Epistemological and practical challenges in using causal inference analyses in social epidemiology : developing a tool to support researchers

Léna Bonin, Hélène Colineaux, Benoît Lepage, Michelle Kelly-Irving

Causality in Epidemiology | Linz | May 4th 2024

Specificity of social epidemiology

An exposure often difficult to capture ….

Occupation

Education

Social class

Occupation status

Income

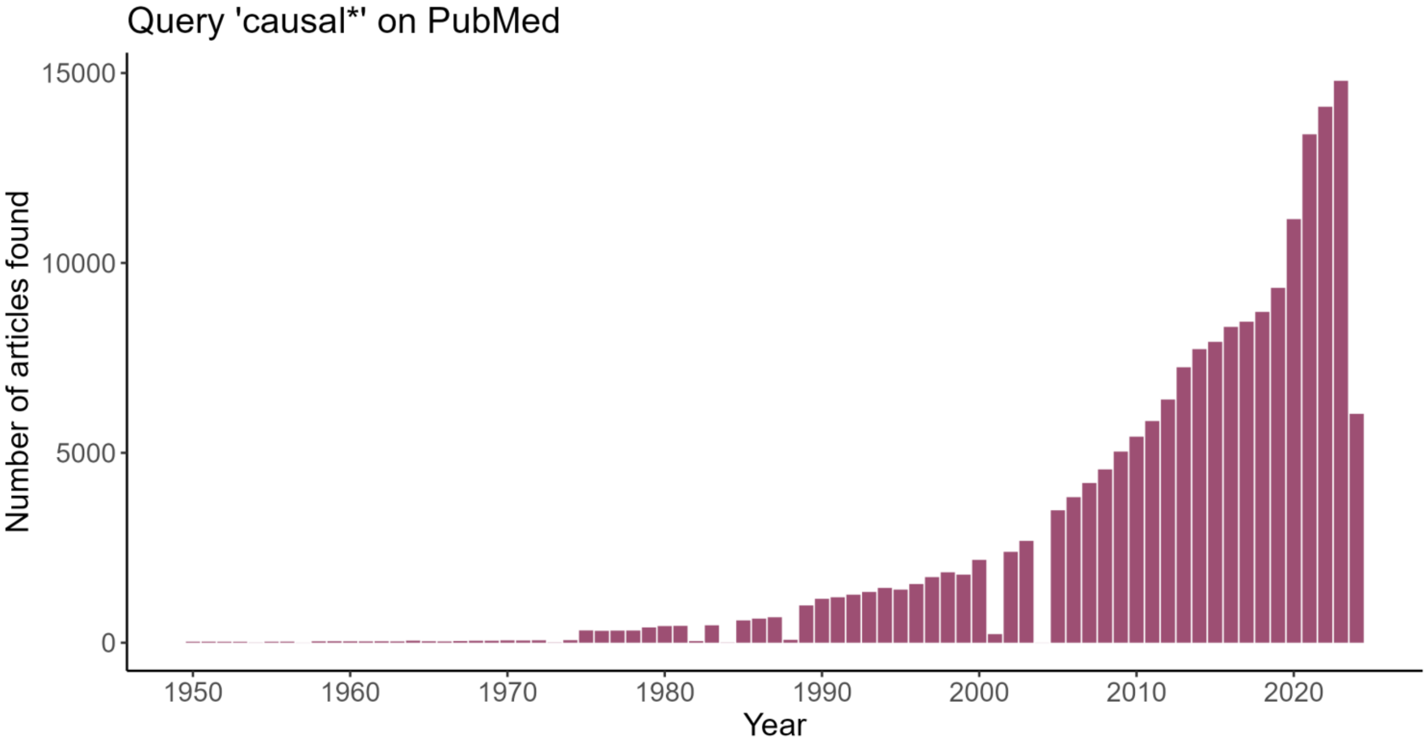
Housing

Partner data ?

An active research topic

A quite recent area of research

Many methods developed recently



Journal of Educational Psychology
1974, Vol. 66, No. 5, 688–701

ESTIMATING CAUSAL EFFECTS OF TREATMENTS IN RANDOMIZED AND NONRANDOMIZED STUDIES[1]

DONALD B. RUBIN[2]
Educational Testing Service, Princeton, New Jersey

A discussion of matching, randomization, random sampling, and other methods of controlling extraneous variation is presented. The objective is to specify the benefits of randomization in estimating causal effects of treatments. The basic conclusion is that randomization should be employed whenever possible but that the use of carefully controlled nonrandomized data to estimate causal effects is a reasonable and necessary procedure in many cases.

Methods

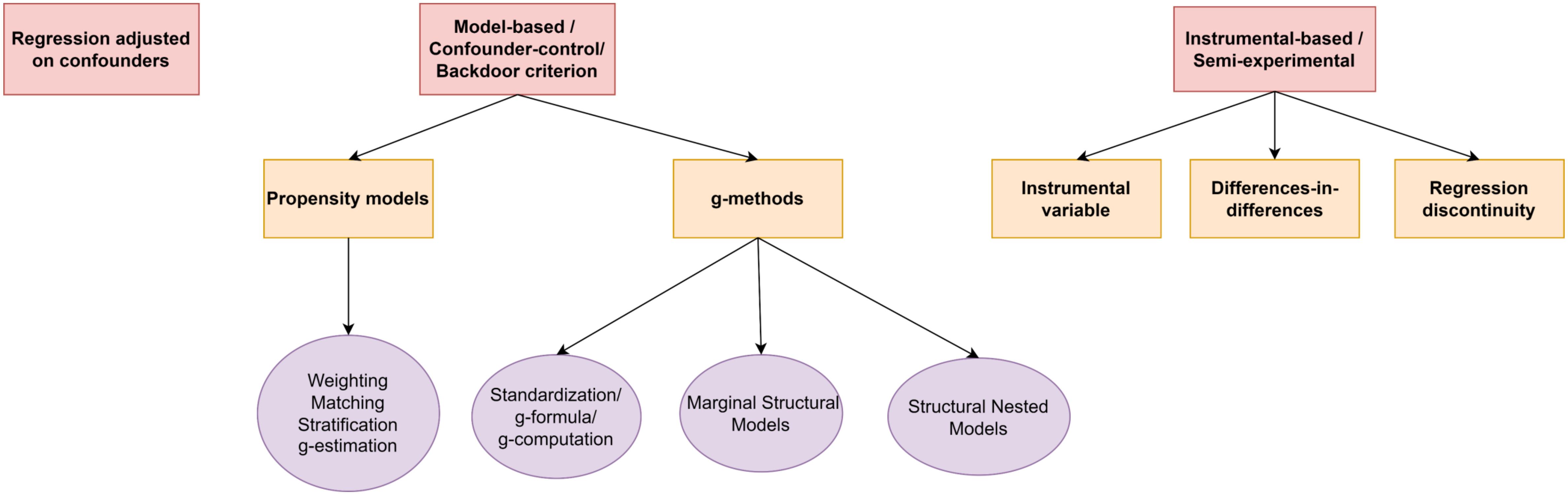
Total effect

- Regressions
- Counterfactual approaches / Backdoor criterion
- Semi-experimental/ IV-based methods
- Front-door criterion

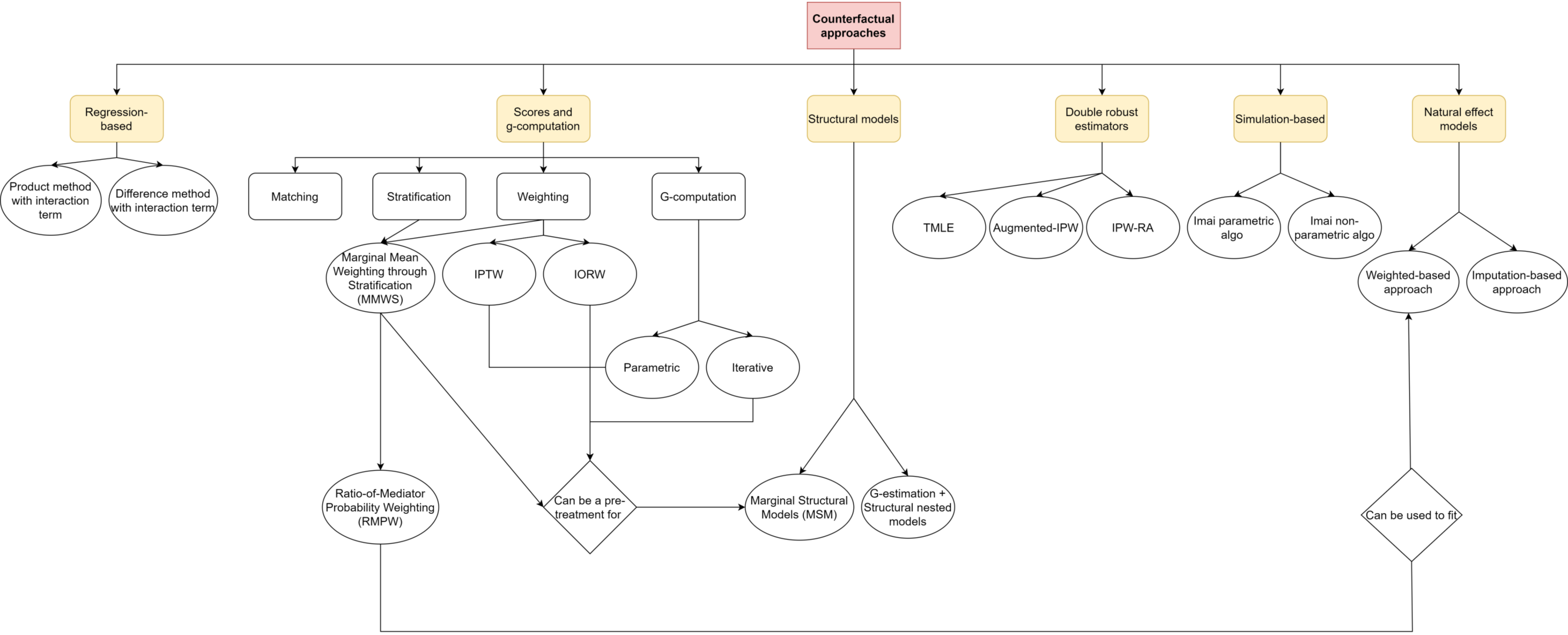
Mediation

- Traditional approaches (Baron and Kenny)
- Structural Equation Models
- Counterfactual approaches
- Semi-experimental/ IV-based methods
- Front-door criterion

Simplified typology of the different methods for estimating a total effect



Typology of methods for mediation analysis



Necessity to develop a tool to help researchers to embrace causal inference

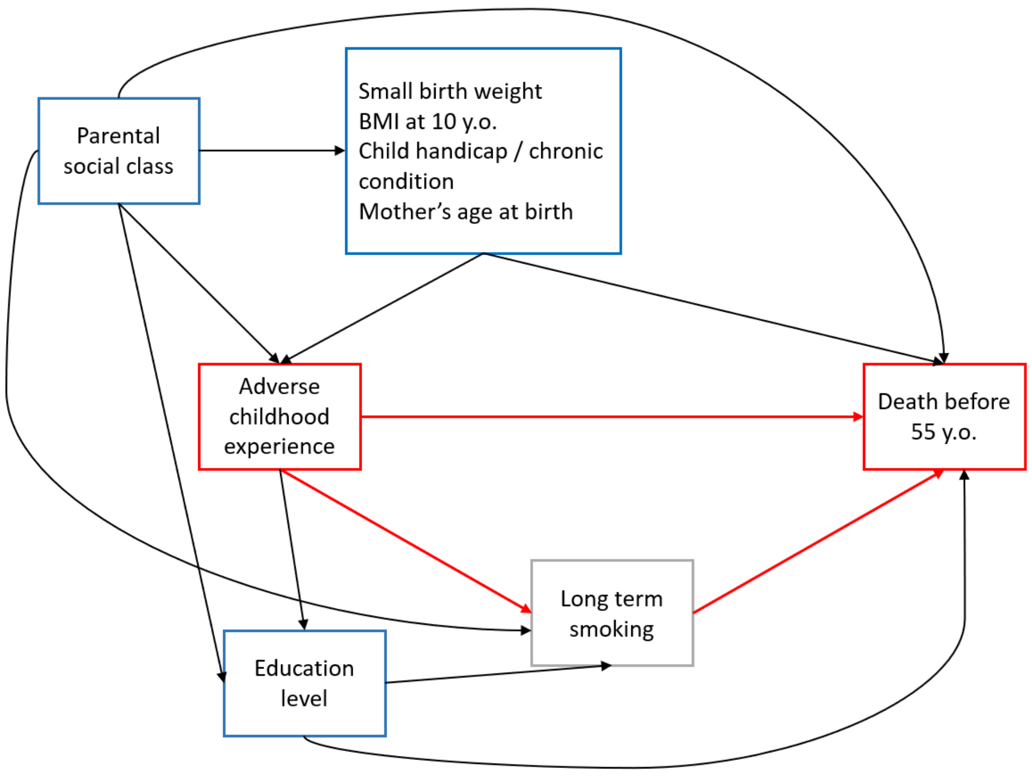
Goal

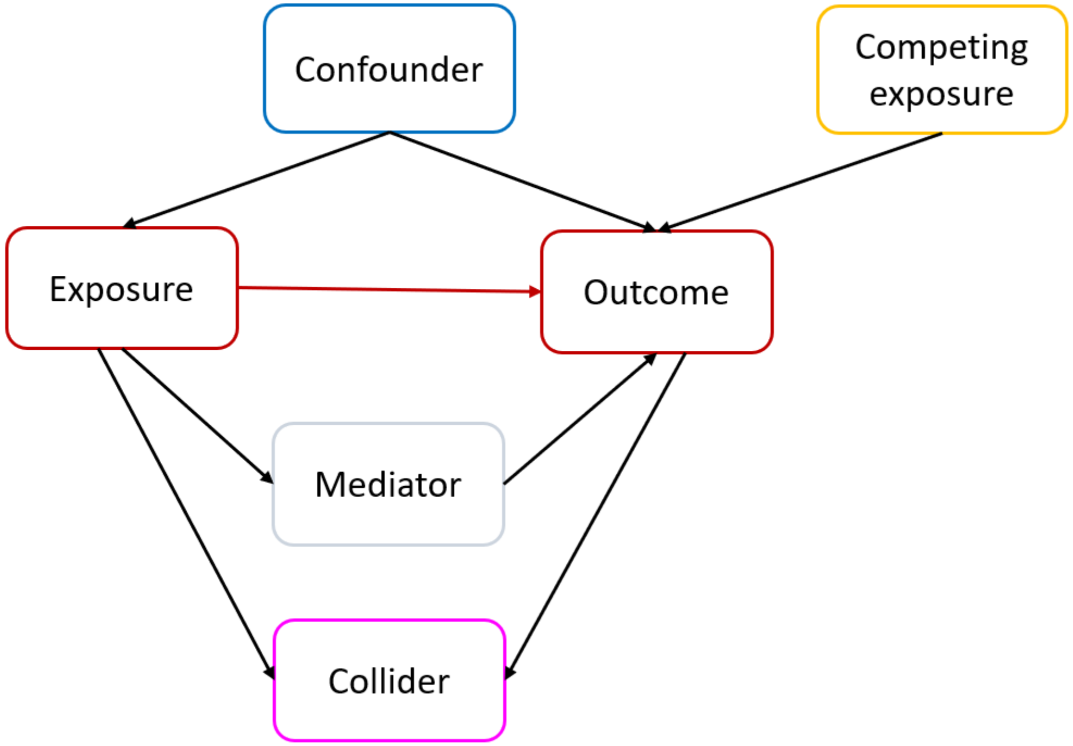
- To determine the estimand of interest
- To provide recommendations on the most appropriate method
- To try to detect obvious assumptions violation
- To provide R functions to perform the analysis

Limitation : We assume that the user has a DAG

Directed acyclic graphs (DAGs)

A visual representation of our hypotheses





Formalizing objectives

Aim :

- To understand the research question

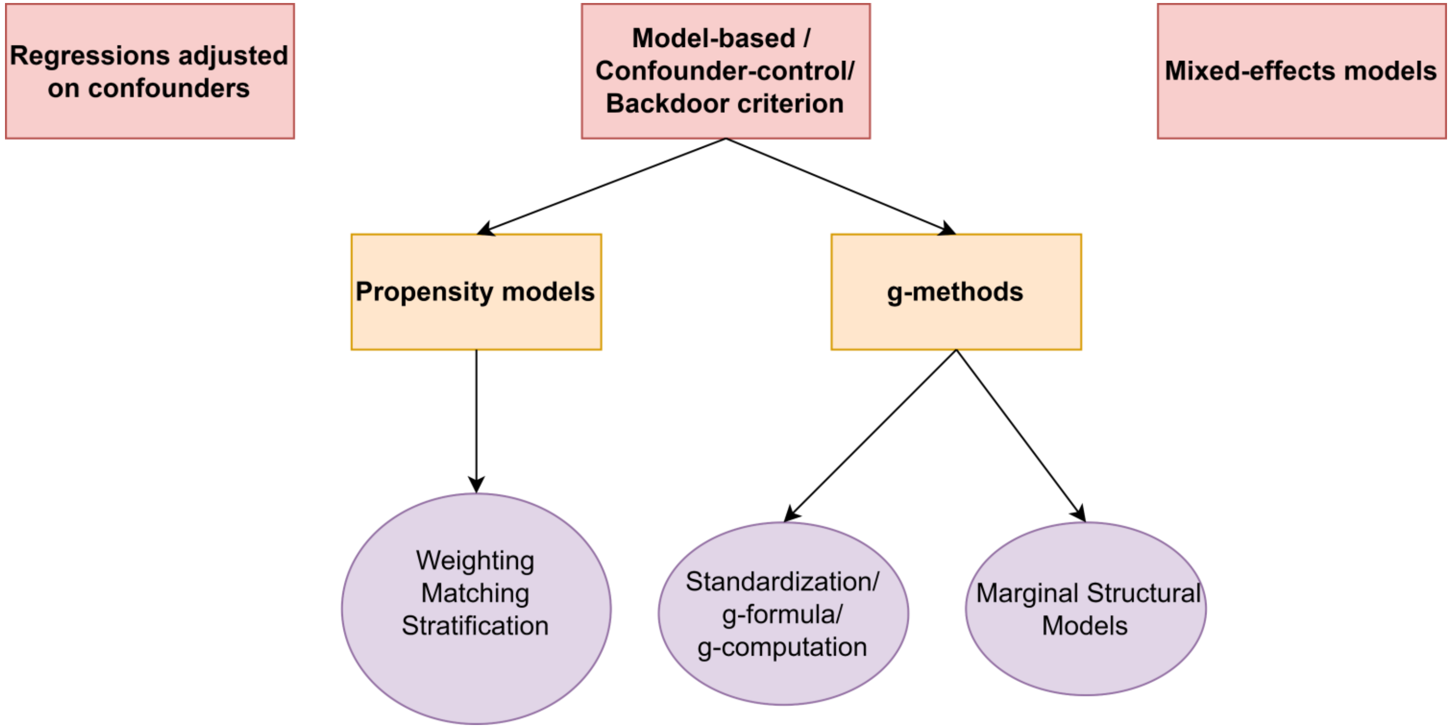
Choices :

- To propose different ways of formulating the same thing
- Not to use the word « mediation »

Average total effect	Controlled direct effect	Eliminated effect
	Pure natural direct effect	Pure natural indirect effect
	Total natural direct effect	Total natural indirect effect

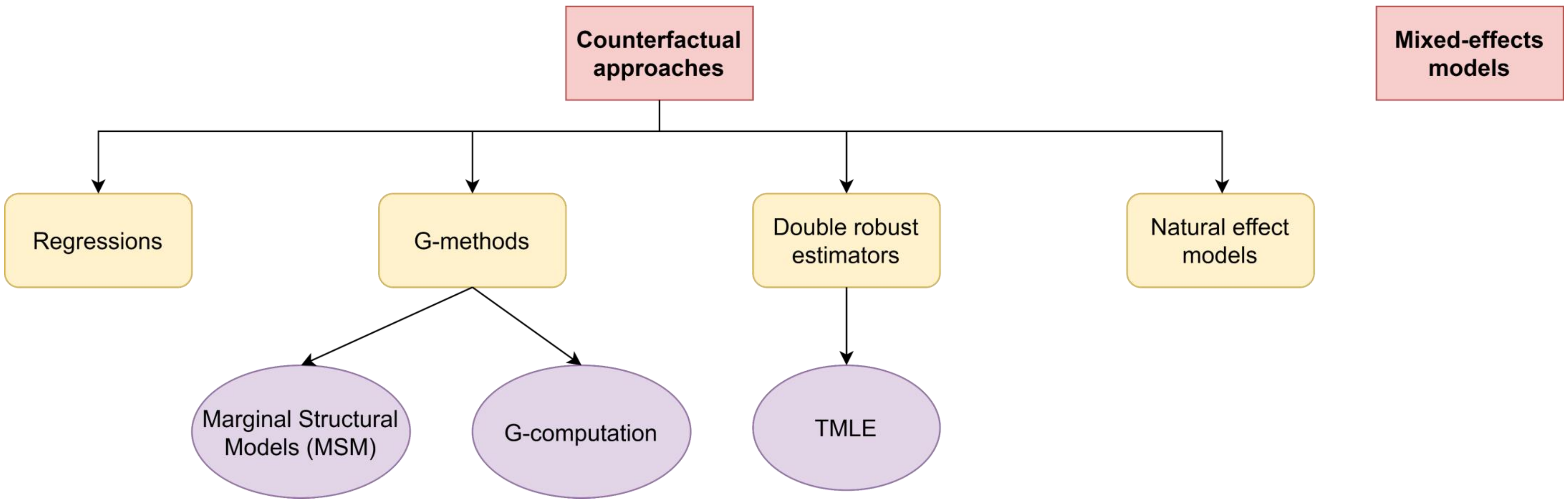
Methods

Estimation of a total effect only



Methods

Estimation of mediated effects



Recommending the most appropriate method

Choice determined by :

- The DAG : is there intermediate confounding, repeated measurements ?
- Variable types
- Assumptions

Goal : go for the simplest / most parsimonious method

Often there is no unique most appropriate method

Choice depends on personnal preferences

Development of a shiny app

Causal app | Home | Questionnaire | Ressources

Welcome

The aim of this application is to guide you in your choice of estimands and methods for carrying out causal or mediation analyses. It has been developed within a framework applied to social epidemiology. For now, the application is based mainly on the counterfactual approach.

Please note that we are assuming that the first stage of the analysis, which involves creating a directed acyclic graph (DAG), has already been completed. If you need help with this step, you will find some references in the Resources tab.

To get started, go to the Questionnaire tab !

The application code is available at the following link: https://github.com/LenaBonin/CausalApp

https://equity-cerpop.shinyapps.io/causalapp_en/

Perspectives

- Add other types of methods, specifically quasi-experimental ones
- Add recommendations for analyses with multiple mediators

Strengths and limitations

Strengths

Efforts to understand the user’s question

Suggestions to check user’s DAG and help them to modify it

Many ressources of different types

Limitations

Assumption that a DAG has already been drawn

Users tend to select all objectives

gendhi gender & health inequalities

Discussion

1) Content

- Should we go deeper in the explanations ?
- How to influence even less the choice of the objective ?

2) Form

- Allow the user to enter their DAG
- What about non R users ?
- Alternative to R package ?

Scan for the app

Thank you !

Contact : lena.bonin@inserm.fr

European Research Council
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N° 856478)

Causal app | Home | Questionnaire | Ressources

Preliminary question

Have you drawn a directed acyclic graph (DAG) or a concept map ?

- (●) Yes
- ( ) No

Next >

No DAG

This application assumes that you have already drawn your DAG.
If not, draw one before starting. If you need help, you'll find references in the 'Resources' tab.

OK

Causal app | Home | **Questionnaire** | Ressources

Is your objective :

(The variables suggested in the following two sentences are only examples, and can be replaced by your own variables later on)

To study the effect of social class on mortality

- (●) Yes
- ( ) No

To study the role of tobacco as an intermediary between social class and mortality

- ( ) Yes
- (●) No

[< Back] [Next >]

Causal app | Home | Questionnaire | Ressources

Variable types

Is your exposure variable :

- (•) Continuous
- ( ) Binary
- ( ) Ordinal
- ( ) Nominal
- ( ) I have several

What is its name ?

Social class

Is your outcome variable :

- (•) Continuous
- ( ) Binary
- ( ) Ordinal
- ( ) Nominal
- ( ) Survival / Time-to-event
- ( ) I have several

What is its name ?

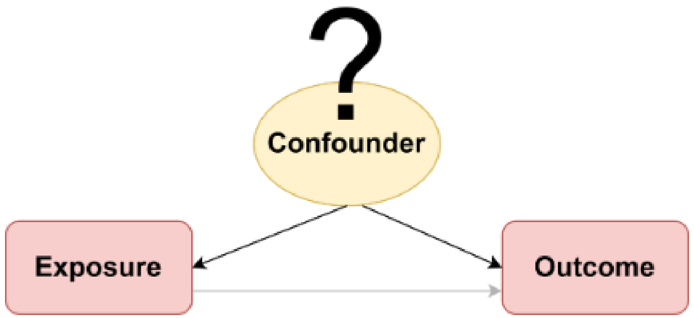
Mortality

< Back | Next >

Causal app | Home | Questionnaire | Ressources

Confounders

Think carefully about all the potential confounders between the exposure and the outcome , i.e. all the variables that affect both exposure and outcome



Do they all appear on your DAG ?

- (•) Yes
- ( ) No

Are any of these confounders not measured in your data ?

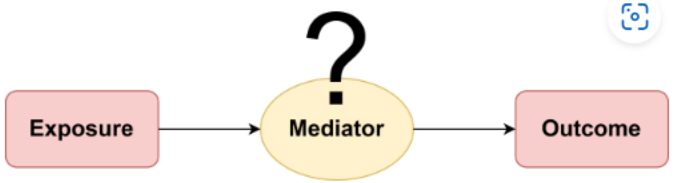
- (•) Yes
- ( ) No

< Back | Next >

Causal app | Home | Questionnaire | Ressources

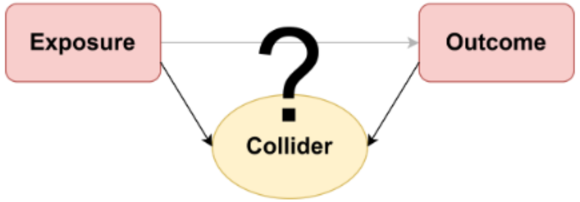
Mediators and colliders

Does your graph contain variables that are affected by the exposure and which in turn affect the outcome , i.e. does it contain mediators between your exposure and your outcome variables ?



- ○ Yes
- ◉ No

Does your graph contain variables that are affected by both exposure and outcome , i.e. does it contain colliders between your exposure and your outcome variables ?



- ○ Yes
- ◉ No

< Back | Next >

Causal app | Home | Questionnaire | Ressources

Intermediate confounders and repeated measurements

Is your exposure measurement repeated over time ?

- (●) Yes
- ( ) No

Is your outcome measurement repeated over time ?

- (●) Yes
- ( ) No

< Back | Next >

Causal app | Home | Questionnaire | Ressources

Positivity

Do you suspect that certain combinations of confounders correspond only to exposed/unexposed individuals (or to specific values of the exposure), i.e are there individuals who cannot be exposed/unexposed because of their characteristics ?

- (•) Yes
- ( ) No
- ( ) I don't know

< Back | Next >

Causal app Home Questionnaire Ressources

Please check that the information provided is correct

Objective

To study the effect of an exposure on an outcome

Variables

Variable	Type
Exposure	Continuous
Outcome	Continuous

Constraints

Criterion	Response
Exposure	Repeated
Confounder(s)	Repeated
Outcome	Repeated
Unmeasured confounders	Yes
Positivity violation	Yes

< Back Confirm

Recommendations

According to the answers you provided

Estimands

Total effect

Estimation method

G-methods or mixed model (fixed effects model)

If the outcome at time *t* is not influenced by exposure at time *t-1* and the outcome at time *t* does not influence exposure at time *t+1*, and if in addition exposure at time *t* does not influence any confounding factors at time *t+1*, you can run a fixed-effects model. Beware though, this is a strong and unlikely assumption.
For more details check Imai, K., & Kim, I. S. (2019). *When should we use unit fixed effects regression models for causal inference with longitudinal data?. American Journal of Political Science*, 63(2), 467-490.
and Imai, K., & Kim, I. S. (2021). *On the use of two-way fixed effects regression models for causal inference with panel data. Political Analysis*, 29(3), 405-415.
Otherwise, consider keeping only the last outcome measure and using g-methods (g-computation or marginal structural models). Given that your exposure is continuous, the easiest way is to use **g-computation**.

The positivity assumption necessary for causal analysis is violated, **so your results will probably be biased**.
If you still want to do the analysis anyway, we recommend that you use **g-computation** , but you will have to be very careful in interpreting the results

Assumptions

1- Positivity: each individual has a non-zero probability of being exposed / non-exposed
2- Ignorability / Exchangeability: the value of the potential (counterfactual) outcome under exposure *a* is independent of the value of the actual value of exposure
3- Non-interference: one individual's outcome is not affected by the value of another individual's exposure
4- Consistency: the potential (counterfactual) outcome of an individual under a certain exposure value corresponds to the value of the outcome it would have actually taken under this exposure

Note 1 : These hypotheses can be formulated conditionally, i.e. by adding 'conditional on confounders'.

Note 2 : These hypotheses are difficult to test in practice.

You indicated that the positivity assumption is probably not verified in your data. **Your results are therefore likely to be biased and inaccurate.**
If you still want to do the analysis, we recommend to use **g-computation** . However, you need to be very careful when interpreting the results. If the positivity problem is due to the fact that a combination is theoretically

Causal app Home Questionnaire Ressources

Is your objective :

(The variables suggested in the following two sentences are only examples, and can be replaced by your own variables later on)

To study the effect of social class on mortality

- (•) Yes
- ( ) No

To study the role of tobacco as an intermediary between social class and mortality

- (•) Yes
- ( ) No

< Back   Next >

In the previous example ('to study the role of tobacco as an intermediary in the relationship between social class and mortality'), we suggest that you replace the variables with your own.

If you do not enter anything, the terms 'exposure', 'intermediate variable' and 'outcome' will be used for the next questions

What variable would you like to replace 'social class' with, i.e. what is your exposure variable ?

Social class

What variable would you like to replace 'tobacco' with, i.e. what is your intermediate variable ?

Tobacco

What variable would you like to replace 'mortality' with, i.e. what is your variable of interest/outcome ?

Mortality

< Back | Next >

Causal app | Home | Questionnaire | Ressources

More specifically, is your objective along the lines of :

To study the effect of the exposure on the outcome after the implementation of an intervention/policy that affects the intermediate variable

- (•) Yes
- ( ) No

To study the effect of the exposure on the outcome if the intermediate variable was completely eliminated

- (•) Yes
- ( ) No

To study the proportion of the effect of the exposure on the outcome that could be eliminated by removing the intermediate variable for all individuals

- (•) Yes
- ( ) No

< Back | Next >

Causal app | Home | Questionnaire | Ressources

and/or of the type :

To study the effect of the exposure on the outcome that goes through the intermediate variable

- (•) Yes
- ( ) No

To study the effect of the exposure on the outcome that does not go through the intermediate variable

- (•) Yes
- ( ) No

To study the effect of the exposure on outcome if all individuals had the intermediate variable value of a given exposure category
e.g. if all individuals had the same value of the intermediate variable as the exposed individuals

- (•) Yes
- ( ) No

To study the proportion of the effect of the exposure on outcome that is due to the effect of the exposure on the intermediate variable

- (•) Yes
- ( ) No

< Back | Next >

variable types

Is your exposure variable :

- (●) Continuous
- ( ) Binary
- ( ) Ordinal
- ( ) Nominal
- ( ) I have several

Is your outcome variable :

- (●) Continuous
- ( ) Binary
- ( ) Ordinal
- ( ) Nominal
- ( ) Survival / Time-to-event
- ( ) I have several

Have you ever tested the total effect of your exposure on the outcome?

- (●) Yes, with my data, I found one
- ( ) Yes, with my data, there isn't any (or it isn't significant), but I still would like to do a mediation analysis.
- ( ) No, but it has been done in the literature
- ( ) No

Is your intermediate variable between exposure and outcome ? (this variable will also be called 'mediator' in the next questions)

- (●) Continuous
- ( ) Binary
- ( ) Ordinal

Causal app | Home | Questionnaire | Ressources

Intermediate confounding and repeated measurements

Is the measurement of your exposure repeated over time in your data ?

- (•) Yes
- ( ) No

Is the measurement of your intermediate variable of interest repeated over time in your data ?

- (•) Yes
- ( ) No

Is the measurement of your outcome repeated over time in your data ?

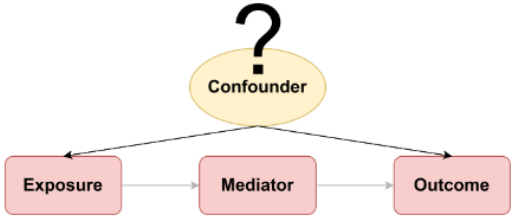
- (•) Yes
- ( ) No

< Back | Next >

Causal app Home Questionnaire Ressources

Confounders

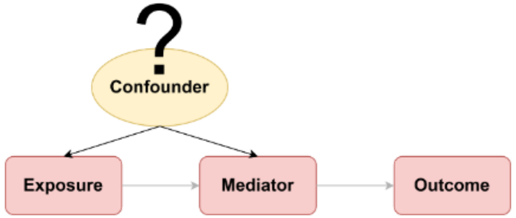
Think carefully about all the potential confounders between the exposure and the outcome , i.e. all variables that affect both exposure and the outcome



Do they all appear on your DAG ? (if there are no confounders for this relationship, check yes)

- (•) Yes
- ( ) No

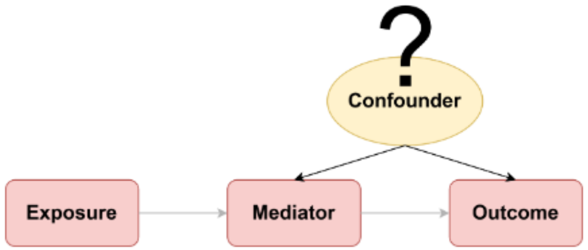
Think carefully about all the potential confounders between the exposure and the intermediate variable (mediator) , i.e. all variables that affect both the exposure and the mediator



Do they all appear on your DAG ? (if there are no confounders for this relationship, check yes)

- (•) Yes
- ( ) No

Think carefully about all the potential confounders between the mediator and the outcome , i.e. all variables that affect both the intermediate variable and the outcome



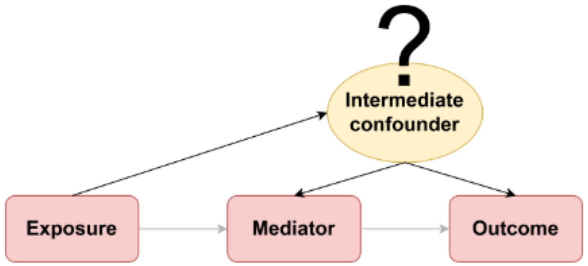
Do they all appear on your DAG ? (if there are no confounders for this relationship, check yes)

- ◉ Yes
- ○ No

Are some confounders not measured in your data ?

- ◉ Yes
- ○ No

Is at least one of the confounders between the mediator and the outcome affected by the exposure , i.e. are there intermediate confounders ?

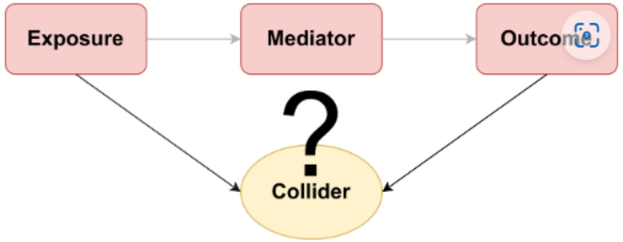


- ○ Yes
- ◉ No

Causal app Home Questionnaire Ressources

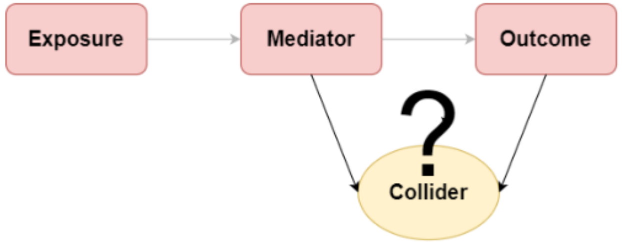
Colliders

Does your graph contain variables that are affected by both the exposure and the outcome , i.e. does it contain colliders between the exposure and the outcome ?



- ◉ Yes
- ○ No

Does your graph contain variables that are affected by both the intermediate variable (mediator) and the outcome , i.e. does it contain colliders between the mediator and the outcome ?



- ◉ Yes
- ○ No

< Back Next >

Causal app | Home | Questionnaire | Ressources

Positivity

Do you suspect that certain combinations of confounders correspond only to exposed/unexposed individuals (or to specific values of the exposure variable), i.e. are there individuals who cannot be exposed/unexposed (are cannot take some values of the exposure variable) because of their characteristics ?

- (●) Yes
- ( ) No
- ( ) I don't know

Do you suspect that certain combinations of confounders and exposure lead systematically to the same value(s) of the intermediate variable
i.e. are there individuals who cannot take certain values of the mediator because of their exposure value ?

- (●) Yes
- ( ) No
- ( ) I don't know

< Back | Next >

Causal app | Home | Questionnaire | Ressources

Exposure-Mediator interaction

Do you want to isolate any interaction term the exposure and the intermediate variable ,
i.e. If there is an interaction between the exposure and the mediator, do you want to highlight it in a separate term ?

- ○ Yes
- ◉ No

If there is actually an interaction, in which effect would you like it to be taken into account ?

- ◉ Direct effect (that does not go through the mediator (intermediate variable of interest))
- ○ Indirect effect (that goes through the mediator)

< Back | Next >

Recommendations

According to the answers provided

Estimands

Effect	Abbreviation	Objective
Total effect	TE	Effect of the exposure on the outcome
Controlled direct effect	CDE	Effect of the exposure on the outcome after implementation of an intervention affecting the mediator Effect of the exposure on the outcome if the mediator was completely removed
Proportion eliminated	PropEliminated	Proportion of the effect of the exposure on the outcome that could be eliminated by removing the mediator for all individuals
Total natural direct effect	TNDE	Effect of the exposure on the outcome that does not go through the mediator
Pure natural indirect effect	PNIE	Effect of the exposure on the outcome that goes through the mediator
Proportion mediated	PropMediated	Proportion of the effect of the exposure on the outcome due to the effect of the exposure on the mediator

Decomposition

2-way decomposition :

$$TE = TNDE + PNIE$$

Proportion eliminated :

Additive scale :

$$\frac{TE - CDE}{TE}$$

Relative risk scale :

$$OR^{TE} - OR^{CDE}$$