In Search of Research Questions for Causal Mediation Analysis

Vanessa Didelez

Leibniz Institute for Prevention Research and Epidemiology – BIPS

Faculty of Mathematics and Computer Science, University of Bremen, Germany

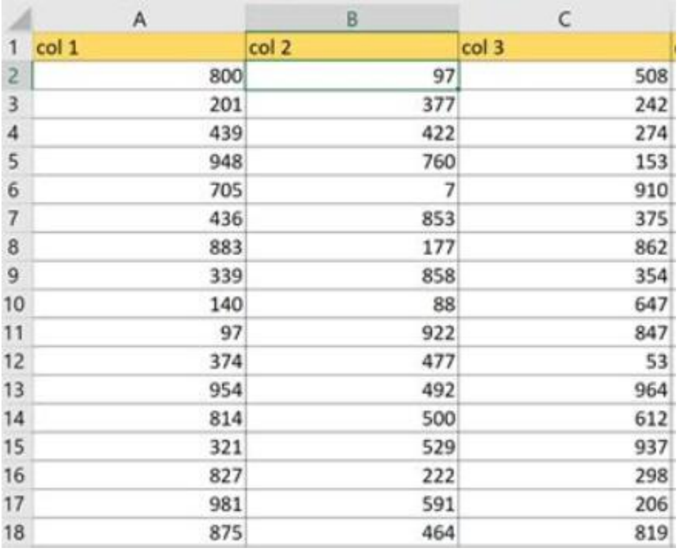
BBS Basel – October 2025

Motivation (?)

Suppose I give you three columns of numbers (*A*,*B*,*C*), (and I may tell you that *A* was randomised)

You can use a standard linear structural equation model (LSEM) to estimate:

- “the total effect” of *A* on *C*
- “the direct effect” of *A* on *C* not via *B*
- “the indirect effect” *A* on *C* via *B*



	A	B	C
1	col 1	col 2	col 3
2	800	97	508
3	201	377	242
4	439	422	274
5	948	760	153
6	705	7	910
7	436	853	375
8	883	177	862
9	339	858	354
10	140	88	647
11	97	922	847
12	374	477	53
13	954	492	964
14	814	500	612
15	321	529	937
16	827	222	298
17	981	591	206
18	875	464	819

This is just maths!

But what does it mean?

Depends on:

- What is the **research question (RQ)**?
- Intended use to inform decision making / actions?
- Under what assumption is the question “**answerable**” *(identification)*
- …and are the **assumptions defendable** in data context?

Recap

1960s mediation analysis

- Linear structural equation models (LSEMs)

→ “Estimands” do not exist outside of the parametric model

  - Other parametric models available – but same issue

Early 2000s causal mediation analysis

(Pearl 2001:UAI, after Robins&Greenland, 1992:Epidemiology)

- Pure / natural (in)direct effects (NDE/NIE)
- Based on nested counterfactuals $E(Y(a,M(a')), a \neq a'$,

→ Cross-world concepts, needing untestable cross-world assumptions

→ Actionable value remains mysterious *(Robins & Richardson, 2011:book.chap)*

Recap

1960s mediation analysis

- Linear structural equation models (LSEMs)

→ "Estimands" do not exist outside of the parametric model

- Other parametric models a

Early 2000s

and, 1992:Epidemiology)

- Pure / natu E/NIE)
- Based on nested counterfactuals *E(Y(a,M(a'))), a≠a'* ,

→ Cross-world concepts, needing untestable cross-world assumptions

→ Actionable value remains mysterious *(Robins & Richardson, 2011:book.chap)*

These are concepts,
not research questions

Notes

Assume throughout

treatment (*A*) is randomised

All examples will be ridiculously simplified to make key points!

“Understanding Mechanisms?”

- “Understanding” is not a research question
  - Need to specify how to verify “understanding”
  - e.g. does it lead to developing actions that improve health?

- “Understanding” should be implementable in the sense of:
  - We will get an idea *what to do* to improve patients’ health
    or
  - We will get *new* ideas for what *else* to do to improve patients’ health

“It is *only* because…”

Example:

Effect of new treatment (compared to standard) is small / large ***“only because”*** of some unintended consequences

- Quality of care, adherence, switching, rescue medication …

This is also not a research question

- How would it affect actions / decisions if we knew that it really is “only because”, or if we knew that it really is not “only because”?
  - Can / would we remove the unintended consequences?

⇒ More transparency if actionable aspects of RQ are made clearer

Interesting Research Questions

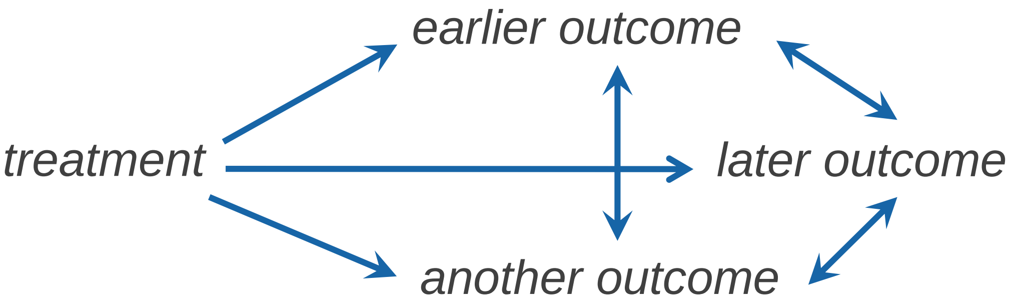
that are not about Mediation

(all identified without cross-world assumptions)

Many Effects of Treatment?

Not a mediation question:

- what are various *different effects* of treatment?



- Example: How does treatment affect adherence, co-interventions, rescue med, adverse events, various biomarkers, primary outcome

Many Effects of Treatment – Notes

- Target several total effects, separately for each outcome
  - Identified under randomised treatment

- Or: target effect on *joint* distribution
  - e.g. does correlation between outcomes depend on treatment?

$\Rightarrow$ No single summary of ‘the treatment effect’…

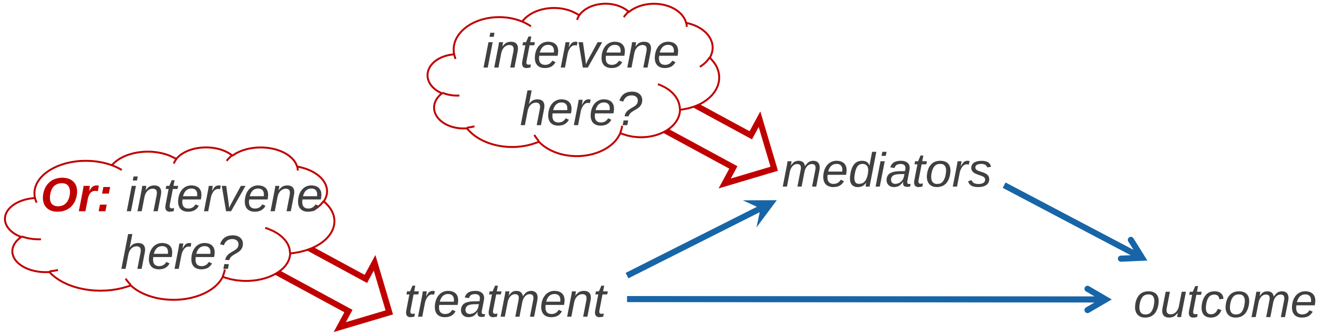
$\Rightarrow$ But provides good insights into mechanisms

- Examples: ???
  - Some recent work on multi-outcomes in causal literature

Where Best to Intervene?

Also not a mediation question:

- *where best* to intervene – on treatment or mediator?



- Example: is it better to increase number of check-ups or to improve care after diagnosis (if you can’t do both)?

Where Best to Intervene – Notes

- Estimate effect of (realistic intervention on) *A* on *Y*
  - And the effect of *A* on *M* – to understand ‘mechanism’

- …then estimate effect of (realistic intervention on) *M* on *Y*

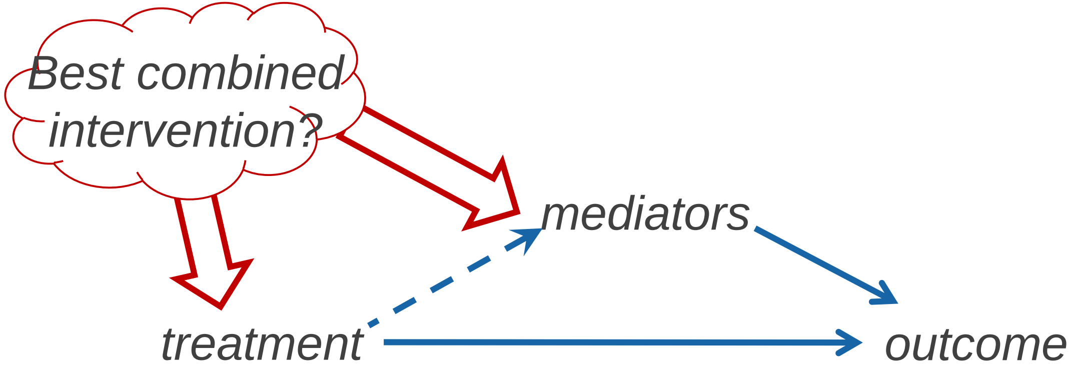
→ Compare two total effects!

- Identified without cross-world assumptions
- Need to control for mediator-outcome confounding

Joint / Sequential / Adaptive Intervention?

Also not a mediation question:

- what is best *joint* (seq./dyn.) intervention on treatment & mediator



- Example: Treatment followed by a fixed rule for when to administer a co-intervention *(adaptive / dynamic intervention)*

Joint / Sequ. / Adaptive Interv. – Notes

Special case: **controlled direct effect (CDE) or “E9-hypothetical”**

- Example: always apply co-intervention – does primary intervention still have an effect?
  i.e. is there a controlled direct effect of the primary intervention?

→ Establish presence of (controlled) direct effect

- Again: control for mediator-outcome confounding required
- But not cross-world independence assumption

Joint / Sequ. / Adaptive Interv. – Notes

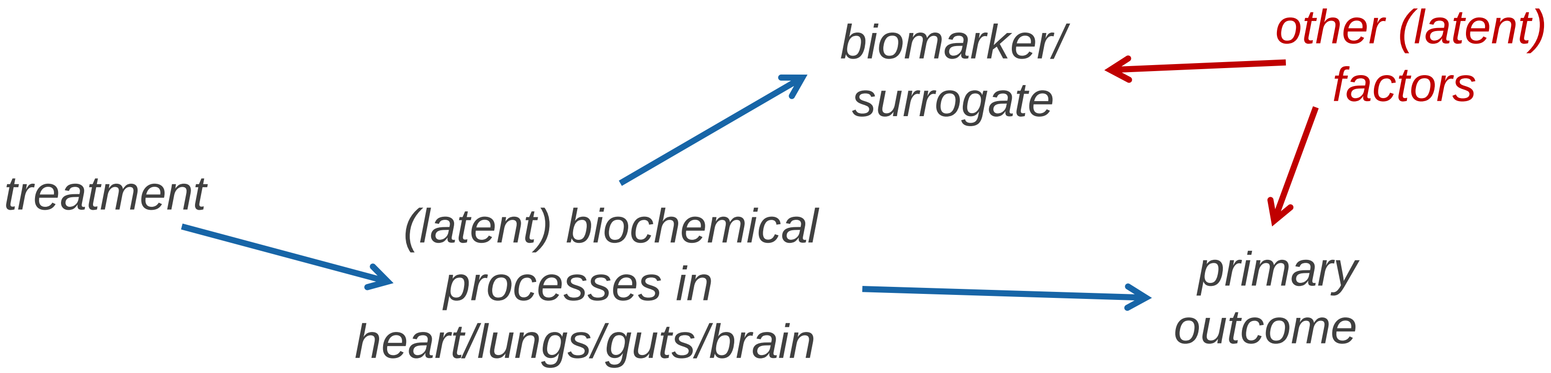
- Identified under weaker assumptions than NDE/NIE
  - Still, need to control for mediator-outcome confounding
  - Still, no cross-world independence assumption

- Especially: can select adaptive intervention to reflect anticipated likely use in practice
  - Perhaps ‘always / never’ not realistic

Biomarker / Surrogate Outcome?

IMHO, also not a mediation question:

- Is *M* a “suitable” surrogate? *What makes it suitable?*

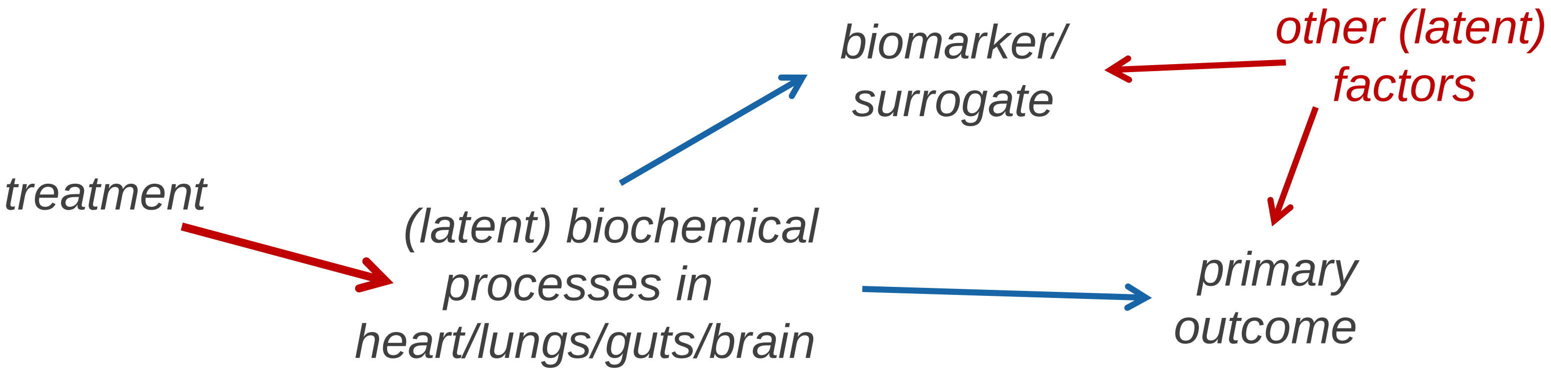


- Exception: if *M* is deterministic measure of latent process
- Example: plenty of examples in clinical research (I think)

Biomarker / Surrogate Outcome?

IMHO, also not a mediation question:

- Is *M* a “suitable” surrogate? *What makes it suitable?*



- **If causal structure of latent processes / factors ignored, analysis may still look like “plausible” mediation**

Interventions on Mediator?

RQ – mediator intervenable:

- Where best to intervene
- Controlled direct effect
- Joint/seq./dyn. interventions
- …and variations

RQ – mediator not intervenable

- Multiple effects on outcomes
  - …or variations thereof
- *Proposed: NDE/NIE ?*
  or principal-strata (e.g. SACE) ?
  - *… but not actionable*

Alternative interventions on *M* ?

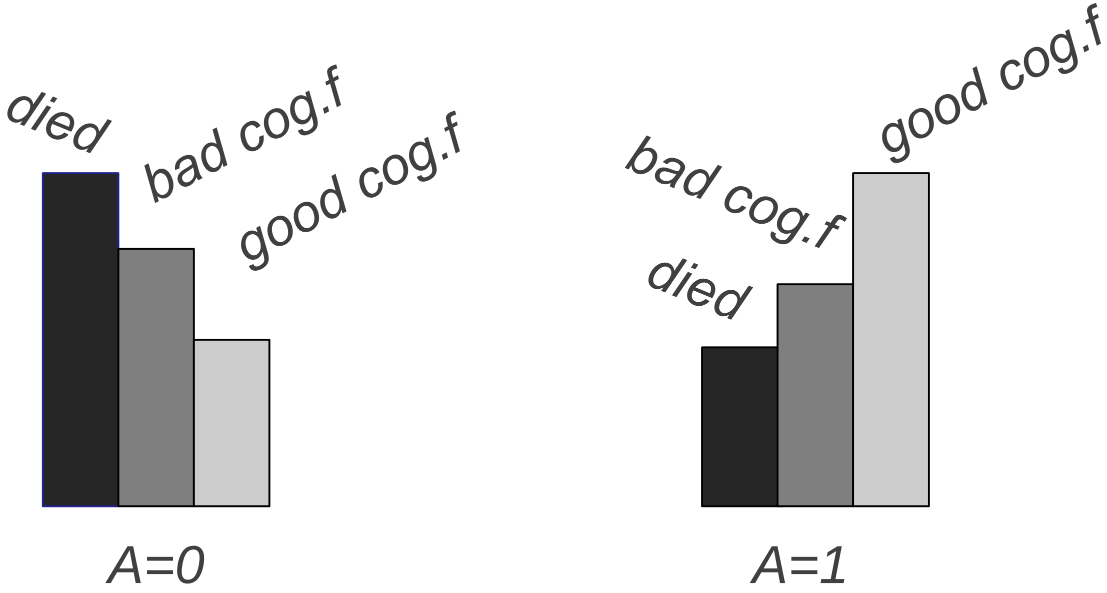
- Adaptive/dyn. Interventions
- Random / shift etc.
- “Randomised interventional (in)direct effects”

Mediator = Survival?

- Example: if outcome *Y* requires patients to be alive
  - e.g., treatment targets cognitive function in elderly
  - Competing events etc.
  - Cannot intervene on survival itself
  - *Plus:* cannot measure *Y* in deceased
- Survival status can only be ignored if you are 500% sure that treatment does not causally affect survival
- Otherwise: consider ‘many effects’ of *A*, on survival and *Y* ?
  - Often deemed unsatisfactory

Mediator = Survival?

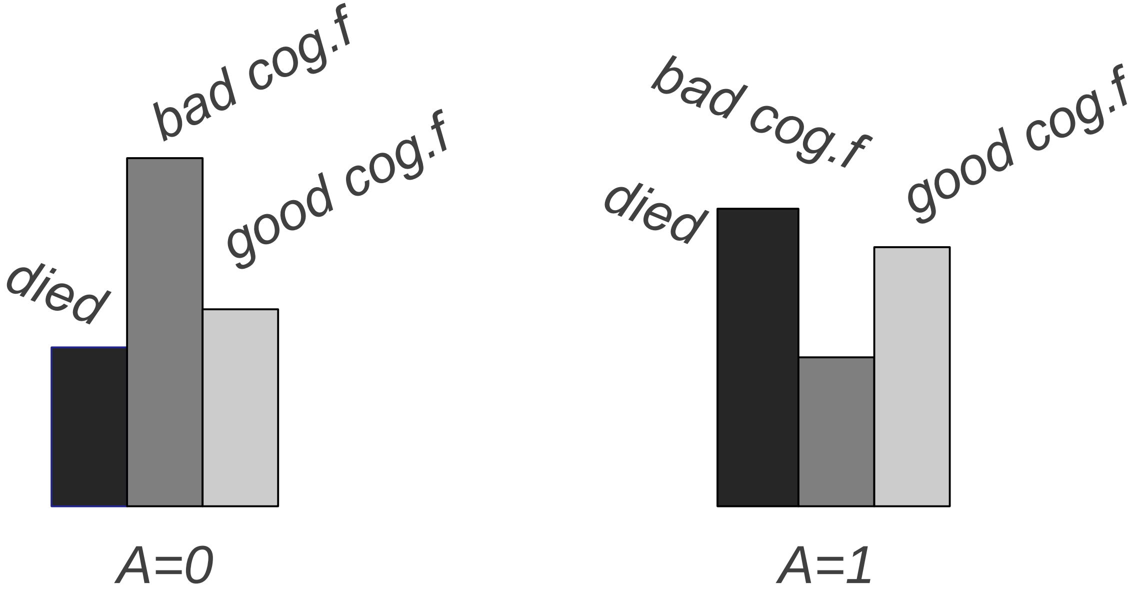
Example: treatment *A* for cognitive function *Y – total effect*



→ Treatment is all-round beneficial

Mediator = Survival?

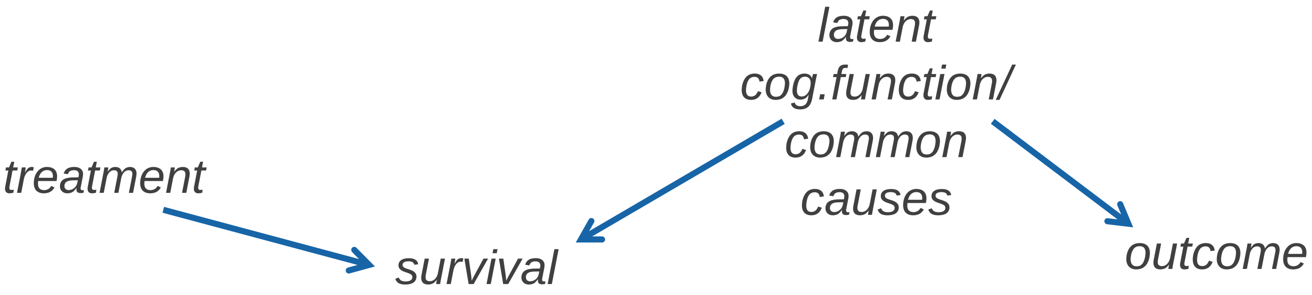
Example: treatment *A* for cognitive function *Y*



→ Treatment: bad for survival, good for cognitive function (total effect)

Mediator = Survival?

Worry: treatment just ‘kills’ those with ‘bad outcomes’



treatment

survival

latent cog.function/ common causes

outcome

This **cannot be resolved** without careful consideration / investigation of common causes affecting survival and outcome

- *No experimental design, as survival cannot be randomised*

Separable Treatments

expanding the story

History of ideas

- *Robins & Richardson (2011:book.chap): ‘manipulable mediation effects’*
- *Didelez (2019:LiDA): use for longitudinal / time-to-event settings*
- *Stensrud et al. (2022:JASA): coined ‘separable effects’ in context of competing events*

Separable Treatments

Example 1: weight-loss programme

- “Wanted: direct and indirect effects of weight-loss programme”
- Why?
- Weight-loss programme has different components:
  - Information about and (intended) change of diet
  - Regular meetings → socialising → improved mental health

⇒ Interested in hypothetical modified programme with or without meetings or with different attendance / frequency of meetings

Separable Treatments

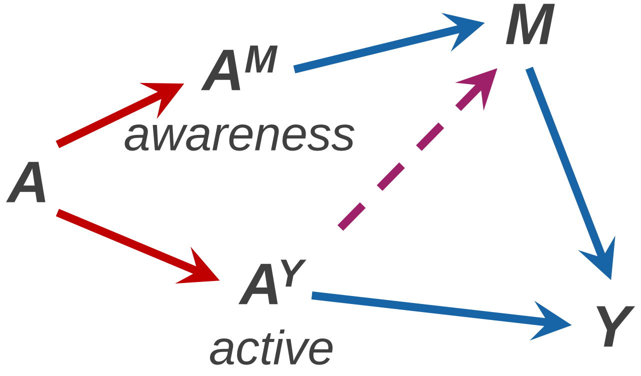
Example 2: placebo controlled randomised trials

- Separate treatment components:

  (1) active ingredient

  (2) psychology of awareness of treatment

- RQ: treatment better than placebo?
  Or: direct effect or active ingredient



Remarkable:

- *Actual* trials – no (cross-world) or other assumptions

But: ‘Separability’ violated, e.g. if strong side-effects induce ‘unblinding’

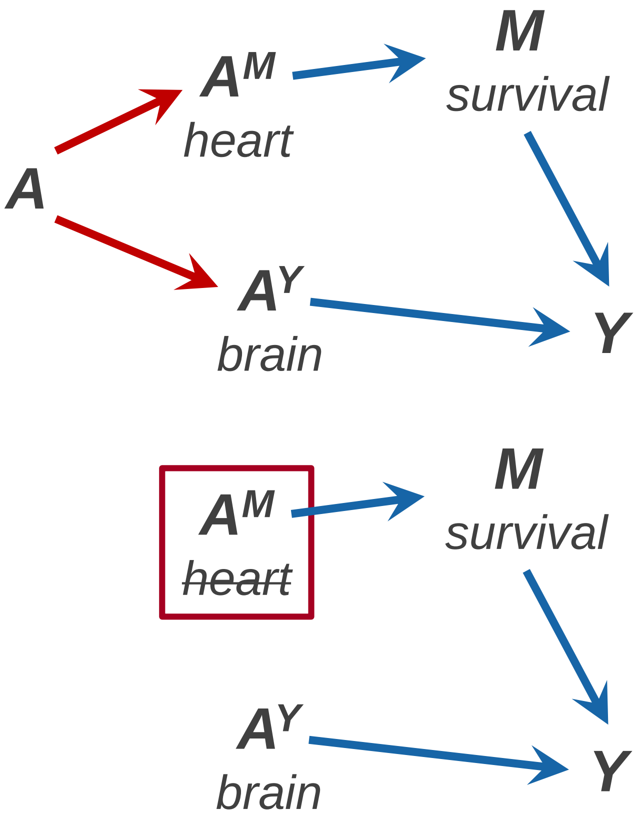
Separable Treatments

Example 3: modified treatments

- Treatment intended to improve cognitive function, but also affects cardiov.system → survival

- RQ: modified treatment with same effect on cogn. function but without affecting the cardiov.system?

⇒ Separable direct effect



Separable Treatments

Remarkable:

- Same identifying functional (mediational g-formula) as NDE/NIE in the simple mediator setting
  - Similar to certain paths-specific effects in longitudinal settings

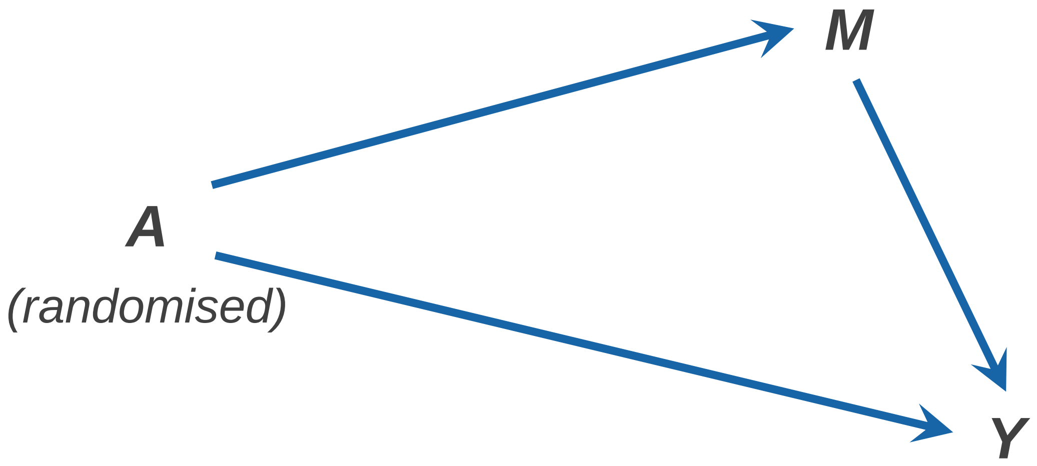
  ⇒ Can use existing methods of estimation
- Similar identification in longitudinal / in competing events settings
- Can view separable treatments as single-world reformulation of “natural” effects concepts (albeit an expanded single world)
- But more conducive to elicit actionable / useful research questions

Assumptions?

RQ answerable under ***defendable*** assumptions?

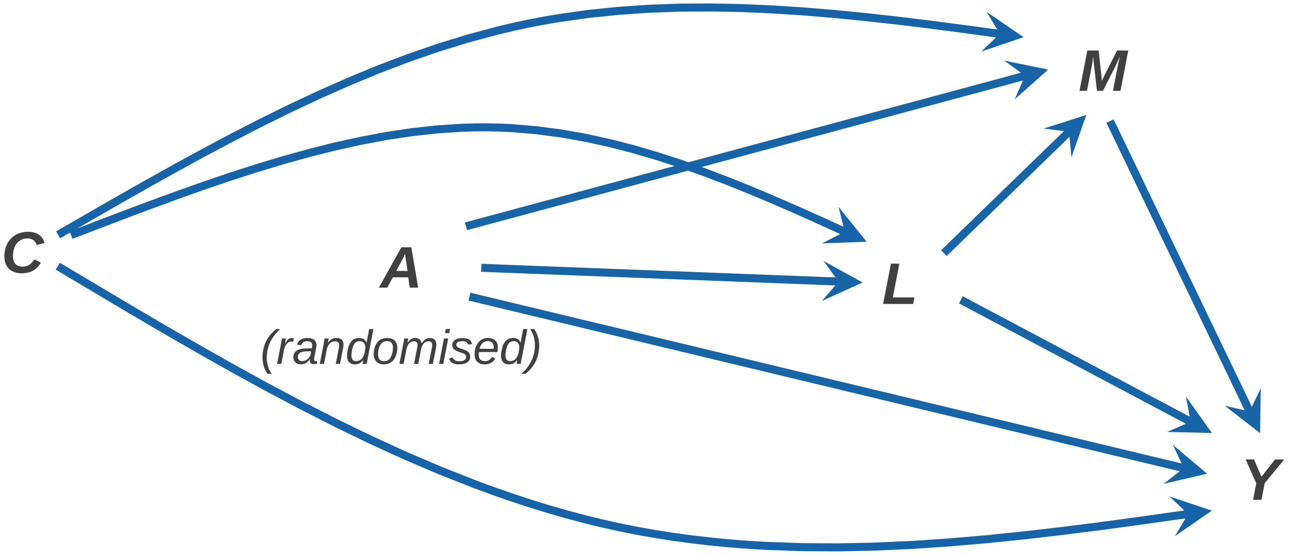
Assumptions

- **Misleading** DAG:



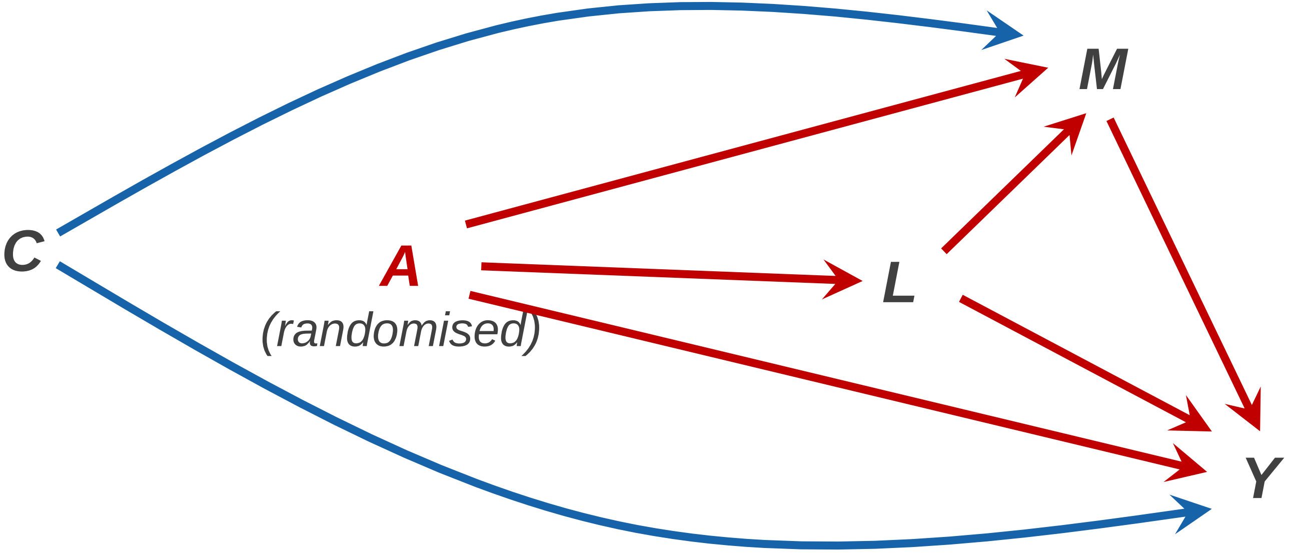
Assumptions

- More realistic DAG



Assumptions

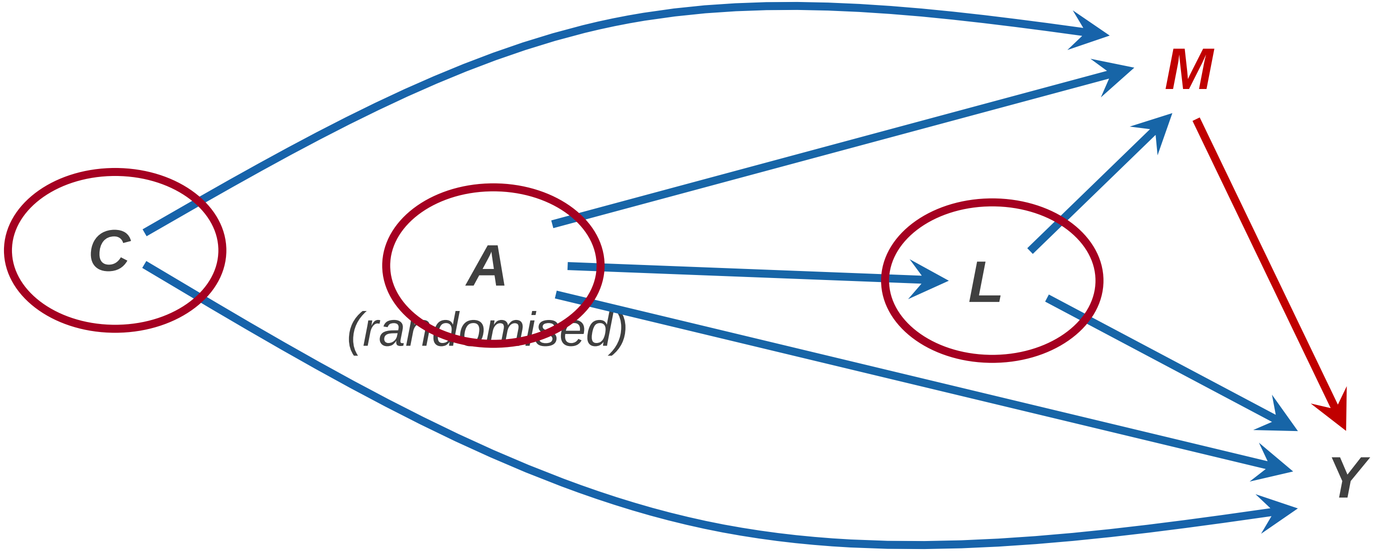
- What is identified?



- Total effects of *A* on *M* and on *Y*, with **no further assumptions**
  - Even in the processes case

Assumptions

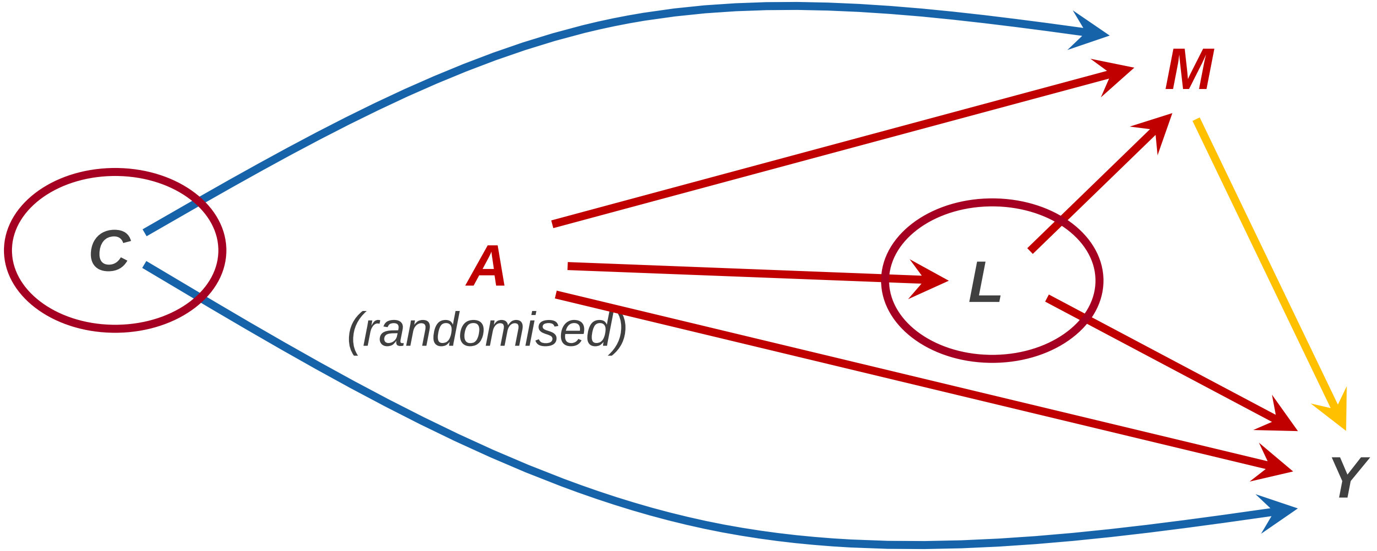
- What is identified?



- Total effect of *M* on *Y* : **must adjust for (*A,C,L*)**

Assumptions

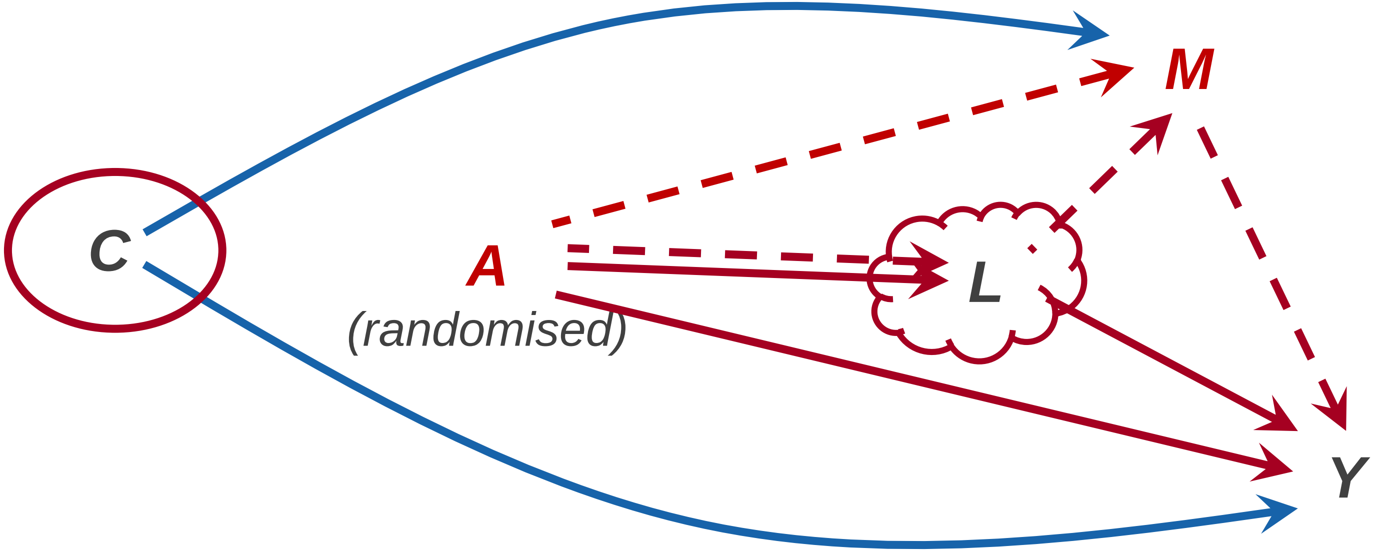
- What is identified?



- Joint / seq. / dyn. / controlled direct effects of *A* involving fixing or (imperfectly) intervening on *M* : **adjust for (*C*,*L*)**

Assumptions

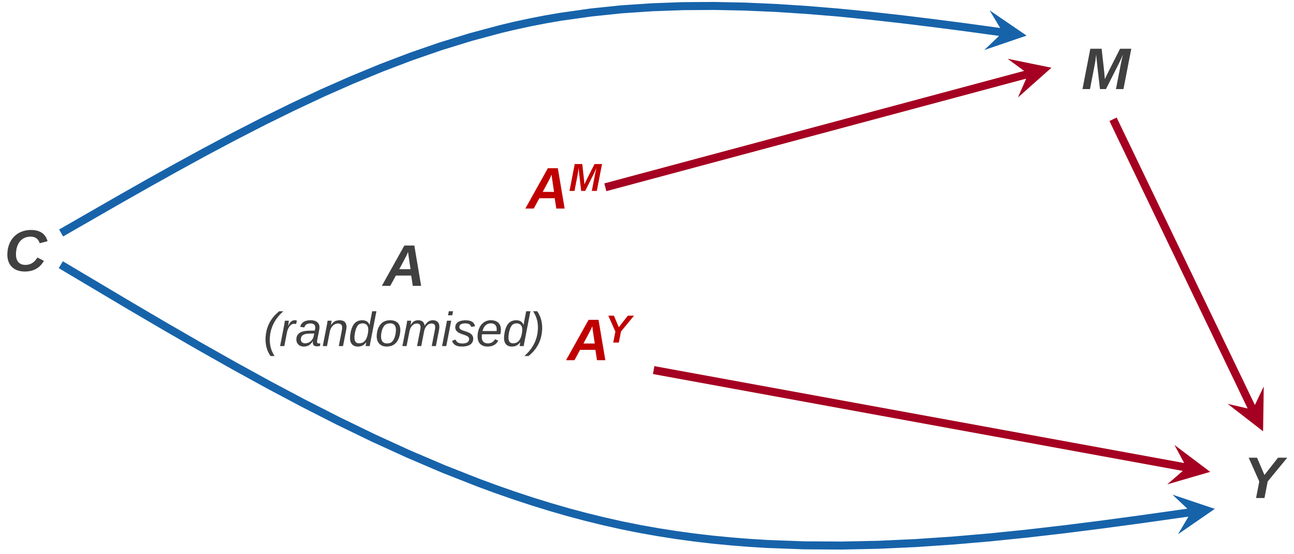
- What is identified?



- NDE / NIE: adjust for *C*;
  but for non-empty *L*: **not identified even if measured**
  - *Cross-world independence* assumption not (entirely) in the DAG

Assumptions

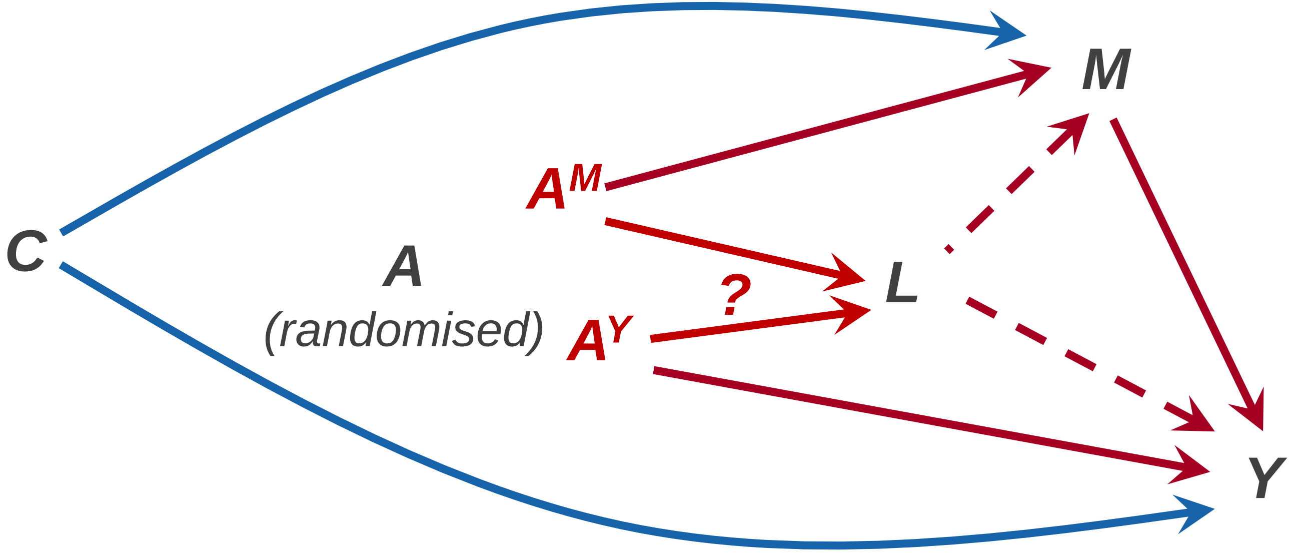
- What is identified?



- Separable effects: separability? Must adjust for *C* (*L* empty)

Assumptions

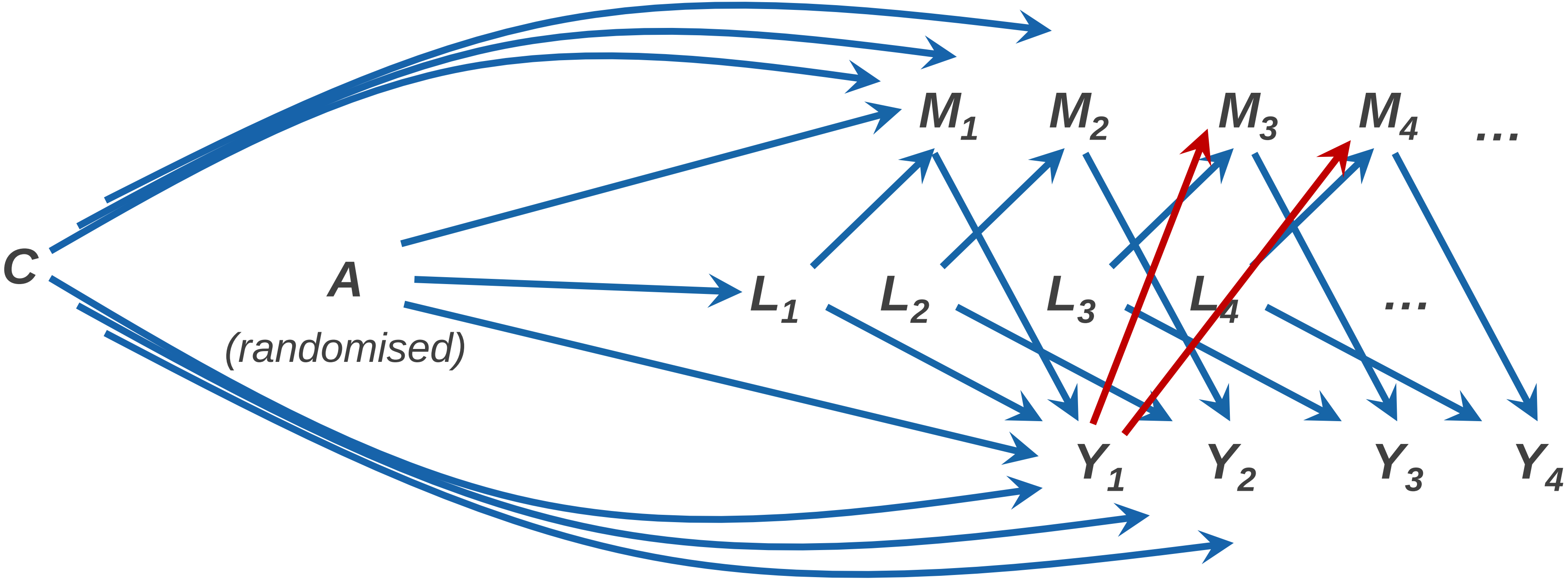
- What is identified?



- Separable effects: with L must clarify relation b/w (A^M,A^Y) and L
  - ***Reformulate research question / estimand!***

Assumptions

- Even more realistic DAG



Summary

- In putative mediation settings: many actionable / useful estimands are not actually about mediation
  - *Multiple total effects? Adaptive (dynamic) interventions?*
  - Single-world assumptions, thus (in principle) empirically verifiable
- Note: practical usefulness of mediational estimands that rely on cross-world assumptions for identification is mysterious – why consider them at all?
- *Sometimes* useful: *expand the story* to hypotheticals like imperfect interventions or separable treatments
  - Can help with eliciting actionable research questions

Thanks for listening!

www.leibniz-bips.de/en

Contact
Vanessa Didelez
Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen
didelez@leibniz-bips.de