

Causal Inference

Vanessa Didelez, Robin Evans, Karla Diaz-Ordaz

BIPS, University of Bremen (Germany), University of Oxford, UCL (UK)

September 2026

APTS — Durham

Part 5b: Causal Discovery

an Example from Epidemiology

- Typically observational (non-interv., non-experimental)
- Cohort studies or panel data: data collected in waves, months or years apart
→ coarse (irregular) discrete time,
some repeated measurements ($\neq$ time series)

Common: missing data

Measurements: heterogenous (from questionnaires to wearables)

- Routinely collected data: electronic health records, registries, claims data

⇒ Many different types of measurements

⇒ Often: incomplete data / missing values

⇒ Typically, information on partial time-ordering available

Systematic/transparent way of **representing the assumed causal structure**

- Illustrate or examine possible sources of bias
 - e.g., due to bad design or analysis choices
 - Typically: **expert-driven** construction of (partial) DAG
- **Identification of causal parameters** via graphical characterization
 - e.g., explicit justification for choice of adjustment sets
 - Popular: backdoor criterion, but also ‘frontdoor criterion’
(*Piccininni et al. 2023 Epidemiology*)

Or: **DAG itself** is object of interest: Causal discovery

⇒ **data-driven** construction of DAG(s) (*Petersen et al., 2023, & Didelez, 2024: AJE*)

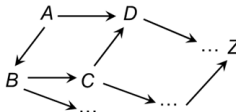
DAG not Known? $\Rightarrow$ Causal Discovery

Input: data

A	B	C	Z
0.3	12	0	...
0.2	13	0	287
0.7	21	1	876
0.6	10	0	326
...	...	...	...



Output: causal DAG



Actually:

→ need **special assumptions**
(faithfulness, causal sufficiency
likelihood, additivity, ...)

Here: constraint-based (PC)
→ output **not a unique** DAG,
but: equivalence class

scientific reports

www.nature.com/scientificreports

OPEN

A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents

Ronja Foraita^{1,2}, Janine Witte^{1,2}, Claudia Böhrhorst¹, Wencke Gwozdz^{3,4}, Valeria Pala⁵, Lauren Lissner⁶, Fabio Lauria⁷, Lucia A. Reisch^{1,8}, Dénes Molnár⁹, Stefaan De Henauw¹⁰, Luis Moreno¹¹, Toomas Veidebaum¹², Michael Tomanitis¹³, Iris Pigeot^{1,2} & Vanessa Didelez^{1,2}

IDEFICS/I.Family Cohort Study



- eight European countries, $\approx$ 16000 children aged 2-9 years at baseline;
- three waves, 2007 – 2017; $n = 5112$ in all waves
- information collected on: health behaviours (diet and physical activity), socioeconomic factors, genetics, medication, peer networks, media consumption, cardiovascular / metabolic health, subjective well-being
 - repeated measures e.g. of BMI, PA etc.
 - at single times: taste pref., puberty stage, smoking etc.

(Ahrens et al., on behalf of the I.Family Consortium, 2017)

Cohort Causal Graph — Analysis



- Methods: PC algorithm with MI (random forest imputation models), various sensitivity analyses PC assumes causal sufficiency
- Efficient use of temporal structure with tPC algorithm
- *Bootstrap* to investigate stability of specific graphical-causal structures
- Apply local and optimal generalised IDA to determine adjustment sets for interesting exposure and outcome pairs

Cohort Causal Graph — Results



<https://bips-hb.github.io/ccg-childhood-obesity/>

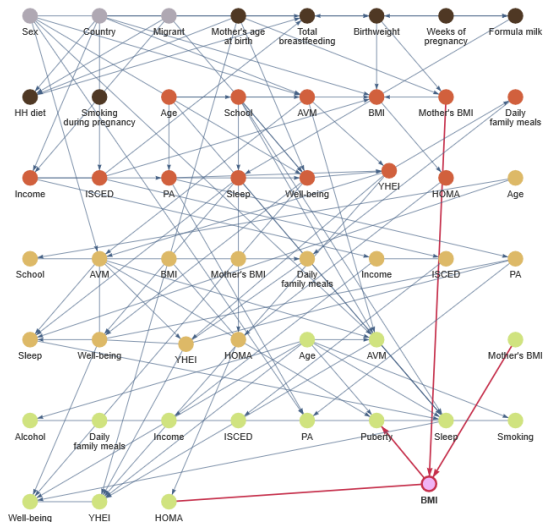
Context

Early life

Baseline

FU1

FU2



Cohort Causal Graph — Stability

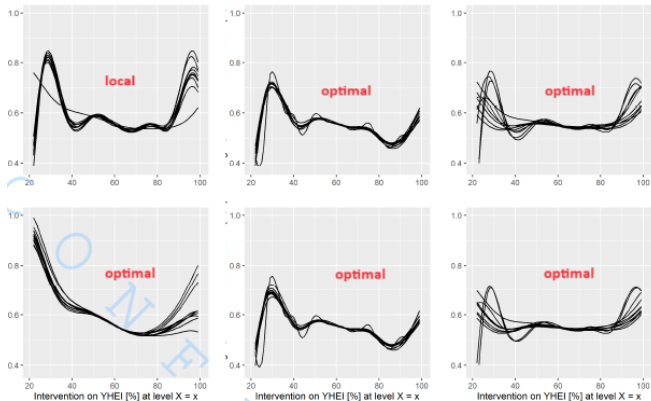


Based on 100 bootstrap graphs:
consider stability of individual (non)edges but also of specific
interesting graphical structures like causal paths

- Of 104 edges (on 51 variables), 36 were stable ($> 80\%$) while 50 were instable ($\leq 50\%$)
- **All** graphs had multiple possibly causal paths from early modifiable behaviours to later BMI
e.g., [youth-healthy eating index \(YHEI\)](#), audio-visual media consumption, sleep-duration, physical activity
- **No** graph had a direct edge from early modifiable behaviours to later BMI
- Cultural, perinatal and familial variables appear more immediate 'causal influences' on obesity than individually modifiable risk factors

- Example here:
estimate causal effect of early YHEI (point exposure) on later BMI (2nd wave)
- *Non-parametric* causal response curves for continuously measured YHEI
- Local adjustment set (least efficient)
- Optimal adjustment sets – *non-unique* in equivalence class
- Nonparametric estimation ('double machine learning') of effects as rough guide (post-selection-inference issues here)

Exposure: healthy-eating-index (baseline);
outcome: BMI at 2nd wave
NP-estimates of average outcome under hypothetical
intervention in exposure
for different adjustment sets and 10 multiply imputed datasets



Thank You!

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

