



Vad är kausal inferens?

Anna Ekman
Arbets- och miljömedicin

Sahlgrenska akademin

Seminariets agenda

- Kausalitet
- Statistiska samband kontra kausalitet
- Konfounding
- DAG – ett grafiskt stöd
- Inverse probability weights

Sahlgrenska akademin



Kausalitet

Relationen mellan en händelse, **orsak**, och en annan händelse, **verkan/effekt**, där den andra händelsen förstås som en konsekvens av den första.

Sahlgrenska akademin



Kausalitet

Counterfactual or potential outcomes

Kausal effekt för en individ i :

$$Y_i^{a=1} \neq Y_i^{a=0}$$

Kausal medeleffekt:

$$E[Y^{a=1}] \neq E[Y^{a=0}]$$

$$P(Y^{a=1}=1) \neq P(Y^{a=0}=1)$$

Sahlgrenska akademin

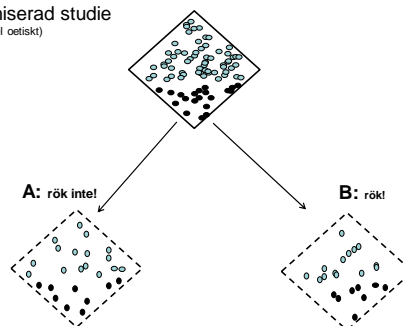


- Kausalitet
- Statistiska samband kontra kausalitet
- Konfounding
- DAG – ett grafiskt stöd
- Inverse probability weights

Sahlgrenska akademin



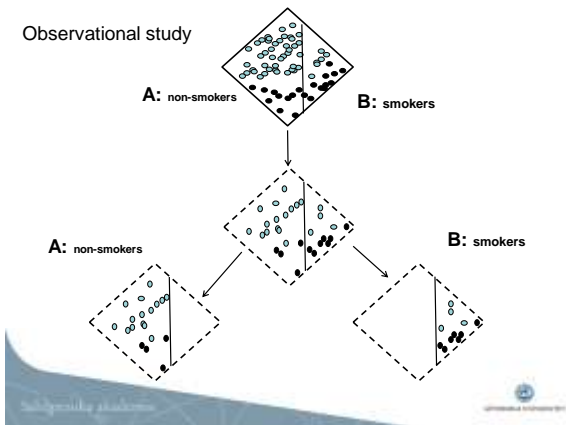
Randomiserad studie
(detta exempel oetiskt)



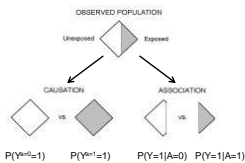
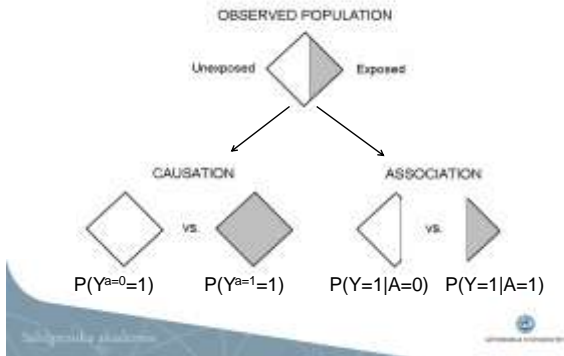
Sahlgrenska akademin



Observational study



Om man nu inte kan göra ett randomiserat experiment, utan får nöja sig med en observationsstudie



Om man i en observationsstudie lyckas få de oexponerade och de exponerade att vara lika i allt **viktigt** utom exponeringen, då har man **utbytbarhet**.

Utbytbarhet ger att $P(Y^a=1=1) = P(Y=1|A=1)$ och $P(Y^a=0=1) = P(Y=1|A=0)$

Frånvaro av utbytbarhet är **konfounding**

Samband kontra **kausalitet**

Säg att vi ser statistisk samband mellan A och B i vår population. Detta kan ske på ett av tre sätt (och ett måste vara sant)

$A \rightarrow B$ $A \leftarrow B$ $A \leftarrow C \rightarrow B$

Men av dessa samband mellan A och B är bara två kausala

$A \rightarrow B$ $A \leftarrow B$ ~~$A \leftarrow C \rightarrow B$~~

Det sista exemplet inte kausalt samband, utan typisk **konfounding**.

Samband kontra **kausalitet**

Statistisk samband mellan A och B innebär **INTE** ett kausalt förhållande mellan A och B, men om

A har samband med B **OCH A händer före B i tid**

Skulle följande kausala samband vara det enda möjliga

$A \rightarrow B$ ~~$A \leftarrow B$~~

Åtgärda konfounding "→" kausalitet



- Smärta en konfunder för sambandet mellan fysisk aktivitet och hälsa
- Stratifiera, dela in i två grupper, smärt =0 och smärta=1



- I ett smärt-strata finns ingen konfounding. Vi har då **betingsad utbytbarhet**.

DAG

Directed Acyclic Graphs

Komplexa modeller kan undersökas för att se vad man bör justera för i sin modell och inte. Alla "regler" för hur dessa grafiska modeller skall hanteras har "matematisk bakgrund".

(Pearl)

DAG - Directed Acyclic Graphs

Kausala diagram

Grafisk metod att se på konfounders och på de ev. kausala samband man vill analysera.

Inga cykler

Pilar skall ha en kausal tolkning, **inte** enbart statistiska samband.

Kräver stark teoretisk förankring.

DAG – symboler/byggstenar

- A orsakar B



- C, en konfounder

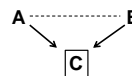


- Justera för C

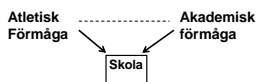


Två variabler som **inte** har samband i populationen

Kan skapa samband i analys genom överjustering!



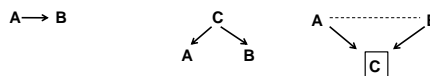
- Antag att C är en effekt av både A och B
- A och B kommer då att ha samband inom strata av C
- Den streckade linjen symboliserar det skapade/framkallade sambandet



- Antag att flera skolor har **selektiv antagning**. Elever antas om de antingen är duktiga **atleter** eller duktiga **akademiskt** eller både och.
- Antag också att atletisk och akademisk förmåga är **oberoende** av varandra i populationen.
- Inom en skola** skapas en negativ association.

Varför? Ta en elev i skolan. Om eleven är dålig atlet, så vet vi att elev är duktig akademiskt, annars hade eleven inte antagits.

Sammanfattning så här långt



- A och B kommer ha samband i **populationen** om:
 - A orsakar B
 - B orsakar A
 - Det finns ett C som är **orsak** till både A och B
- A och B kommer ha samband i **sub-population**, definierad av C, om C är en **effekt** av både A och B.
- Detta är byggstenarna i kausala diagram.

DAG

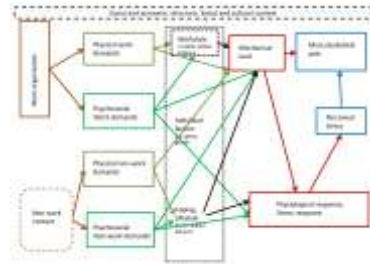
Pearls regler

Följer man Pearl's regler får vi en "slut-graf" som visar om vi har konfounding eller inte (givet att utgångsmodellen var korrekt).

Låt oss se ett exempel ...

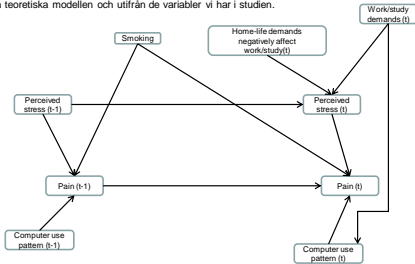
Forskningsfråga: Är upplevd stress en *short-term* risk faktor för att ha smärta nu? Är även datoranvändningsmönster en *short-term* risk faktor?

En teoretisk modell som vi är vana.



Forskningsfråga: Är upplevd stress en *short-term* risk faktor för att ha smärta nu? Är även datoranvändningsmönster en *short-term* risk faktor?

En DAG utifrån teoretiska modellen och utifrån de variabler vi har i studien.



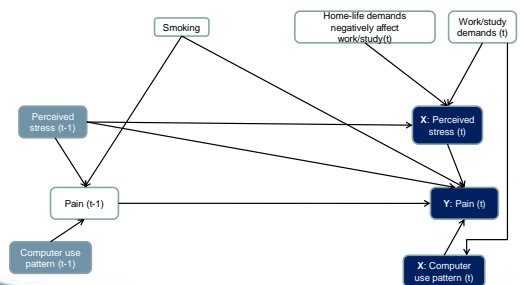
- A. Vilka variabler måste vi justera med för att inte ha konfounding? Gratis programvara.
- B. Fria från konfounding om vi justerar för *Perceived stress (t-1)* och *Computer use pattern (t-1)*?

Vi skall utgå från fråga B ...

Forskningsfråga: Är upplevd stress en *short-term* risk faktor för att ha smärta nu? Är även datoranvändningsmönster en *short-term* risk faktor?

Regression:

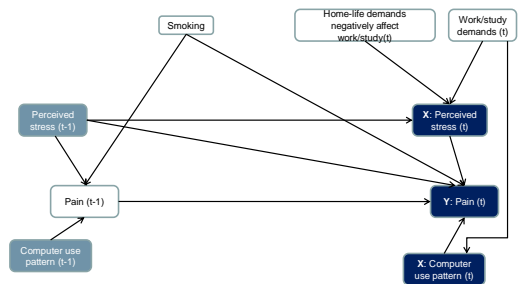
$$Pain(t) = \text{Perceived stress}(t) + \text{Computer use pattern}(t) + \text{Perceived stress}(t-1) + \text{Computer use pattern}(t-1)$$



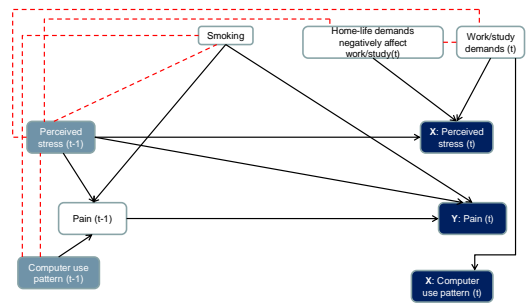
Ta bort alla pilar som går

från exponeringen (*perceived stress (t)*, *Computer use pattern (t)*)

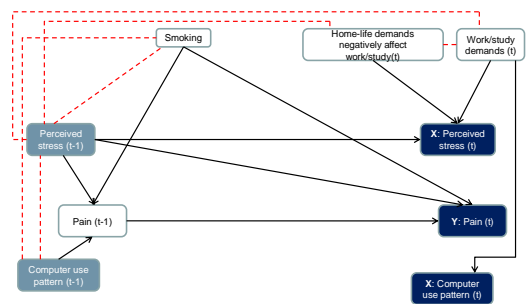
till utfallet (*pain (t)*).



Bind samman alla "föräldrar",
som har ett gemensamt "barn".



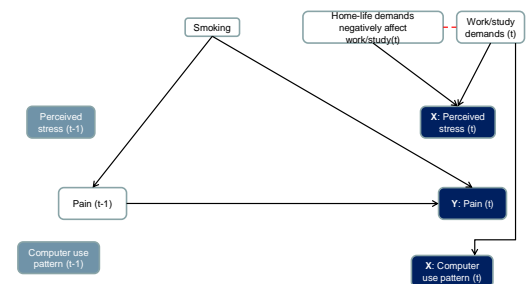
Ta bort alla linjer/pilar från eller till
de eventuella konfunder-variablerna
(*perceived stress (t-1)*, *Computer use pattern (t-1)*).



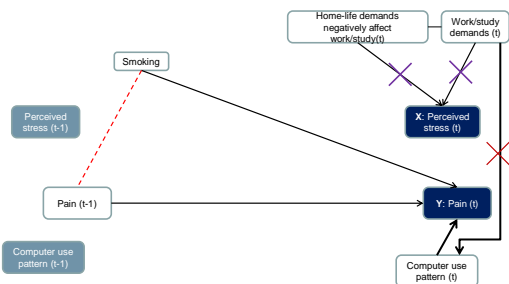
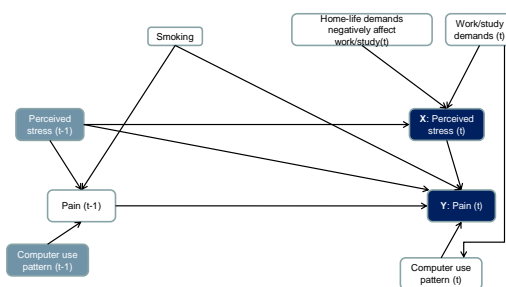
Om nu
exponering och utfall är separerade
har vi justerat bra. Vår analys är inte
konfounded,

Givet att den kausala modellen vi började
med är korrekt .

(Justerat för *perceived stress (t-1)*, *Computer use pattern (t-1)*.)



Låt oss se en variant av vårt exempel om inte tar med *Computer use pattern (t)* som exponering i vår planerade modell.



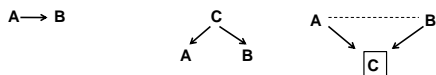
X och Y är nu inte separerade och vi har därför konfounding.
 Justera för *Computer use pattern (t)* så stängs denna väg, d.v.s. ingen konfounding kvar.
 Eller
 Justera för *Home-life demands negatively affect work/study(t)* och *Work/study demands (t)*

I komplicerade modeller är proceduren lite stökig eller om man har fråga A (vilka konfounders bör jag justera för) så finns gratis programvara att ladda ner till hjälp.

www.dagitty.net

OBS! Vi kan skapa konfounding genom att över-justera! (Skolexemplet)

Sammanfattning



- Kausalitet – formaliserat
- Konfounding – att skapar samband, justera därför
- Kan överjustera – själv skapa samband
- DAG: Kausala diagram
 - Antaganden – medvetandegörs och kan lättare diskuteras
 - Se vad som bör justeras för och vad man INTE skall justera för

Inga kausala antaganden in → inga kausala slutsatser ut (CARTWRIGHT, 1989)

Kausala antaganden kan INTE uttryckas i den traditionella statistikkens matematiska språk.

Hur kan vi "justera" i en analys?

- Inkludera 'konfounders' i regressionsmodell som x-variabler.
- Vikta:

Kausal-effekt

- > Standardisera till mål-populationens fördelning av konfounders (average of individual causal effect in the target population)
- > Standardisera till de exponerades fördelning av konfoundervariabler
- > Standardisera till de oexponerades fördelning av konfoundervariabler

Vanlig justering för konfounders?

$$\ln \frac{P(Y=1)}{1-P(Y=1)} = a + b_{stress} * x_{stress} + b_{nobreak} * x_{nobreak}$$

Effekten av stress i gruppen med bra pausmönster (nobreak=0, 1)

Effekten av stress i gruppen med dåligt pausmönster (nobreak>1)

"Medel" av dessa två effekter.

Effekten mycket stor i en ovanlig grupp – justerade effekten blir stor, men inträffar sällan i den aktuella populationen.

Inference based on counterfactuals

IPTW – Inverse probability treatment weights

En exponering i fokus.

Viktat observationerna, tex

- Exponerade: $1/(\text{sannolikheten för exp givet confounders})$
- Oexponerade: $1/(\text{sannolikheten för oexp givet confounders})$

Detta ger exponeringens effekt i en population med samma fördelning av justerings-variablerna som i datamängden (målpop).

IPTW – en konfunder

Nackont: Y = 1 om ja Y = 0 om nej

Stress: $x_{stress} = 1$ om ja $x_{stress} = 0$ om nej

Konfounder: **Pausmönster**

Endast denna konfunder inte troligt, men för att följa exemplet från tidigare.

IPTW – en konfunder

Modellerar först exponeringen, stress, som Y och konfoundern, pausmönster, som x i en logistisk regression.

Kan sedan, för varje observation individ räkna fram

- Sannolikheten att vara exponerad, givet värdet för ens egna konfoundervärden.
- Sannolikheten att vara oexponerad, givet värdet för ens egna konfoundervärden.

IPTW – en konfunder

Räknar sedan ut en vikt som varje individs utfall multipliceras med.

Dessa vikter kan se olika ut.

Standardisering till populationens fördelning av konfounders:

$$\frac{1}{P(\text{stress}=1|\text{nobreak}_{\text{indiv}})} \text{ för de exponerade}$$

$$\frac{1}{P(\text{stress}=0|\text{nobreak}_{\text{indiv}})} \text{ för de oexponerade}$$

Modellerar sedan utfallet, ont i nacken, med enbart exponeringen, stress, som x i en logistisk regression.

IPTW – en konfounder

Konfounderfördelning								
Målpopulation			Exponerad population			Oexponerad population		
Stress	Prevalens %	Pga stress	Stress	Prevalens %	Pga stress	Stress	Prevalens %	Pga stress
0	14		0	14		0	14	
1	28	14	1	29	14	1	28	14

	Mål	Exp	Oexp
OR	2,4	2,4	2,4
PR	2,0	2,0	2,0
PD	14	14	14

Vanlig konfounder justering
OR=2,5

IPTW – en exponering, många konfounders

Nackont: Y = 1 om ja Y = 0 om nej

Stress: X_{stress} = 1 om ja X_{stress} = 0 om nej

Konfounders: Förutom pausmönster vid datoranvändning BMI, frukostvanor, motion, arbetstimmar, krav i studier för stora, kraven i studierna påverkar hem/familjelivet, kraven hemifrån påverkar studierna, astma, diabetes, hjärt.

IPTW – en exponering, många konfounders

```
proc genmod data=work.woment2 descending;
class nobreak/order=internal descending param=ref;
model stress = nobreak overw2 smoke2 onlysnu2 fruko2di brachef2
brakam2 kravpav2 pavarb2 astma2 diabet2 hjarta2/ link=logit dist=binomial
type3;
output out=vikt_women pred=vikt ;
run;
*** Calculate vikterna, Standardized to the distribution in the total cohort;
data work.vikt_women;
set work.vikt_women;
if stress=1 then iptw=1/vikt;
if stress=0 then iptw=1/(1-vikt);
run;
```

IPTW – en exponering, många konfounders

Vikterna som vi räknade ut på förra sidan används nu i en princip vanlig analys.

```
proc genmod data=work.vikt_women descending;
class stress/order=internal descending param=ref;
WEIGHT iptw ;
model nuneck = stress/ link=logit dist=binomial type3;
run;
```

IPTW – en exponering, många konfounders

Konfounderfördelning								
Målpopulation			Exponerad population			Oexponerad population		
Stress	Prevalens %	Pga stress	Stress	Prevalens %	Pga stress	Stress	Prevalens %	Pga stress
0	18		0	21		0	14	
1	28	10	1	29	7	1	28	14

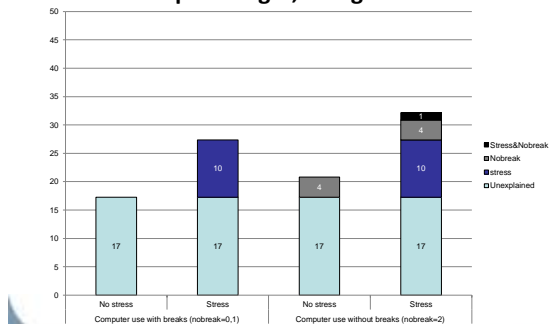
	Mål	Exp	Oexp
OR	1,8	1,5	2,4
PR	1,6	1,3	2,0
PD	10	7	14

Vanlig konfounder justering
OR=2,2

IPTW – två exponeringar, många konfounders

Exponering		OR	PR	PD
Stress	Pausmönster=0,1	1,8	1,6	10
	Pausmönster=2	1,8	1,5	11
Pausmönster	Stress=0	1,3	1,2	4
	Stress=1	1,3	1,2	5

IPTW – två exponeringar, många konfounders



Inference baserad på 'counterfactuals'

Positivt

- Rena, tydliga resultat
- Lätt att räkna ut prevalens, prevalens-kvot och prevalens-skillnad
- SAS : Proc Genmod med weight-commando

Negativt

- Många justeringsvariabler med mycket partiellt bortfall kan ge mycket missing.
- Använder skattade vikter som om de var de sanna värdena (bias).

Inference based on counterfactuals

Time-varying exposures

Adjustment for unobserved confounding and other biases

- Kausalitet
- Statistiska samband kontra kausalitet
- Konfounding
- DAG – ett grafiskt stöd
- Inverse probability weights