Causal Inference

Vanessa Didelez and Robin Evans
BIPS, University of Bremen (Germany), and University of Oxford (UK)

September 2024

APTS — Oxford

Part 5b: Causal Discovery

an Example from Epidemiology

Data in Epidemiology

- Typically observational (non-interv., non-experimental)
- Cohort studies or panel data: data collected in waves, months or years apart
  $\longrightarrow$ 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

$\Rightarrow$ Many different types of measurements

$\Rightarrow$ Often: incomplete data / missing values

$\Rightarrow$ Typically, information on partial time-ordering available

Use of Causal DAGs in Epidemiology

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

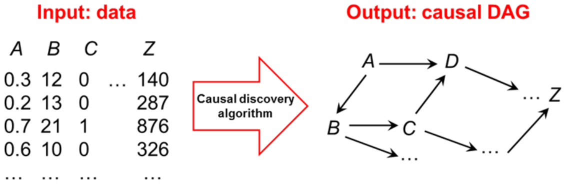
- Illustrate or examine possible sources of bias
  – e.g., due to bad design or analysis choices
  – Typically: expert-driven construction of (partial) DAG
- **Identification of causal parameters** via graphical characterization
  – e.g., explicit justification for choice of adjustment sets
  – Popular: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

(Piccininni et al. 2023 Epidemiology)

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

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

DAG not Known? $\Rightarrow$ Causal Discovery



Actually:

$\rightarrow$ need special assumptions (faithfulness, causal sufficiency likelihood, additivity, ...)

Here: constraint-based (PC)
$\rightarrow$ output not a unique DAG,
but: equivalence class

Application

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], Janine Witte[1,2], Claudia Börnhorst[1], Wencke Gwozdz[3,4], Valeria Pala[5], Lauren Lissner[6], Fabio Lauria[7], Lucia A. Reisch[1,8], Dénes Molnár[9], Stefaan De Henauw[10], Luis Moreno[11], Toomas Veidebaum[12], Michael Tornaritis[13], 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.

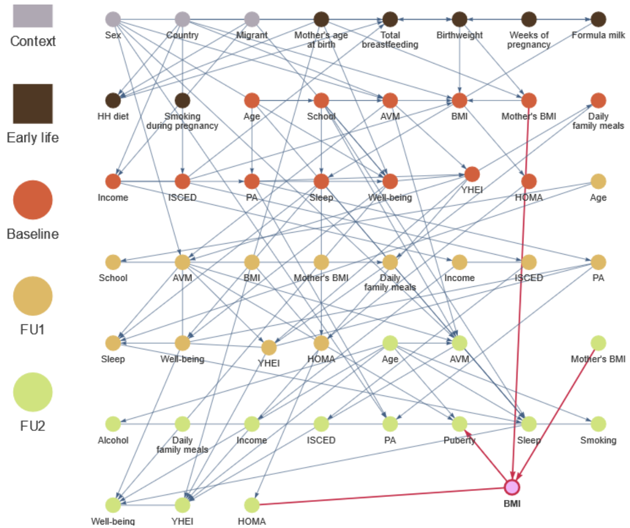
(Ahrens et al., on behalf of the I.Family Consortium, 2017)

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
Sex
Country
Migrant
Mother's age at birth
Total breastfeeding
Birthweight
Weeks of pregnancy
Formula milk
HH diet
Smoking during pregnancy
Age
School
AVM
BMI
Mother's BMI
Daily family meals
Income
ISCED
PA
Sleep
Well-being
YHEI
HOMA
Age
School
AVM
BMI
Mother's BMI
Daily family meals
Income
ISCED
PA
Sleep
Well-being
YHEI
HOMA
Age
AVM
Mother's BMI
Alcohol
Daily family meals
Income
ISCED
PA
Puberty
Sleep
Smoking
Well-being
YHEI
HOMA
BMI

Cohort Causal Graph — Stability

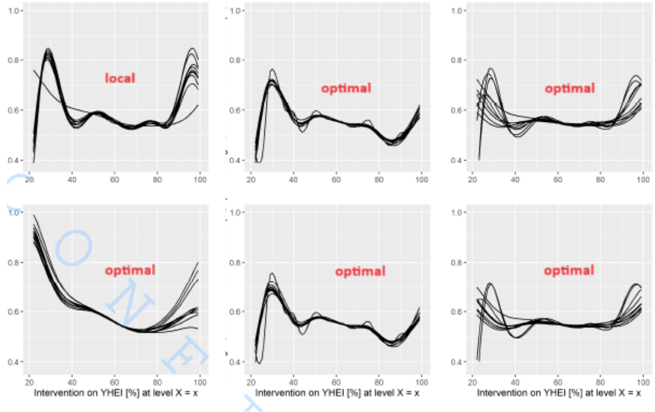
Based on 100 bootstrap graphs:
consider stability of individual (non)edges but also of specific interesting graphical structures like causal paths

- 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-healty 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

Estimated Effects

- 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