Heva

Facteurs de confusion non mesurés : quels designs et méthodes adopter

Eléonore Herquelot, Nada Assi, Clémentine Vabre, Fanny Raguideau, Arnaud Panes

Heva, Lyon

Congrès EMOIS 2023 - 16-17 Mars - Nancy

DOCAPOSTE
l'avenir devient plus simple

Introduction

Unmeasured confounding factor

Unidentifiable confounding factor within the database → source of bias in studies based on medico-administrative database

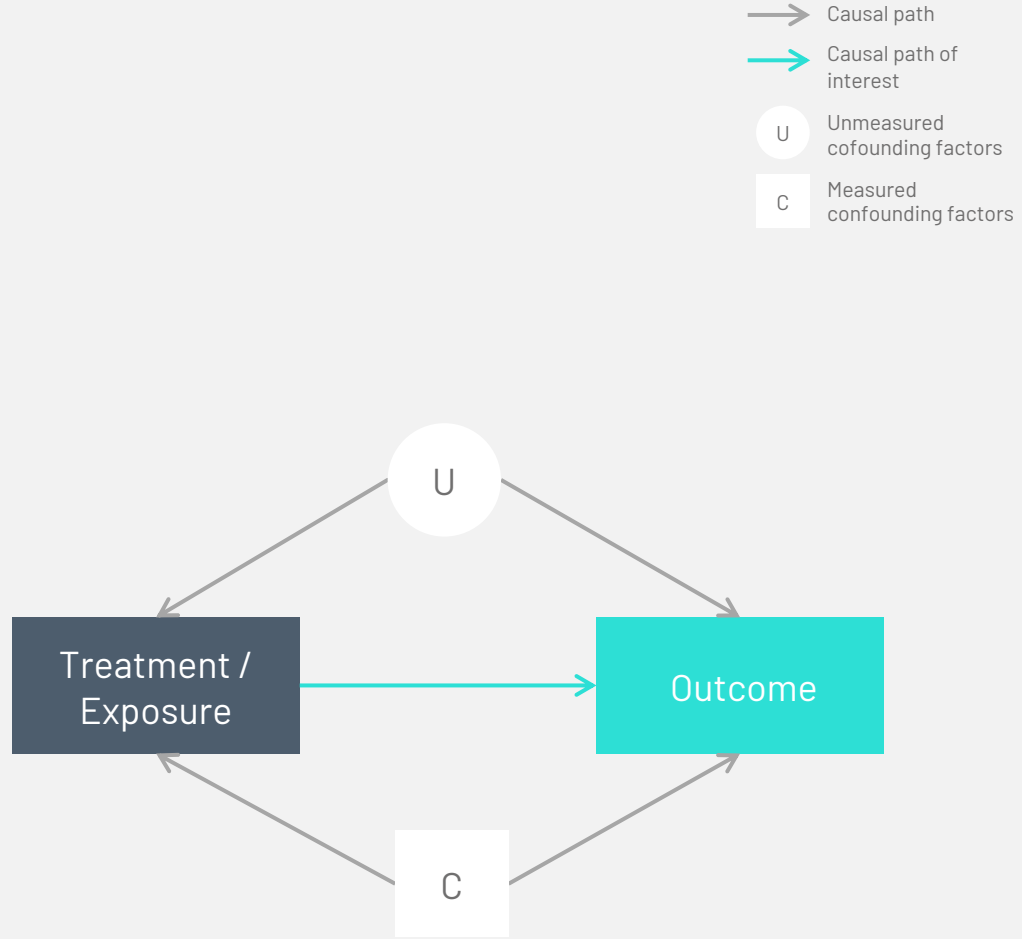
Taken into account during :

Study design

Self-controlled designs

Statistical analyses

Several possibilities: E-value, instrumental variable, negative control outcome, etc.



C = Confounding factors (influence both the dependent variable and independent variable)

Self-controlled designs

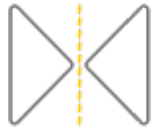
General information

Comparison within the same subject

Only subject with the event are included

Increase statistical power (less variability)

Cancel out confounding factors (measured and unmeasured) that are stable over time, e.g., sex, genetic factors, socioeconomic position...



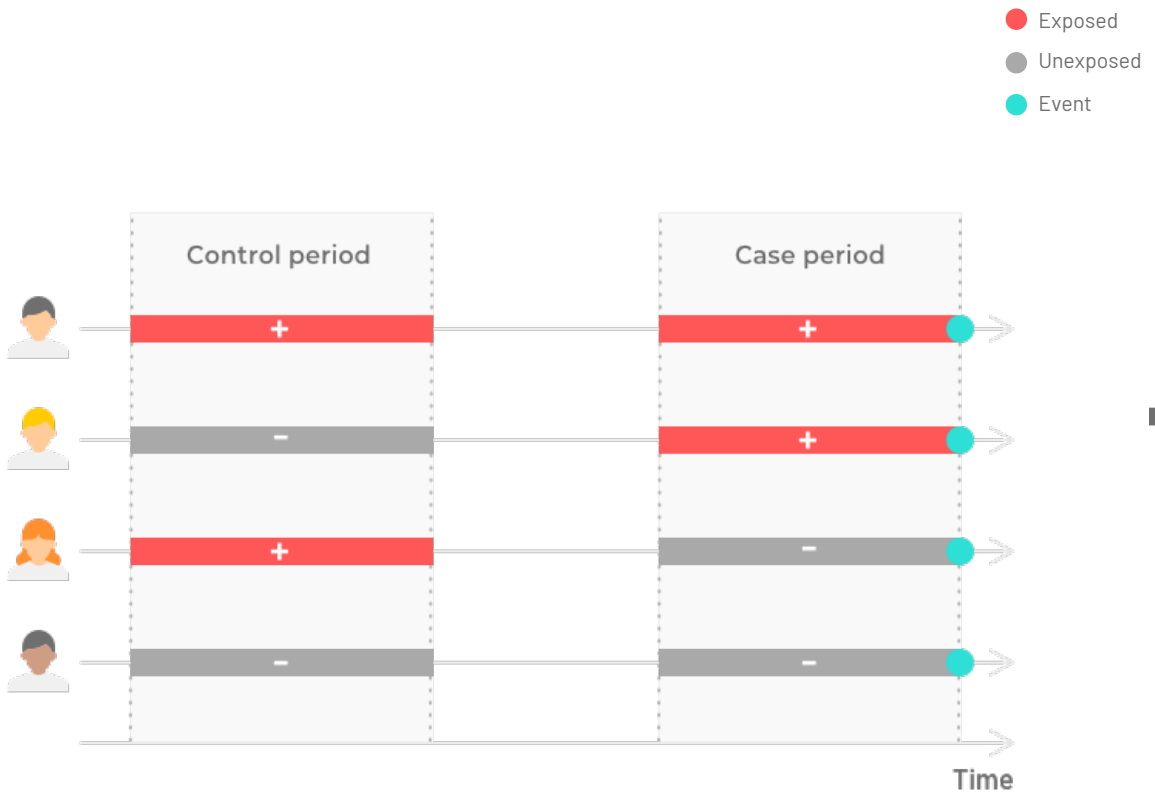
Two types of complementary designs:

Self-controlled design, **for intermittent exposure** and **acute event**

Cohort design, for **chronic exposure**

Self-controlled designs

Case-crossover (CCO)



Estimation of **odds-ratio** with **discordant pairs**

$$OR = \frac{\text{Number of subjects exposed during case period, non exposed during control period}}{\text{Number of subjects non exposed during case period, exposed during control period}}$$

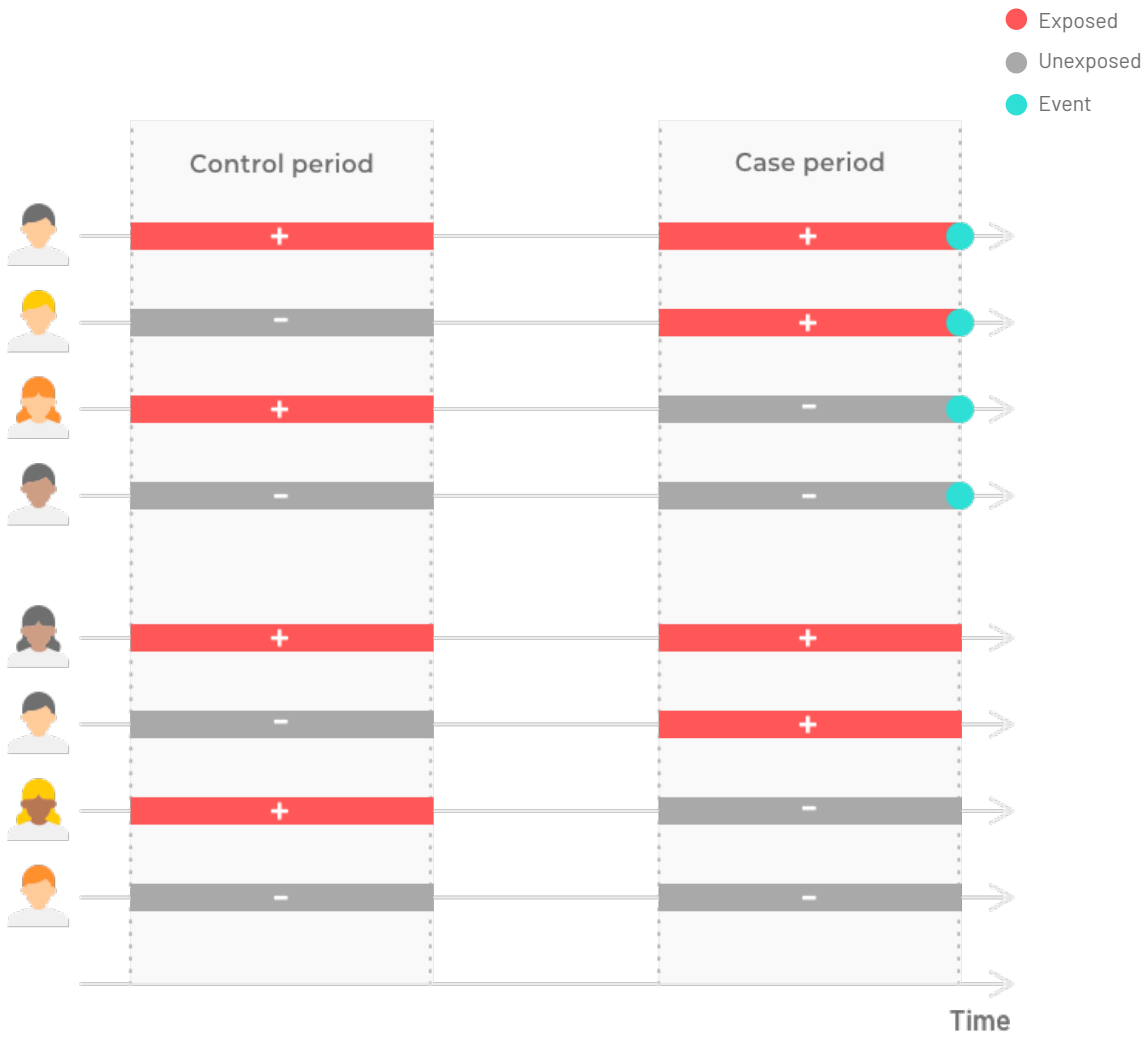
Assumptions

- Events are rare
- Disease risk is stable over time
- **Probability of being exposed are stable over time (no trend in exposure)**

MACLURE, Malcolm. The case-crossover design: a method for studying transient effects on the risk of acute events. American journal of epidemiology, 1991, vol. 133, no 2, p. 144-153.

Self-controlled designs

Case-time-control (CTC)



What if the probability of exposure is not stable over time?

Case-time-control design

Selection of a control group, with subjects who don't have the event

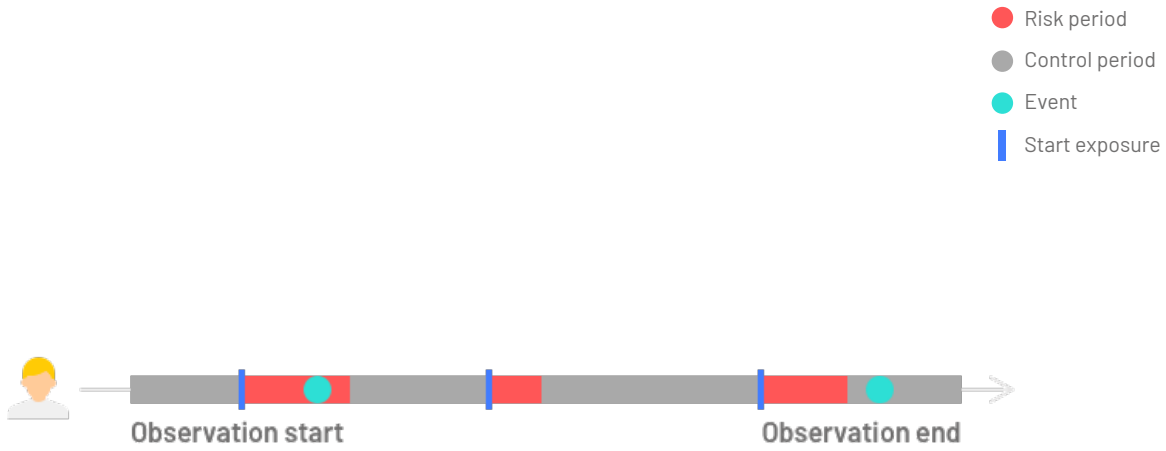
$$\text{OR} = \frac{\text{Odds ratio estimated for cases (with an event)}}{\text{Odds ratio estimated for controls (without an event)}}$$

If the trend exposure is the same for cases and controls, then **the bias generated by trend exposure is deleted**

SUISSA, Samy. The case-time-control design. Epidemiology, 1995, p. 248-253.

Self-controlled designs

Self-controlled case series (SCCS)



→ Estimation of the **Incidence Rate Ratio (IRR)** (risk period incidence vs control period incidence)

What to do when most of the population is exposed?

Self-controlled case series

Comparison of different period

- **Risk period** following the exposure → fixed time (to be defined)
- **Control period**: all remaining time

Assumptions

- **Occurrence of an event should not (appreciably) affect subsequent exposures**
- Event rates are stable within each defined period
- Events should be recurrent or rare (< 10% of onset)

FARRINGTON, C. P. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics, 1995, p. 228-235.

E-value

What is it?

How strong would the unmeasured confounding have to be to negate the observed results?

→ **E-value**

The minimum strength of association for an unmeasured confounder to fully explain away a specific treatment-outcome association.

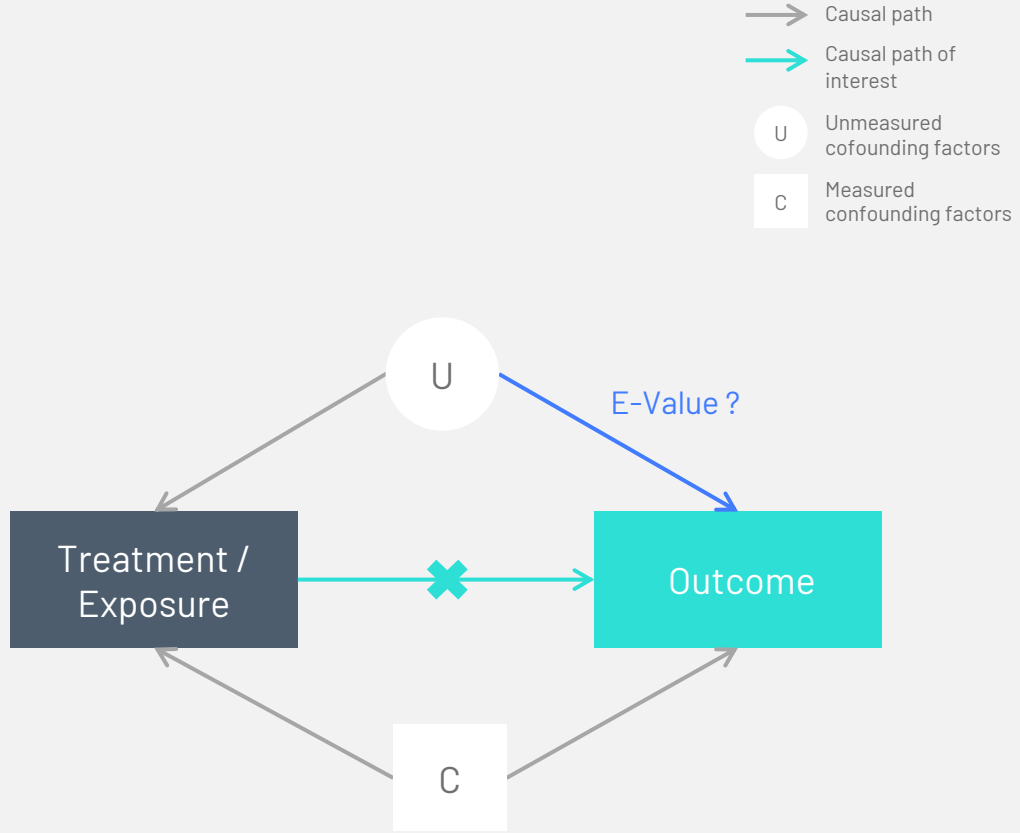
→ **Useful for interpretation and assessing robustness**

Strengths

- Simple sensitivity analysis
- **Does not require assumptions**
- **Simple way to discuss the unmeasured confounding** in articles and submitted study protocol with quantifiable argument

Limitations

- **No cutoff** to validate the results at your own discretion: compare whether or not the E-value is higher than the RRs in the study or in the literature
- Second indicator to be validated (p-value and E-value)



	E-value
RR > 1	$RR + \sqrt{RR * (RR - 1)}$
RR < 1	$\frac{1}{RR} + \sqrt{\frac{1}{RR} * (\frac{1}{RR} - 1)}$

E-value

Example

Occurrence of a clinical event after intervention according to the type of treatment

Logistic regression model adjusted with treatment A as reference (multivariate model)

	OR (95% CI)	P-VALUE MODALITY	E-VALUE	LOWER LIMIT
Treatment B vs A	0.81 (0.76-0.86)	<.0001	1.46	1.36
Treatment C vs A	0.93 (0.90-0.96)	<.0001	1.23	1.18

→ **The results may be prone to unmeasured confounding** (because it is likely to find a confounding factor with an association with the clinical event of 1.2-1.4)

Instrumental variables (IV)

What is it?

Idea of IV:

Find a **new variable highly correlated** with Treatment/Exposure not associated with U to remove the bias due to unmeasured confounding

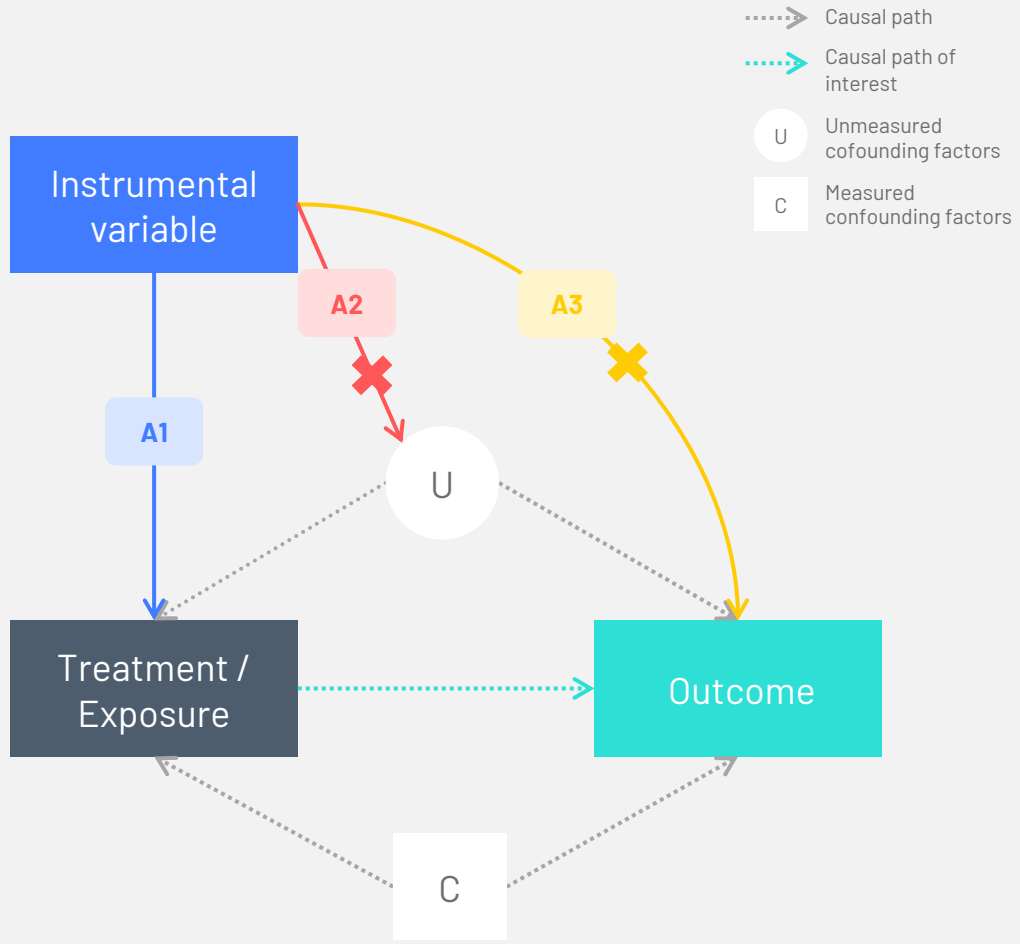
Assumptions

External measured variable used to remove the bias due to unmeasured confounding

A1 Associated with the treatment/exposition of interest
Testable

A2 Not associated with unmeasured confounding factors after conditioning on measured factors
Untestable

A3 Not directly associated with outcome (excepted through the treatment)
Untestable



Instrumental variables

Methodology

Example of IV in SNDS studies

Geographic: distance between the ZIP code and nearest facilities, geographic area

Time: Time-based characteristic of treatment (year of prescription)

Historical: Prescription preference of physician or facility (based on previous records)

Lagged: Previous therapy

Modeling

Two-stage Least Squares Models (2SLS)

Initially applied in economy with linear outcome

2 steps:

- Model for treatment
- Model for outcome with prediction of treatment from previous model

Extension GMM IVA (Johnston *et al.* 2008) for nonlinear models as GLM model

Results of the IV analysis: Local Average Treatment Effect (LATE)

Causal effect on compliers e.g., those who have the expected treatment according to their IV (given the previous assumptions)

The estimation is valid even in case of unmeasured confounding

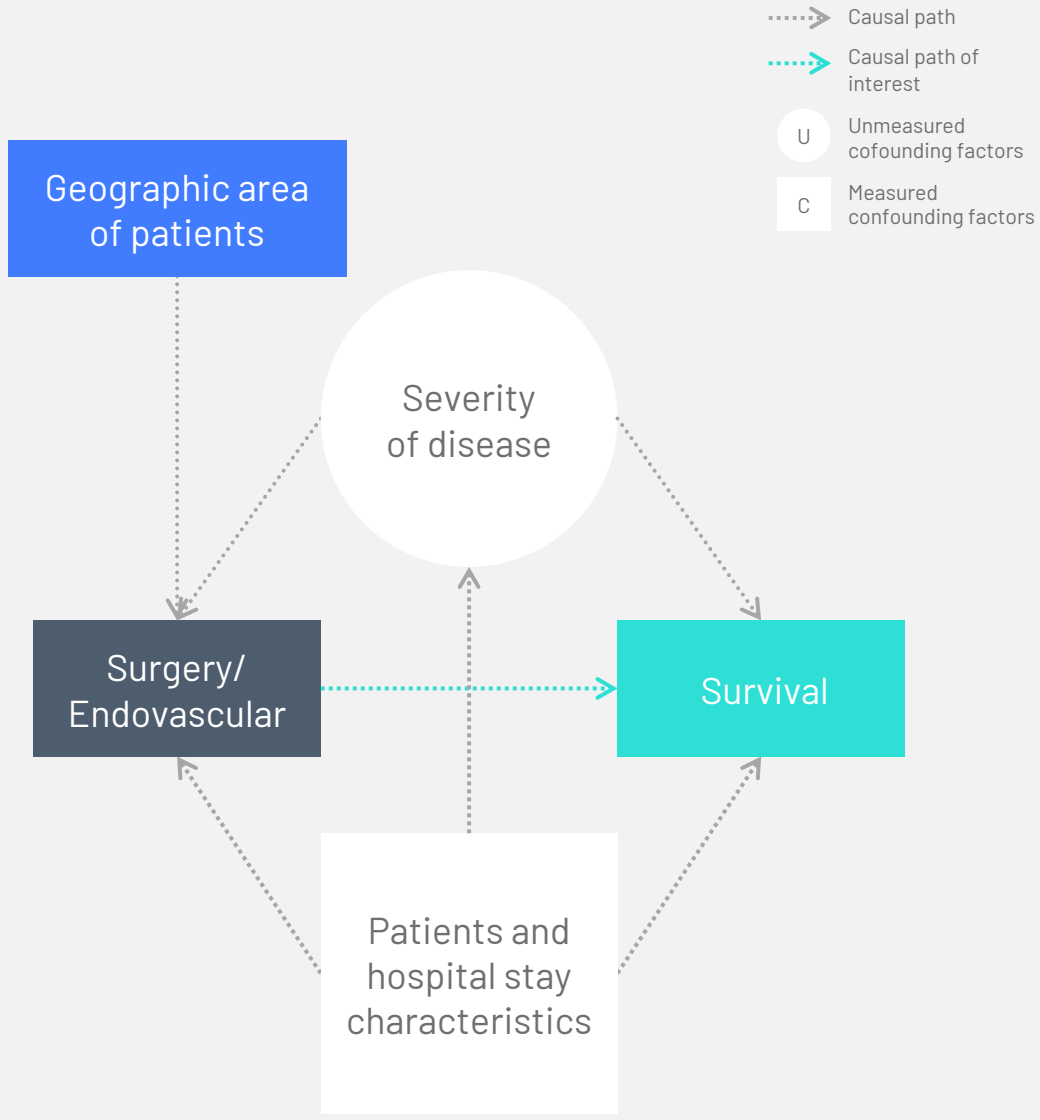
Example of IV application

Example in SNDS

Study of death at one year in lower limb acute ischemia and effect of type of treatment (Surgery or Endovascular)

Patients with surgery have **higher severity** than patients with endovascular treatment

Issue: The **severity of disease may be not properly captured** by covariates measured in SNDS



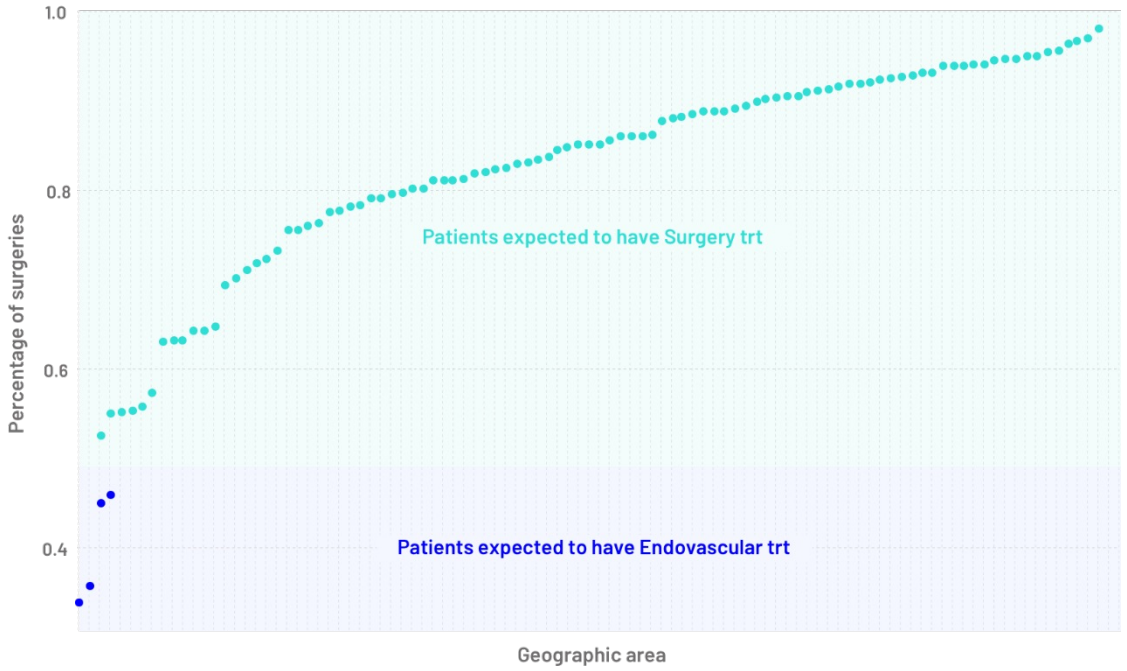
Example of IV application

Verification of assumption A1

Estimation of 42% of compliers*

Moderately strong association between treatment and geographic area

F-statistic = 50 (to use with caution)



*Compliers in this study = Patients having the treatment expected (Surgery or Endovascular) according to the geographic area

Example in SNDS

MODEL	OR (95% CI)	P-VALUE
Regression adjusted model	1.51 (1.40-1.62)	<0.0001
IV model	2.33 (1.86-2.91)	<0.0001

It is expected to have higher effect in IV model.

→ The **results of IV analysis are consistent** with the regression model.

Conclusion

Self-controlled designs can help control for unmeasured confounding .

Statistical methods can **mitigate against unmeasured confounding** and help evaluate how much the results can be trusted (E-value, NCO, IV…).

Sensitivity analyses can be applied using several other statistical methods to **evaluate the impact of unmeasured confounders**, these include but are not limited to:

- High Dimensional Propensity Score
- Negative Control Exposure
- Difference in Difference analysis

Unmeasured confounders are a frequent occurrence in epidemiological studies, analyses can still be conducted in their presence, only care should be taken during interpretation, rest assured it will not make the study go to waste !

NCO: Negative Control Outcome; IV: Instrumental variables

Heva

Thank you for your attention

hevaweb.com

@hevaweb

linkedin.com/company/heva

DOCAPOSTE

l'avenir devient plus simple