Causal Inference

Jon Michael Gran

Oslo Centre for Biostatistics and Epidemiology,
Oslo University Hospital and University of Oslo

CDDF Workshop, London, July 7th, 2016

j.m.gran@medisin.uio.no

UiO : University of Oslo

Outline

1 Introduction

Causal inference

- A very active field in statistics (and especially towards medicine, epidemiology and public health) over the last decades, formalizing a **calculus of interventions**; what is the outcome if I give *a treatment* compared to if I give *another treatment*
- Traditionally the relationship between statistics and causality has been difficult – **"association is *not* causation"** – this has now changed
- Motivated by a wish for analysing **observational data** (non-experimental) like they where randomized controlled trials
- Causal inference = statistical models + **causal assumptions**

Recent development

- The **methodological development** is to a large degree driven by the novel work of:
  - Donald **Rubin** (potential outcomes, propensity scores)
  - Judea **Pearl** (graphical models, non-parametric structural equations)
  - Jamie **Robins** (counterfactuals, time-dependent confounding)

  in the 1980's and later
- The work by these three and their colleagues form three slightly different schools of causal inference, both in substance and form, but they are all **closely connected**

Origin

- The use of counterfactuals (or potential outcomes) goes back to **Neyman (1923) and Fisher (1935)**, who applied them in experimental studies (randomized trials)
- **Rubin (1974)** formalized them for observational studies; later extended by **Robins (1986)**
- The use of graphical models is largely due to **Pearl (1986)**, but also parallel work of many others
- Graphical modelling and Pearl have roots in econometric literature and structural equation modelling, going back to Norwegian Nobel price winner **Trygve Haavelmo (1943)**

How to identify causal effects

- **The gold standard** for identifying causal treatment effects are randomized controlled trials; **RCT's**
- In observational studies on the other hand, data are collected without the researcher being able to affect treatment assigment; **the causal inference approach** is then:
  1. Specify a **well-defined intervention**
  2. build a **model** that adjust for the lack of randomization
  3. List the **assumptions** needed for the estimate to be causal
- The three general assumptions for causal inference: 1) **No unmeasured confounding**, 2) positivity and 3) consistency

Why analyse observational data

- **RCT's can be difficult**: expensive, time consuming, unethical or to risky - e.g. in children or among pregnant women
- Even when RCT's exist well-performed observational studies can **add to the overall evidence**; confirming or contradicting previous studies, subgroup analyses
- Most often **RCT's also have observational data in them** (post exposure measurements); can be used to understand droput, non-complience or the mechanisms behind how treatment works

Why analyse observational data

- **RCT's can be difficult**: expensive, time consuming, unethical or to risky - e.g. in children or among pregnant women
- Even when RCT's exist well-performed observational studies can **add to the overall evidence**; confirming or contradicting previous studies, subgroup analyses
- Most often **RCT's also have observational data in them** (post exposure measurements); can be used to understand droput, non-complience or the mechanisms behind how treatment works

Of course, causal inference is not about replacing RCT's,
but **complimenting** them

Two causal frameworks

- The Pearl **framework of graphical models** and non-parametric structural equations:
  - The graphical properties makes for a well-suited tool for applied researchers
- **The framework of counterfactuals** (a.k.a. potential outcomes)
  - May seem mathematically stringent, but having a well-defined framework for looking at interventions has lead to various methodological breakthroughs

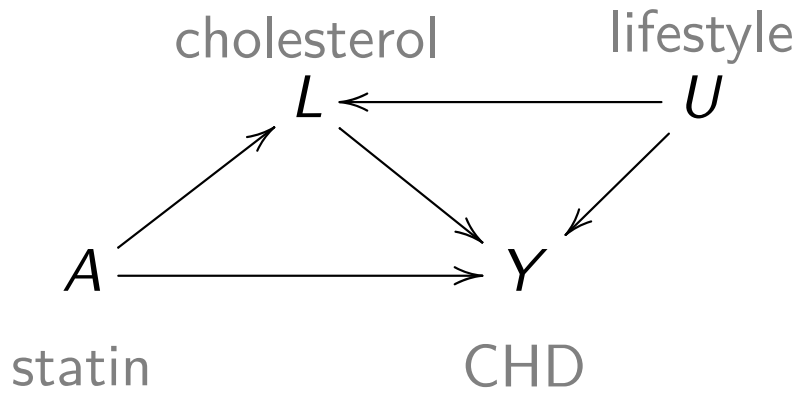
  Most fundamental concepts are equivalent in both frameworks

Let us look at the basic concepts in these two frameworks

2 Grapical models

Causal directed acyclic graphs (DAGs)

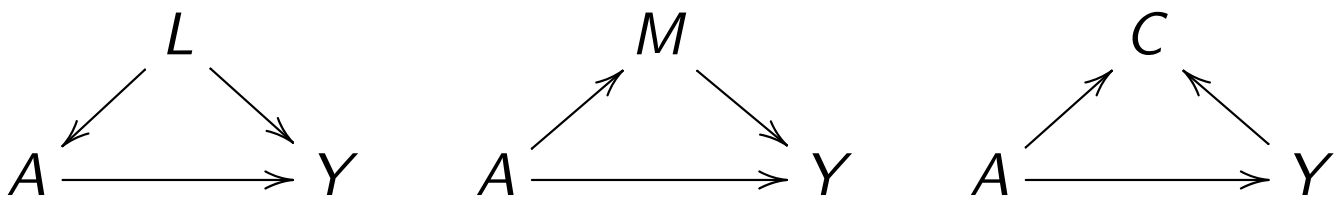
- Draw **nodes for all the relevant variables** with lack of an arrow meaning absence of direct causal effect, for example:



- A graph alone is "just dots and arrows"; the strength of causal DAGs is it's **underlying connection to conditional probabilities**
- Basicly; causal DAGs **can help identify which variables to condition on** when estimating the effect of A on Y

Causal structures

- The **three basic causal structures** between a treatment variable A, an outcome Y and a third variable:



Confounder **Mediator** **Collider**

- **Concept**:
  1. **Adjust for confounders**
  2. Only adjust for mediators if you want direct effect
  3. **Never adjust for colliders**

- In larger DAGs **look at all possible paths** from A to Y (formally based on d-separation or the adjustment criterion)

Pros of DAGs

- Despite the mathmatical theory behind; **an intuitive and algebra-free approach** to complicated causal questions
- Helpful in **clarifying advanced confounding problems** and various epidemiological "paradoxes" (collider biases)

Cons

- DAGs **only tell us the presence or absence of bias**, not how much or in what direction
- Whether a DAG capture reality is **not testable**
- Nodes in a DAGs represent variables, but **researchers often think in terms of processes** when drawing them

3 Counterfactual models

The essentials

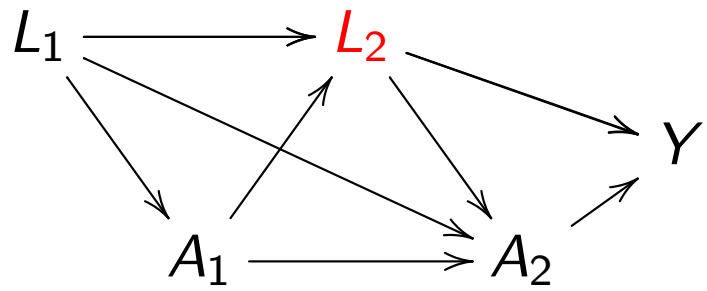
- Want to model counterfactual scenarios: **what would have happened if everyone was on treatment 1?** And if everyone was on treatment 0?
- Want to find the **average causal effect**

$$\text{ACE} = E(Y^1) - E(Y^0),$$

  where Y^a are counterfactual random variables
- Related to **missing data problems**; there are always missing counterfactuals
- **Fixed treatments**: propensity score methods through the work of Donald Rubin (calls Y^a potential outcomes)
- **Time-varying treatments**: addressing time-dependent confounding through the work of Jamie Robins

Time-dependent confounding (feedback)

- Present when **covariates *L*, affected by past exposure A, both affects future exposure and the outcome *Y***

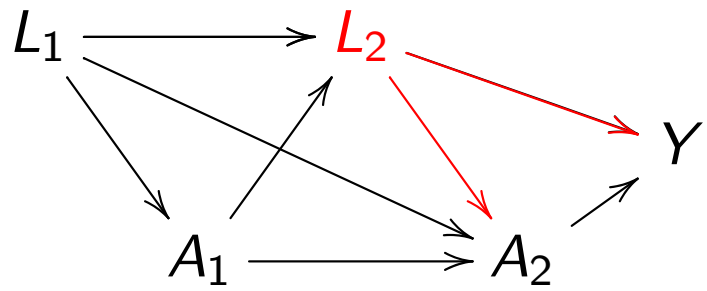


Simplified DAG with only two time points.

- Classical example: the effect of **HIV treatment *A*** on time to AIDS or death Y, confounded by CD4 cell count L
- Can't both adjust and *not* adjust for L_2 – **need advanced methods** to handle this problem

Time-dependent confounding (feedback)

- Present when **covariates *L*, affected by past exposure A, both affects future exposure and the outcome *Y***

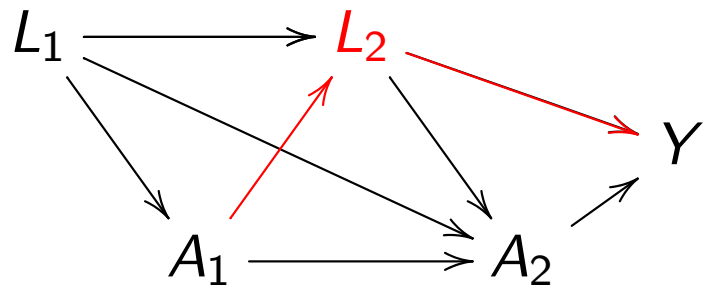


Simplified DAG with only two time points.

- Classical example: the effect of **HIV treatment *A*** on time to AIDS or death Y, confounded by CD4 cell count L
- Can't both adjust and *not* adjust for L_2 – **need advanced methods** to handle this problem

Time-dependent confounding (feedback)

- Present when **covariates *L*, affected by past exposure A, both affects future exposure and the outcome *Y***

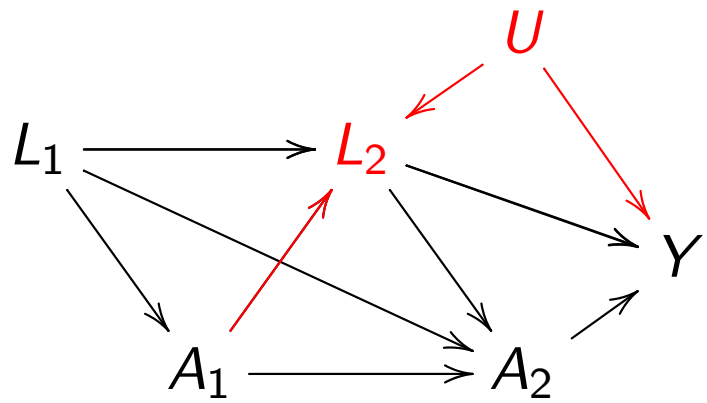


Simplified DAG with only two time points.

- Classical example: the effect of **HIV treatment *A*** on time to AIDS or death Y, confounded by CD4 cell count L
- Can't both adjust and *not* adjust for L_2 – **need advanced methods** to handle this problem

Time-dependent confounding (feedback)

- Present when **covariates *L*, affected by past exposure A, both affects future exposure and the outcome *Y***



Simplified DAG with only two time points.

- Classical example: the effect of **HIV treatment *A*** on time to AIDS or death Y, confounded by CD4 cell count L
- Can't both adjust and *not* adjust for L_2 – **need advanced methods** to handle this problem

G-methods

- Starting with his 1986 paper, Robins have revolutionized how we analyse observational data proposing **three very general methods for dealing with time-dependent confounding**;

  1. **G-computation**
  2. **Inverse probability of treatment weighting**
  3. **G-estimation**

  the two first are so-called marginal structural models (MSMs)
- **Extensions** include doubly robust methods, mediation analysis, dynamic treatment regimes and optimal treatment regimes (personalized medicine)

The classical example

- Two papers analyzing HIV cohort data in **Epidemiology 2000 and Lancet 2005** with respect to effects of early HAART treatment on time to AIDS or death

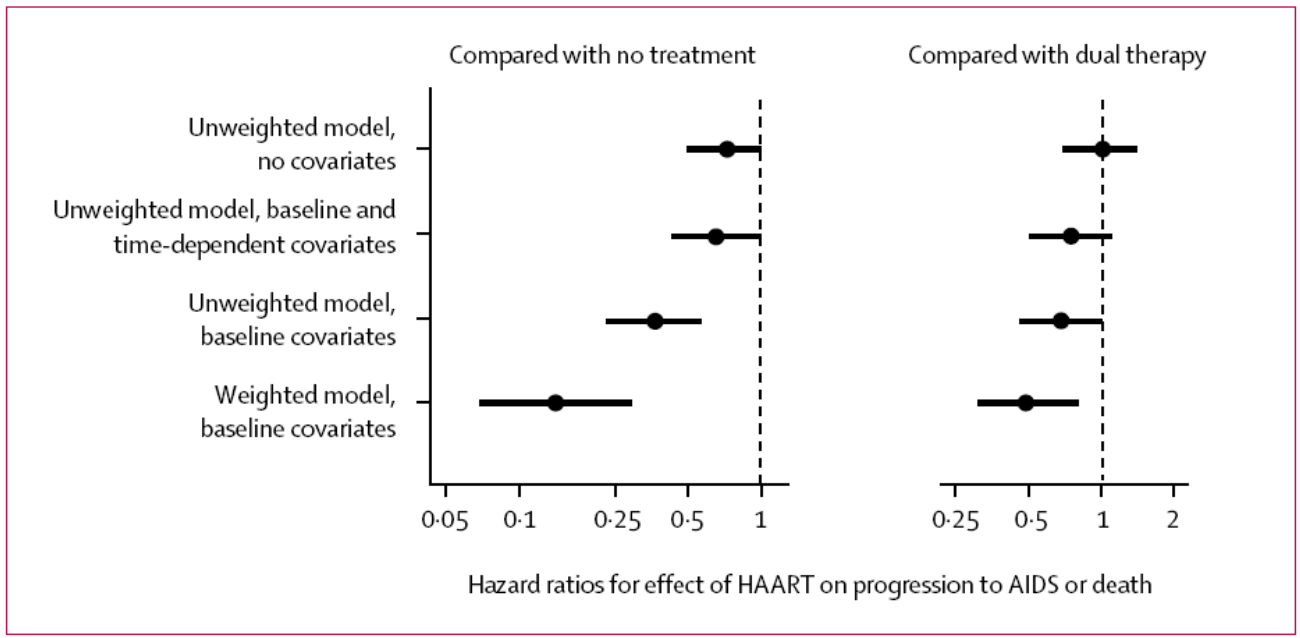


Figure: **Estimated effect of HAART from unweighted (standard) and weighted Cox models**
Weighted model with baseline covariates estimates parameters of marginal structural model. Weights adjust for confounding due to measured time-dependent covariates.

Results from Sterne et al. Lancet 2005.

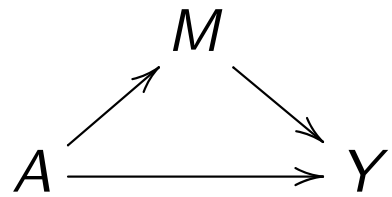
- **Detailed cohort data was collected** at introduction in 1996

Examples with cancer

- The 1982-88 aspirin component of the **Physicians' Health Study**, a randomized trial of aspirin on cardiovascular disease and cancer was stopped due to an extreme reduction in first myocardial infarction; analysis of mortality endpoint analyzed using MSMs in Am J Epidemiol 2002
- Also various studies using the **Nurses' Health Study**; e.g. Smoking and survival after breast cancer diagnosis, Int J Cancer 2007, Mammographic Screening and Risk Factors for Breast Cancer, Am J Epidemiol 2009
- Cancer is one possible outcome when looking at the potential in Scandinavian **registry studies** through data linkage

Mediation analysis

- Want to quantify the effect through M (**the mediating effect**) and the effect *not* going through M (**the direct effect**)



- **Drastical improvement from traditional mediation analysis** through well-defined (natural and controlled) direct and indirect effects based on counterfactuals; e.g.:

$$Y^{A=0,M=m^{a1}} - Y^{A=0,M=m^{a0}} \quad \text{(natural indirect effect)}$$

- From Baron and Kennys 1986 paper ($>$ 60.000 cites) to the new "bible" for mediation by **Tyler VanderWeele, 2015**

4 Summary

Causal inference

- Causal modelling can be theoretically heavy, but when stuck, the beauty of it is that one **can separate completely between the advanced method and an easy to interpret causal contrast**;

"what is the hypothetical trial?"

- **Major impact** in: 1) Graphical criteria for confounder control 2) methods for time-dependent confounding that where not around two decades ago, 3) a formalized language for analyzing interventions, mediating effects and more, and hence;
  4) improved utilization of observational data
- Our **causal inference group in Oslo** has a particular focus on time-to-event outcomes and processes
- Not mentioned: **adjusting for unobserved confounding**