Mediation analysis

ehsan.karim@ubc.ca

Oct 22, 2020

SPPH 504/007

Ref

- Reference for reading

[HTML] Applied mediation analyses: a review and tutorial

T Lange, KW Hansen, R Sørensen… - Epidemiology and …, 2017 - ncbi.nlm.nih.gov

In recent years, mediation analysis has emerged as a powerful tool to disentangle causal pathways from an exposure/treatment to clinically relevant outcomes. Mediation analysis has been applied in scientific fields as diverse as labour market relations and randomized clinical trials of heart disease treatments. In parallel to these applications, the underlying mathematical theory and computer tools have been refined. This combined review and tutorial will introduce the reader to modern mediation analysis including: the mathematical

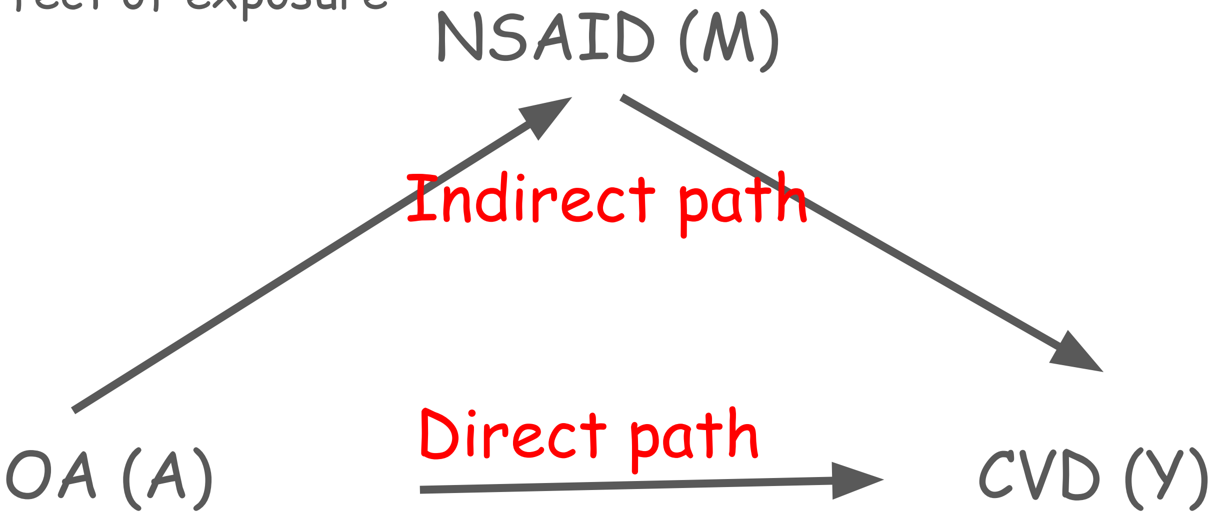
Cited by 13 Related articles All 14 versions Web of Science: 5 Import into

Notations & An Example

- **Exposure group (A)**: osteoarthritis (OA)
- **Control group**: Non-osteoarthritis (non-OA)
- **Outcome (Y)**: Cardiovascular disease (CVD)
- **Mediator (M)**: Pain medication (Nonsteroidal anti-inflammatory drugs / NSAID)
- **Confounder (C)**: Age, sex, BMI, SES, comorbidity

Total effect

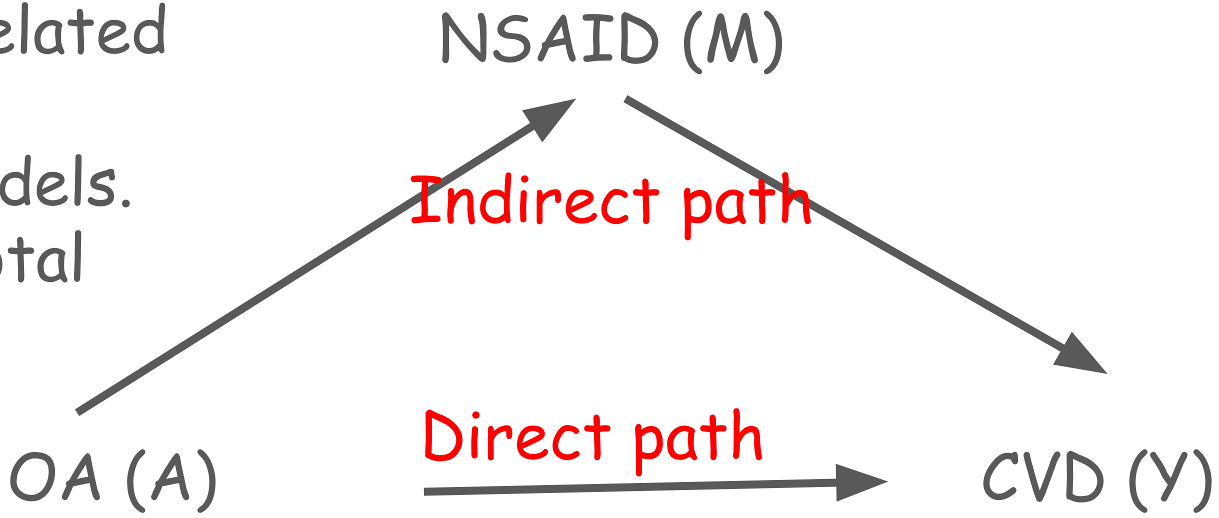
- Outcome model
  - CVD ~ intercept + b * OA (assuming no confounder present)
  - NSAID is not controlled. Why?
  - b = total effect of exposure



(Statistical) Mediation analysis

DAG representation:

- ***Translate*** loose causal path-related concepts to statistical models.
- ***Decompose*** total effects to
  - Direct
  - Indirect



(Statistical) Mediation analysis

- Long history
  - Path analysis
  - Structural equation modelling
- Baron, Kenny paper from 1986

The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.

RM Baron, DA Kenny - Journal of personality and social ..., 1986 - psycnet.apa.org

In this article, we attempt to distinguish between the properties of moderator and mediator variables at a number of levels. First, we seek to make theorists and researchers aware of the importance of not using the terms moderator and mediator interchangeably by carefully elaborating, both conceptually and strategically, the many ways in which moderators and mediators differ. We then go beyond this largely pedagogical function and delineate the conceptual and strategic implications of making use of such distinctions with regard to a ...

Baron and Kenny (1986) approach I

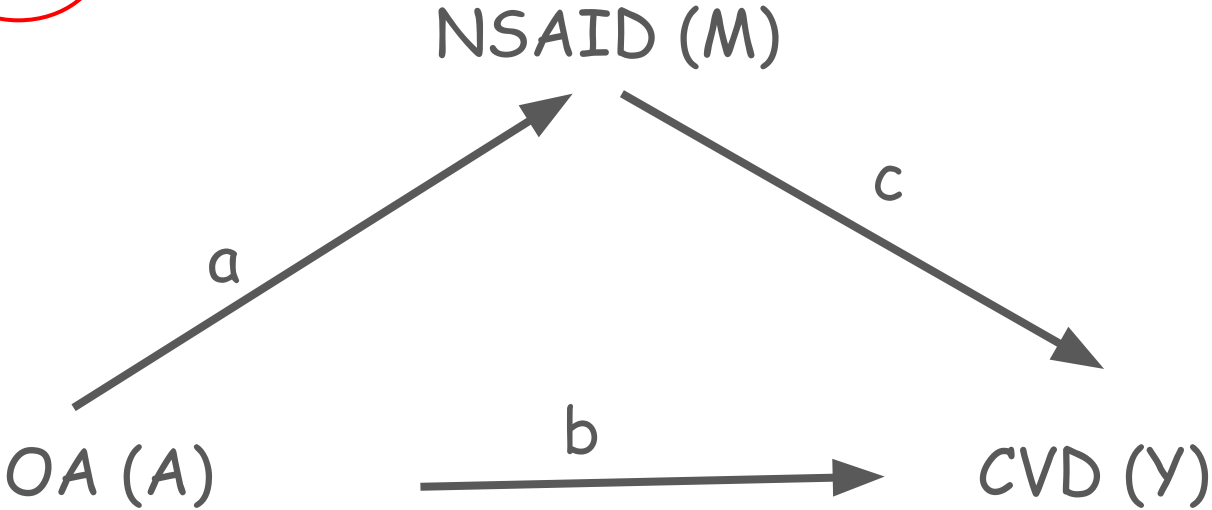
$$Y = \text{intercept} + b_{YAM} \cdot A + c_{YAM} \cdot M$$

$$Y = \text{intercept} + b_{YA} \cdot A$$

Direct effect = b_{YAM} (M-adj)

Total effect = b_{YA}

Indirect effect = $b_{YA} - b_{YAM}$



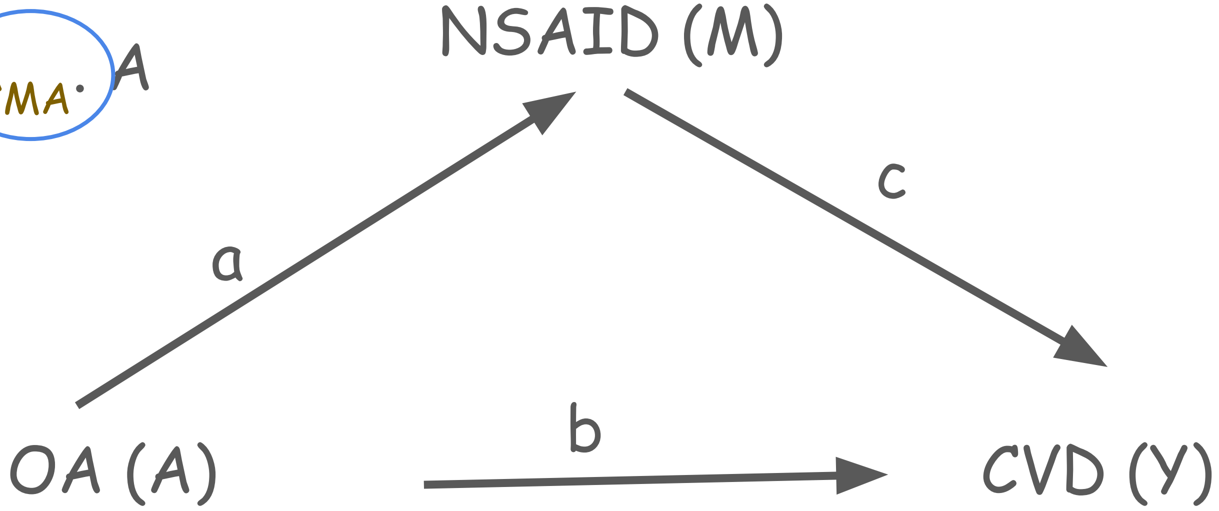
Baron and Kenny (1986) approach 2

$$Y = \text{intercept} + b_{YAM} \cdot A + c_{YAM} \cdot M$$

$$Y = \text{intercept} + b_{YA} \cdot A$$

$$M = \text{intercept} + a_{MA} \cdot A$$

$$\boxed{\begin{array}{l} \text{Indirect effect} = a_{MA} * c_{YAM} \\ b_{YA} - b_{YAM} = a_{MA} * c_{YAM} \end{array}}$$



Limitations of Baron & Kenny (1986) approach

1. Non-collapsibility (OR / HR)
    a. Change-in-estimate approach does not work
    b. Product of coefficient methods hard to interpret
    c. Only continuous outcomes/linear model work (beta-coef)
        i. Generally not sure what we are really estimating
        ii. In particular, when confounding exists
2. How to address confounding?
    a. Not clear
    b. Need modern mediation methods based on counterfactual

Counterfactual definition

Exposed:

OA ($A=1$)

NSAID ($M=1$)

What is the effect of OA on CVD?

Control:

Non-OA ($A=0$)

No NSAID ($M=0$)

Total effect

Counterfactual definition

Exposed:	Counterfactual to exposed:	Control:
OA ($A=1$)	OA ($A=1$)	Non-OA ($A=0$)
NSAID ($M=1$)	No NSAID ($M=0$)	No NSAID ($M=0$)

Counterfactual definition

Exposed:

OA (A=1)

NSAID (M=1)

Counterfactual to exposed:

OA (A=1)

No NSAID (M=0)

Control:

Non-OA (A=0)

No NSAID (M=0)

Direct effect ($A = 1$ vs $0 | M$)

Counterfactual definition

Exposed:	Counterfactual to exposed:	Control:
OA ($A=1$)	OA ($A=1$)	Non-OA ($A=0$)
NSAID ($M=1$)	No NSAID ($M=0$)	No NSAID ($M=0$)

Indirect effect ($M = 1$ vs $0 \mid A$)

Counterfactual definition

Potential outcomes for 1 person:

1. Y (A=1) = CDV status when OA = 1
2. Y (A=0) = CDV status when OA = 0 / non-OA

- Total effect for a group = E[Y(A=1)] vs. E[Y(A=0)]

  (ratio for binary such as CDV = 0 vs. 1; then E[Y] is replaced by Probability Pr(CVD = 1); difference for continuos Y)

Counterfactual definition

Potential outcomes when mediator (M) is present:

1. Y (A=1, M=0) = CDV status when OA = 1, M = 0 (no NSAID)
2. Y (A=0, M=0) = CDV status when OA = 0, M = 0 (no NSAID)
3. Y (A=1, M=1) = CDV status when OA = 1, M = 1 (uses NSAID)
4. Y (A=0, M=1) = CDV status when OA = 0, M = 1 (uses NSAID)

- **Direct effect** = E[Y(A=1, M =0)] vs. E[Y(A=0, M=0)]
- **Direct effect** = E[Y(A=1, M =1)] vs. E[Y(A=0, M=1)]

Direct effect is generally known as NDE (**fixed M**).

Counterfactual definition

Potential outcomes when mediator (M) is present:

1. Y (A=1, M=0) = CDV status when OA = 1, M = 0 (no NSAID)
2. Y (A=0, M=0) = CDV status when OA = 0, M = 0 (no NSAID)
3. Y (A=1, M=1) = CDV status when OA = 1, M = 1 (uses NSAID)
4. Y (A=0, M=1) = CDV status when OA = 0, M = 1 (uses NSAID)

- **Indirect effect** = E[Y(A=1, M =1)] vs. E[Y(A=1, M=0)]
- **Indirect effect** = E[Y(A=0, M =1)] vs. E[Y(A=0, M=0)]

Indirect effect is generally known as NIE (**fixed A**).

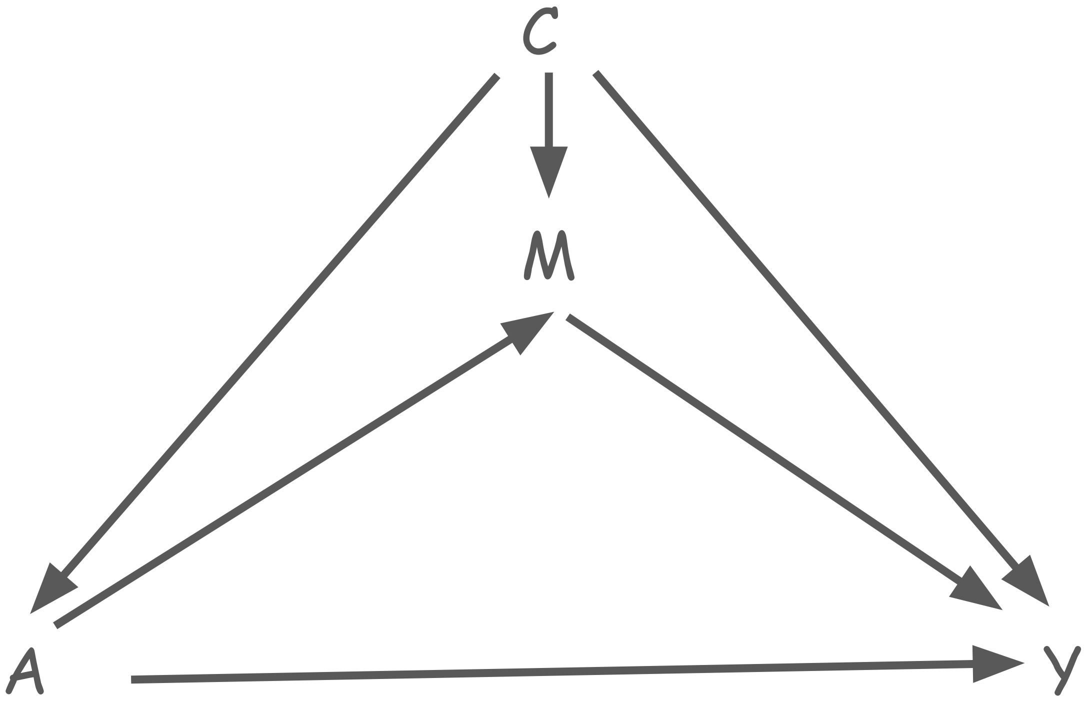
Modelling in RCT: Adjusting for C necessary?

In RCT,

- A is randomized.
- But M is not.
- In a mediation analysis, we are essentially trying to estimate effect of <u>2 exposure group</u> (A and M).
- Hence, we necessarily need to adjust for <u>confounders</u> C in both
    - Y~A and M~A relationships.
    - Not much different than observational case.

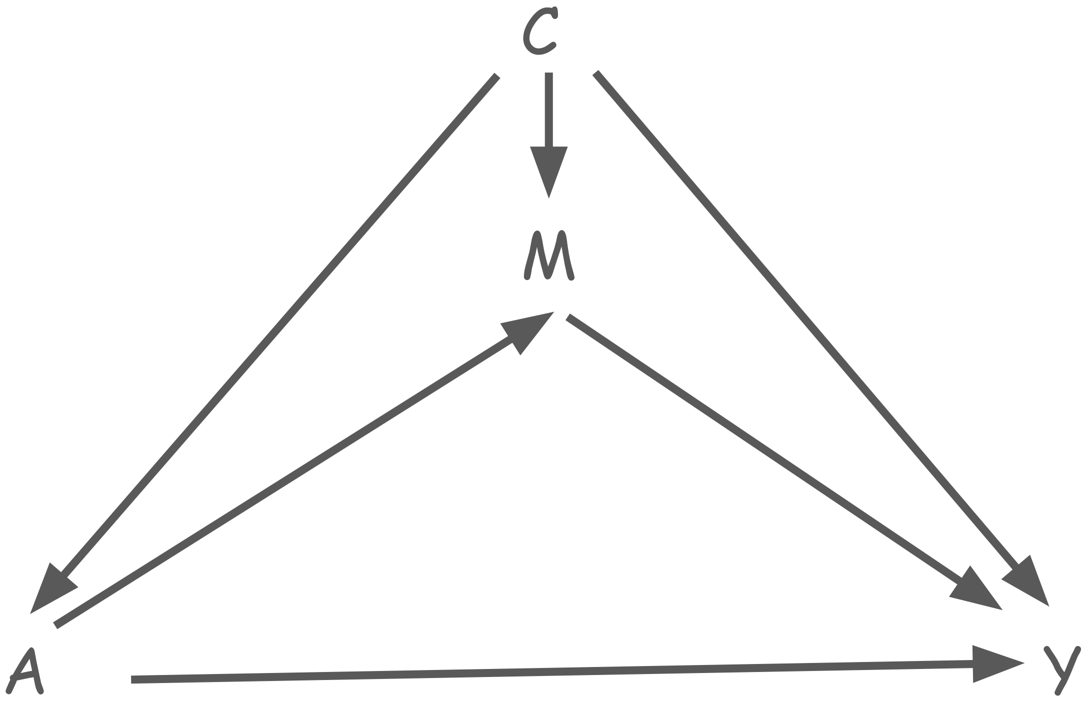
Mediation analysis

DAG with a confounder



Mediation analysis

The main problem with the **counterfactual** approach implementation is that we <u>do not observe both counterfactuals</u>:

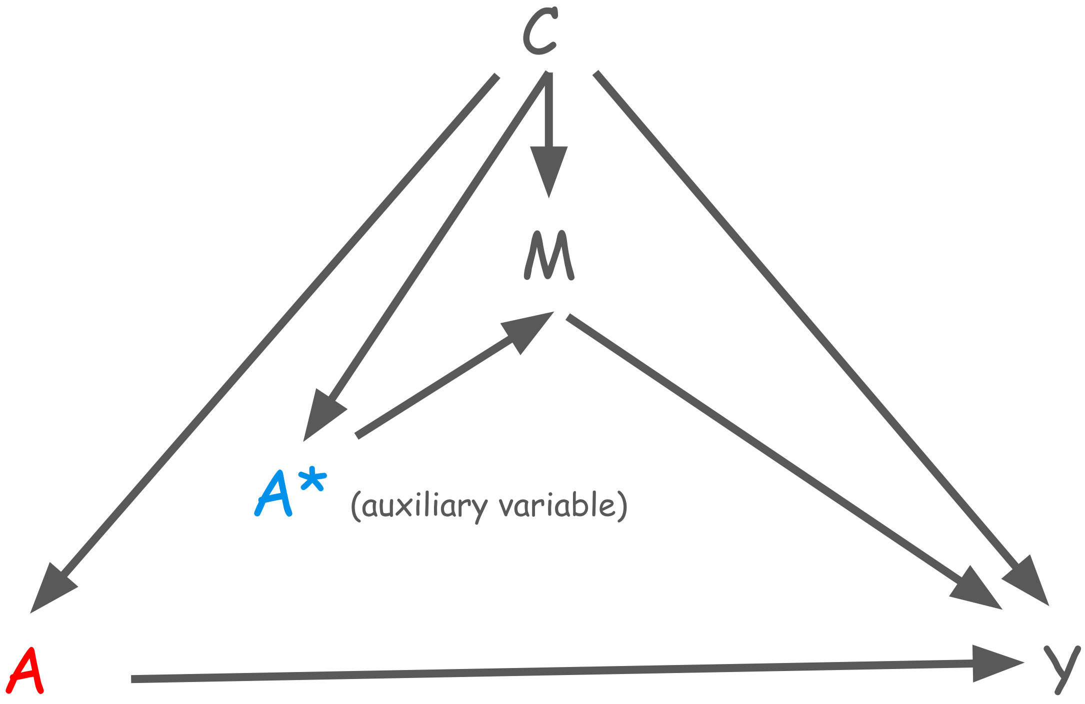


- Either observe Y(A=1, M=1) or Y(A=0, M=1): hence <u>*can't contrast*</u> to obtain NDE
- How can we get direct effect if we can't have same M (fixed M in 2 individuals will allow NDE estimation)

Mediation analysis

Modified DAG to understand modelling better:

Decomposing direct vs. indirect parts of exposure (OA) to the outcome (CVD).



- This will somewhat allow us to contrast if we had 2 copies of *A*

Mediation analysis: Mechanism

- Step 0:
  - Include Y, A, M in the data and necessary C (C could be more than 1)
- Step 1:
  - Replicate exposure A with same exposures A* ('facts')
- Step 2:
  - Replicate exposure A with alternative exposures A* ('alternative facts')
- Step 3: (2 approaches)
  - Impute Y ~ A + M + C or Model M ~ A+C vs. A*+C for weighting
- Step 4: (2 approaches)
  - Fit outcome model Y ~ A + A* + C on the imputed/weighted data

Mediation analysis: Mechanism

Assuming Y is continuous for the moment, our original data should look like this (**step 0**):

ID	C	M	A	Y
1	1	0	1	100
2	0	1	0	50
...				

Mediation analysis: Mechanism

Now we add another variable $A^* = A$ (step 1):

ID	C	M	A	A*	Y
1	1	0	1	1	100
2	0	1	0	0	50
...					

Mediation analysis: Mechanism

- Now we add another row where $A^* = \text{not } A$ (step 2):
- But don't impute Y yet in this new rows.

ID	C	M	A	A*	Y	W
1	1	0	1	1	100	1
1	1	0	1	0	?	?
2	0	1	0	0	50	1
2	0	1	0	1	?	?
...						

- Add column of W. W = 1 in original, W = ? in new rows.

Mediation analysis: approach 1 - Imputation

- (step 3a) Fit $Y \sim A + M + C$ using the original rows/data
- Impute missing Ys (using new data with A^*) = $E[Y|A=A^*,C=C]$

ID	C	M	A	A*	Y	W
1	1	0	1	1	100	1
1	1	0	1	0	?	?
2	0	1	0	0	50	1
2	0	1	0	1	?	?
...						

Note that fitting and imputing happening in different parts of the data.

Mediation analysis: approach 1 - Imputation

- After imputing Y: (**step 4a**) Fit Y ~ A + A* + C

ID	C	M	A	A*	Y	W
1	1	0	1	1	100	1
1	1	0	1	0	**70**	?
2	0	1	0	0	50	1
2	0	1	0	1	**60**	?
...						

Coef of A = direct, Coef of A* = indirect

Mediation analysis: approach 2 - weighting

- (step 3b) Fit: M ~ A + C, using the original rows/data
- Use fit to predict M~A*+C & M~A+C in **all data** (new + old)

ID	C	M	A	A*	Y	W
1	1	0	1	1	100	1
1	1	0	1	0	?	?
2	0	1	0	0	50	1
2	0	1	0	1	?	?
...						

Calculate W = (fitted values from model with A*) / (fitted values from model with A)

Mediation analysis: approach 2 - weighting

- (step 4b) Fit Y ~ A + A* + C, when W is the model weight
- Keep original Y for the new rows

ID	C	M	A	A*	Y	W
1	1	0	1	1	100	1
1	1	0	1	0	**100**	**1.5**
2	0	1	0	0	50	1
2	0	1	0	1	**50**	**0.7**
...						

Coef of A = direct (NDE), Coef of A* = indirect (NIE)

Mediation analysis: SE?

How to get correct SE as we are dealing with double observations (new + old):

1. We can find the robust SE
    - by including a cluster(ID) option in the final model.
2. We can simply bootstrap
    - b=large # of replications.

Mediation analysis: Sensitivity analysis

- Mediation model
  - Non-linear relationships
    - Polynomials
    - Interactions between A and C

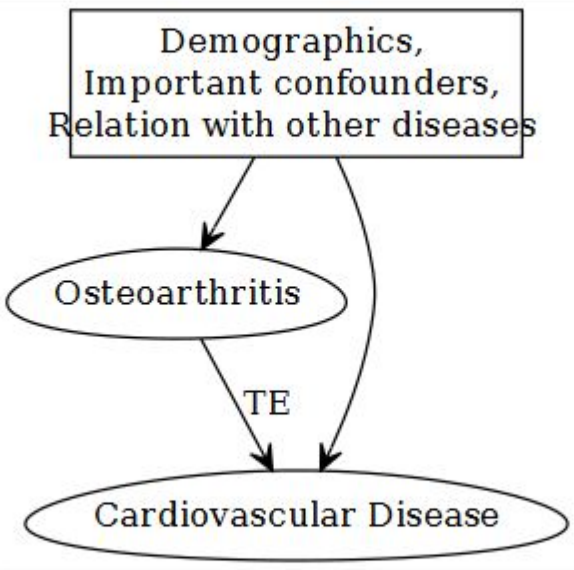
Mediation analysis: PM

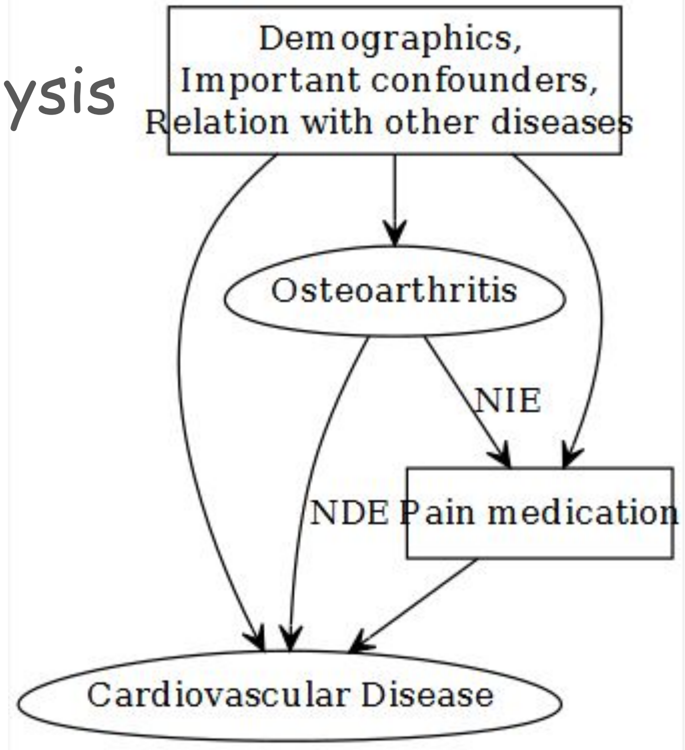
Proportion mediated (PM):

1. the proportion of the effect (in A-Y) that is being mediated via the mediator
2. **PM** = indirect effect/total effect
3. Possible calculate confidence intervals for PM

Our example

Omitting mediation vs mediation analysis



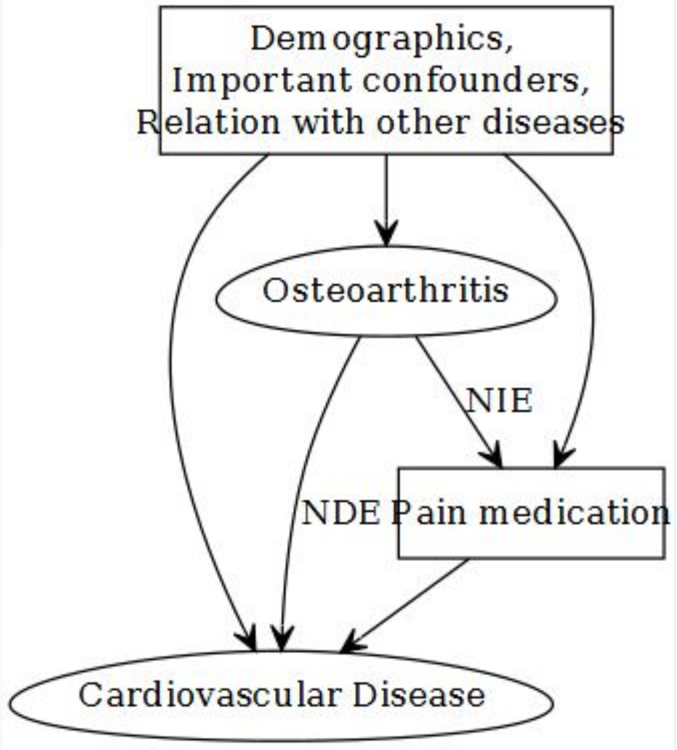


Our example

Mediation analysis justification:

- Check mediator model

Variable	Units	OddsRatio	CI.95	p-value
exposureTemp		2.43	[2.06;2.86]	< 1e-04
age	20-29 years	Ref		
	30-39 years	1.00	[0.88;1.13]	0.9442989
	40-49 years	0.93	[0.82;1.06]	0.2651302
	50-59 years	0.66	[0.58;0.76]	< 1e-04
	60-64 years	0.61	[0.51;0.72]	< 1e-04
	65 years and over	0.61	[0.52;0.71]	< 1e-04
sex	Female	Ref		
	Male	0.50	[0.46;0.55]	< 1e-04



Exposure is a significant predictor for the mediator.

Our example

Mediation analysis (after following steps): Bootstrap!

```
# Total Effect
c(bootresBin$t0[1], bootci1b$percent[4:5])
```

```
##        TE
## 1.544694 1.293208 1.894417
```

```
# Direct Effect
c(bootresBin$t0[2], bootci2b$percent[4:5])
```

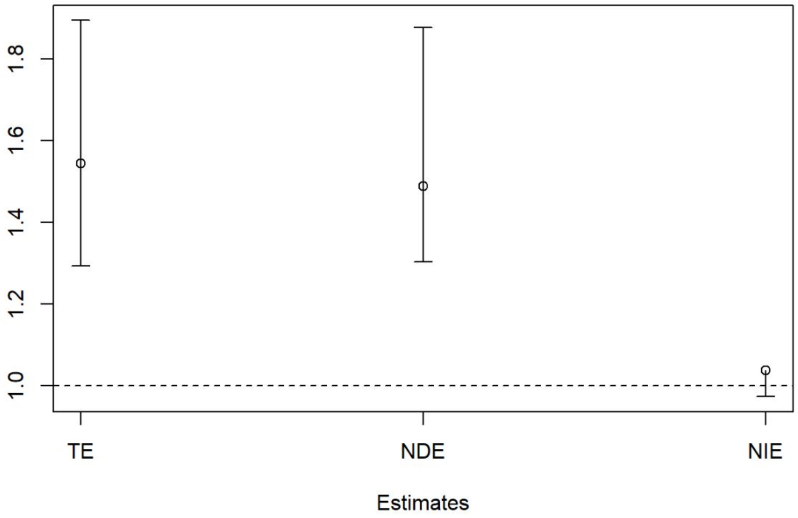
```
##        DE
## 1.488554 1.303554 1.876916
```

```
# Indirect Effect
c(bootresBin$t0[3], bootci3b$percent[4:5])
```

```
##         IE
## 1.0377144 0.9738072 1.0093246
```

```
# Proportion Mediated
c(bootresBin$t0[4], bootci4b$percent[4:5])
```

```
##            PM
##  0.08513902 -0.08360848  0.01655013
```



The proportion mediated through pain medication was about 8.5% on the log odds ratio scale.

Mediation analysis using survey data

- Outcome model needs to incorporate survey features
  - Strata
  - Cluster
  - weights
- Not clear if the mediator model need to include survey features
  - Same issue within the propensity score literature
  - We will incorporate the same idea
    - Mediator weights calculated omitting weights
    - Outcome regression will incorporate weights

Assumption - 1

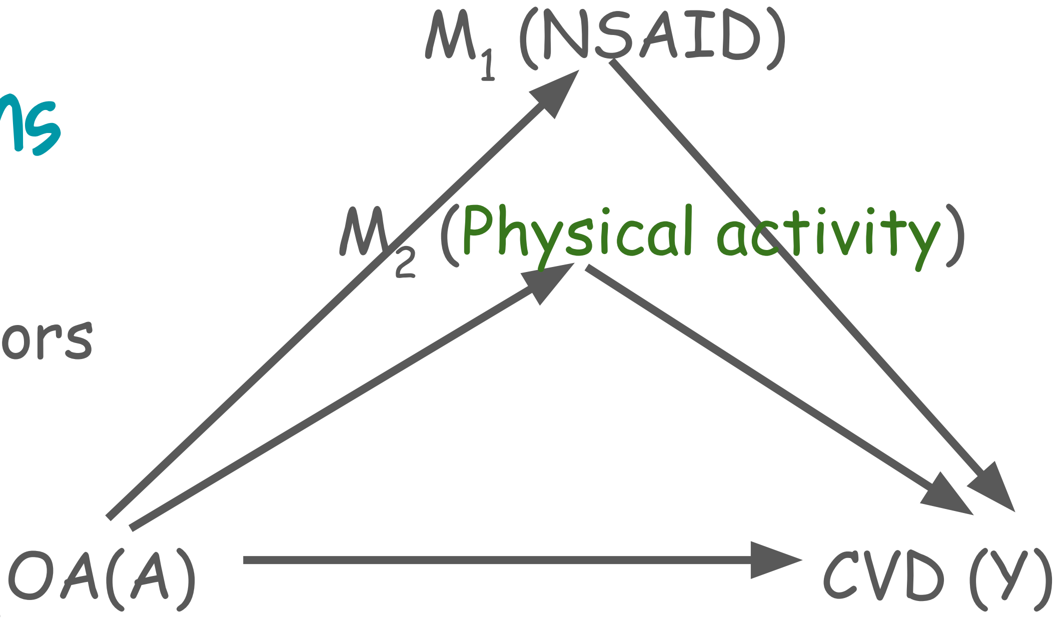
- C is sufficient to address confounding. No uncontrolled confounding in:
    - exposure-outcome associations
        - Y(A=a, M(a)) independent of A assignments given C
    - exposure-mediator associations
        - M(a) independent of A assignments given C
    - mediator-outcome associations
        - Y(A=a, M(a)) independent of M assignments given C
- One related idea is model-misspecification
    - Generally good to consider realistic/plausible interactions between
        - Exposure * covariate; or Mediator * covariate; or covariate * covariate

Assumptions - 2, 3 & 4

- Positivity
  - All exposure values have non-zero probability for any values of C
    - $P(A=a|C=c) > 0$ for all a and c
  - All mediator values have non-zero probability for any values of A & C
    - $P(M=m|A=a, C=c) > 0$ for all m, a and c
- Causal Consistency
  - Observed values are realistic
  - No multiple version of A or M
- No exposure-mediator interactions

Methodologic Extensions

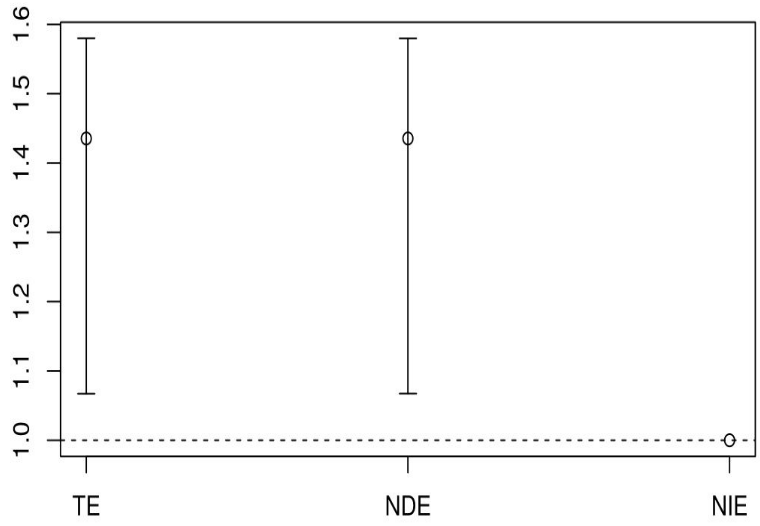
- More mediators
- Multicategory mediators
  - Active
  - Moderate
  - Inactive
- Additional extensions
  - Survival outcome
  - Additive vs multiplicative effects
  - Marginal vs conditional estimates
  - Non-compliance
  - Sensitivity analyses

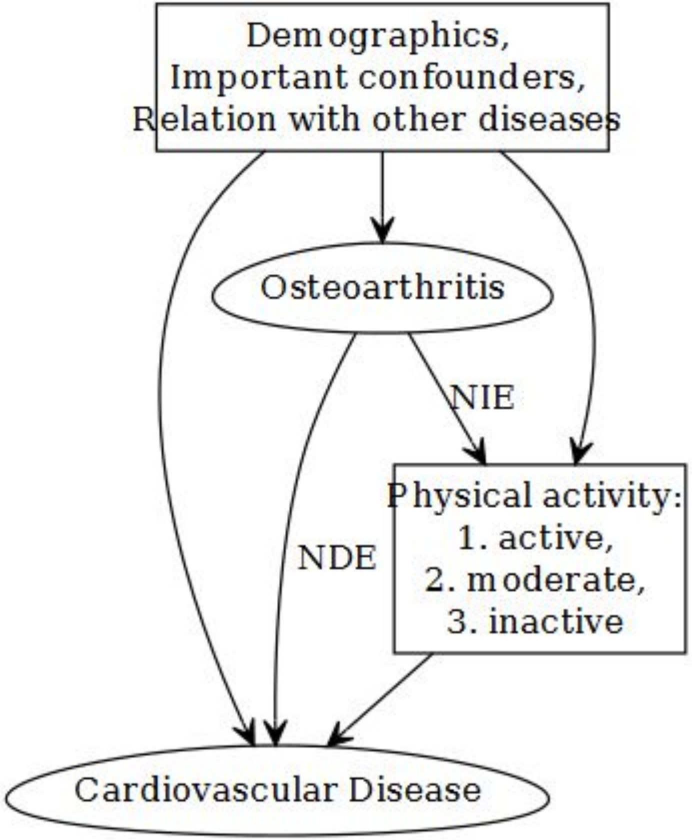


Our example

Multi-category mediator

```
##   TE   DE   IE   PM
## 1.44 1.44 1.00 0.00
```

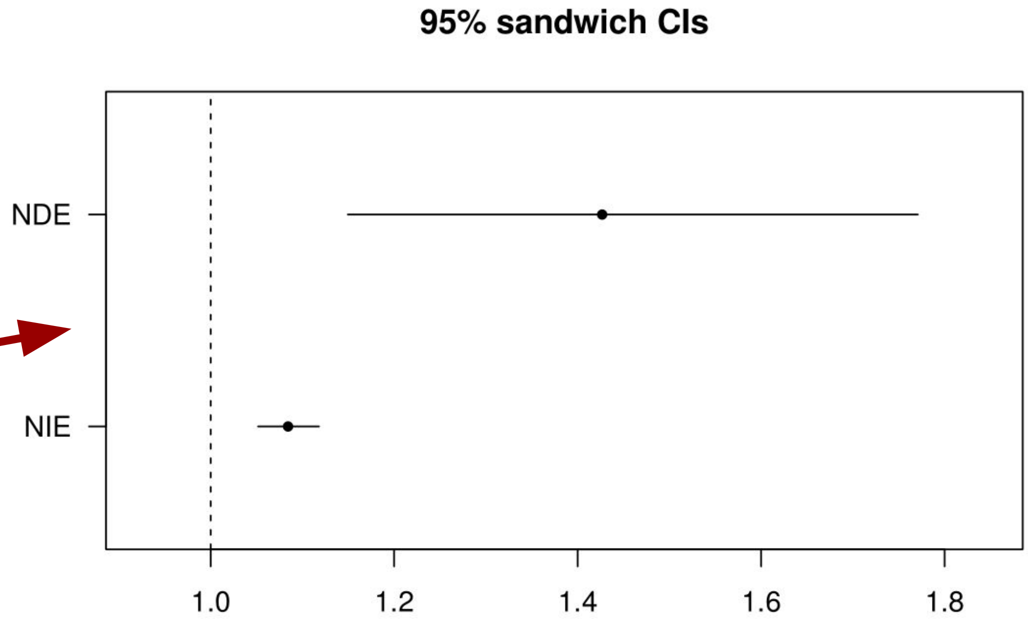




The proportion mediated through physical activity was about 0% on the log odds ratio scale!

Software

- R extensions
  - Mediation
  - MedFlex
  - MMA
  - GEEmediate
  - IORW (code)
- SAS & Stata have some.



References / workshops

- 'Mediation analysis using R' by Theis Lange, Stijn Vansteelandt, ISCB Conference 2019, Leuven
- 'Applied Mediation Analysis' by Theis Lange; see his teaching website
- 'Causal Mediation Analysis' by Tyler VanderWeele via statistical horizons

References / workshops

[HTML] Mediation analysis of the relationship between institutional research activity and patient survival

J Rochon, A du Bois, T Lange - BMC medical ..., 2014 - bmcmedresmethodol.biomedcentral ...

Recent studies have suggested that patients treated in research-active institutions have better outcomes than patients treated in research-inactive institutions. However, little attention has been paid to explaining such effects, probably because techniques for mediation analysis existing so far have not been applicable to survival data. We investigated the underlying mechanisms using a recently developed method for mediation analysis of survival data. Our analysis of the effect of research activity on patient survival was based on ...

Cited by 38 Related articles All 15 versions Web of Science: 25 Import into BibTeX

In particular, look at supplementary materials

Thanks!

ehsan.karim@ubc.ca
www.ehsankarim.com