

Mediation

Jon Michael Gran

Oslo Centre for Biostatistics og Epidemiology (OCBE),
Department of biostatistics, Institute of basic medical sciences, UiO

j.m.gran@medisin.uio.no

March, 2022



UiO : University of Oslo

Mediation

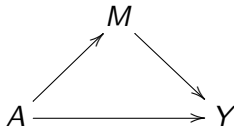
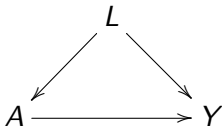
- Consider a **treatment A** , **outcome Y** , mediator **M** :



- Just establishing the effect of **A on Y** is **challenging**
- Estimating **the mediating effect** of A on Y via M is *very ambitious*

Mediation

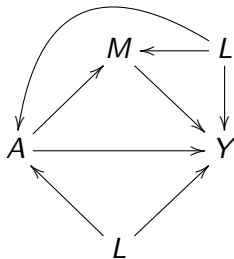
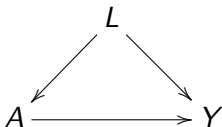
- Consider a **treatment A** , **outcome Y** , **mediator M** and set of possible confounders L :



- Just establishing the effect of **A on Y** is **challenging**
- Estimating **the mediating effect** of A on Y via M is *very ambitious*

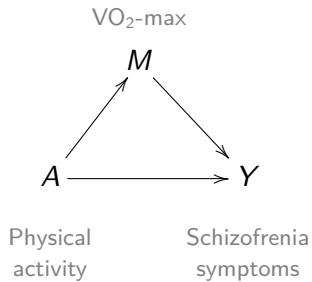
Mediation

- Consider a **treatment A** , **outcome Y** , mediator **M**

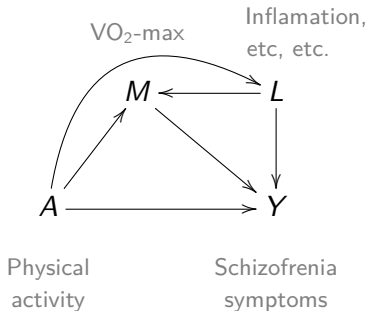


- Just establishing the effect of **A on Y** is **challenging**
- Estimating **the mediating effect** of A on Y via M is *very ambitious*

For example

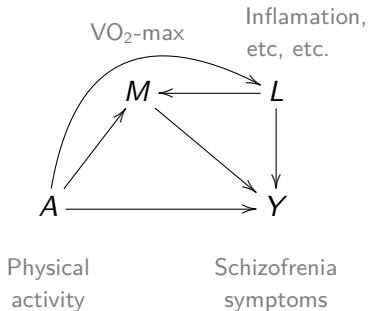


For example



- Even though A is randomised, a study of mediation through M is by all means **an observational study**

For example



- Even though A is randomised, a study of mediation through M is by all means **an observational study**
- **Timing of measurements** matter

A brief history of mediation analysis

- Goes back to **Wright (1921, 1934)**
- Originally using **linear models** without interactions, fitting

$$\begin{aligned}Y &= \alpha_0 + \alpha_1 A + \alpha_2 M \\ M &= \beta_0 + \beta_1 A,\end{aligned}$$

where α_1 would be called the direct effect and $\beta_1\alpha_2$ the indirect effect (with various extensions)

- Popularised in social science by **Baron, Kenny (1986)**, currently with 110 000 citations in Google scholar
- **Causal inference** focus following Robins, Greenland (1992) and Pearl (2001); see VanderWeele (2015)

Causal inference

- ① **Formulate the research question**
(in terms of a hypothetical randomised trial?)
- ② **Formulate the research question mathematically**
(the estimand)
- ③ **Lay out the assumptions** you need
- ④ **Estimate** from data

Causal inference

- ① **Formulate the research question**
(in terms of a hypothetical randomised trial?)
- ② **Formulate the research question mathematically**
(the estimand)
- ③ **Lay out the assumptions** you need
- ④ **Estimate** from data

Starting at step 4 is not a good idea...

Estimands

- **Natural direct effects:**

$$E(Y^{a=1,M^1}) \text{ vs. } E(Y^{a=0,M^1})$$

- **Natural indirect effects:**

$$E(Y^{a=0,M^1}) \text{ vs. } E(Y^{a=0,M^0})$$

Estimands

- **Natural direct effects:**

$$E(Y^{a=1, M^1}) \text{ vs. } E(Y^{a=0, M^1})$$

- **Natural indirect effects:**

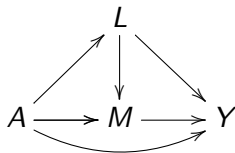
$$E(Y^{a=0, M^1}) \text{ vs. } E(Y^{a=0, M^0})$$

$Y^{a=0, M^1}$ and alike are nested counterfactual variables.

Also called pure indirect and direct effects.

Direct effect + indirect effect = total effect.

- Consider a **treatment A** , **outcome Y** and **mediator M** , with a so-called intermediate confounding variable L :



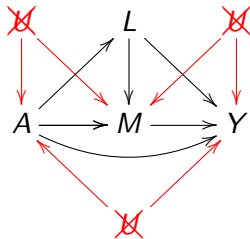
Assumptions

- **No unmeasured exposure-outcome confounding**
- **No unmeasured mediator-outcome confounding**
- **No unmeasured exposure-mediator confounding**
- The **cross-world assumption** (implying no intermediate confounding)

Assumptions

- **No unmeasured exposure-outcome confounding**
- **No unmeasured mediator-outcome confounding**
- **No unmeasured exposure-mediator confounding**
- The **cross-world assumption** (implying no intermediate confounding)

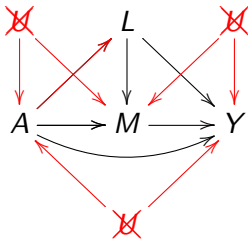
In addition to: consistency, positivity, no interference, no measurement error and no model misspecification.



- The **three first assumptions** is about no unmeasured confounding^a
- The cross-world assumption basically mean **no intermediate confounders**^b

^aThe no unmeasured mediator-outcome confounding assumption largely ignored by Baron, Kenny (1986) and followers.

^bMultiple mediators cause similar problems. Assumption can be relaxed, but alternative still very strong and not possible to verify, not even hypothetically.

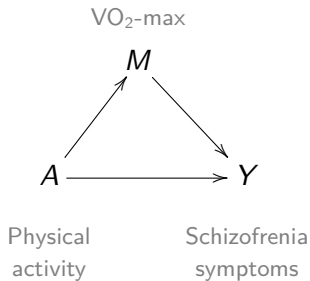


- The **three first assumptions** is about no unmeasured confounding^a
- The cross-world assumption basically mean **no intermediate confounders**^b

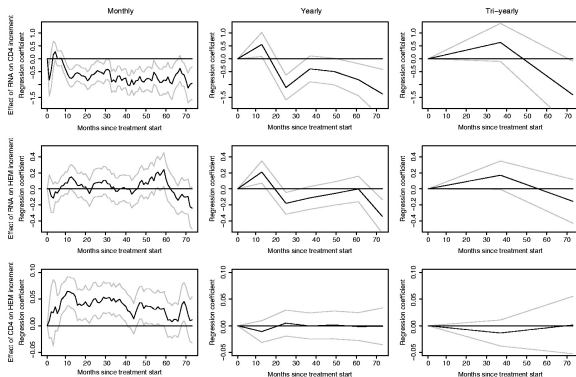
^aThe no unmeasured mediator-outcome confounding assumption largely ignored by Baron, Kenny (1986) and followers.

^bMultiple mediators cause similar problems. Assumption can be relaxed, but alternative still very strong and not possible to verify, not even hypothetically.

Longitudinal mediators



- *M* is really **a processes**, interacting with other processes, what is captured with one measurement? Or two? Or five?
What does "mediated through *M*" mean?



- Representing a process with a single, or a few, measurements leads to potential huge **measurement error**^a
- At best, **affects interpretation** of "mediation"

^aSee Aalen et al. (2014) and e.g. Maxwell, Cole (2007).

Mediation analysis and research questions

- Sometimes, "mediation" questions are **not answered with mediation analysis**
- Helpful to re-formulating the question in terms of **hypothetical RCTs**
- *For example*, for physical activity $\rightarrow$ **VO₂-max** $\rightarrow$ Schizophrenia symptoms, can we consider a treatment regime of training until VO₂-max reach a specific level and staying there, vs a regime without VO₂-max improvement?
If not performing the trial, do we have the data to model it? Is VO₂-max measured often enough?

Mediation analysis and research questions

- Sometimes, "mediation" questions are **not answered with mediation analysis**
- Helpful to re-formulating the question in terms of **hypothetical RCTs**
- *For example*, for physical activity $\rightarrow$ **VO₂-max** $\rightarrow$ Schizophrenia symptoms, can we consider a treatment regime of training until VO₂-max reach a specific level and staying there, vs a regime without VO₂-max improvement?

If not performing the trial, do we have the data to model it? Is VO₂-max measured often enough?

If it is hard to describe a conceptual randomised trial that capture the mediating effect of VO₂-max, is it reasonable to rely on "methods" to do better?

Final remarks

- With **causal inference** we have better understood the requisites for doing mediation analysis

Mediation analysis demands a lot from your data
(in what's being measured, how/when, and in sample size)

Knowledge about mediation also key also when *discussing* observed behaviour of measured, possibly mediated, variables

- More **plug and play software available**, for various types of mediators and outcomes (f.ex. medflex and mediation in R, paramed in Stata)

Should we do *more* mediation analysis?

- Not talked about **estimation** (beyond software)
For linear models without interactions, it is basically down to Baron and Kenny (1986)
- Due to the complexity of mediation problems, we should **be extra careful**, also with the conclusions, and (extra) critical when reading applied literature
- Assessing **interaction** can also shed light on *mechanisms* (VanderWeele, 2015)
- Can do **sensitivity analysis** to f.ex. put bounds on mediating effects and study how big unmeasured confounding effects need to be to explain away effects

References

-  Vanderweele T. Explanation in Causal Inference. Methods for Mediation and Interaction. Oxford University Press, 2015.
-  Aalen O, Røysland K, Gran JM, Kouyos R, Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. SMMR, 2014.
-  Ding P, VanderWeele T. Mediation E-values: approximate sensitivity analysis for unmeasured mediator-outcome confounding. Epidemiology, 2019.
-  Maxwell SE, Cole DA. Bias in Cross-Sectional Analyses of Longitudinal Mediation. Psychological methods, 2007.



Didelez V. Defining Causal Mediation With a Longitudinal Mediator and a Survival Outcome. LIDA, 2019.



Stensrud MJ, Young JG, Didelez V, Robins JM, Hernan, MA. Separable Effects for Causal Inference in the Presence of Competing Events. JASA, 2022.