



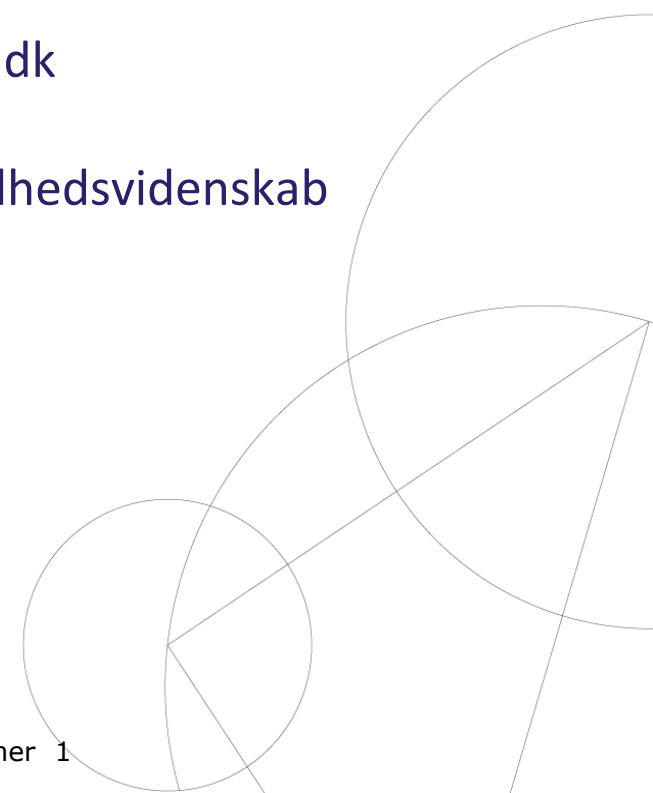
Faculty of Health Sciences



# Selektionsbias

Mads Kamper-Jørgensen, lektor, [maka@sund.ku.dk](mailto:maka@sund.ku.dk)

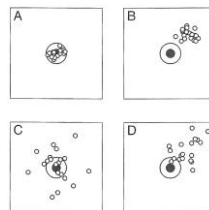
Afdeling for Social Medicin, Institut for Folkesundhedsvidenskab



# Sidste gang

## Vi snakkede om

- Præcision: tilfældige fejl
- Bias: systematiske fejl
- Sensitivitet: klassificering af syge som syge
- Specificitet: klassificering af raske som raske
- Kilder til misklassifikation
- Non-differentiel misklassifikation: misklassifikation er lige sandsynlig i sammenligningsgrupperne, fortyndning af effekt
- Differentiel misklassifikation: misklassifikation er ikke lige sandsynlig i sammenligningsgrupperne, informationsbias
- Typiske informationsbias: recall-bias, detektionsbias, interviewerbias

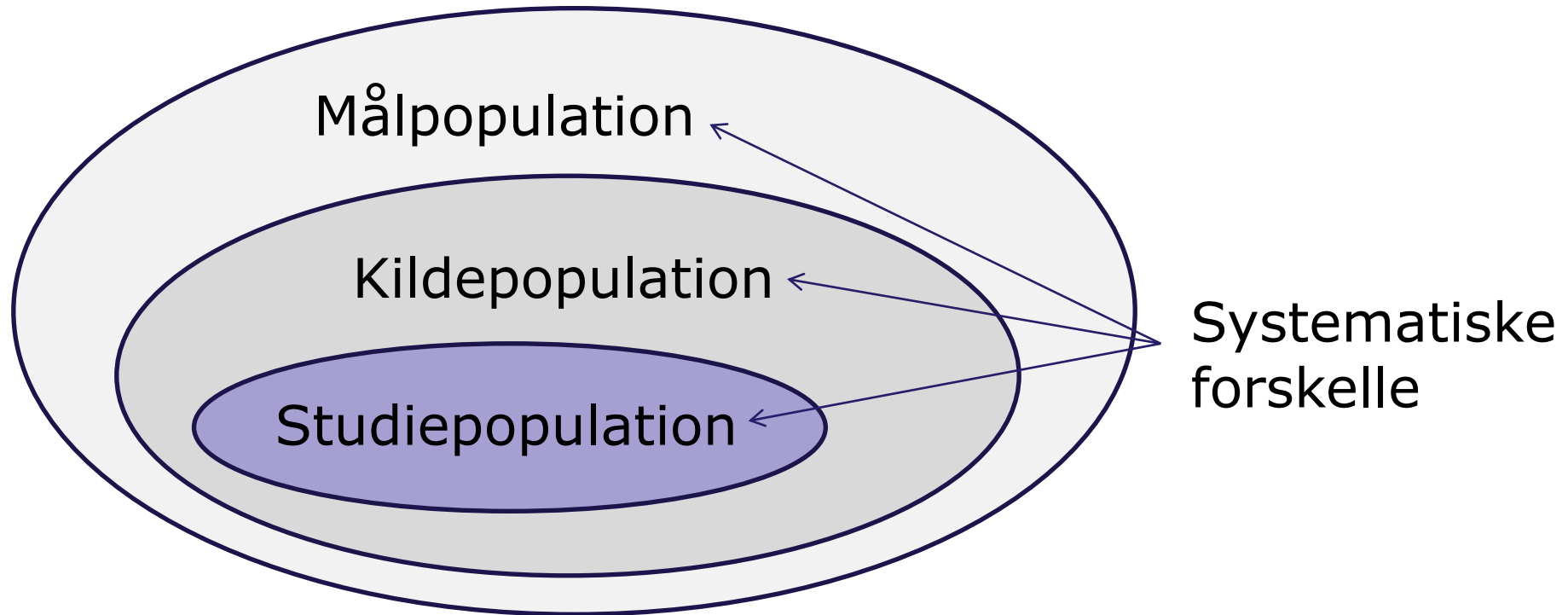


# Hvorfor selektionsbias?

- Fordi vi kan komme til at over- eller undervurdere hyppigheder og associationer, og drage forkerte konklusioner hvis deltagerne i studiepopulationen ikke repræsenterer målpopulationen
- Indtil videre har vi forudsat at dem der indgår i vores undersøgelser, og dem der ikke indgår i vores undersøgelser er sammenlignelige: Det er ikke altid tilfældet
- Selektionsbias er fejl der skyldes *systematisk* forskel på dem som indgår i vores undersøgelse og dem der ikke indgår i vores undersøgelse
- Selektion ind til kohorten, og loss to follow-up (attrition)



# Selektionsbias



# Et eksempel

## Målpopulation

- Gravide kvinder i Danmark

## Kildepopulation

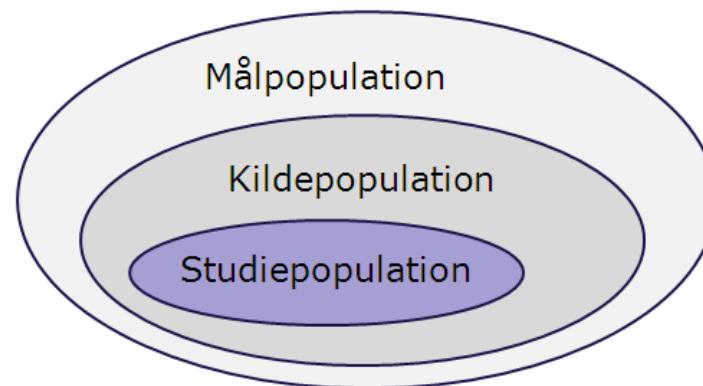
- Gravide kvinder hos udvalgte praktiserende læger

## Studiepopulation

- Deltagere i Bedre Sundhed for Mor og Barn (BSMB): deltagelse var afhængig af om kvinden ville være med i studiet

## Selektionsbias?

- Er studiepopulationen systematisk forskellig fra kildepopulationen, og er kildepopulationen systematisk forskellig fra målpopulationen?



# Det kommer an på

## BSMB kvinder er anderledes

- De drikker mindre, de er bedre uddannet, de spiser sundere, de bruger mindre medicin osv.

## Det videnskabelige spørgsmål

- *Hyppighed* af medicinforbrug: Ja, formentlig selektionsbias
- *Association* mellem folsyreindtag og neuralrørsdefekter: Nej, formentlig ikke selektionsbias

## Fordi

- Både eksponeringen og udfaldet skal være associeret med sandsynligheden for at deltage i studiet *i sammenlignende undersøgelser*



# Validitet

## Intern validitet

- Vedrører resultaternes gyldighed for målpopulationen
- Trues af selektionsbias, informationsbias og confounding

## Ekstern validitet/generaliserbarhed

- Vedrører gyldighed udover målpopulationen
- Forudsætter intern validitet
- Kvalitativt udsagn om en effekts retning og styrke



# Er resultaterne biased af selektion?

- Vi ved (oftest) ikke om forekomsten eller associationen er biased af selektion, da vi (oftest) ikke har svarene fra dem der ikke ville deltage i studiet
- Risikoen for selektionsbias må overvejes afhængig af det videnskabelige spørgsmål, studiedesignet, og de anvendte data
- Texansk studie af prævalens af HIV-infektion





# Hvad kan man gøre?

## Indsamling af data

- Maksimere responsraten gennem rykkerprocedurer, konkurrencer, betaling osv.
- Responsraterne falder og har gjort det de seneste 25 år
- Snowball sampling (svært tilgængelige grupper)
- Respondent driven sampling (korrektion for afhængighed)
- Nationale registre uden selektion

EPIDEMIOLOGY

**When an Entire Country Is a Cohort**

Lone Frank<sup>2</sup>

Lone Frank writes from Copenhagen, Denmark.

Denmark has gathered more data on its citizens than any other country. Now scientists are pushing to make this vast array of statistics even more useful

**Science**



# Hvad kan man gøre?

## Analyse af data

- Bortfaldsanalyser: Adskiller deltagere sig fra ikke-deltagere?
- Intention-to-treat analyser: fastholde den oprindelige eksponering uanset compliance



# Stofmisbrug og leverbetændelse

Dagens afstemning på

- [www.madskamper.dk/afstemning](http://www.madskamper.dk/afstemning)



# Selektionsbias, eksempler

## Randomiserede kliniske-, og kohortestudier

*Selektionsbias er generelt ikke et stort problem i disse studiedesigns fordi selektionen skal være relateret til både eksponering og udfald (som sker i fremtiden)*

- Attrition bias. F.eks. ny anti-depresiva og depression. Undervurderer betydningen af ny anti-depresiva fordi de mest deprimerede på den gamle medicin dropper ud
- Length-time bias. F.eks. kræftscreening og overlevelse. Overvurderer betydningen af screening fordi patienter med langsom progression har bedre overlevelse



# Selektionsbias, eksempler

## Case-kontrol studier

- Diagnostisk bias. F.eks. skiløb og benbrud. Overvurderer betydningen af skiløb fordi cases hovedsageligt samples blandt personer med en kendt risikofaktor
- Valg af kontroller. F.eks. bugspytkirtelkræft og kaffe. Overvurderer betydningen af kaffe fordi kontroller er blevet rådet til ikke at drikke kaffe

## Tværsnitsstudier

- Self-selection bias. F.eks. prævalens af motion blandt gravide overvurderes fordi motionerende kvinder oftere deltager
- Survival bias. F.eks. rygning og KOL. Underestimerer betydningen af rygning fordi rygere med KOL har stor risiko for at dø



# Kan selektionsbias forklare associationen?

1000 personer var inviteret til at deltage i en undersøgelse af sammenhængen mellem køn og hårtab. I alt deltog 650 personer (65%)

$$OR = (100/200) / (50/300) = 3.00 \text{ (95\% CI 2,04-4,40)}$$

	+ Hårtab	- Hårtab
Mand	100	200
Kvinde	50	300

Vi mistænker mænd, der taber håret, for at være mere interesserede i at deltage i undersøgelsen, end de andre grupper



# Kan selektionsbias forklare associationen?

Vi antager, at alle mænd med hårtab deltager (100%) mens de øvrige grupper har en ensartet lavere svarprocent (61%)

	+ Hårtab	- Hårtab
Mand	100 (100%)	200 (61%)
Kvinde	50 (61%)	300 (61%)

$$\text{ObserveretOR} = \frac{\text{svar\%(a)} / \text{svar\%(c)}}{\text{svar\%(b)} / \text{svar\%(d)}} \times \text{SandOR}$$

$$3 = \frac{100 / 61}{61 / 61} \times \text{SandOR}$$

$$\text{Sand OR} = 1,83 \text{ (95\% CI 1,32 - 2,53)}$$



# Næste gang

## Confounding

- Dvs. at de associationer vi observerer kan skyldes forveksling af effekter pga. uombyttelighed mellem de sammenlignede grupper

