Directed Acyclic Graphs (DAGs) Jönköping 2014-19-16

Liisa Byberg and Karl Michaëlsson

Department of Surgical Sciences and Uppsala Clinical Research Center (UCR) Uppsala University

Does the use of bicycle helmets reduce the risk of internal head injuries?





How to observe?

• Compared to what, who is not exposed?

 Ideal situation: to compare each individual during a period of exposure and during the same period with no exposure → "counterfactual situation"





Foto: Shannon Pifko

Randomized clinical trial



Cohort study





Common definition of confounding

- Associated with the exposure
- Associated with the outcome
- Not in the "causal pathway" (not an intermediate)!



Example

Exposure

Physical Activity

Outcome

– Diabetes type 2

Covariates

- Smoking
- Overweight
- Blood Pressure

Results from 4 models

RR for diabetes type 2, high vs. low physical activity

	mod 1	mod 2	mod 3	mod 4
Physical Activity	0.53	0.72	0.85	0.92
Smoking		•	•	•
Overweight			•	•
Blood Pressure				•

- Best model?
 - Likelihood ratio tests or Akaike criteria \Rightarrow mod 4
 - All changes in PA estimate considered important \Rightarrow mod 4
 - P<0.05 for association with exposure and outcome \Rightarrow mod 4
 - Stepwise regression
 - Expert opinion, tradition
- Model choice can not be based on data only

The real problem is to identify appropriate covariates

Not too few and not too many...

.....selection of variables in our study with the use of DAGs

Does physical exercise lead to a lower risk of diabetes?

How do we select confounding factors?

Covariates? Suggestions?



DAG

Directed Acyclic Graph

Think chain reactions!



Exposure-Disease influenced by C?



- C can be:
 - Confounder
 - Intermediate
 - Collider
 - Effect modifier

Confounder
Common cause for E and D



- Collider
 - Common effect of E and D



New view – with DAG

Confounding factor



New view – with DAG

Collider



New view – with DAG

Intermediate



.....selection of variables in our study with the use of DAGs

Does physical exercise lead to a lower risk of diabetes?

How do we select covariates/confounding factors?

Simple example



Results from 4 models

RR for diabetes type 2, high vs. low physical activity

	mod 1	mod 2	mod 3	mod 4
Physical Activity	0.53	0.72	0.85	0.92
Smoking		•	•	•
Overweight			•	•
Blood Pressure				•

Which model is most correct?



www.dagitty.net

<u>http://dagitty.net/</u>

<u>http://dagitty.net/mkB5AU</u>

<u>http://dagitty.net/mJXX3</u>



Legend

- 🕑 exposure
- 🕕 outcome
- ancestor of exposure
- ancestor of outcome
- ancestor of exposure and outcome
- \bigcirc adjusted variable
- unobserved (latent)
- other variable
- 🛑 causal path
- biasing path

😻 Summary

exposure(s) PA outcome(s) diabetes covariates 9 causal paths 8

Adjustment for total effect

Minimal sufficient adjustment sets for estimating the total effect of PA on diabetes:

- {SES, Smoking}
- Adjustment for direct effect

Minimal sufficient adjustment sets for estimating the direct effect of PA on diabetes:

• {BMI, Fitness, Lipids, SES}