The use of the causal inference algorithm in life course epidemiology

Dr. Sacha la Bastide-van Gemert

Prof. dr. R.P. Stolk

Dr. V. Fidler

Department of Epidemiology, UMCG

Groningen, the Netherlands

Overview

- Life course epidemiology
- Causality and causal diagrams (example)
- Searching for causal models using various algorithms
- Causal diagrams: concluding remarks
- Literature

Life course epidemiology (LCE)

- LCE: studying temporal and causal interrelationships of longitudinally measured variables and outcome variable(s), across individual life course as well as across generations
- Dealing with: confounding, intermediate variables, interaction, different roles played by risk factors at different times in the same model, ...
- Searching for `the’ mechanism underlying the observational data:

→ *How does it work?*

What are the possible causal roles of the different variables?

Causality: a cause for debate

- Karl Pearson (1857-1936): association is all there is!
- Sewall Wright (1889-1988): the introduction of path analysis (1921)
- Ronald Fisher (1890-1962): randomisation and experimental control are the only reliable ways of obtaining causal knowledge

Causality: a cause for debate

- Karl Jöreskog (1935-): structural equation modelling (LISREL-model)
- Donald Rubin (1943-): the Rubin causal model (using potential outcomes and matching)
- Judea Pearl (1936-): formalization of causal reasoning and structural equation modelling (using directed acyclic graphs)

Life course epidemiology and causal inference

Theoretical framework (Pearl; Spirtes; Shipley;...):

looking for causal effects in observational data:

- causal inference
- structural equations models (SEM)
- graphical models (causal diagrams)

Practical framework:

using PREVEND-data (*Prevention of REnal and Vascular ENdstage Disease*), examining the mechanism (causal relationships) between factors influencing overweight, plasma-glucose-levels, etc... in a longitudinal data setting

Causal diagrams

Causal diagram: *a graphical representation of causal relationships between variables*

$$A \longrightarrow B$$

- Causal diagrams and the associated theory can help translate (dependencies within) a causal model into an observational (statistical) model, hence providing a way to test the plausibility of hypothesized models with observational data (*SEM, path analysis*)
- Also the other way around: the theory provides a way to search for possible (classes of) causal models consistent with given observational data using search algorithms (PC, CI, ...)

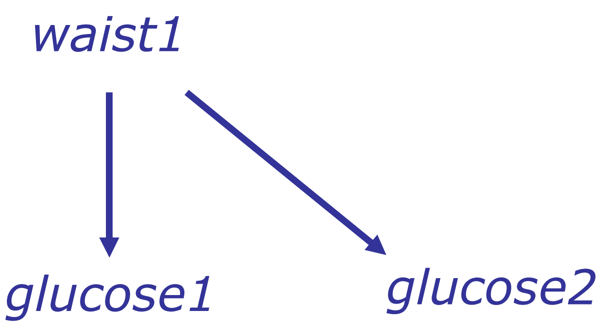
Causal diagrams: an example using Prevend data

Let’s consider the effect of *waist circumference* on *glucose level* assuming there is a direct effect, using measurements from three different examinations (here at examination 1):

$$\begin{tikzcd} \textit{waist1} \arrow[d] \\ \textit{glucose1} \end{tikzcd}$$

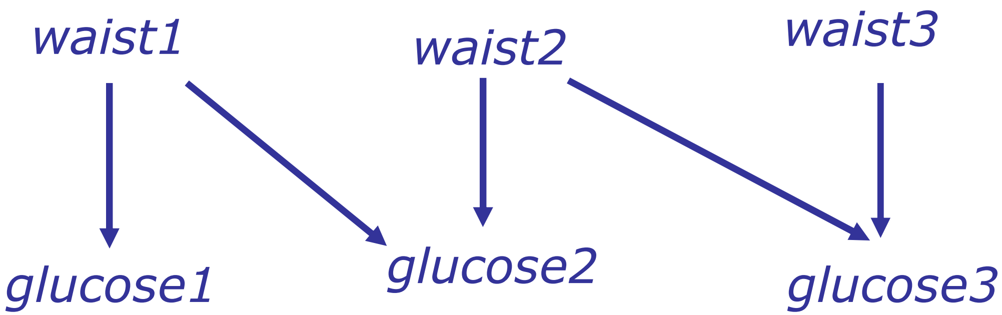
Causal diagrams: an example using Prevend data (continued)

... and an added delayed effect of *waist circumference* (at examination 1) on *glucose level* (at examination 2):



Causal diagrams: an example using Prevend data (continued)

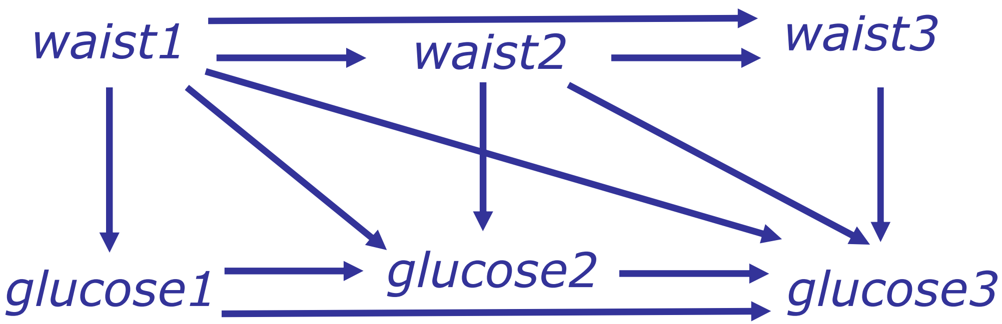
We have measurements of all variables at three moments in time, so:



This gives rise to various causal effects due to repeated measurements…

Causal diagrams: an example using Prevend data (continued)

… resulting in:

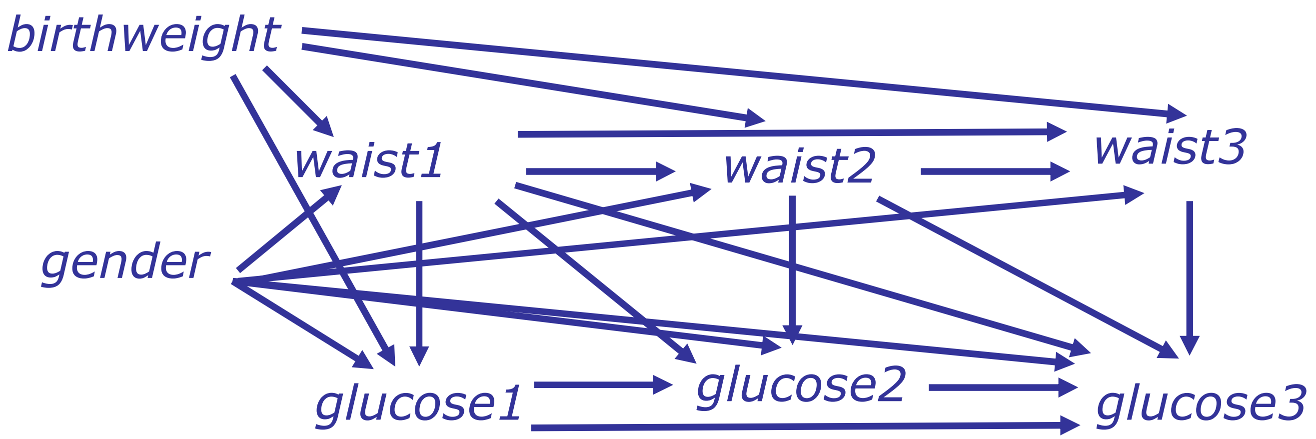


What about the background variables?

Causal diagrams: an example using Prevend data (continued)

Adding background variables like *gender*, *birthweight, etc*…

results in:



Let’s turn things around:

What causal diagrams are consistent with the given data?

Searching for causal models using Prevend data

Searching for plausible causal models consistent with the given data: applying the Causal Inference (CI) algorithm (Spirtes et al./Shipley)

Constructing graphs, consisting of the variables from the Prevend-data:

- waist-measurements (3)
- plasma-glucose levels (3)
- birthweight, gender, ...

Searching for causal models: the Causal Inference (CI) algorithm (Shipley 2000)

- Step 1: two variables in a graph are connected, when the association between them remains non-zero when conditioned on any subset of the other variables (under some assumptions)

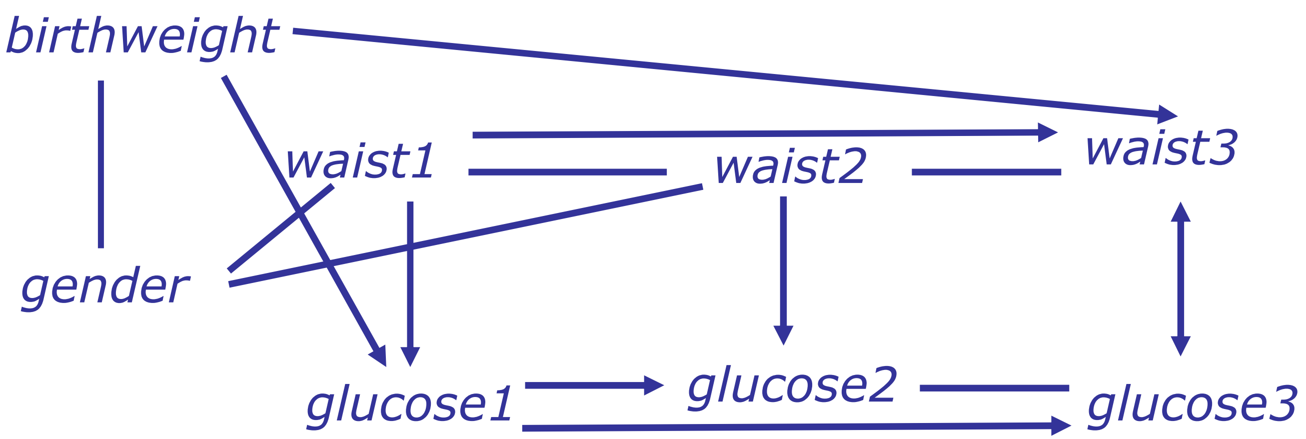
  → *undirected dependency graph*

- Step 2: orienting the edges as far as possible (using the theory of d-separation)

  → (*partially) oriented graph*

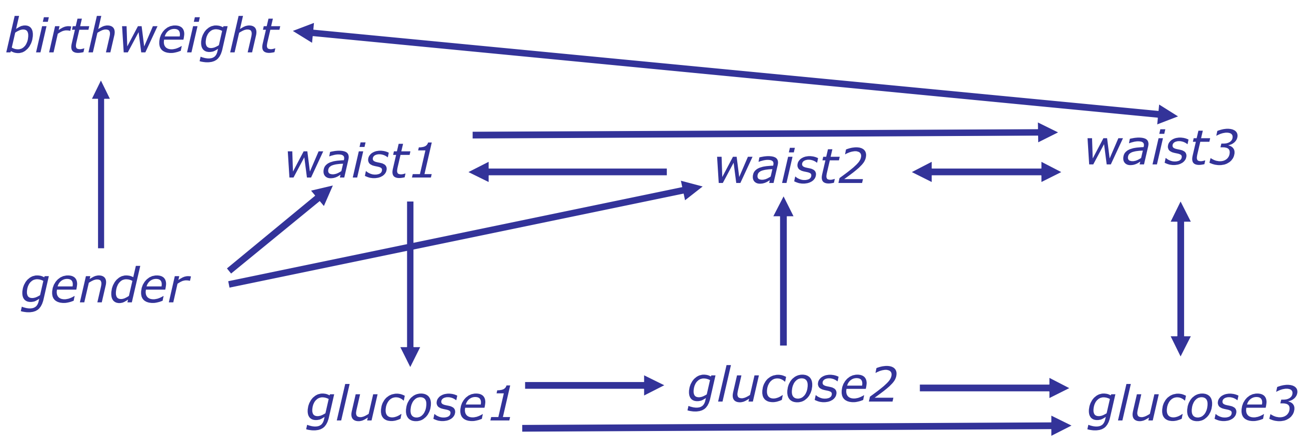
Searching for causal models using Prevend data (CI algorithm-Shipley)

Applying the CI algorithm (Shipley) results in a partially oriented graph:



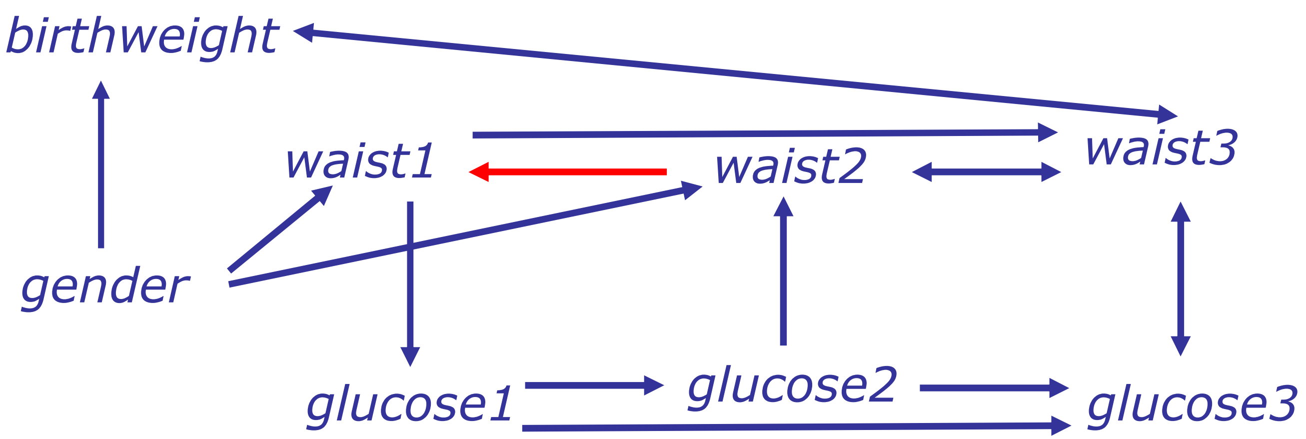
Searching for causal models using Prevend data (PC algorithm-TETRAD)

Using Tetrad (a program for creating and searching for causal models) the PC algorithm results in:



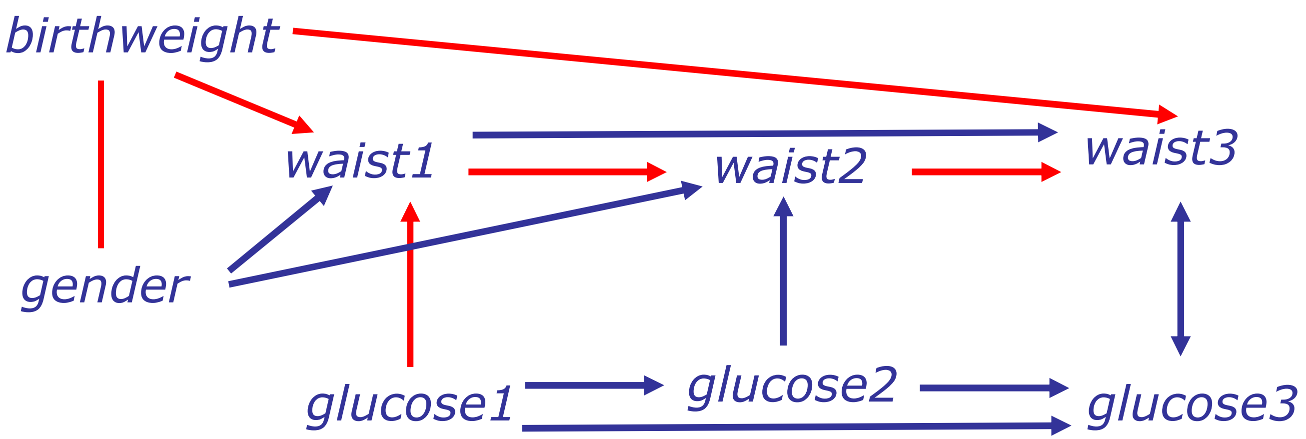
Searching for causal models using Prevend data (PC algorithm-TETRAD)

Using Tetrad (a program for creating and searching for causal models) the PC algorithm results in:



Searching for causal models using Prevend data (PC algorithm-TETRAD)

The PC algorithm with additional constraints results in:



Using causal diagrams – concluding remarks

Causal diagrams and associated theory:

- require to make implicit causal assumptions explicit
- provide a way to test hypothesized models
- as an exploratory method, they help to identify causal models compatible with the data

Questions that need further attention:

- Cyclic graphs? Non-normal data? Missing data?...
- How well can an algorithm detect the underlying causal relationships?
- When/how should additional constraints on the data be added?

Literature

- S. Greenland, J. Pearl, J.M. Robins, `Causal diagrams for epidemiologic research', *Epidemiology* 10 (1999), 37-48
- M.A. Hernán, S. Hernández-Diaz, J.M. Robins, `A structural approach to selection bias', *Epidemiology* 15 (2004), 615-625
- J. Pearl, *Causality. Models, reasoning and inference* (Cambridge 2000)
- J. M. Robins, `Data, design and background knowledge in etiologic inference', *Epidemiology* 11 (2001), 313-320
- K.J. Rothman, S. Greenland, `Causation and causal inference in epidemiology', *American Journal of Public Health* 95 (2005), S144-S150
- B. Shipley, *Cause and correlation in biology* (Cambridge, 2000)
- P. Spirtes, C. Glymour, R. Scheines, *Causation, prediction and search* (New York, 1993)
- T.J. VanderWeele, J.M. Robins, `Four types of effect modification. A classification based on directed acyclic graphs', *Epidemiology* 18 (2007), 561-568

Thank you

CI-Shipley-model vs PC-model
(slightly altered, using AMOS 16.0)

	CI-Shipley-model	PC-model
χ^2	50.5	323
p	0.000	0.000
GFI	0.997	0.981
CFI	0.998	0.981
RMSEA	0.025	0.073
BIC	234	506
CAIC	256	528

Searching for causal models using Prevend data (FCI algorithm-TETRAD)

The FCI-algorithm with additional constraints results in: