining only elements that change the mediator?
  iii. some other way?

Example

- *Real-world* questions: what if the intervention (college prep program for high school students) is modified,
  i. removing elements that change the mediator (self awareness)?
  ii. retaining only elements that change the mediator?
  iii. some other way?

- Rough translation to *ideal* interventions
  i. one that sets exposure to 1 and mediator dist. to $d_{M(0)|C}$
     - not require the modified intervention to give everyone their own $M(0)$ values, just the same distribution

Example

- *Real-world* questions: what if the intervention (college prep program for high school students) is modified,
  i. removing elements that change the mediator (self awareness)?
  ii. retaining only elements that change the mediator?
  iii. some other way?

- Rough translation to *ideal* interventions
  i. one that sets exposure to 1 and mediator dist. to $d_{M(0)|C}$
     - not require the modified intervention to give everyone their own $M(0)$ values, just the same distribution
  ii. one that sets exposure to 0 and mediator dist. to $d_{M(1)|C}$
  iii. ???

Example

- *Real-world* questions: what if the intervention (college prep program for high school students) is modified,
  i. removing elements that change the mediator (self awareness)?
  ii. retaining only elements that change the mediator?
  iii. some other way?

- Rough translation to *ideal* interventions
  i. one that sets exposure to 1 and mediator dist. to $d_{M(0)|C}$
     - not require the modified intervention to give everyone their own $M(0)$ values, just the same distribution
     - IDE($\cdot 0$)
  ii. one that sets exposure to 0 and mediator dist. to $d_{M(1)|C}$
     - IIE($0\cdot$)
  iii. ???

Interventional (in)direct effects: a couple of challenges

- What if we wish to use the same comparison condition?
  - $\mathrm{E}[Y(1, \mathcal{M}(1|C))] - \mathrm{E}[Y(0)]$ and $\mathrm{E}[Y(0, \mathcal{M}(0|C))] - \mathrm{E}[Y(0)]$ are interventional effects
  - in this situation, more appropriate than IDE($\cdot 0$), IIE($0\cdot$)
  - not standard IDE/IIEs, but who cares?
- Also, are iii. questions completely off limits?

Interventional Effects

more generally

Larger class of effects, more possibilities

Let the scientific interest guide the definition of the effects!

Conditions contrasted are

- interventions that set one or more variables to specific values, or distributions, that are priorly determined
- or the null intervention – doing nothing

Interventional effects include but are not limited to IDE/IIEs

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

Sexual minority adolescents experience higher levels of bullying and lower levels of several well-being measures, compared to sexual majority (heterosexual) adolescents.

$$\text{disparity in bullying experience: } \mathrm{E}[M|A=1] > \mathrm{E}[M|A=0]$$
$$\text{disparity in well-being: } \mathrm{E}[Y|A=1] < \mathrm{E}[Y|A=0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\mathrm{E}[Y(1)|A=1] - \mathrm{E}[Y(0)|A=0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\mathrm{E}[Y(1, M(1))|A = 1] - \mathrm{E}[Y(0, M(0))|A = 0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\mathrm{E}[Y(1, M(1))|A = 1] - \mathrm{E}[Y(0, M(0))|A = 0]$$

For current purpose, don't need to consider $Y(0), M(0)$ for sexual minority adolescents or $Y(1), M(1)$ for sexual majority adolescents (or decide whether those variables exist or what they mean)

– I am not opposed to imagining those, but that's another topic

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A = 1] - \mathrm{E}[Y|A = 0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A = 1] - \mathrm{E}[Y|A = 0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

Well-being depends on sexual minority status and bullying experience, $Y = Y(1, m)$. If bullying experience changes, well-being may change.

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A = 1] - \mathrm{E}[Y|A = 0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

Take the distribution of bullying experience in sexual majority adolescents (conditional on demographic/contexttual covariates), $d_{M(0)|C}$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A = 1] - \mathrm{E}[Y|A = 0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

Take the distribution of bullying experience in sexual majority adolescents (conditional on demographic/contexttual covariates), $d_{M(0)|C}$

Imagine that each sexual minority adolescent, instead of their own bullying experience, would experience $\mathcal{M}_{0|C}$, a random draw from $d_{M(0)|C}$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

Take the distribution of bullying experience in sexual majority adolescents (conditional on demographic/contexttual covariates), $d_{M(0)|C}$

Imagine that each sexual minority adolescent, instead of their own bullying experience, would experience $\mathcal{M}_{0|C}$, a random draw from $d_{M(0)|C}$

Then they would have well-being level $Y(1, \mathcal{M}_{0|C})$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

$$\mathrm{E}[Y|A=1] - \mathrm{E}[Y(1,\mathcal{M}_{0|C})|A=1] \; + \; \mathrm{E}[Y(1,\mathcal{M}_{0|C})|A=1] - \mathrm{E}[Y|A=0]$$

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

$$\text{total-disparity} = \mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$$

Question 1: How much of the disparity in well-being would be removed if we could reduce the level of bullying experience of sexual minority adolescents down to the level experienced by sexual majority adolescents?

$$\underbrace{\mathrm{E}[Y|A=1] - \mathrm{E}[Y(1,\mathcal{M}_{0|C})|A=1]}_{\text{disparity-removed}} + \underbrace{\mathrm{E}[Y(1,\mathcal{M}_{0|C})|A=1] - \mathrm{E}[Y|A=0]}_{\text{remaining-disparity}}$$

$\downarrow$

an interventional effect on the sexual minority group
contrasting an intervention & the null intervention condition

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

sexual minority well-being: $\mathrm{E}[Y|A = 1]$

Question 2: How much improvement in the well-being of sexual minority adolescents would be achieved if a school anti-bullying intervention – being considered by the school board – could bring their bullying experience down to halfway between the levels of the two groups?

Example

A: sexual minority (LGB) status (yes/no)
M: bullying experience
Y: well-being (incl. mental health, life satisfaction, etc.)

sexual minority well-being: $\mathrm{E}[Y|A=1]$

Question 2: How much improvement in the well-being of sexual minority adolescents would be achieved if a school anti-bullying intervention – being considered by the school board – could bring their bullying experience down to halfway between the levels of the two groups?

Relevant now is the half-half mixture of $d_{M(0)|C}$ and $d_{M(1)|C}$. Denote a random draw from this mixture distribution by $\mathcal{M}_{0.5|C}$

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

is the effect that such anti-bullying intervention might have on sexual minority adolescents' well-being.

Comments on the broader class of interventional effects

- Helps better tune effect definition to question of interest
  - admit interventions that intervene only on exposure
  - null intervention
  - broader range of mediator interventions, not just setting to $d_{M(0)|C}$ or $d_{M(1)|C}$

Comments on the broader class of interventional effects

- Helps better tune effect definition to question of interest
  - admit interventions that intervene only on exposure
  - null intervention
  - broader range of mediator interventions, not just setting to $d_{M(0)|C}$ or $d_{M(1)|C}$
- Very flexible
  - intervention on either variable (A or M) can be deterministic (setting to values) or stochastic (setting to distributions)

Comments on the broader class of interventional effects

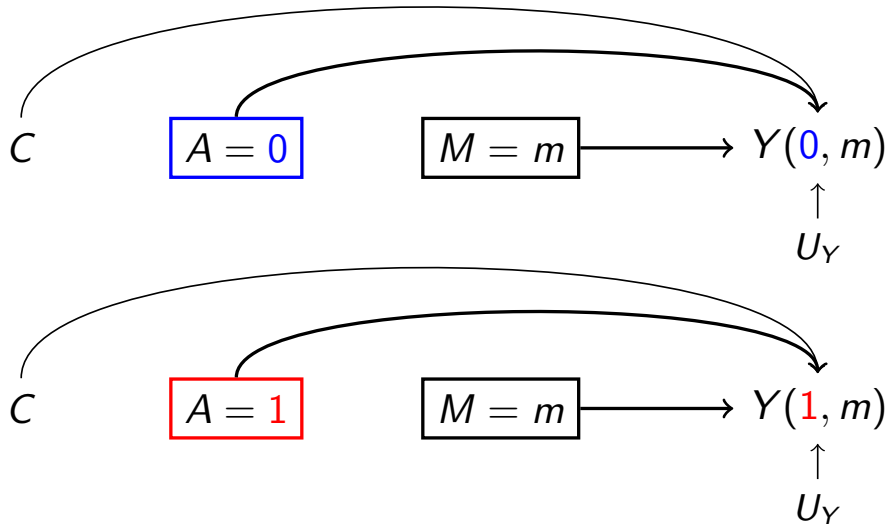
- Helps better tune effect definition to question of interest
  - admit interventions that intervene only on exposure
  - null intervention
  - broader range of mediator interventions, not just setting to $d_{M(0)|C}$ or $d_{M(1)|C}$
- Very flexible
  - intervention on either variable (A or M) can be deterministic (setting to values) or stochastic (setting to distributions)

  Special case: generalized interventional direct effects (GIDEs)
  - $\text{GIDE}(\cdot\mathscr{D}) = \text{E}[Y(1, \mathcal{M}_{\mathscr{D}})] - \text{E}[Y(0, \mathcal{M}_{\mathscr{D}})]$ for med. dist. $\mathscr{D}$
  - include both IDEs and controlled direct effects
  - except for IDEs, are not paired with indirect effects

Controlled Direct Effects

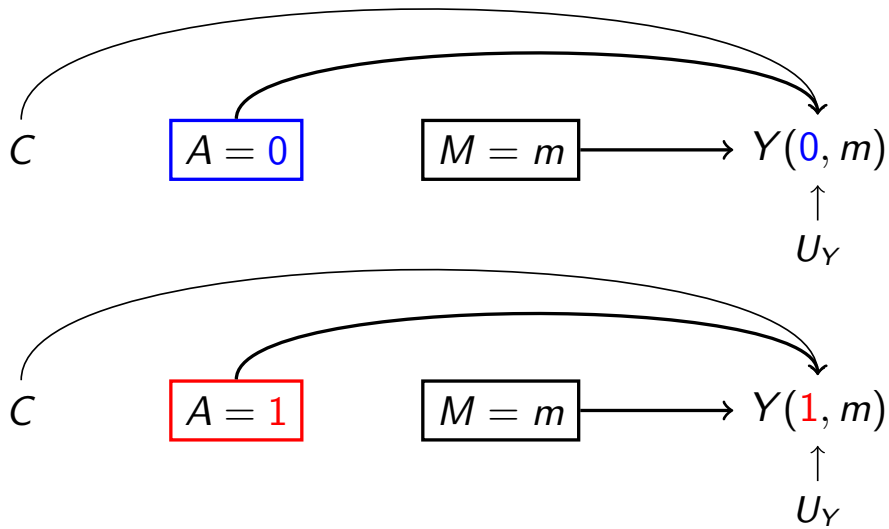
Controlled direct effects

Effect of exposure on outcome if mediator were fixed at a specific level



Controlled direct effects

Effect of exposure on outcome if mediator were fixed at a specific level

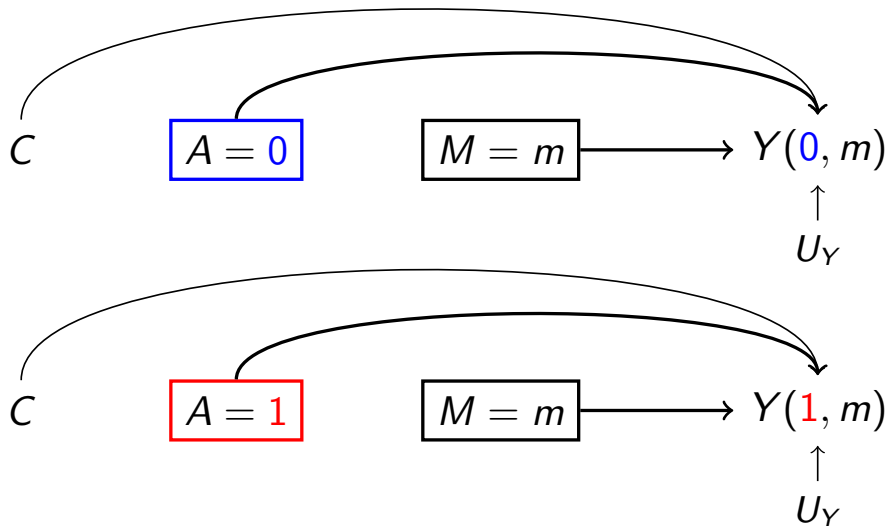


For mediator control level m,

- individual effects: $CDE_i(m) = Y_i(1, m) - Y_i(0, m)$
- average effect: $\text{CDE}(m) = \text{E}[Y(1, m)] - \text{E}[Y(0, m)]$

Controlled direct effects

Effect of exposure on outcome if mediator were fixed at a specific level



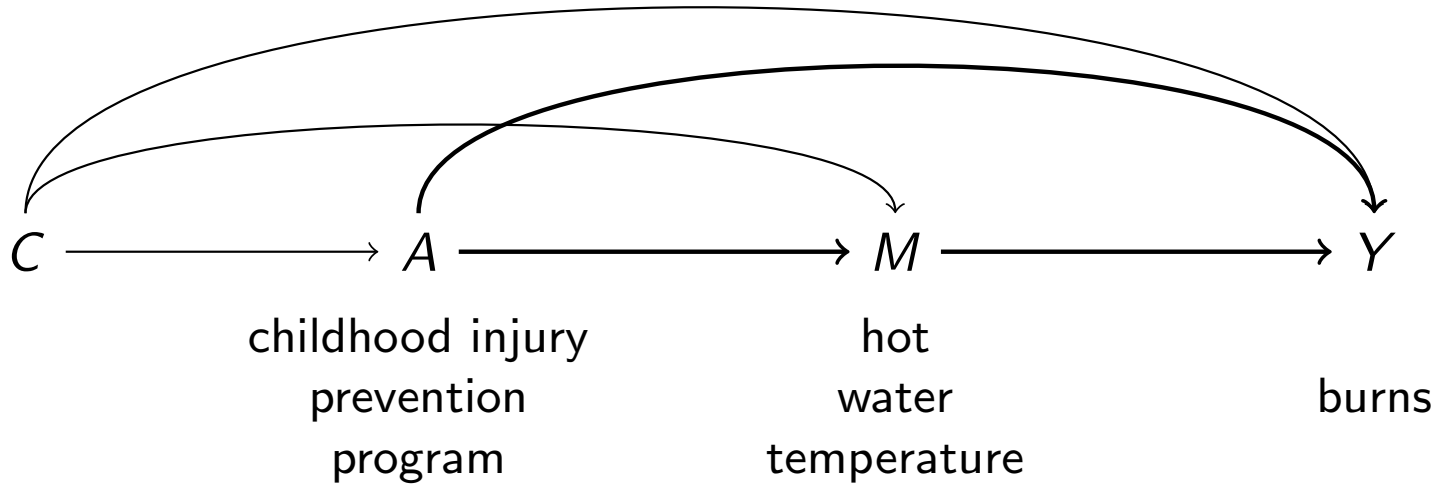
CDE(m) is an interventional effect – a special type of GIDEs

- relevant when m is a desirable level for M, and a feasible and ethical intervention exists to set M to m

CDEs not paired with indirect effects – like other GIDEs that are not IDEs

Example

The city has an effective childhood injury prevention program



It now has a structural intervention on home water heating

and wants to know the effectiveness of the injury prevention program

$$\text{CDE}(\text{water temperature} = 120)$$

$$\text{GIDE}(\text{a water temperature dist. w/ 100-125 range})$$

Summary of Effect Types

- Natural (in)direct effects – explaining TE
  - two TE decompositions
  - select based on research question

- Interventional effects – asking *what if* questions
  - based on research question, define each of the intervention conditions to be contrasted
    - an intervention condition sets variables to priorly determined values/distributions
    - also null intervention
  - flexible
    - special types: IIEs and IDEs, CDEs, GIDEs, TE, OE
    - and a lot more options