Education corner Session 3

Back to school: Confounding and selection bias

Elena Dudukina, MD, MSc, PhD
Pharmacoepidemiologist @ DAC
ELDU@dkma.dk

25 September 2023

LÆGEMIDDELSTYRELSEN

Objective of epidemiologic research

- Valid, precise, and generalizable estimate
- Internal validity: precision and validity of the effect estimate in the analytic sample
  - Analytic sample is the observed portion of the study sample
- External validity: correspondence between the analytic sample and the target population
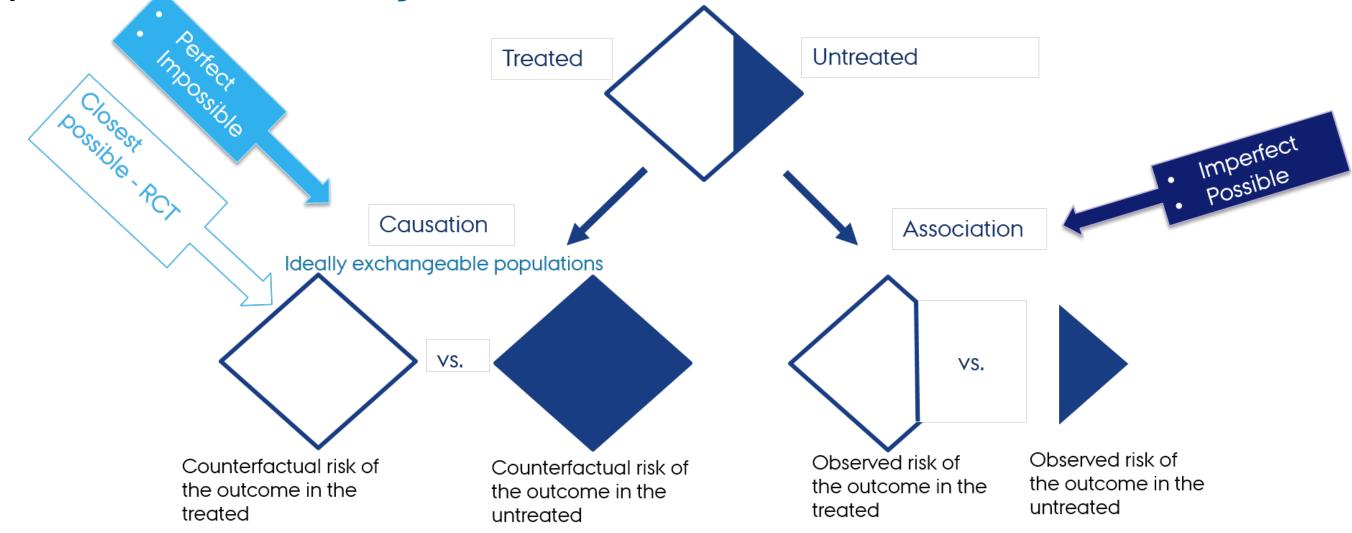
target population

study sample*

analytic sample

*It may be possible to extrapolate (transport) study results to a target population that is partially or completely nonoverlapping with the study sample.

T. Lash, M. Fox. Applying Quantitative Bias Analysis to Epidemiologic Data

LÆGEMIDDELSTYRELSEN

Counterfactual vs factual 'worlds'

- Counterfactual comparisons of exposed and unexposed → perfect and **impossible**

- RCTs → closest possible to the counterfactual comparison of exposed vs unexposed → **not always feasible**
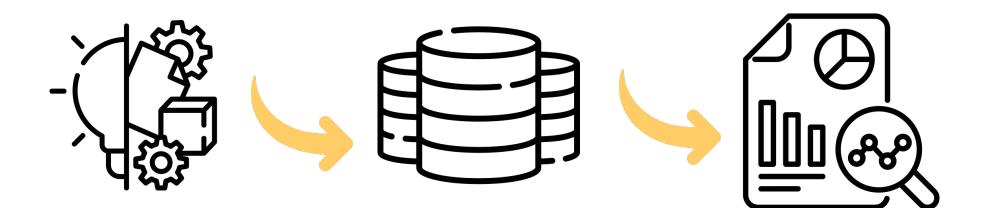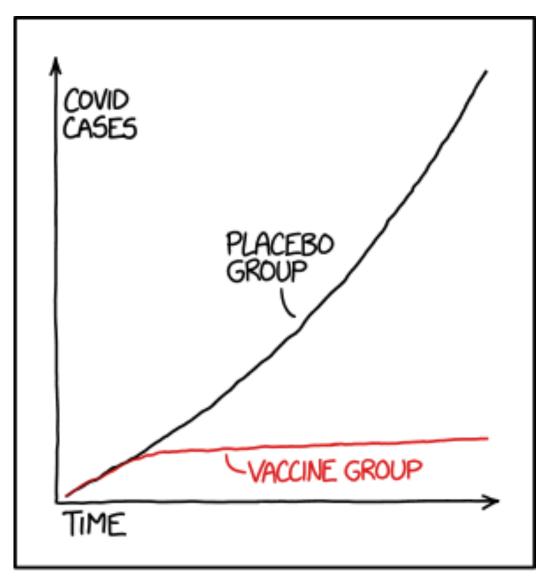

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

Estimating magnitude of effect in pharmacoepidemiologic studies
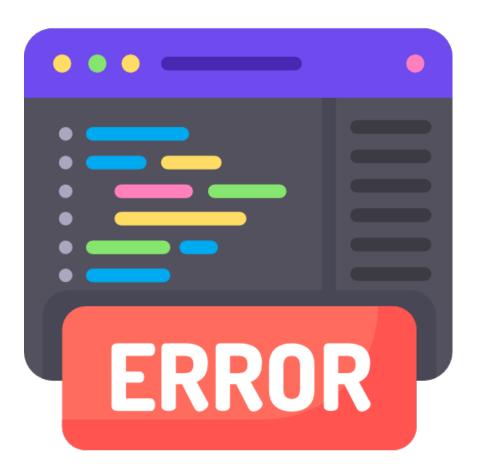
- Design
- Data
- Statistical analyses

COVID CASES

PLACEBO GROUP

VACCINE GROUP

TIME

STATISTICS TIP: ALWAYS TRY TO GET DATA THAT'S GOOD ENOUGH THAT YOU DON'T NEED TO DO STATISTICS ON IT

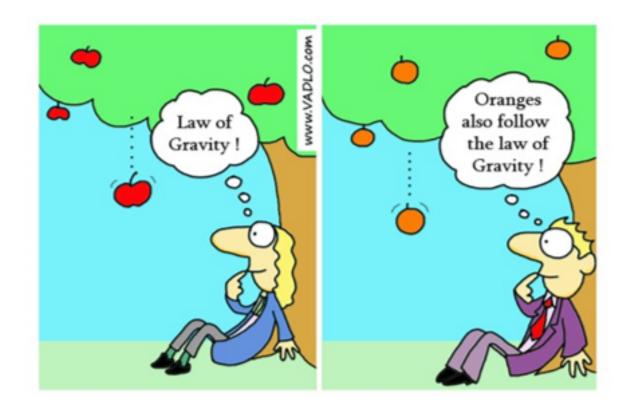LÆGEMIDDELSTYRELSEN

Error in epidemiologic research

- Random error
  - Sampling (random) variability
  - Chance
- Systematic error (bias)
  - Selection bias
  - Confounding
  - Measurement (information) bias

Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012-4

LÆGEMIDDELSTYRELSEN

Selection bias

- Strict inclusion/exclusion criteria or sampling from a subset of a population
- Not representative of population as a whole
  - "Representativeness does not, in itself, deliver valid scientific inference"
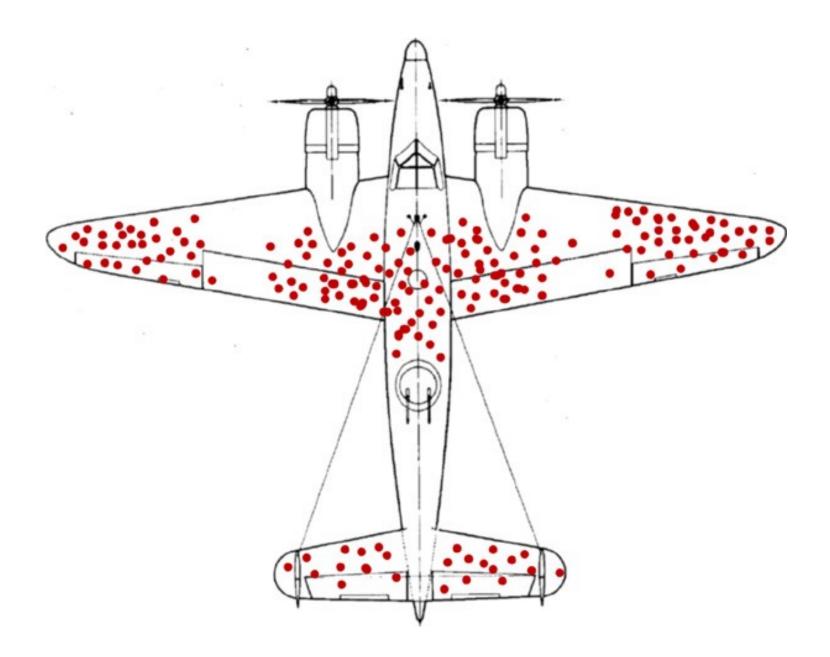- May enhance internal validity
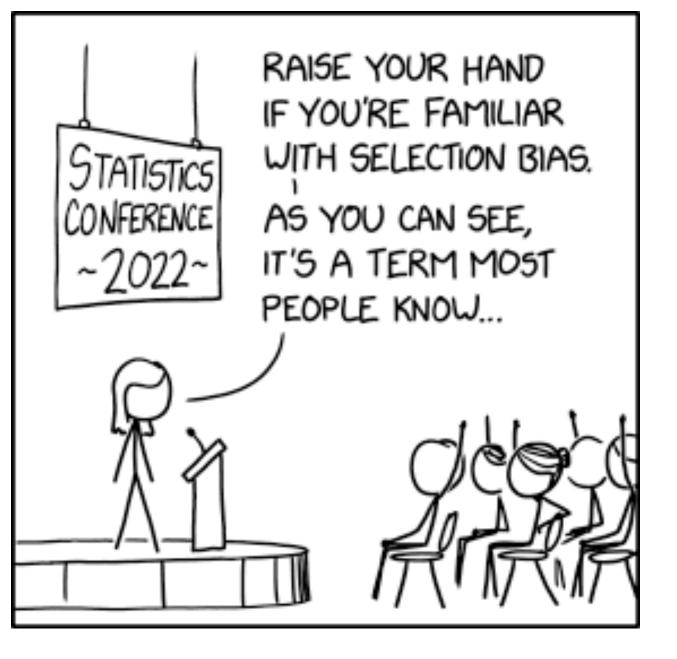
Selection bias

- Risk of the outcome in exposed and unexposed in all eligible:
  - 100/1000 = 10%

- Exposure-Outcome association in all eligible:
  - RR = (100/1000) / (100/1000) = 1.0

- Exposure-Outcome association in participants
  - RR = (80/480) / (60/260) = 0.72

	Outcome+		Outcome-	
	Exposure +	Exposure-	Exposure+	Exposure-
Participants	80	60	400	200
Non-participants	20	40	500	700
All eligible	100	100	900	900

T. Lash, M. Fox. Applying Quantitative Bias Analysis to Epidemiologic Data

LÆGEMIDDELSTYRELSEN

Selection bias

- Association is measured among participants only
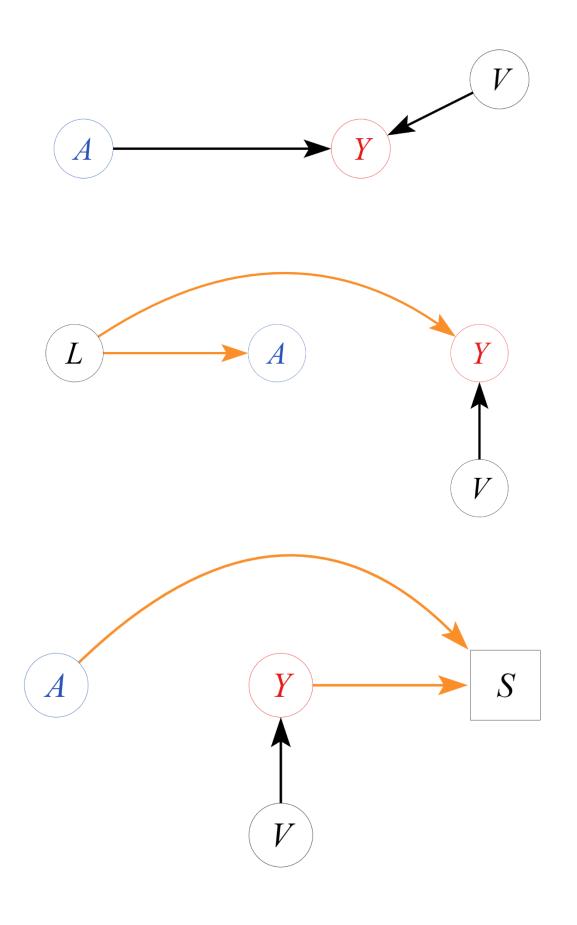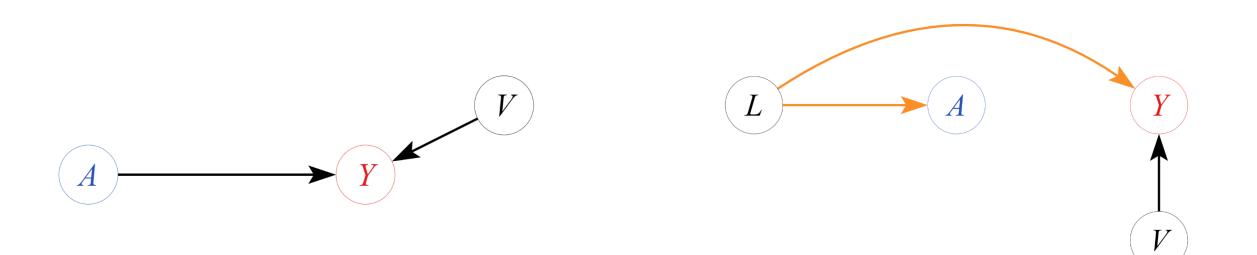 - Survivorship bias

DAGs

- **D**irected **A**cyclic **G**raphs (causal graphs)
- Drawing causal graphs: how world works
- Very (very-very-very) quick introduction
  - **A:** exposure/treatment status
  - **Y:** outcome
  - **L:** confounder
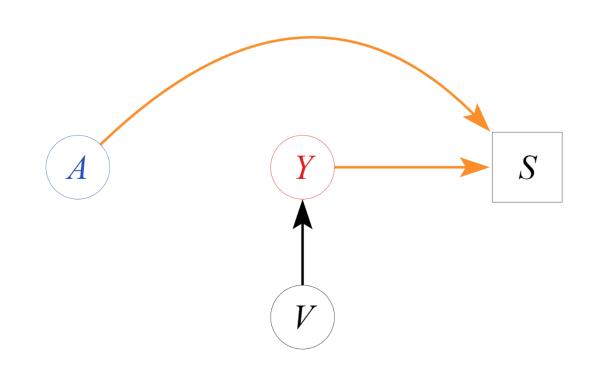 - **S:** selection/collider variable
  - **V:** effect measure modifier

DAGs

- Variables are associated in the data when:

  1. One causes another: **A → Y**

  2. They have a common cause (confounding): **A ← L → Y**

  3. They have a common consequence that has been conditioned on (selection bias/collider stratification bias): **A → [S] ← Y**
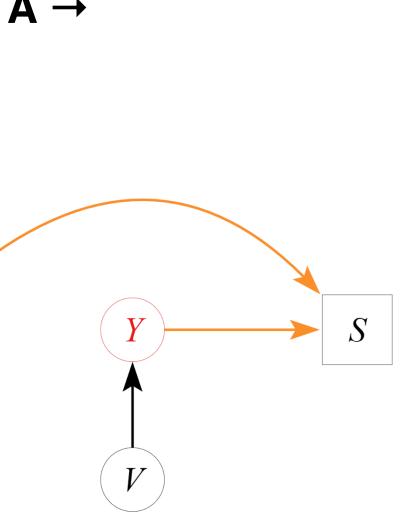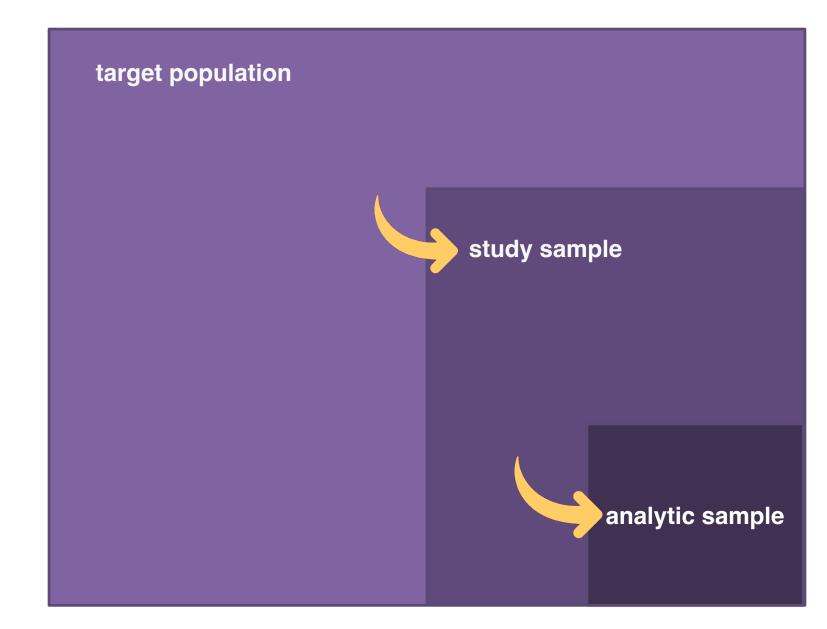
Collider stratification bias

- Variables are associated in the data when:
  - They have a common consequence that has been conditioned on (selection bias/collider stratification bias): **A → [S] ← Y**
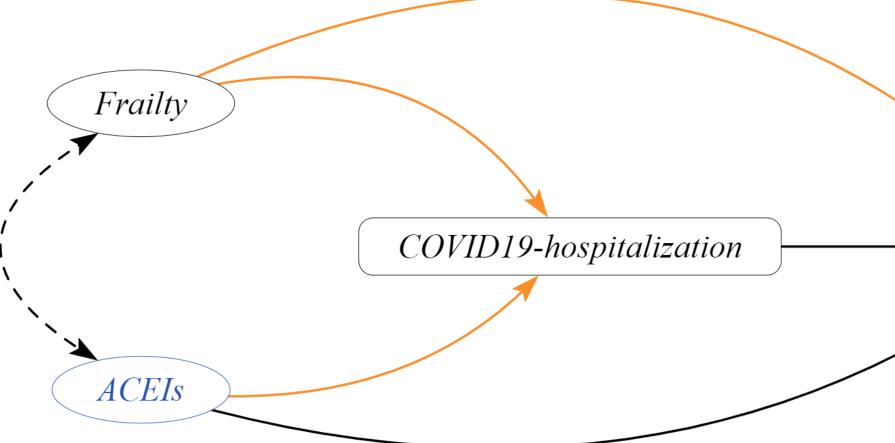
Selection bias

- Bias away from the true effect in the underlying population, due to selecting the sample of it.

Selection bias with colliders

- "Type 1" selection bias
  - Restricting to one or more level(s) of a **collider** (or a descendant of a collider)
  - Affects internal validity
  - May be accounted for analytically

Lu H, Cole SR, Howe CJ, Westreich D. Toward a Clearer Definition of Selection Bias When Estimating Causal Effects. Epidemiology. 2022;33(5):699-706. doi:https://doi.org/10.1097/ede.0000000000001516

Selection bias without colliders

- "Type 2" selection bias
  - Restricting to one or more level(s) of an **effect measure modifier**
  - Affects external validity
    - May affect internal validity
  - Does not occur when there is **no effect** of exposure on the outcome
  - Scale dependent (additive vs multiplicative)
  - May be accounted for analytically

28% of exposed and 28% of unexposed were censored

S: censoring from the longitudinal study

study sample
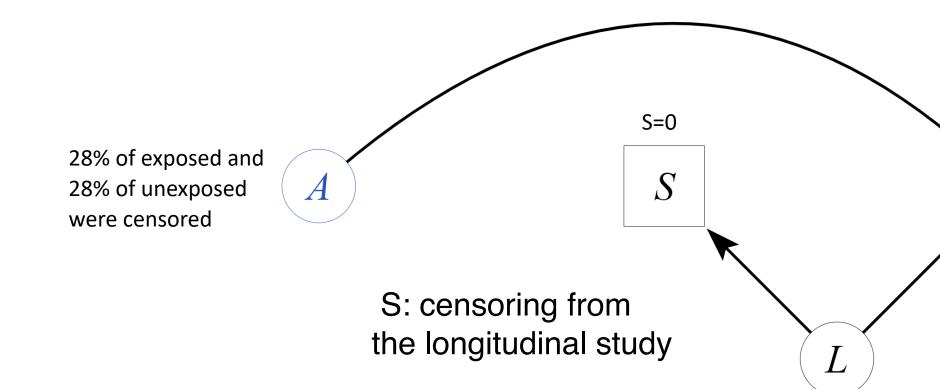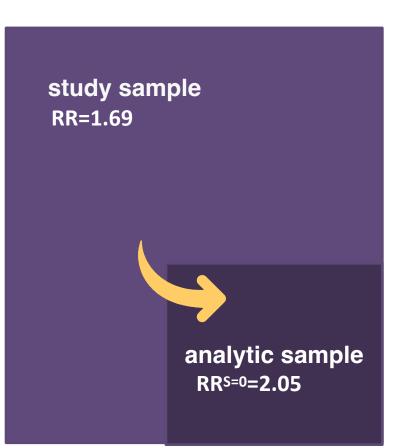RR=1.69

analytic sample
RR^{S=0}=2.05

Lu H, Cole SR, Howe CJ, Westreich D. Toward a Clearer Definition of Selection Bias When Estimating Causal Effects. Epidemiology. 2022;33(5):699-706. doi:https://doi.org/10.1097/ede.0000000000001516
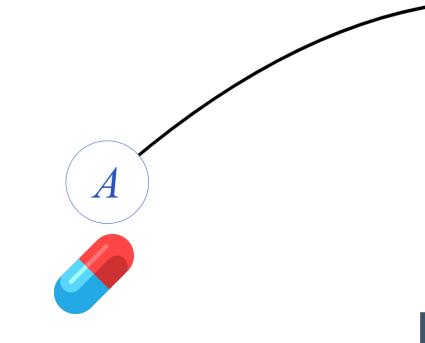
LÆGEMIDDELSTYRELSEN

Selection bias without colliders

- **A:** Medical treatment
- **Y:** Cardiovascular mortality
- **L:** Quality of healthcare
- **S:** Costs of treatment

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

Selection bias with versus without colliders

Selection bias	Type 1	Type 2
Definition	Restricting to one or more level(s) of a collider (or a descendant of a collider)	Restricting to one or more level(s) of an effect measure modifier
Other names	Collider stratification bias; collider restriction bias	generalizability bias
Can occur on sharp null?	Yes	No
Bias in the referent population?	Yes	Yes
Bias in the selected sample?	Yes	No
Can affect internal validity?	Yes	Yes
Can affect external validity?	No	Yes
Effect measure scale dependent?	No	Yes

- Simultaneously adjusting for confounding and adjusting for type 1 and type 2 selection bias may induce new collider bias

Lu H, Cole SR, Howe CJ, Westreich D. Toward a Clearer Definition of Selection Bias When Estimating Causal Effects. Epidemiology. 2022;33(5):699-706. doi:https://doi.org/10.1097/ede.0000000000001516
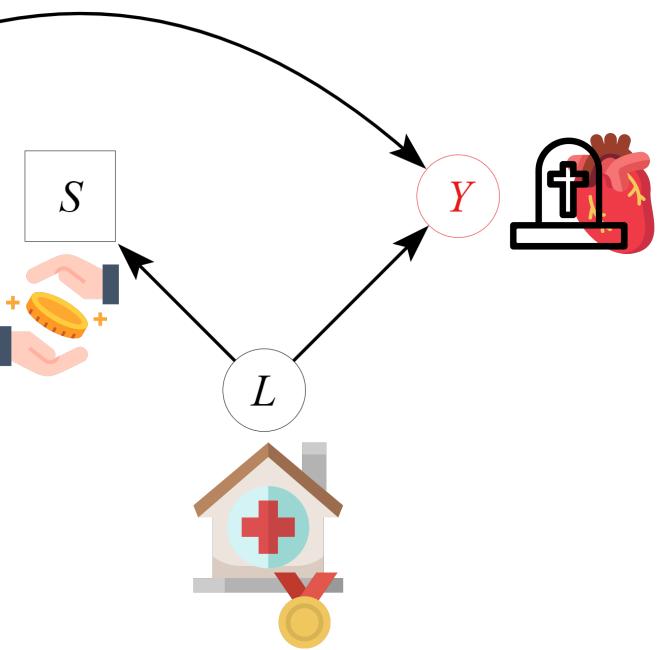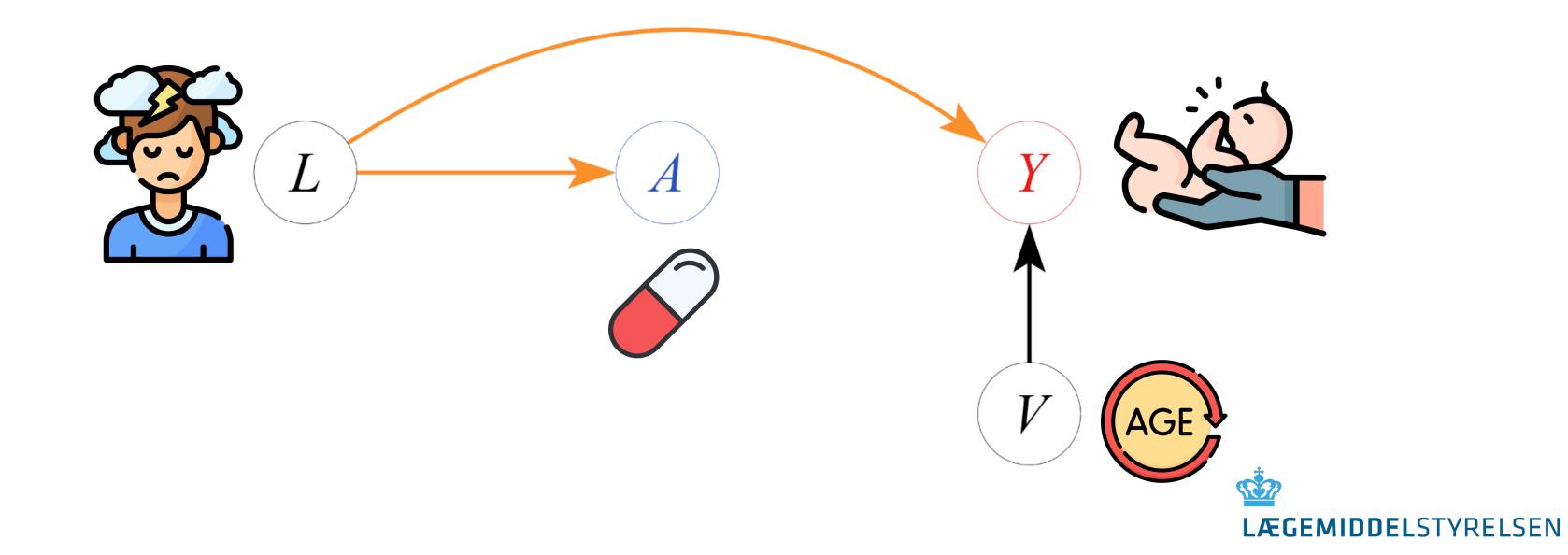
LÆGEMIDDELSTYRELSEN

Confounding

- Affects internal validity
  - When unaddressed, inferences about exposure-outcome association are limited if at all possible

Selection vs confounding bias
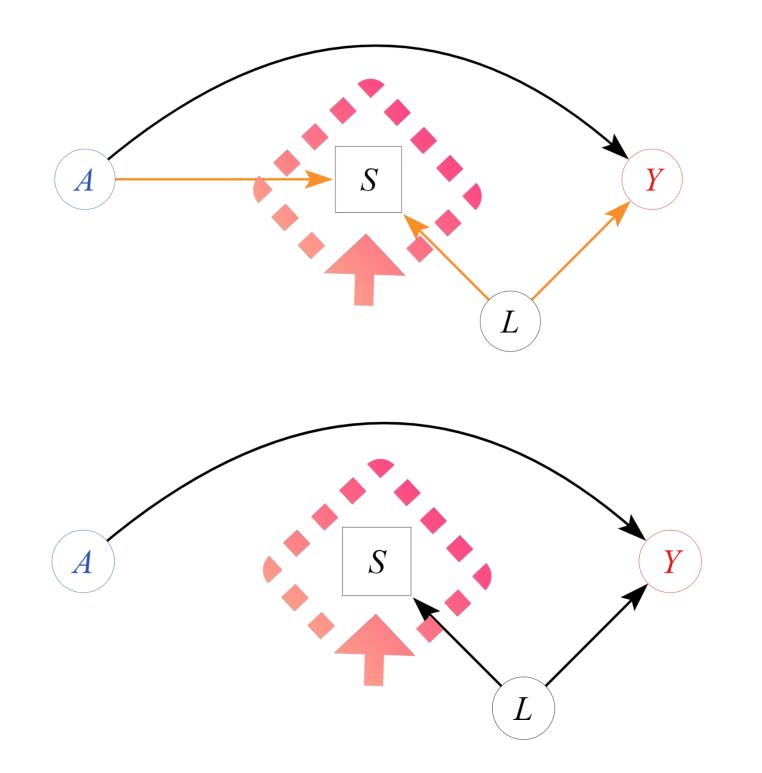
- **Selection bias**
  - May affect internal and external validity
  - May be addressed by
    - Avoiding conditioning (adjusting, restricting, stratifying, or matching) on the colliders on the path between the exposure and the outcome
    - Inclusion/exclusion criteria (design and analysis-motivated)
- May be addressed analytically
  - Data quality
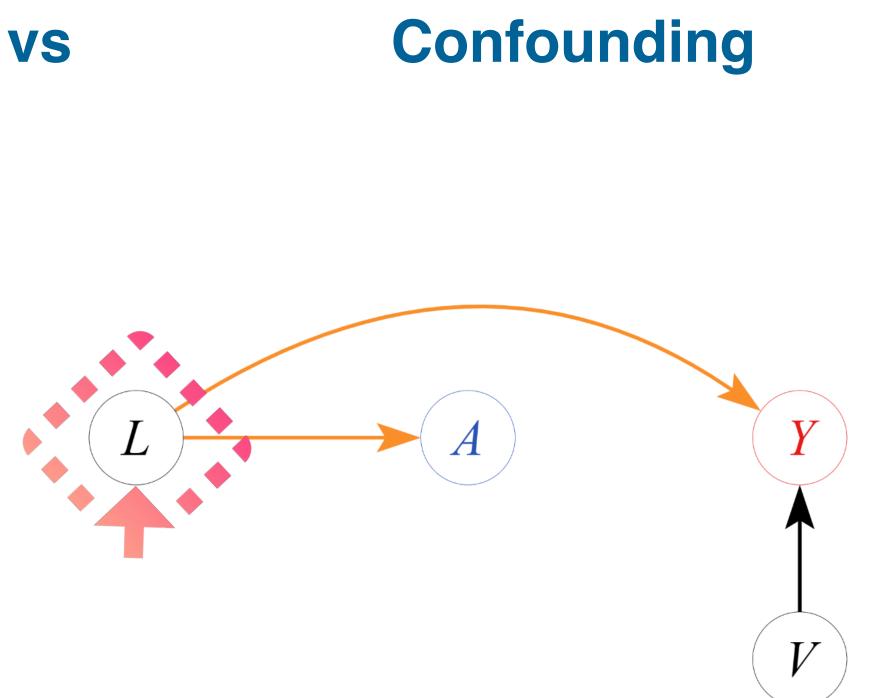 - Availability of prognostic and participation-related variables

- **Confounding**
  - Affects internal validity
    - External validity relevance is low in the absence of internal validity
  - May be addressed by design:
    - Randomization
  - May be addressed analytically (adjustment, restriction, stratification, standardization, matching)
    - Data quality
    - Availability of common causes of exposure and outcome
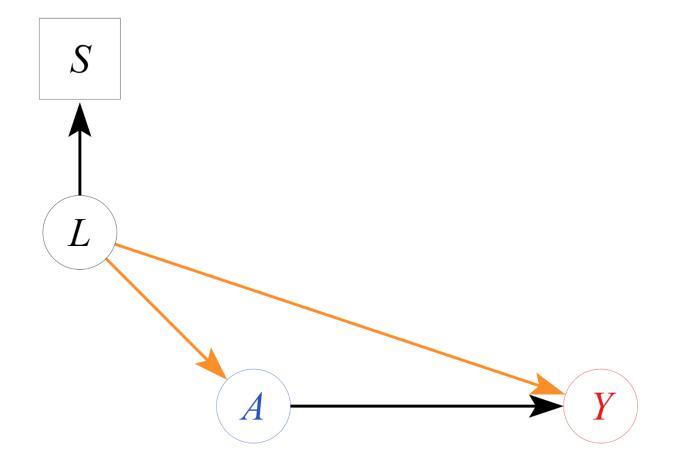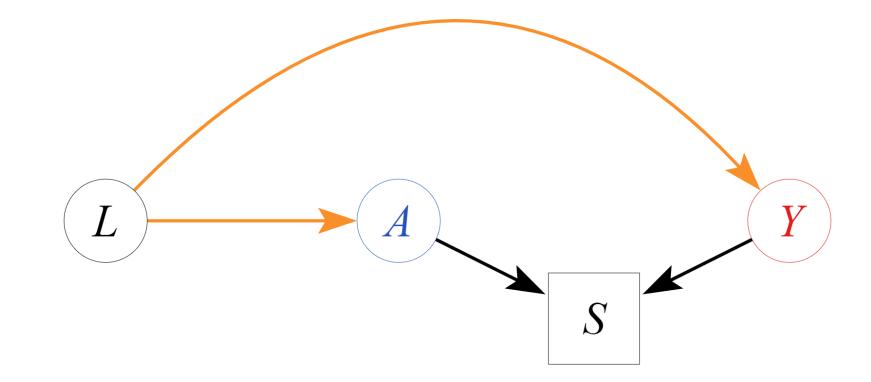   - Strong untestable assumptions

Selection bias   vs   Confounding

Selection and confounding bias

Estrogen/progestin therapy and coronary heart disease (CHD) in women

- Women's Health Initiative randomized trial
  - 68% greater in the first two years after **initiation vs no intiation**, 24% greater after an average of 5.6 years
- Observational studies were based on the Nurses' Health Study:
  - 32% reduced risk of CHD among postmenopausal hormone **users vs never users**
- Confounding vs selection bias?
  - Users vs never users analyses:
    - No adjustment for non-adherence
    - Selection bias due to "attrition of susceptibles"
  - "Confounding for the effect of therapy initiation in the NHS seems to play little role"

Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19:766–779.
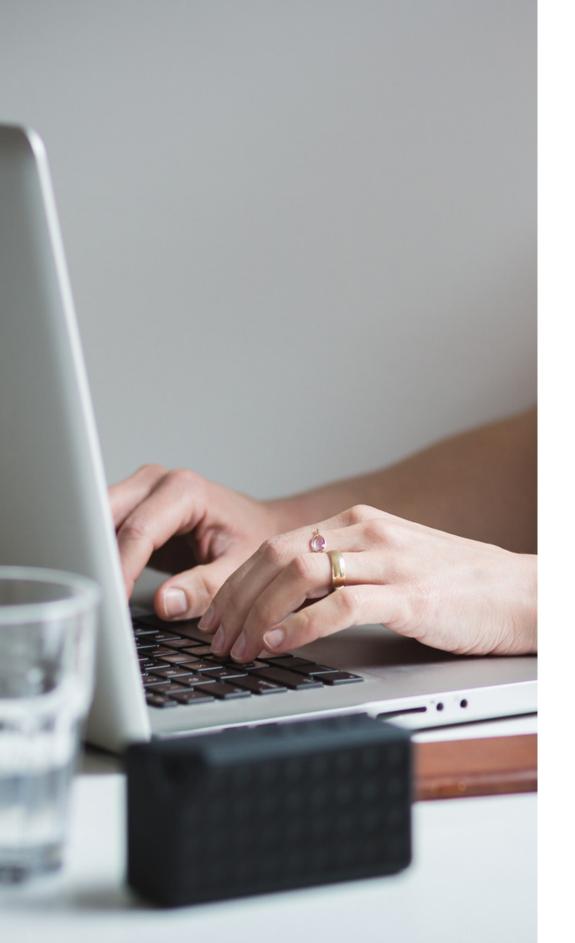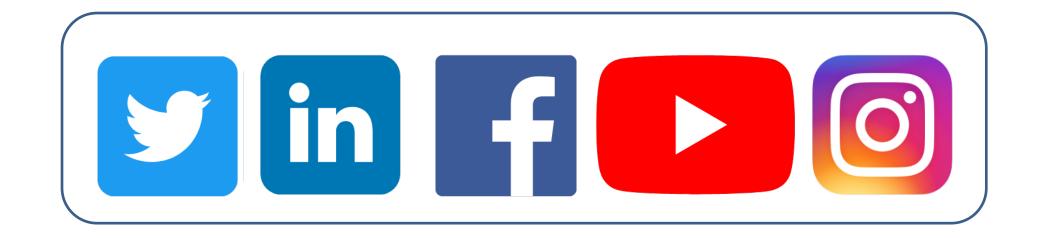
LÆGEMIDDELSTYRELSEN

Summary

- Selection bias and confounding have distinct definitions

- Selection bias and confounding are threats to internal validity

- Selection bias arises from restricting to one or more level(s) of a collider (or a descendant of a collider) or of an effect measure modifier

- Selection bias is a threat to the external validity

- Simultaneous presence of selection bias and confounding (and measurement error) and their joint impact on the study results should be considered
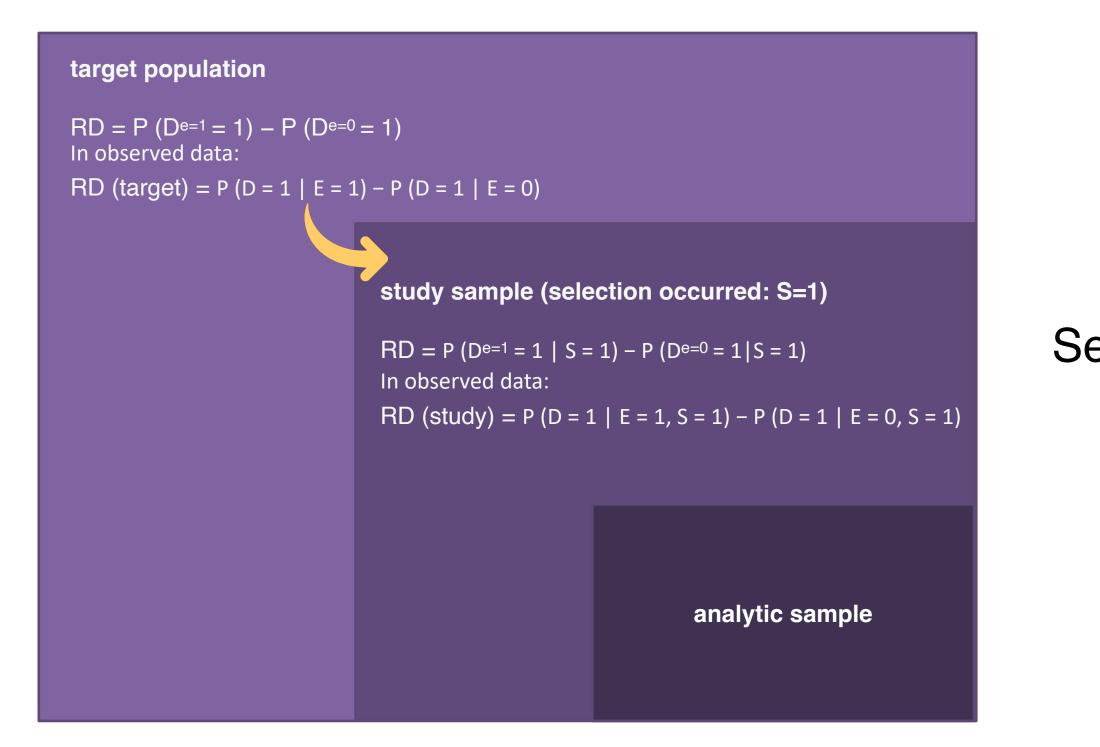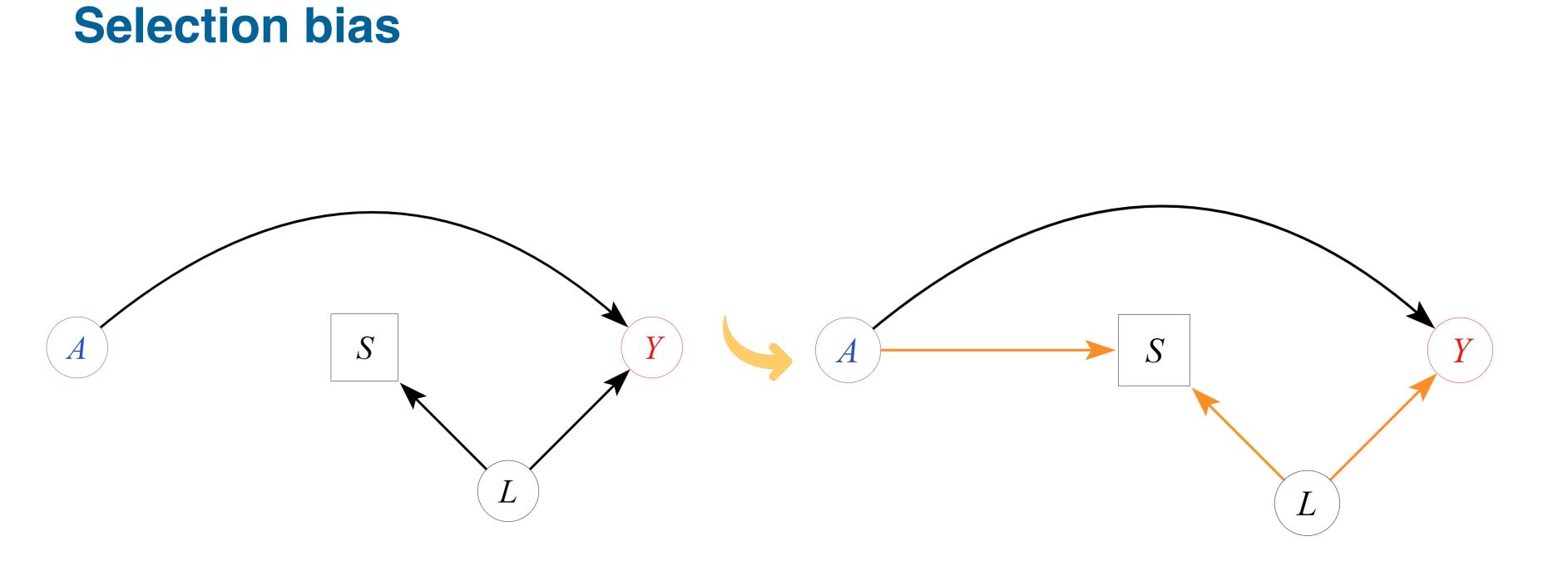
@laegemiddelstyrelsen

Selection bias formulas

Selection bias is defined as the difference between the true causal effect in the underlying (referent) population and the effect estimate in the selected sample.

target population

$RD = P (D^{e=1} = 1) - P (D^{e=0} = 1)$
In observed data:
$RD (target) = P (D = 1 \mid E = 1) - P (D = 1 \mid E = 0)$

study sample (selection occurred: S=1)

$RD = P (D^{e=1} = 1 \mid S = 1) - P (D^{e=0} = 1 \mid S = 1)$
In observed data:
$RD (study) = P (D = 1 \mid E = 1, S = 1) - P (D = 1 \mid E = 0, S = 1)$

analytic sample

Selection bias = RD (target) - RD (study)

LÆGEMIDDELSTYRELSEN

Selection bias

"Type 2" selection bias (without collider)

"Type 1" selection bias (collider stratification bias)

LÆGEMIDDELSTYRELSEN