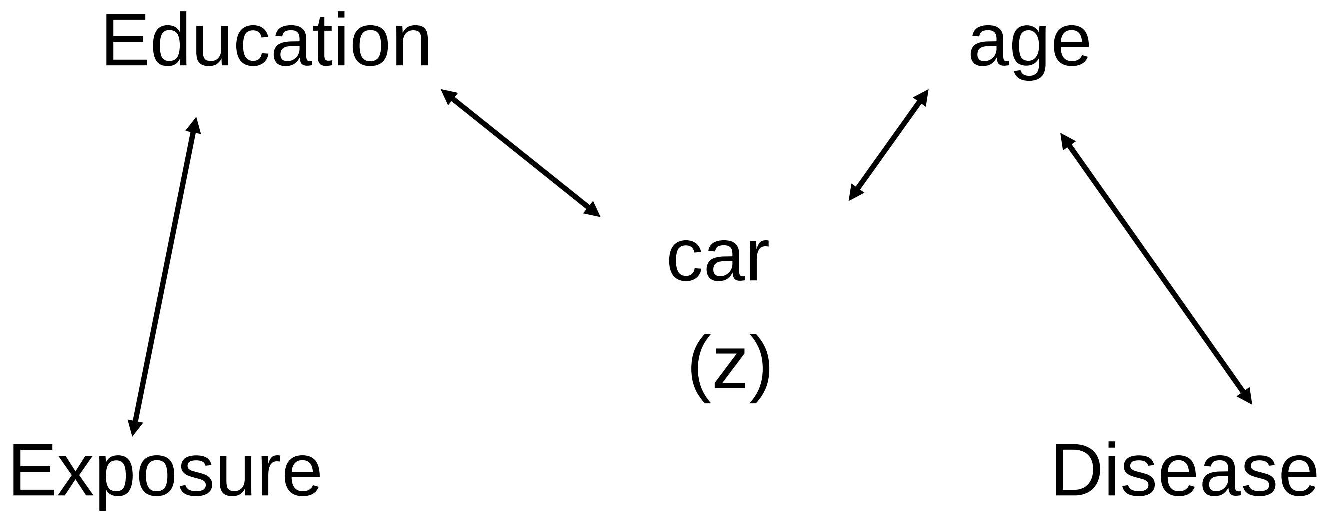
Directed Acyclic Graphs (DAGS)

Pearl, 2000. “Causality”*

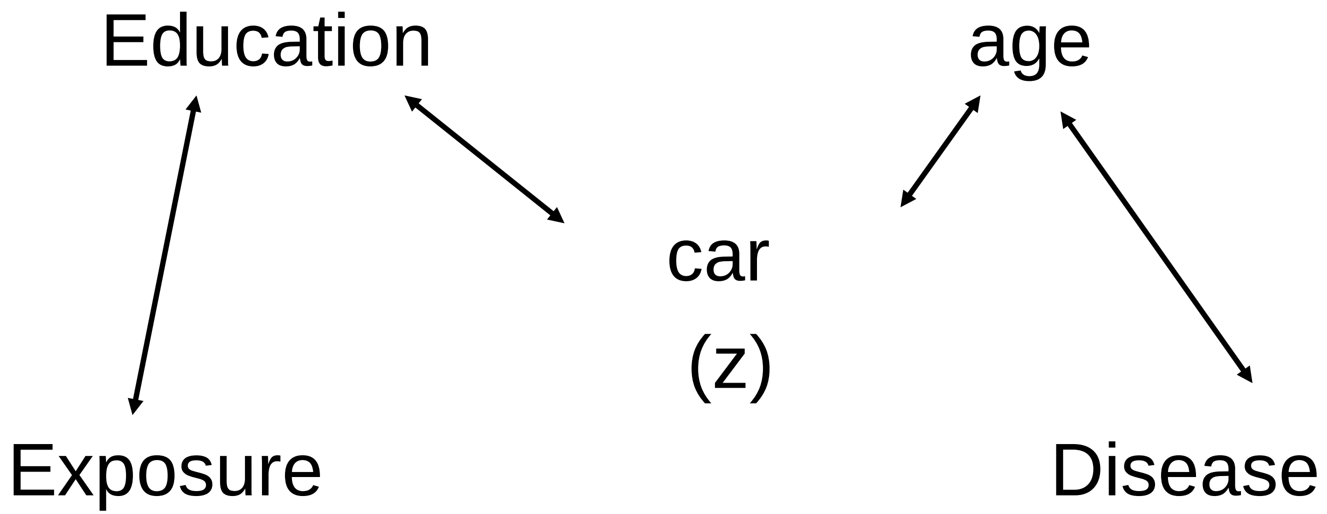


If you are interested in the effect of the exposure on the disease, and Z is the potential confounder:

1) Z is associated with the exposure
2) Z is associated with the disease
3) Z is not an intermediate variable

*Pp:186

Pearl, 2000. “Causality”



If you are interested in the effect of the exposure on the disease, and Z is the potential confounder:

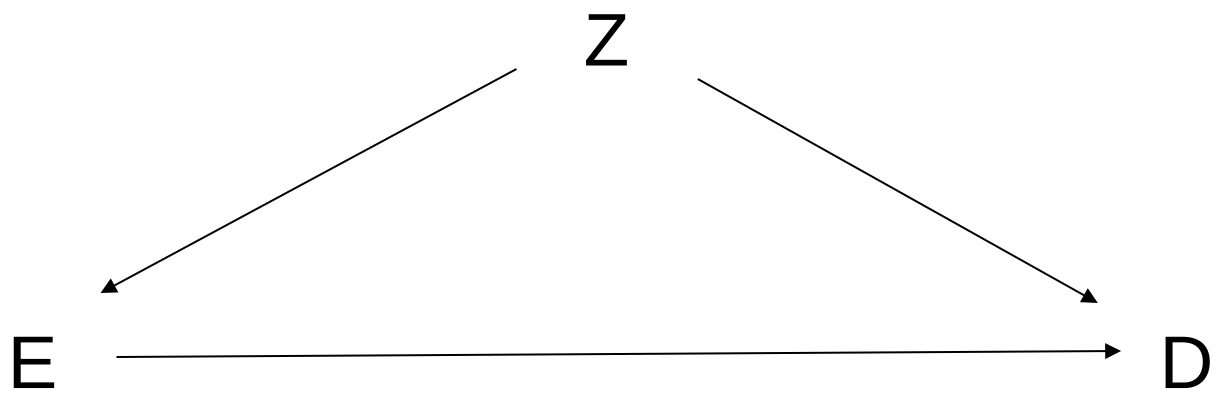
1) Z is associated with the exposure

2) Z is associated with the disease

3) Z is not an intermediate variable

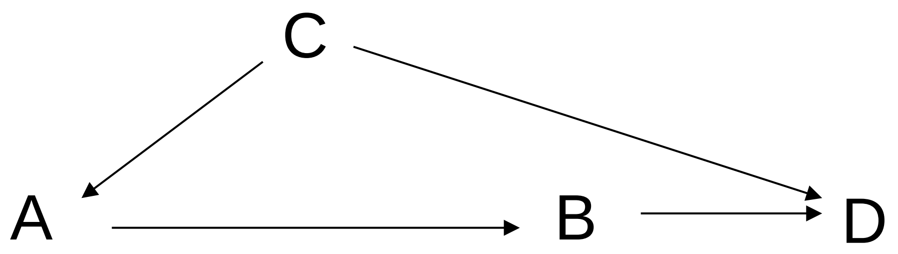
Definition of a confounder using causal diagrams

Directed acyclic graphs (DAGs) permit solving limitations of previous definitions of a confounder



Causal graphs: terminology (1)

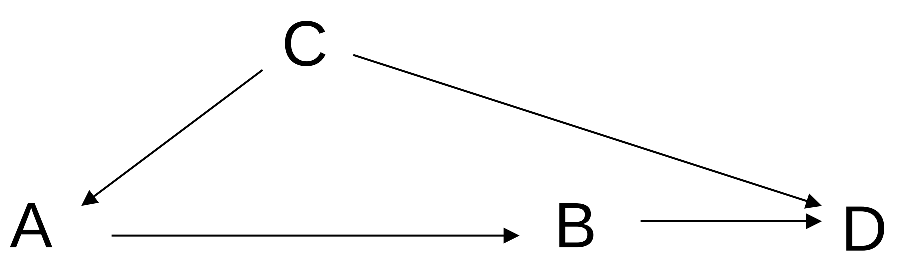
- An arrow between A and B means that A <u>may</u> cause B, but not the other way around
- The lack of an arrow between A and B means that A <u>does not</u> cause B
- Variables (“knots”) connected by an arc are named “adjacent” or “neighbor”
- If A causes B, A is a parent or ancestor of B and B is a child or descendent of A



Causal graphs: terminology (2)

- The lack of a common parent of two variables (say A and C) assumes that the two variables do not have any common causes

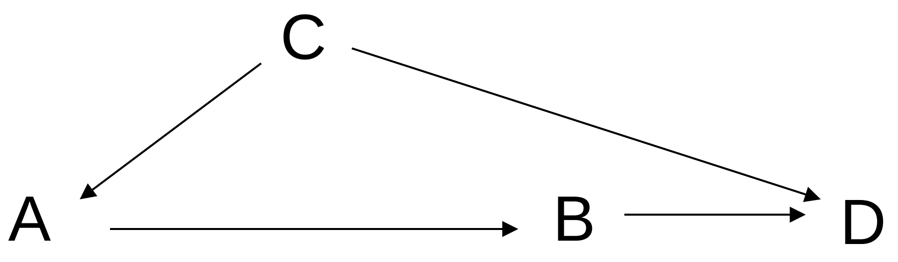
- A **path** between two variables is a sequence of arcs connecting the two variables, regardless the direction of the arrowheads (e.g. A $\leftarrow$ C $\rightarrow$ D)



Causal graphs: terminology (3)

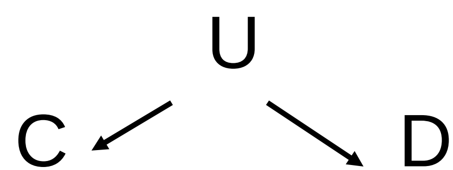
- Backdoor path: a path between X and Y is termed backdoor (relative to X) if it starts with an arrow pointing into X

- If at least 2 arrows point towards the same variable, that variable is called a **“collider”** (e.g. C $\rightarrow$ **D** $\leftarrow$ B) in that path. The collider D is a common effect of B and C

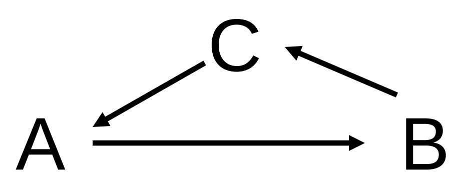


Directed Acyclic Graphs: terminology (4)

- A *directed path* is a sequence of arrows such that the child in the sequence is the parent in the next step

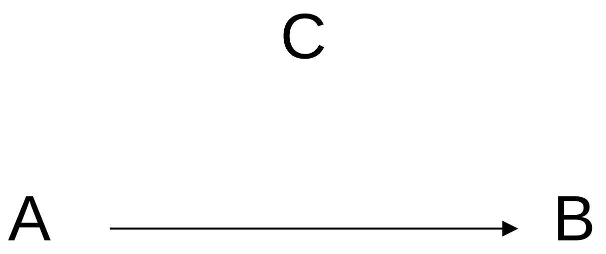


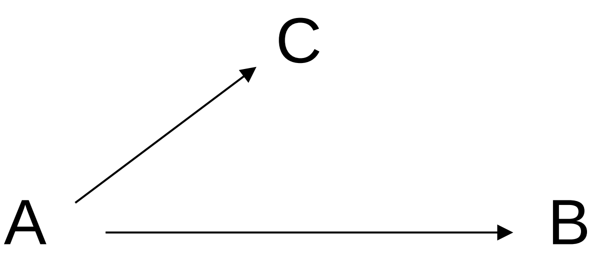
- A graph is *acyclic* if no directed path forms a closed loop



DAGs vs. counterfactuals

- When we draw a DAG:





- implies that:

$Pr(B^{a=1}) \neq Pr(B^{a=0})$

$Pr(B^{c=1}) = Pr(B^{c=0})$

$Pr(B^{a=1}) \neq Pr(B^{a=0})$

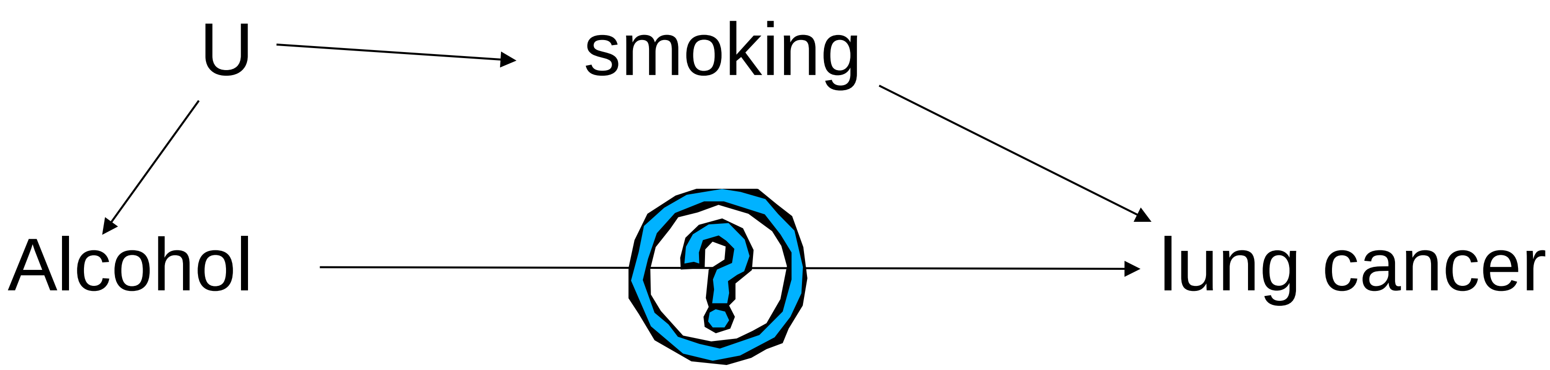
$Pr(B^{c=1}) = Pr(B^{c=0})$

$Pr(B=1|C=1) \neq Pr(B=1|C=0)$

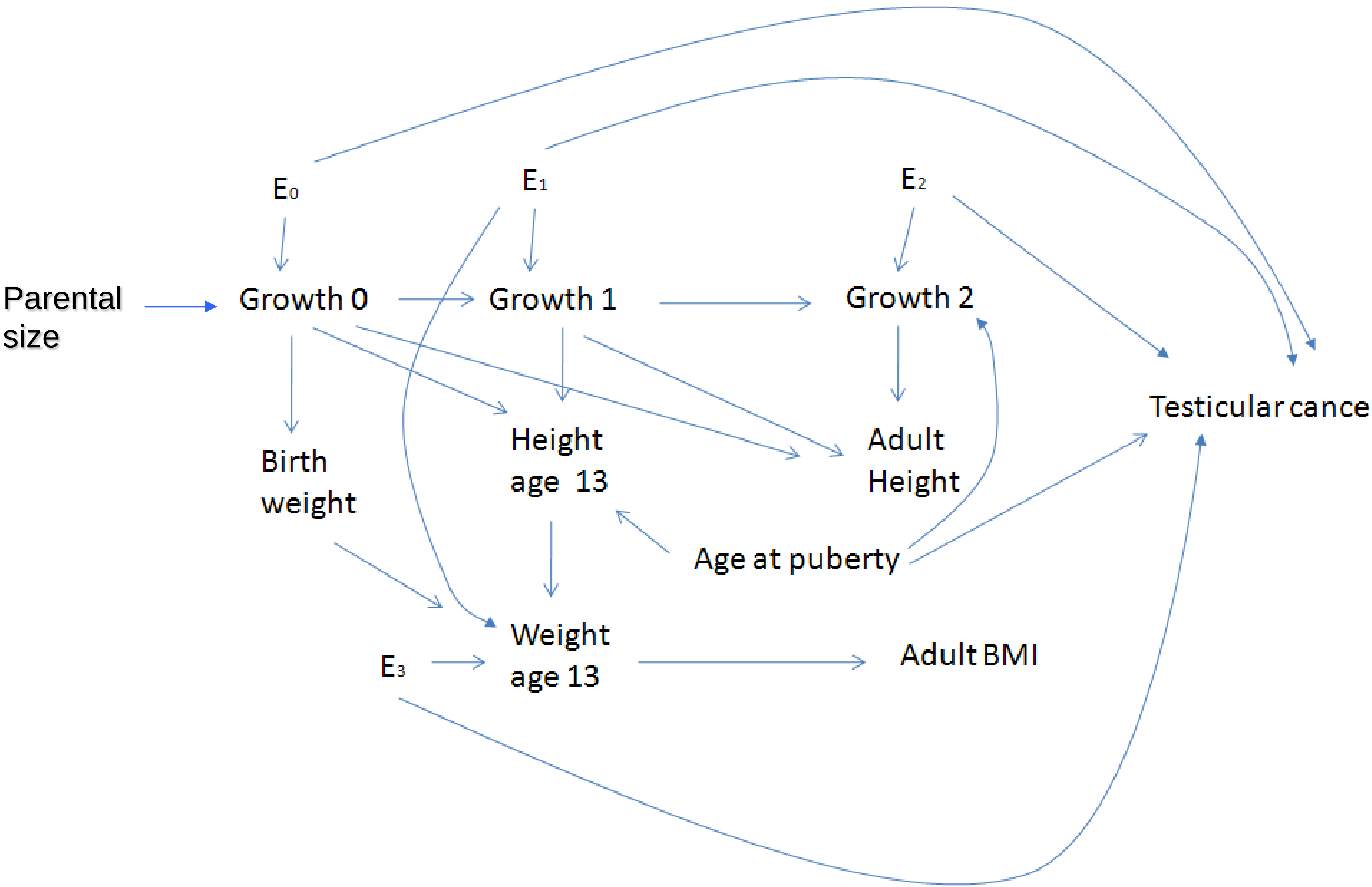
(lack of exchangeability)

Example

DAG for a study of alcohol and lung cancer, in which smoking is a potential confounder



Growth and testicular cancer risk

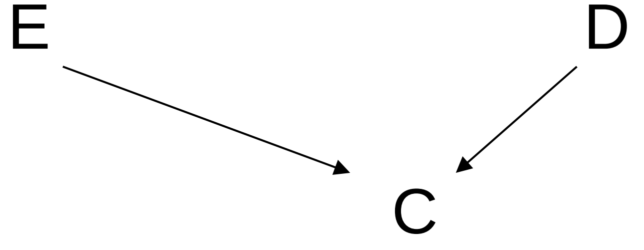


Graphical rules to understand if two variables are independent or associated ("d-separation" method)

- Two variables are d-separated (independent) if all paths between the two variables are blocked (otherwise they are d-connected)
- They can be marginally or conditionally independent
- Note: lack of random error is assumed → association may arise as a consequence of purely random events

D-separation rules

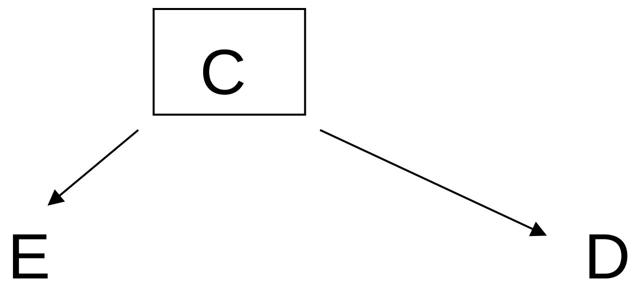
1) If there are no variables being conditioned on, a path is blocked if and only if two arrowheads on the path collide



C is a collider. E and D are marginally independent

D-separation rules

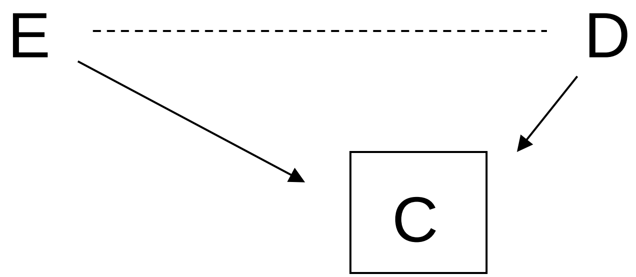
2) Any path that contains a noncollider that has been conditioned on is blocked.



E and D are conditionally independent. (NOTE: If C is not conditioned on, the path is open and E and D are associated)

D-separation rules

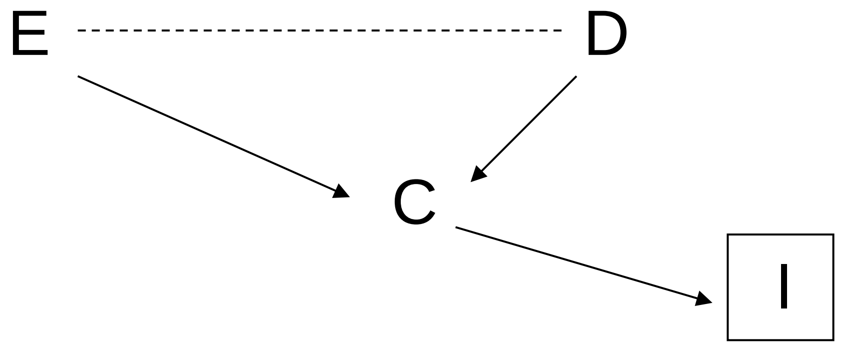
3) A collider that has been conditioned on does not block a path.



E and D are d-connected: they are associated

D-separation rules

4) A collider that has a descendant that has been conditioned on does not block a path

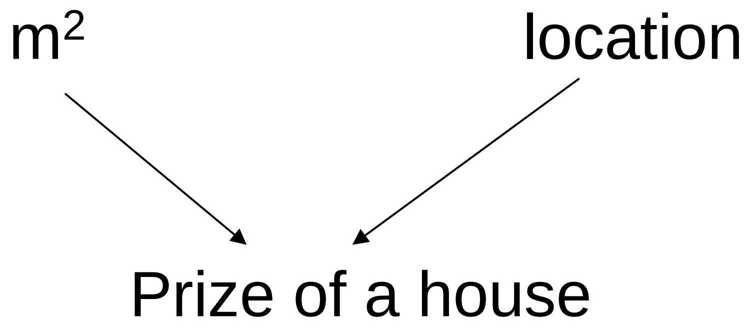


E and D are d-connected.

For example, E represents “smoking”, D represents “chronic bronchitis”, C represents “cough” and I is “intake of cough sedatives”

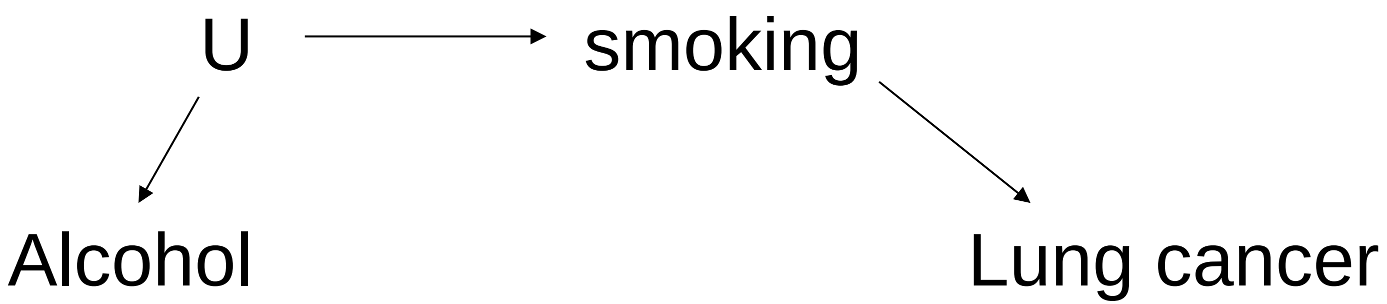
Example of conditioning on a collider

- Let’s suppose that in a city, the size of a house is marginally independent from its location
- Also, both the size and the location positively affect the prize of the house
- For a given prize → size and location will be associated; the association will be inverse (for the same prize you can buy a small house in the city centre or a large house in the suburb)

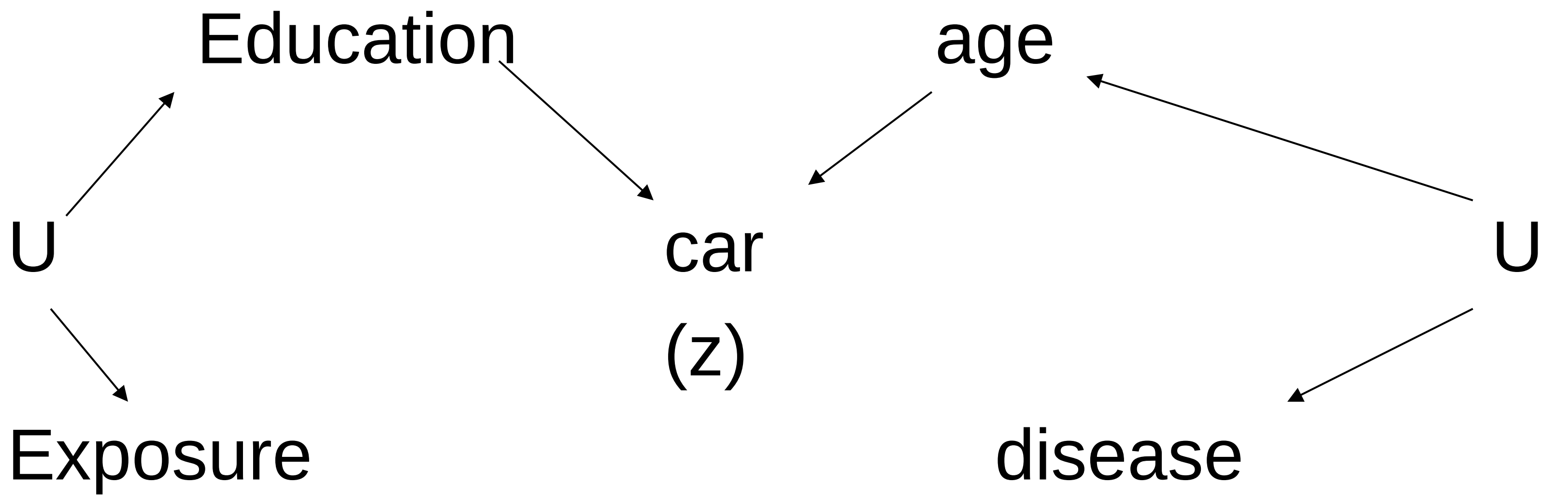


Definition of a confounder (3)

- There is confounding when the exposure and the outcome have common causes
- A confounder is a variable that can be used to block a backdoor path

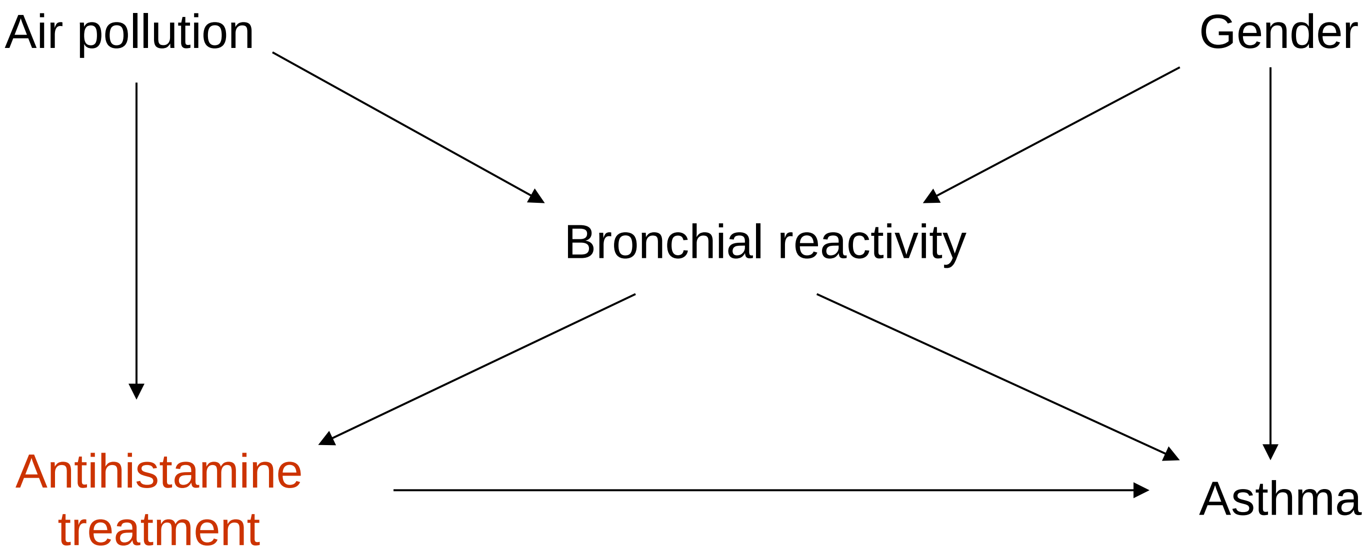


Previous example



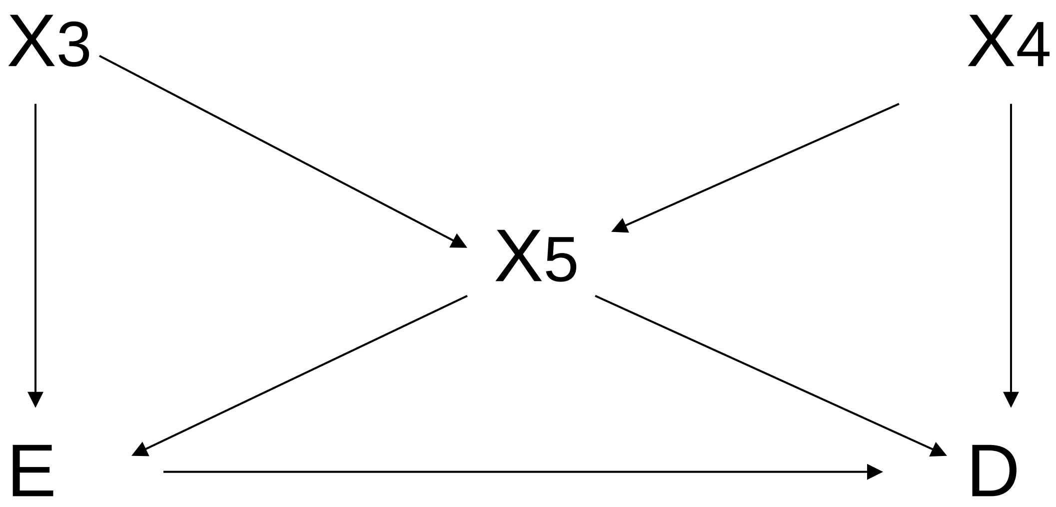
A well known example (Greenland 1999)

Which variable should we control for?



Backdoor path method

How many unblocked backdoor paths?



1. E ← X5 → D

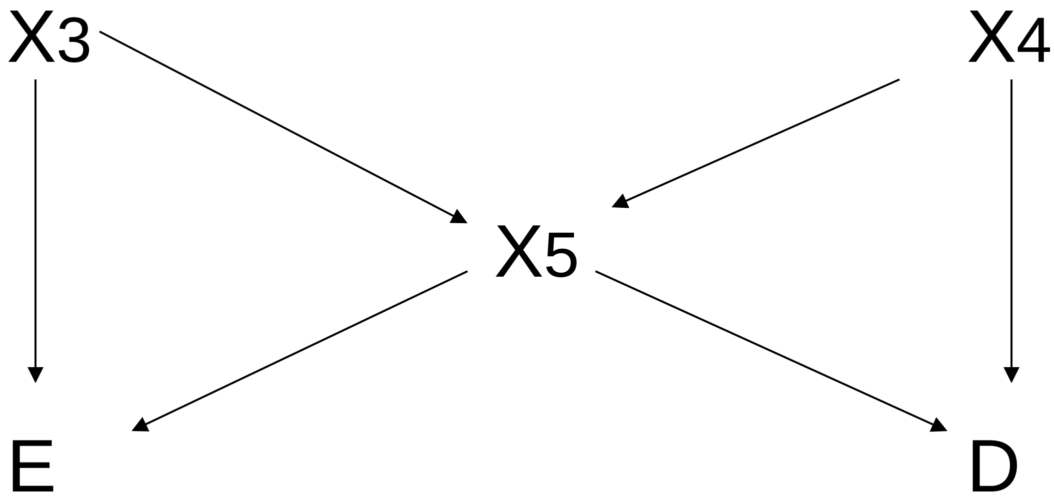
2. E ← X3 → X5 → D

3. E ← X3 → X5 ← X4 → D

4. E ← X5 ← X4 → D

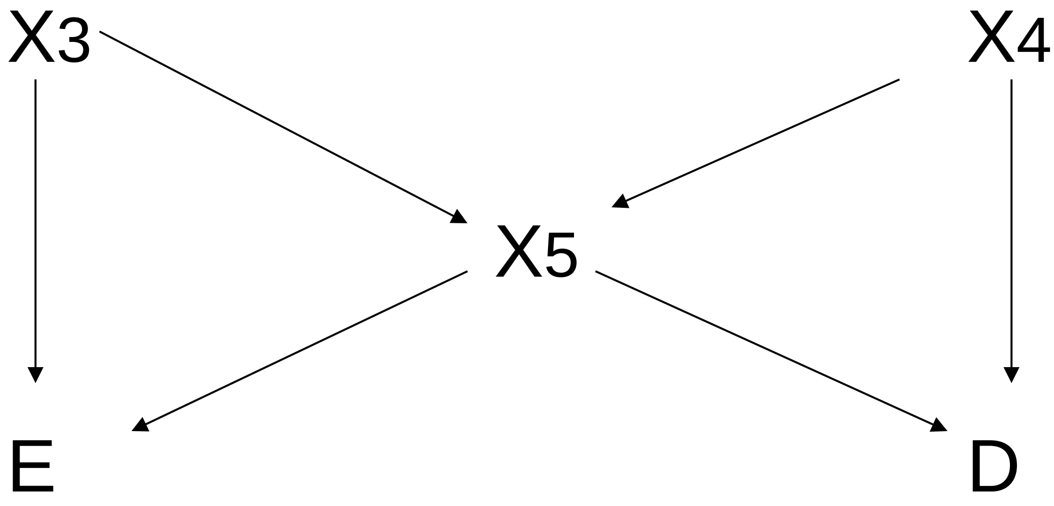
Backdoor path method

1) Cancel the arcs which connect directly the exposure to the disease



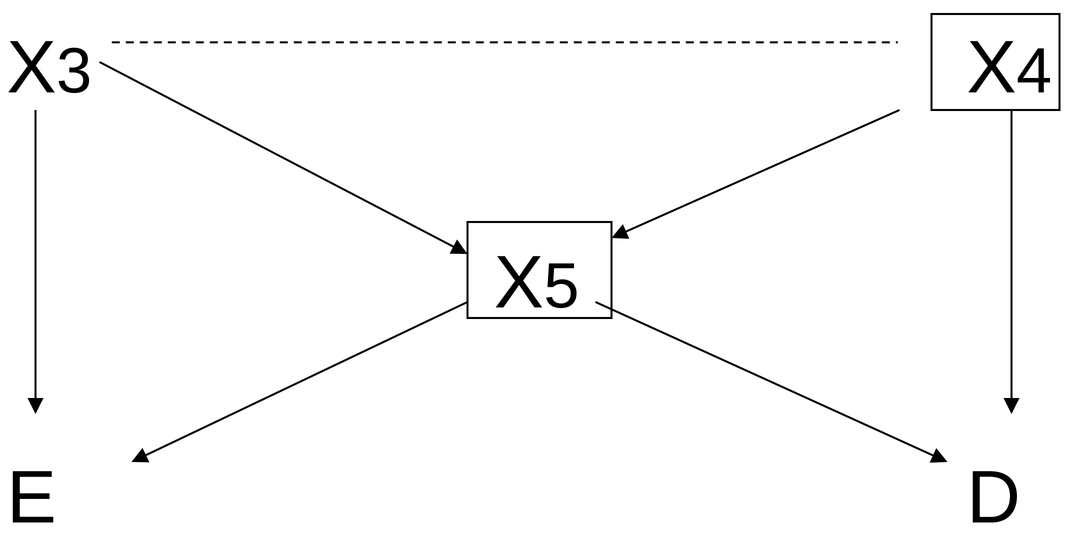
Backdoor path method

2) Check all the remaining paths from the exposure to the disease



Backdoor path method

3) Condition on a non-collider if the path including that variable is opened

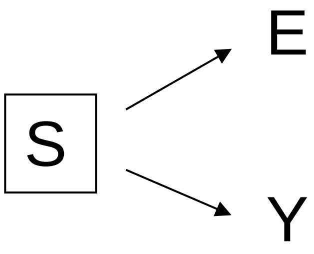


The backdoor criterion

- Delete all exposure effects (arrows from the exposure)
- Observe all remaining paths from E to D
- Select a variable on each open backdoor path as a potential variable to be controlled
- Link any parents of variables that have been selected
- Check again that, if there are new backdoor paths, they are blocked
- If new paths are open, block them with another variable and then check again if new paths are open

Compatibility

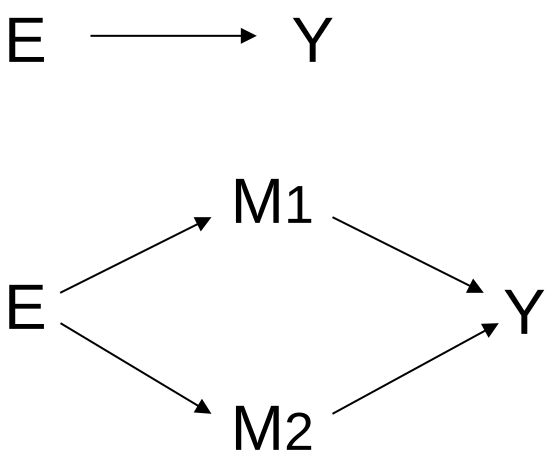
- Two sets of variables are independent conditional on a third set whenever the two sets are d-separated given the third (Mansournia et al . Int J Epidemiol 2013)



This implies that E and Y are independent for any possible value of S, E and Y

Faithfulness

- Two sets of variables are associated conditional on a third set whenever the two sets are d-connected given the third(Mansournia et al . Int J Epidemiol 2013)



If the DAG indicates an association that does not occur in fact there is <u>unfaithfulness:</u>

by chance→ e.g. E-M1-Y and E-M2-Y balance out

by design → e.g. matching in cohort studies

Example

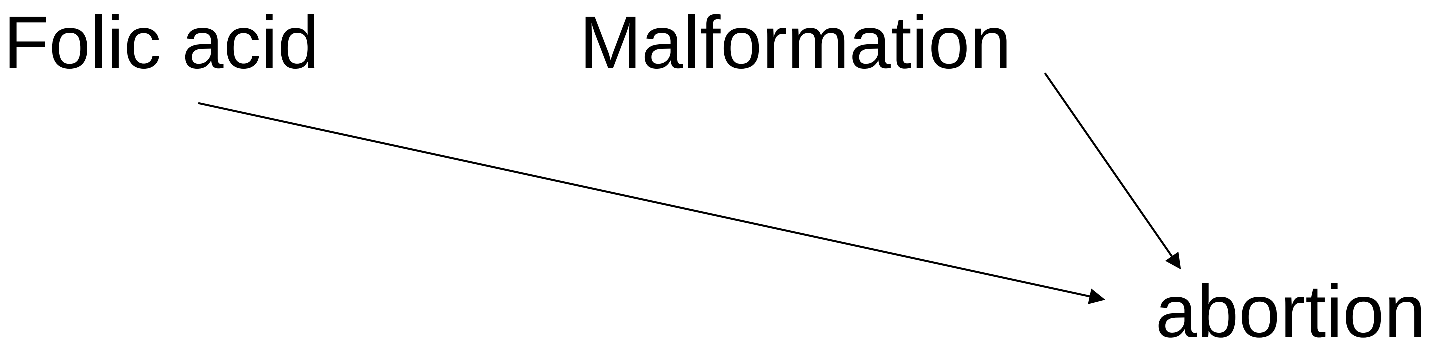
Case-control study on intake of folic acid during pregnancy and risk of malformations

Are stillbirth and therapeutic abortion confounders?

Hernan et al, Am J Epidemiol 2002;155:176-84

Example

Case-control study on intake of folic acid during pregnancy and risk of malformations.



LIMITS

- This approach does not take into account:
  - a. Sampling variation
  - b. Problems of model complexity
- Causal relationships between variables should be specified. Different DAGs can lead to different models
- The magnitude and the direction of the associations are not taken into account → QUALITATIVE APPROACH
- It is difficult to specify effect modification

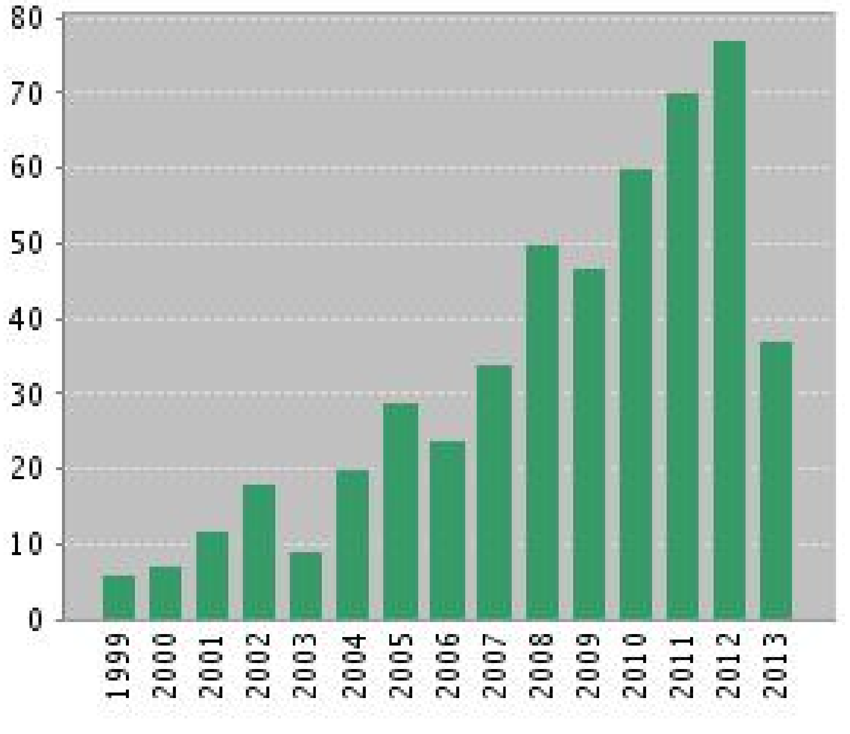
ADVANTAGES

- Causal relationships between variables should be specified
- It is possible to solve very complex problems
- Decisions are taken a-priori → very useful when the study is designed

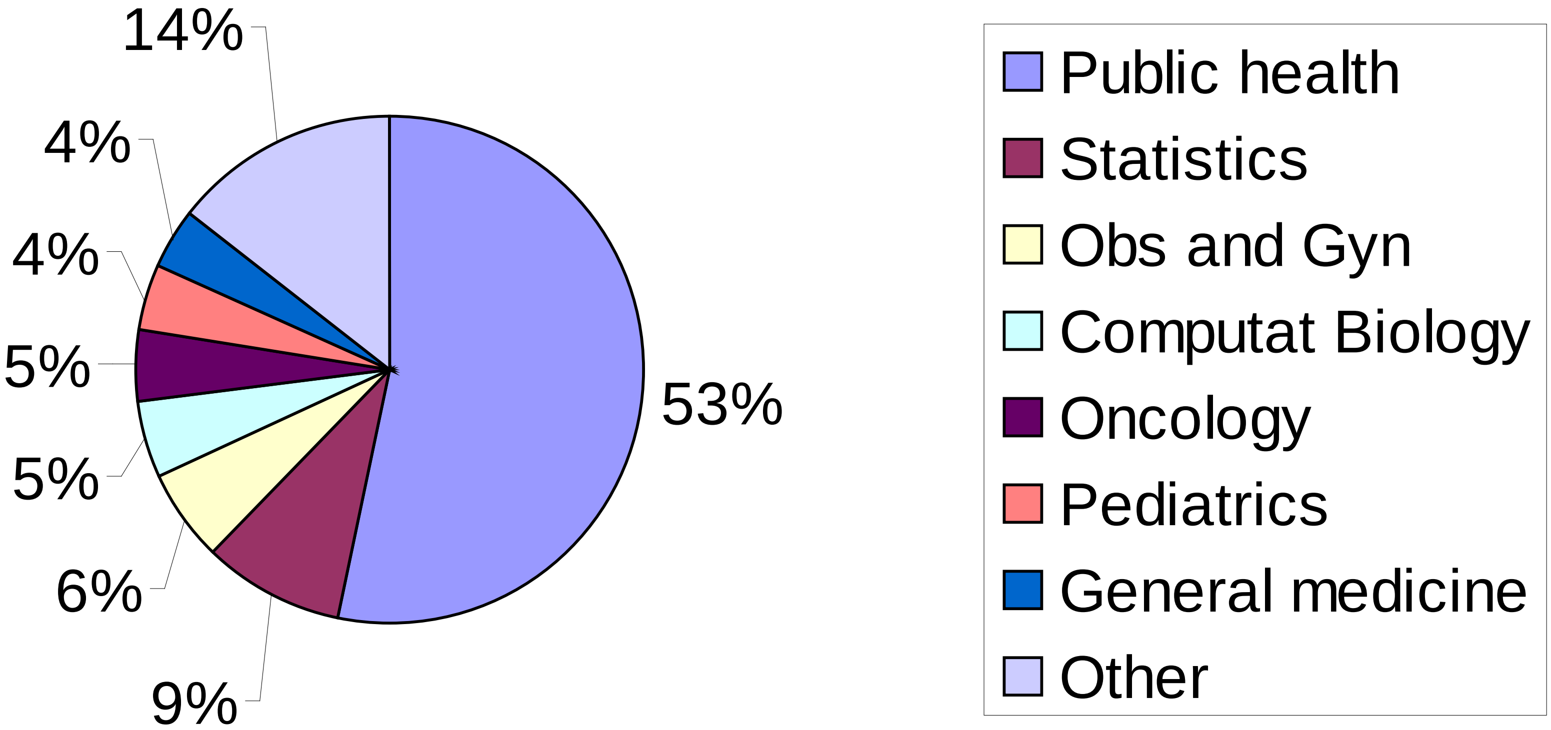
Casual diagrams are becoming increasingly common in medical science

Number of citations per year of the paper

Greenland S et al. Causal diagrams for epidemiologic research. Epidemiology 1999



Citations of the paper Greenland S et al Epidemiology 1999



DAGs and BIAS

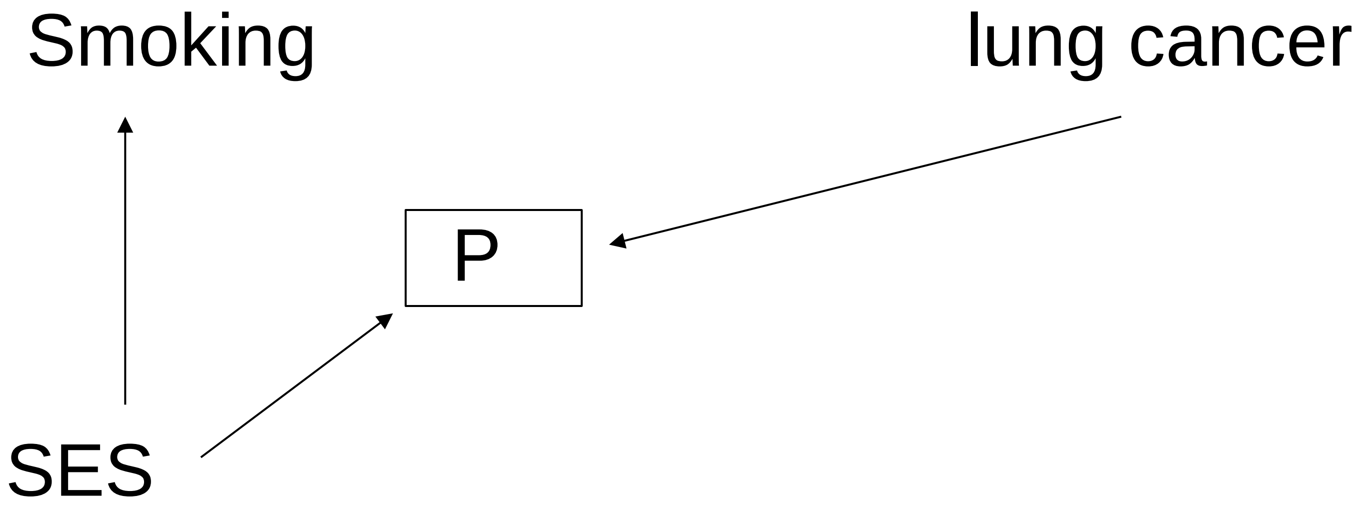
Hernan et al, Epidemiology 2004;15:615-25

An association between an exposure and an outcome can be produced by 3 casual structures:

1) Cause and effect: E causes D or D causes E (If D causes E → INFORMATION BIAS)
2) Common causes: E and D share a common cause (CONFOUNDING)
1) Common effects: E and D have a shared effect (SELECTION BIAS)

Non-response bias

Case-control bias:



Berkson's bias

Berkson's bias

"the relative frequency of disease in a group of patients who are hospitalised is biased when compared to the population served by the hospital" (1946)

Empirical demonstration: Survey on 2784 patient (257 hospitalized)

Circ. diseases	Resp. diseases		
	yes	no	Tot
Yes	7	29	36
No	13	208	221
tot	20	237	257

OR: 3.9

Circ. diseases	Resp. diseases		
	yes	no	Tot
Yes	22	171	193
No	202	2389	2591
tot	224	2560	2784

OR: 1.5

Berkson's bias numerical examples

Table 1. Association in the general population as reported in Berkson[2]

	Exposed	Unexposed	Total
Cases	3000	97 000	100 000
Non-cases	297 000	9 603 000	9 900 000
Total	300 000	9 700 000	10 000 000

Odds ratio = 1.0.

Table 2. Association using hospitalized cases and general population controls

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	29 700	960 300	990 000
Total	30 290	965 611	995 901

Odds ratio = 3.59.

Table 3. Association using hospitalized cases and controls from patients hospitalized for any disease, with a 0.2 population prevalence and a 0.025 probability of hospitalization for any disease other than D1 (exposure) or D2 (cases)

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	45 812	48 015	93 827
Total	46 402	53 326	99 728

Odds ratio = 0.12.

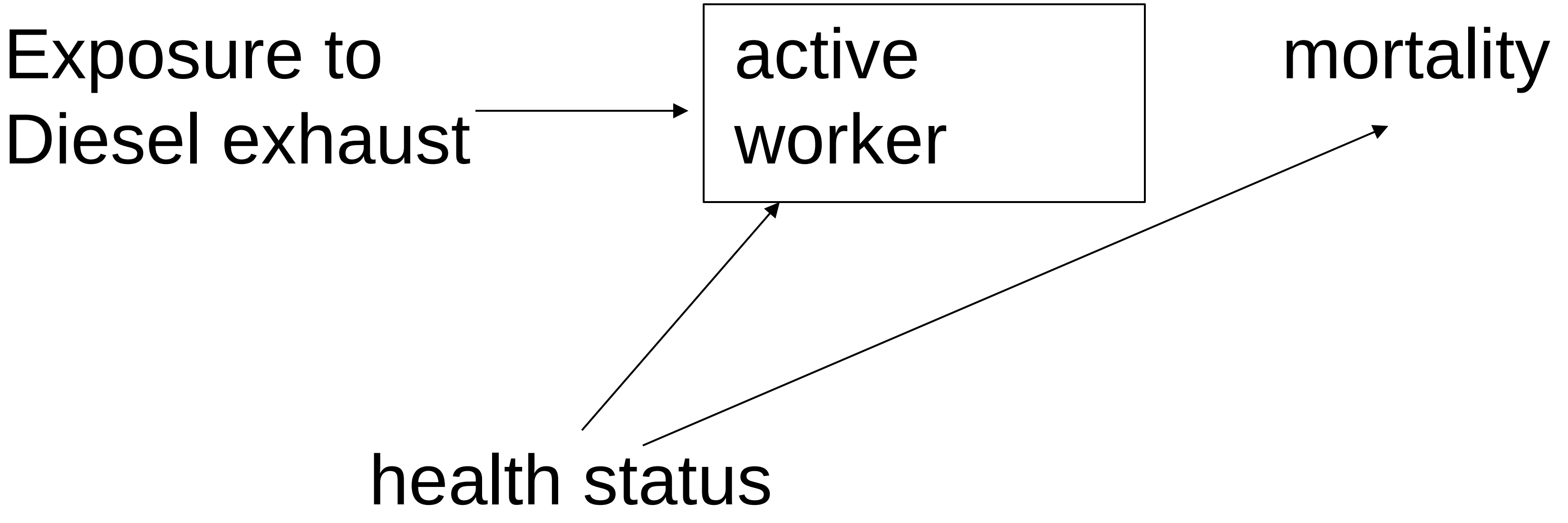
Table 4. Association using controls hospitalized with a particular disease, with a 0.005 population prevalence and a 0.20 probability of hospitalization for the control disease

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	480	9757	10 237
Total	1070	15 068	16 138

Odds ratio = 2.26.

Pearce N et al. Int J Epidemiol 2014; 43:521-4

Healthy worker effect



Recall bias

Case control study on malformations and drug use during pregnancy

Drugs Malformation

recall

Reverse causality

Case-control study on BMI and colon cancer risk

BMI

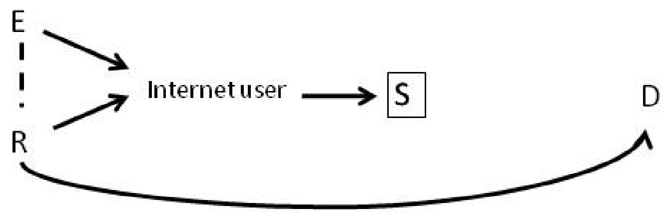
colon cancer

altered

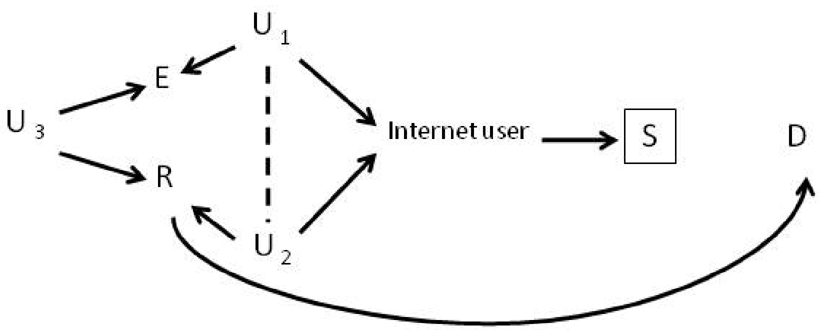
metabolism

Baseline selection

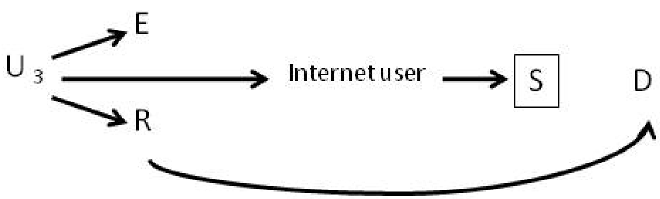
E= exposure, R=risk factor
S=selection, D= disease



Internet recruitment induces bias



Internet recruitment may decrease or introduce bias



Internet recruitment decreases bias