

# *What are graphical models?*

Sara Geneletti

Department of Epidemiology and Public Health, Imperial College

02/07/2008

# Outline

1. Introduction
2. What is a DAG?
3. What can it do?
4. What does it mean?
5. Heuristic tool
6. Formal tool
7. Causality

# *Introduction*

## *Uses*

- ▶ Physics
- ▶ Genetics
- ▶ Psychology - Path analysis, Structural equation models
- ▶ Statistics
- ▶ Causal inference

# Introduction

## Uses

- ▶ Physics
- ▶ Genetics
- ▶ Psychology - Path analysis, Structural equation models
- ▶ Statistics
- ▶ Causal inference

## Types

- ▶ Directed
- ▶ Directed Acyclic
- ▶ Unidirected
- ▶ Chain graphs

## *What is a DAG?*

DAGs are *directed acyclic graphs*

- ▶ All arrows have direction

## *What is a DAG?*

DAGs are *directed acyclic graphs*

- ▶ All arrows have direction
- ▶ No cycles  $A \rightarrow B \rightarrow A$

## What is a DAG?

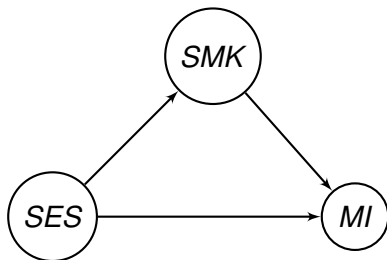
DAGs are *directed acyclic graphs*

- ▶ All arrows have direction
- ▶ No cycles  $A \rightarrow B \rightarrow A$
- ▶ Arrows are *not* causal unless extra assumptions made - time ordering, intervention

## What is a DAG?

DAGs are *directed acyclic graphs*

- ▶ All arrows have direction
- ▶ No cycles  $A \rightarrow B \rightarrow A$
- ▶ Arrows are *not* causal unless extra assumptions made - time ordering, intervention



## *What does it do?*

DAGs are used to encode **conditional independence statements**

- ▶ In words if we know about  $C$ , knowing about  $A$  gives us no extra clues about  $B$  (and vice-versa)

## *What does it do?*

DAGs are used to encode **conditional independence statements**

- ▶ In words if we know about  $C$ , knowing about  $A$  gives us no extra clues about  $B$  (and vice-versa)
- ▶ Formally, we write  $A \perp\!\!\!\perp C | B$  [1]
- ▶ which means  $p(A, C | B) = p(A | B)p(C | B)$

## *What does it do?*

DAGs are used to encode **conditional independence statements**

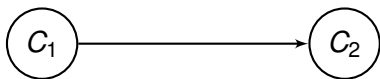
- ▶ In words if we know about  $C$ , knowing about  $A$  gives us no extra clues about  $B$  (and vice-versa)
- ▶ Formally, we write  $A \perp\!\!\!\perp C | B$  [1]
- ▶ which means  $p(A, C | B) = p(A | B)p(C | B)$
- ▶ Although DAGs have arrows, they DO NOT automatically mean causal relationships

## *What does it do?*

DAGs are used to encode **conditional independence statements**

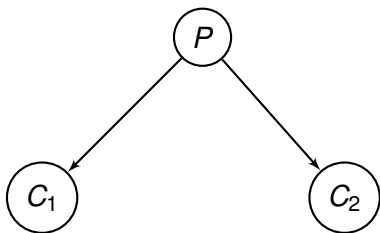
- ▶ In words if we know about  $C$ , knowing about  $A$  gives us no extra clues about  $B$  (and vice-versa)
- ▶ Formally, we write  $A \perp\!\!\!\perp C | B$  [1]
- ▶ which means  $p(A, C | B) = p(A | B)p(C | B)$
- ▶ Although DAGs have arrows, they DO NOT automatically mean causal relationships
- ▶ rather an arrow means dependence/association and lack of an arrow means independence/no association

## Simple example - inheritance



1. Two children are siblings
2. If you know the DNA of one, you know something about the DNA of the other
3. they are **associated**

## Simple example - inheritance



1. Two children are siblings
2. If you know the DNA of one, you know something about the DNA of the other
3. they are **associated**
4. If you know their parents' DNA however
5. knowing about one child tells you nothing new about the other
6. they are **independent GIVEN the parents**

## *Qualitative approach*

- ▶ DAGs can be constructed to make sense of a particular set of relationships

## *Qualitative approach*

- ▶ DAGs can be constructed to make sense of a particular set of relationships
- ▶ Make it easier for - **qualitative and quantitative researchers** to understand one another

## *Qualitative approach*

- ▶ DAGs can be constructed to make sense of a particular set of relationships
- ▶ Make it easier for - **qualitative and quantitative researchers** to understand one another
- ▶ Pictorial representation can **highlight uncertainty and bias**

### *Caveats*

- ▶ a DAG that expresses assumptions about relationships (i.e. pre-data analysis) does not necessarily correspond to reality

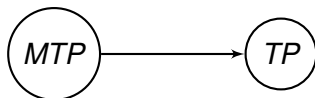
## *Qualitative approach*

- ▶ DAGs can be constructed to make sense of a particular set of relationships
- ▶ Make it easier for - **qualitative and quantitative researchers** to understand one another
- ▶ Pictorial representation can **highlight uncertainty and bias**

### *Caveats*

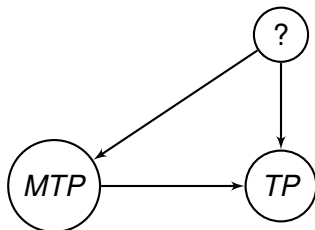
- ▶ a DAG that expresses assumptions about relationships (i.e. pre-data analysis) does not necessarily correspond to reality
- ▶ Putative associations/causal relations need to be tested against data where possible and assessed carefully

## Constructing a DAG



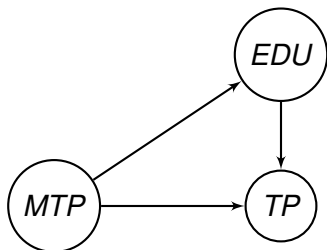
- ▶ A teenager whose mother had children as a teenager is more likely to have children herself

## Constructing a DAG



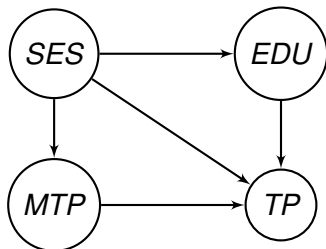
- ▶ A teenager whose mother had children as a teenager is more likely to have children herself
- ▶ BUT there are factors that influence both these events

## Constructing a DAG



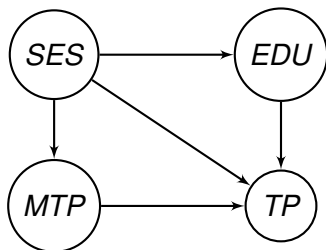
- ▶ A teenager whose mother had children as a teenager is more likely to have children herself
- ▶ BUT there are factors that influence both these events
- ▶ Education (full-time vs school leaver) is one of these
- ▶ But surely that is influenced in its own way by?? Anyone?

## Constructing a DAG



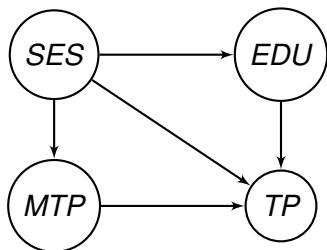
- ▶ A teenager whose mother had children as a teenager is more likely to have children herself
- ▶ BUT there are factors that influence both these events
- ▶ Education (full-time vs school leaver) is one of these
- ▶ But surely that is influenced in its own way by?? Anyone?
- ▶ SES

## Constructing a DAG



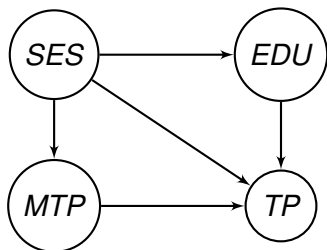
- ▶ This is a simple example - could add
  - ▶ Ethnicity

## Constructing a DAG



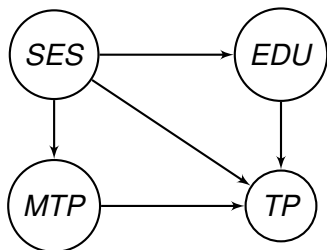
- ▶ This is a simple example - could add
  - ▶ Ethnicity
  - ▶ Low-self esteem

## Constructing a DAG



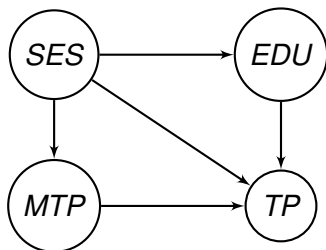
- ▶ This is a simple example - could add
  - ▶ Ethnicity
  - ▶ Low-self esteem
  - ▶ Substance abuse

## Constructing a DAG



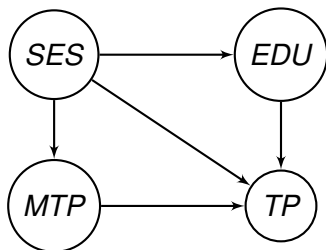
- ▶ This is a simple example - could add
  - ▶ Ethnicity
  - ▶ Low-self esteem
  - ▶ Substance abuse
  - ▶ history of violence

## Constructing a DAG



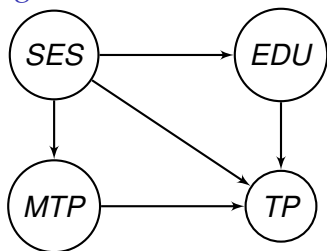
- ▶ This is a simple example - could add
  - ▶ Ethnicity
  - ▶ Low-self esteem
  - ▶ Substance abuse
  - ▶ history of violence
- ▶ Some of these could be unobserved or reported with bias

## Constructing a DAG



- ▶ This is a simple example - could add
  - ▶ Ethnicity
  - ▶ Low-self esteem
  - ▶ Substance abuse
  - ▶ history of violence
- ▶ Some of these could be unobserved or reported with bias
- ▶ e.g. low-self esteem or substance abuse

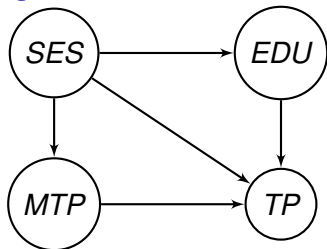
## *Incorporating data*



$$\Pr(TP|MTP) = \sum_{SES, EDU} \Pr(TP|MTP, SES, EDU) \Pr(EDU|SES) \Pr(MTP|SES) \Pr(SES)$$

can use frequencies from contingency tables to estimate  $\Pr(TP|MTP)$  and Odds Ratio

## Incorporating data

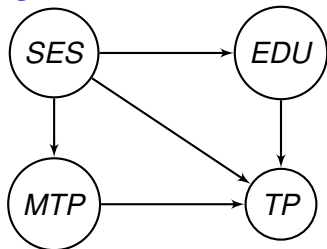


$$Pr(TP|MTP) = \sum_{SES, EDU} Pr(TP|MTP, SES, EDU)Pr(EDU|SES)Pr(MTP|SES)Pr(SES)$$

can use frequencies from contingency tables to estimate  $Pr(TP|MTP)$  and Odds Ratio

- ▶ The graph tells us how to **factorise** the distribution of variables into smaller simple parts

## Incorporating data

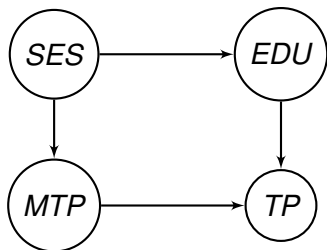


$$Pr(TP|MTP) = \sum_{SES, EDU} Pr(TP|MTP, SES, EDU)Pr(EDU|SES)Pr(MTP|SES)Pr(SES)$$

can use frequencies from contingency tables to estimate  $Pr(TP|MTP)$  and Odds Ratio

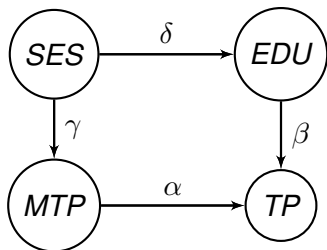
- ▶ The graph tells us how to **factorise** the distribution of variables into smaller simple parts
- ▶ Helps to estimate using a **modular** approach - see later

## Incorporating data



- ▶ We can do a path analysis [2] by assuming linear relationships between the variables
- ▶ For example, if we think that the influence of SES on TP is **mediated only** by MTP and EDU
- ▶ i.e.  $TP \perp\!\!\!\perp SES | (MTP, EDU)$  then

## Incorporating data



- ▶ We can do a path analysis [2] by assuming linear relationships between the variables
- ▶ For example, if we think that the influence of SES on TP is **mediated only** by MTP and EDU
- ▶ i.e.  $TP \perp\!\!\!\perp SES | (MTP, EDU)$  then
- ▶  $TP = \mu_1 + \alpha MTP + \beta EDU + \epsilon$
- ▶  $MTP = \mu_2 + \gamma SES + \epsilon$  and  $EDU = \mu_3 + \delta SES + \epsilon$

## *Does the DAG correspond to reality?*

*True?*

- ▶ So you have a DAG that represents your belief about the relationships

## *Does the DAG correspond to reality?*

*True?*

- ▶ So you have a DAG that represents your belief about the relationships
- ▶ Does it fit with observed data?

1. What conditional independences does DAG encode?

## *Does the DAG correspond to reality?*

### *True?*

- ▶ So you have a DAG that represents your belief about the relationships
- ▶ Does it fit with observed data?

1. What conditional independences does DAG encode?
2. Moralisation criteria (see next slide)
3. Use e.g.  $\chi^2$  or Mantel-Haenszel test (or Bayesian network software) to determine if hold in data

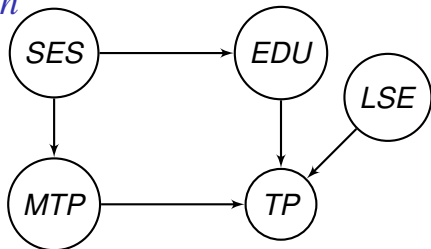
## *Does the DAG correspond to reality?*

### *True?*

- ▶ So you have a DAG that represents your belief about the relationships
- ▶ Does it fit with observed data?

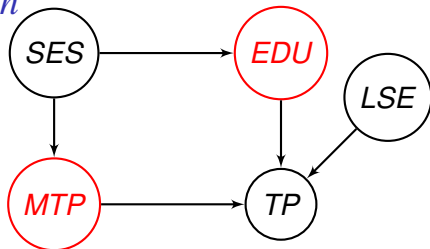
1. What conditional independences does DAG encode?
2. Moralisation criteria (see next slide)
3. Use e.g.  $\chi^2$  or Mantel-Haenszel test (or Bayesian network software) to determine if hold in data
4. Regressions - if adding a variable to reg makes no difference to the outcome - maybe there is no dependence (not 100%).

## *Moralisation*



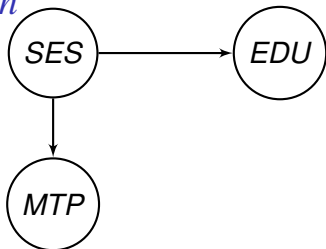
- ▶ Say you care about relationship between EDU and MTP

## Moralisation



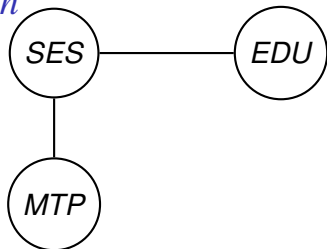
- ▶ Say you care about relationship between EDU and MTP
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP -only SES here

## Moralisation



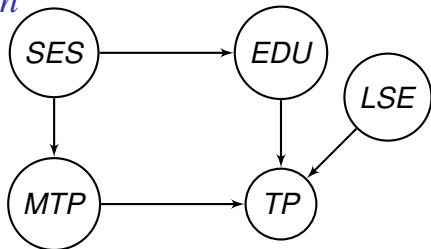
- ▶ Say you care about relationship between EDU and MTP
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP -only SES here
- ▶ Join (marry - hence moralise) parents of common children (none here)
- ▶ remove direction from arrows

## Moralisation



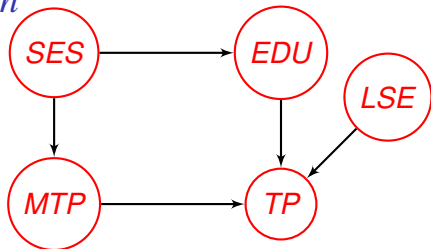
- ▶ Say you care about relationship between EDU and MTP
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP -only SES here
- ▶ Join (marry - hence moralise) parents of common children (none here)
- ▶ remove direction from arrows
- ▶ all paths from EDU and MTP go through SES -  
 $MTP \perp\!\!\!\perp EDU \mid SES$
- ▶ i.e. mother being a teen mum is only associated to daughter's education via SES - makes sense?

## *Moralisation*



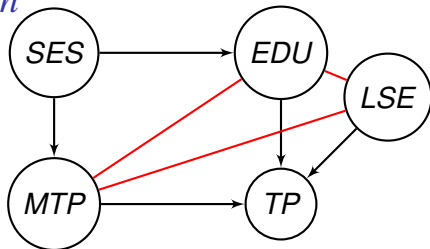
- ▶ Say you care about relationship between TP and SES

## *Moralisation*



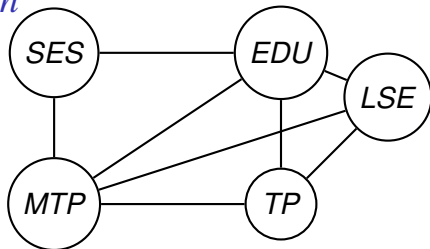
- ▶ Say you care about relationship between TP and SES
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP - all variables are ancestors of TP

## Moralisation



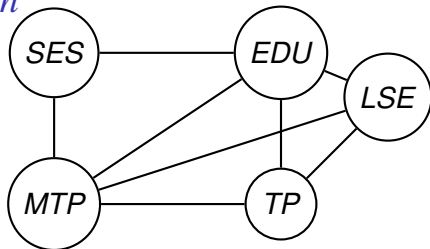
- ▶ Say you care about relationship between TP and SES
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP - all variables are ancestors of TP
- ▶ Join parents of common children

## Moralisation



- ▶ Say you care about relationship between TP and SES
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP - all variables are ancestors of TP
- ▶ Join parents of common children
- ▶ remove direction from arrows

## Moralisation



- ▶ Say you care about relationship between TP and SES
- ▶ Exclude all variables that are not **ancestors** of EDU and MTP - all variables are ancestors of TP
- ▶ Join parents of common children
- ▶ remove direction from arrows
- ▶ all paths from SES and TP go through EDU and MTP -  $TP \perp\!\!\!\perp SES \mid (MTP, EDU)$
- ▶ i.e. being a teen mum is only associated to SES via mother's teen mum status and education - not plausible, need more confounders!

# Getting DAGs from data

## *Data mining*

- ▶ There are various methods for extracting DAGs from data
- ▶ Most ask what the conditional independences are between variables (using e.g.  $\chi^2$  tests) and construct a series of DAGs

# Getting DAGs from data

## *Data mining*

- ▶ There are various methods for extracting DAGs from data
- ▶ Most ask what the conditional independences are between variables (using e.g.  $\chi^2$  tests) and construct a series of DAGs
- ▶ There are also loads of computer programmes that take data and turn it into DAGs

## Simple example

Political affiliation (PA), abuse as a child (AC) and abusive parent (AP) [3]

### Contingency table

Obs		PA			
AC	AP	l	s	r	tot
1	1	<b>12</b>	<b>27</b>	<b>58</b>	
	0	<b>7</b>	<b>28</b>	<b>30</b>	
0	1	<b>9</b>	<b>5</b>	<b>9</b>	
	0	<b>19</b>	<b>15</b>	<b>18</b>	
	tot				

## Simple example

Political affiliation (PA), abuse as a child (AC) and abusive parent (AP)

*Contingency table*

Obs		PA			
AC	AP	l	s	r	tot
1	1	<b>12</b>	<b>27</b>	<b>58</b>	97
	0	<b>7</b>	<b>28</b>	<b>30</b>	65
		19	55	88	162
0	1	<b>9</b>	<b>5</b>	<b>9</b>	23
	0	<b>19</b>	<b>15</b>	<b>18</b>	52
	tot	28	20	27	75

## Simple example

Political affiliation (PA), abuse as a child (AC) and abusive parent (AP)

### Contingency table

Obs		PA					Exp		PA				
AC	AP	l	s	r	tot	AC	AP	l	s	r	tot		
1	1	<b>12</b>	<b>27</b>	<b>58</b>	97	1	1	<b>12</b>	<b>33</b>	<b>53</b>	97		
	0	<b>7</b>	<b>28</b>	<b>30</b>	<b>65</b>		0	<b>8</b>	<b>22</b>	<b>35</b>	65		
		19	55	88	162			19	55	88	162		
0	1	<b>9</b>	<b>5</b>	<b>9</b>	23	0	1	<b>9</b>	<b>6</b>	<b>8</b>	23		
	0	<b>19</b>	<b>15</b>	<b>18</b>	52		0	<b>19</b>	<b>14</b>	<b>19</b>	52		
	tot	28	20	27	75		tot	28	20	27	75		

## Simple example

Political affiliation (PA), abuse as a child (AC) and abusive parent (AP)

*Contingency table*

Obs		PA				Exp		PA			
AC	AP	l	s	r	tot	AC	AP	l	s	r	tot
1	1	<b>12</b>	<b>27</b>	<b>58</b>	97	1	1	<b>12</b>	<b>33</b>	<b>53</b>	97
	0	<b>7</b>	<b>28</b>	<b>30</b>	<b>65</b>		0	<b>8</b>	<b>22</b>	<b>35</b>	65
		19	55	88	162			19	55	88	162
0	1	<b>9</b>	<b>5</b>	<b>9</b>	23	0	1	<b>9</b>	<b>6</b>	<b>8</b>	23
	0	<b>19</b>	<b>15</b>	<b>18</b>	52		0	<b>19</b>	<b>14</b>	<b>19</b>	52
	tot	28	20	27	75		tot	28	20	27	75

## Simple example

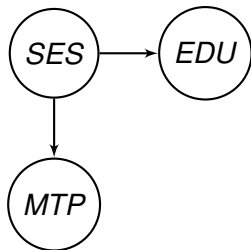
Political affiliation (PA), abuse as a child (AC) and abusive parent (AP)

Contingency table

Obs		PA				Exp		PA			
AC	AP	l	s	r	tot	AC	AP	l	s	r	tot
1	1	<b>12</b>	<b>27</b>	<b>58</b>	97	1	1	<b>12</b>	<b>33</b>	<b>53</b>	97
	0	<b>7</b>	<b>28</b>	<b>30</b>	<b>65</b>		0	<b>8</b>	<b>22</b>	<b>35</b>	65
		19	55	88	162			19	55	88	162
0	1	<b>9</b>	<b>5</b>	<b>9</b>	23	0	1	<b>9</b>	<b>6</b>	<b>8</b>	23
	0	<b>19</b>	<b>15</b>	<b>18</b>	52		0	<b>19</b>	<b>14</b>	<b>19</b>	52
	tot	28	20	27	75		tot	28	20	27	75

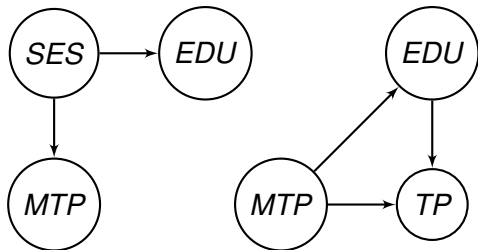
The two tables are very similar and “say” that  $PA \perp\!\!\!\perp AP \mid AC$

## *DAGs are modular*



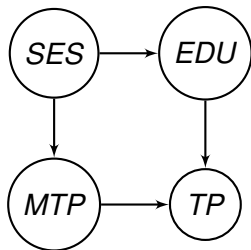
- ▶ Data source 1: SES,EDU, MTP

## *DAGs are modular*



- ▶ Data source 1: SES, EDU, MTP
- ▶ Data source 2: MTP, EDU and TP

## *DAGs are modular*



- ▶ Data source 1: SES, EDU, MTP
- ▶ Data source 2: MTP, EDU and TP
- ▶ Can join two sources to make inference about SES and TP!

# Causal inference

## Types

- ▶ Potential outcomes/Counterfactuals (Rubin [4], Pearl [5])
- ▶ Causal Graphs (Pearl [5], Greenland, Robins [6])
- ▶ Decision theory (Dawid [7], Geneletti [8], Didelez [9])

## General issues

- ▶ **no causation w/out manipulation**
- ▶ Means need to be careful about observational data
- ▶ typically there are unobserved confounders, reporting bias etc
- ▶ Causality is an external assumption

## *Advert!*

- ▶ One day course in “How to use graphical models to understand relationships between variable”
- ▶ Date TBA - April 2009
- ▶ email me [s.geneletti@imperial.ac.uk](mailto:s.geneletti@imperial.ac.uk) for further info!

## BIBLIOGRAPY

- [1] A. P. Dawid. Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 41(1):1–31, 1979.
- [2] D. Kaplan. *Structural Equation Modeling: Foundations and Extensions*. SAGE, 2000.
- [3] S.L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, 1996.
- [4] Donald B. Rubin. Bayesian Inference for Causal Inference: The Role of Randomization. *Annals of Statistics*, 6(1):34–58, 1978.
- [5] Judea Pearl. *Causality*. Cambridge University Press, 2000.
- [6] J. Robins and S. Greenland. Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology*, 3:143–155, 1992.
- [7] A. P. Dawid. Causal Inference without Counterfactuals (with comments and rejoinder). *Journal of American Statistical Association*, 95(450):407–448, 2000.
- [8] S. Geneletti. Direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society Series B*, 69(2):199–215, 2007.
- [9] N. Sheehan and V. Didelez. Mendelian randomisation as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16, 2007.