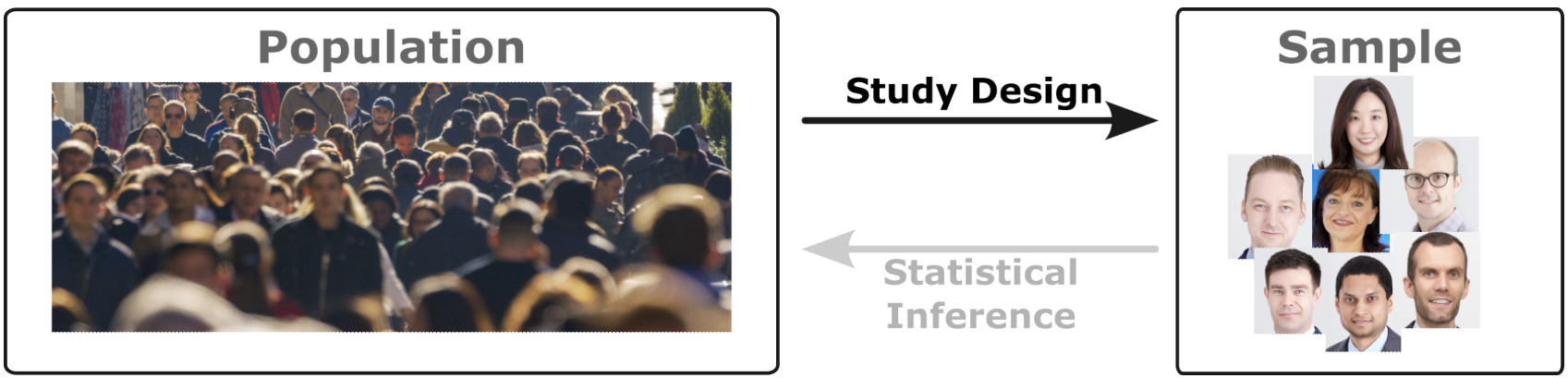
Probability and Statistics

Chapter 2: Study Design

Joris Vankerschaver

Goal of this chapter

- Understand how we can draw conclusions about an entire population by data gathered from a smaller sample, through an appropriately designed experiment.
- How we choose that sample is the subject of this chapter.



Outline

Example: Captopril

Does the medicine Captopril reduce blood pressure?

- **Population**: Hypothetical group of all current and future patients, with high blood pressure, for whom we want to draw conclusions
- **Sample**: Subset of the population for which we will register observations
- **Experimental design**:
  - How to draw a sample in such a manner that it is representative for population and contains as much information as possible concerning the research question
  - What information should we measure, how and when?

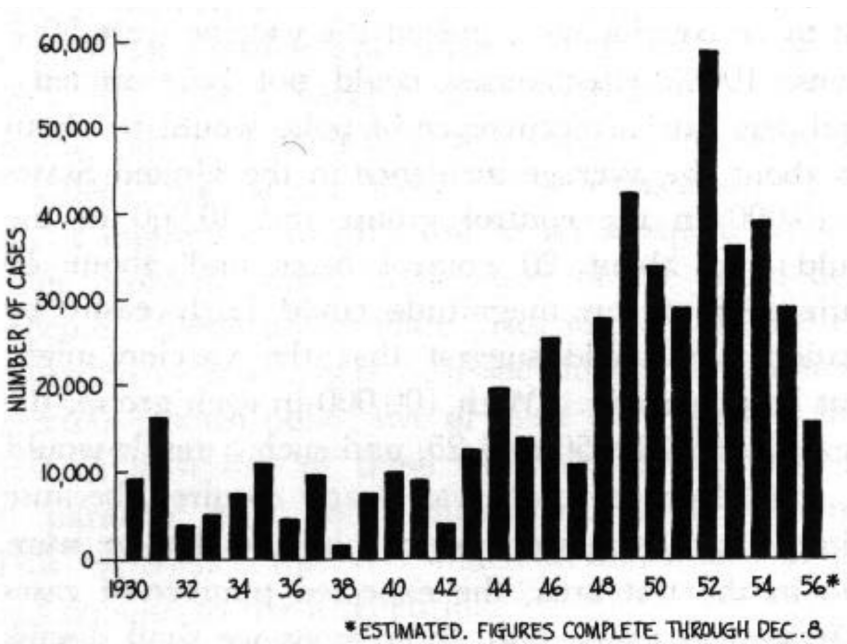
Section 1

The Salk vaccine field study

Outline

The Polio disease

- Highly infectious
- Endemic for thousands of years
- Symptoms: headaches, stiffness, debilitating paralysis
- 20th century: epidemics with 1000s of deaths



Testing the efficacy of the polio vaccine

Vaccine history:

- 1930s: Early vaccines, but hasty/shoddy testing:
  - Several deaths, paralytic cases
  - Anaphylactic shock
- By 1950s: several vaccines (Salk, Sabin, . . .) under development

But is the vaccine safe?

- 1954: National Foundation for Infantile Paralysis (NFIP) decides to run a field study to evaluate effect of vaccine.
  - **100,000s take part**
  - First double blind randomized control trial for vaccines

We will revisit some of the characteristics of this study.

The NFIP field study

- Largest medical trial (over 1,000,000 children participated)
- Mix of parallel and randomized, double-blind study
- **Proved efficacy of vaccine**

The need for a large study

- Prevalence of polio (1950s): 50 per 100,000.
- Assume vaccine 50% effective
- Run a trial comparing vaccinated with control group

Participants	Treatment	Control
10k/10k	5	3
100k/100k	**50**	**25**

Trial with 10,000 participants in treatment/control group may be underpowered to see effect from vaccination!

What if we vaccinate everybody?

- Reasonable if goal is to eradicate disease as quickly as possible.
- **Impossible to decide if vaccination is effective.**
  - Disease may be going away by itself
  - Other factors (e.g. nourishment) may be at play.

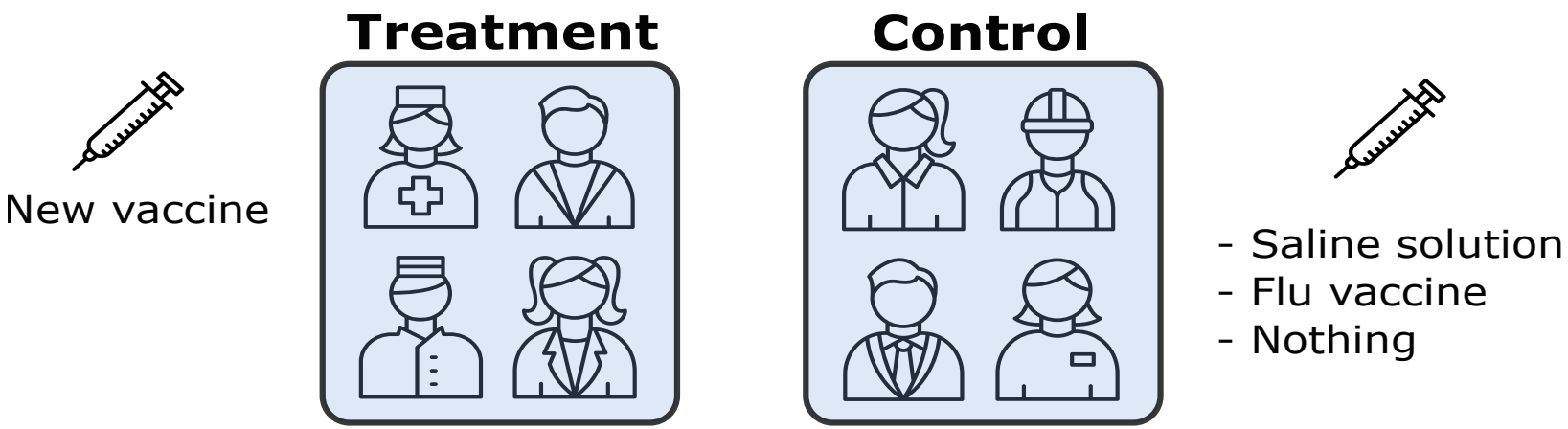
We need to compare with a controlled experiment!

The need for control

Controlled experiment

Design with reference group where observations are obtained given certain conditions that are controlled by researcher.

- Placebo control (e.g. saline solution)
- Standard-of-care (e.g. flu vaccine)
- ...



How do we choose these groups?

Historical control

Historical control

Compare vaccinated children with unvaccinated children from another era, when no vaccine existed.

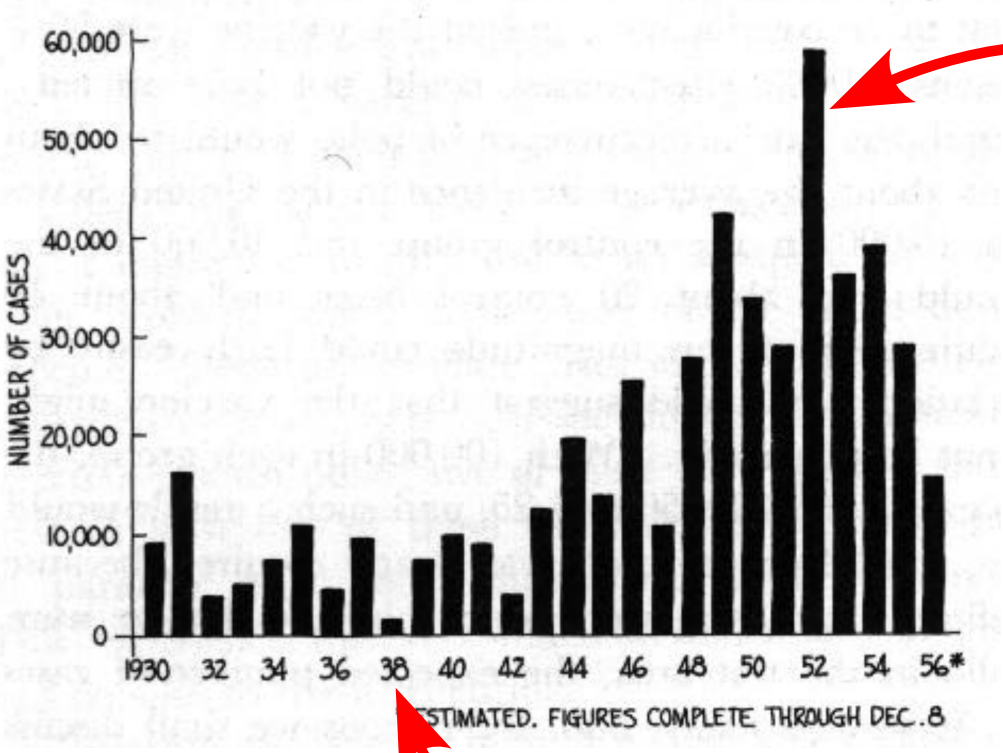
Advantages

- Cost-effective: data often already exists
- Can be impossible/unethical to recruit new control group

Can the groups actually be compared?

- 1954: not an epidemic year as opposed to 1953
- Detection of polio ameliorated between 1953 and 1954
- ...

Historical control



Or here?

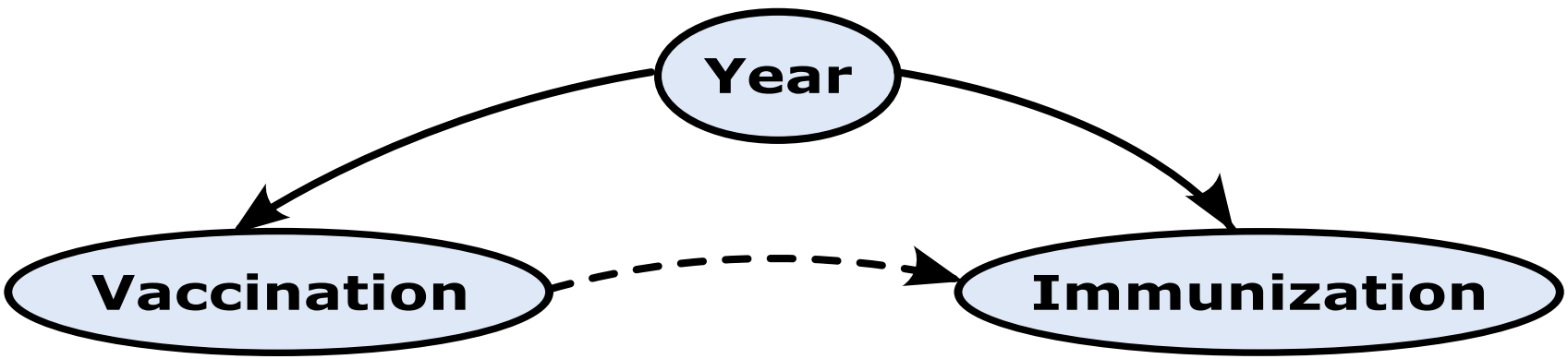
What if we introduced the vaccine here?

Problem: confounding

Confounding

- Happens when a variable (confounder) influences both the treatment and the outcome, but is influenced by neither.
- Confounders obscure the relation between treatment and outcome!

- Year of testing is a confounder
- We will see many more examples!



Parallel studies

Parallel study

To avoid confounding (by year), compare test and control simultaneously.

NFIP study is parallel:

- School districts and age groups that are most vulnerable in study
- Researcher decides who gets vaccine:
  - 350,000 from second grade vaccinated (225,000 test group, 125,000 refusals)
  - 750,000 from first and third grade unvaccinated (control group)

Results of controlled NFIP study

	Number	Incidence (per 100,000)
Vaccine	225,000	25
Control	725,000	54
No consent	125,000	44

But are the groups really comparable?

- Are children from the second grade more susceptible to polio?
- Are children who had consent for vaccination more susceptible to polio?

Confounding in NFIP design

Even in a parallel study, **confounding** can take place.

Socio-economic status

Children from well-off families ...

- Tend to sign up for vaccination
- Have less acquired immunity to polio

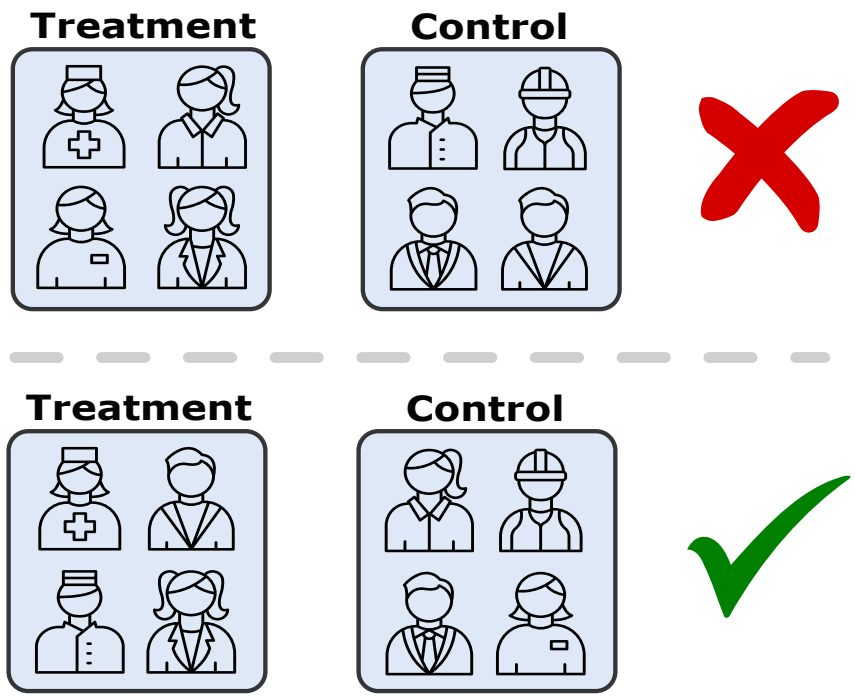
Even worse

Not all confounders are known!

Randomized experiments

Avoid confounding through **randomization**.

- Random allocation creates comparable groups
- Systematic differences between groups can 'only' be caused by difference in intervention



Results Salk vaccine

Controlled NFIP study

	Number	Incidence (per 100,000)
Vaccine	225,000	25
Control	725,000	54
No consent	125,000	44

Double blind randomized controlled study

	Number	Incidence (per 100,000)
Vaccine	200,000	28
Control	200,000	71
No consent	350,000	46

Section 2

Randomization techniques

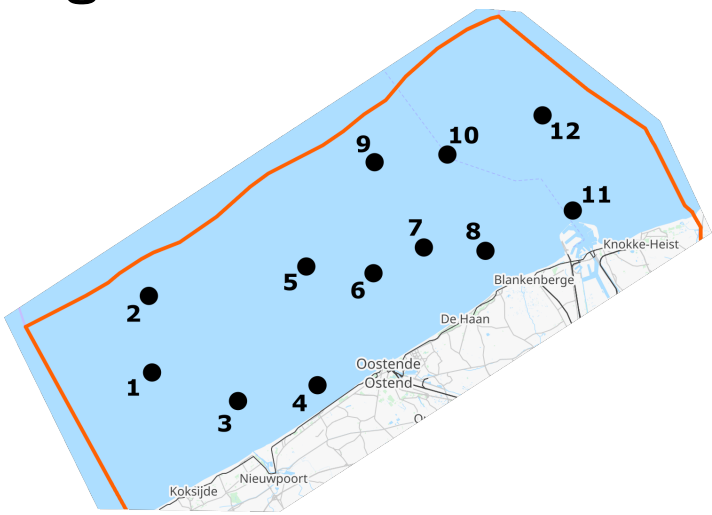
Outline

1. The Salk vaccine field study
    1. Controlled randomized trials
2. **Randomization techniques**
    1. Simple randomization
    2. Block randomization
    3. Stratified randomization
3. Other experimental designs
    1. Pretest/posttest design
    2. Crossover design
    3. Factorial design
4. Observational studies
5. Simpson's paradox (optional)

Case study: effect of predator fish on seabed habitats

- 12 random areas on seabed
- 6 cages that retain predators (R)
- 6 control cages (C)
- Seabed habitats studied and compared between both groups

Where to place the experimental cages?

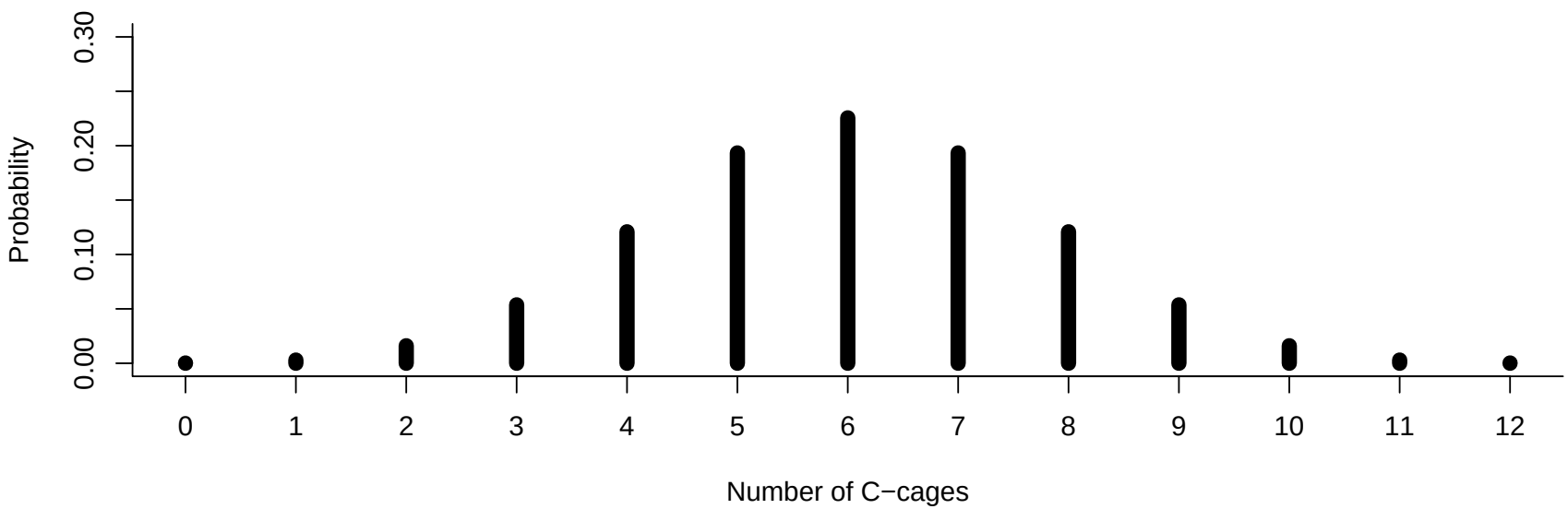


Simple randomization

Flip a coin, and place an R-cage for heads, C-cage for tails.

Problem: imbalance

- Only 23% chance of getting exactly 6 C-cages
- Even 0 or 12 C-cages is possible



Balanced design

Balanced design

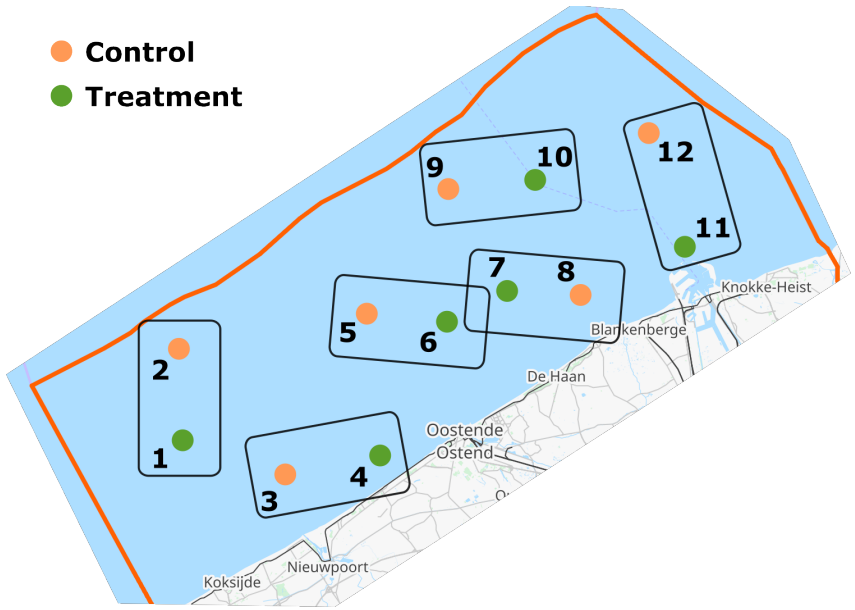
Experimental units (subjects, areas, ...) are divided in smaller blocks of equal size and within each block the same number of units is assigned to the different treatment groups (here: R-cage or C-cage)

- Ensures equal numbers of treatments/controls
- Still does not guarantee that treatment/control groups are completely comparable

Blocks of size 2

- Block size = 2: (1) RC, (2) CR
- Flip a coin: H T T H T H

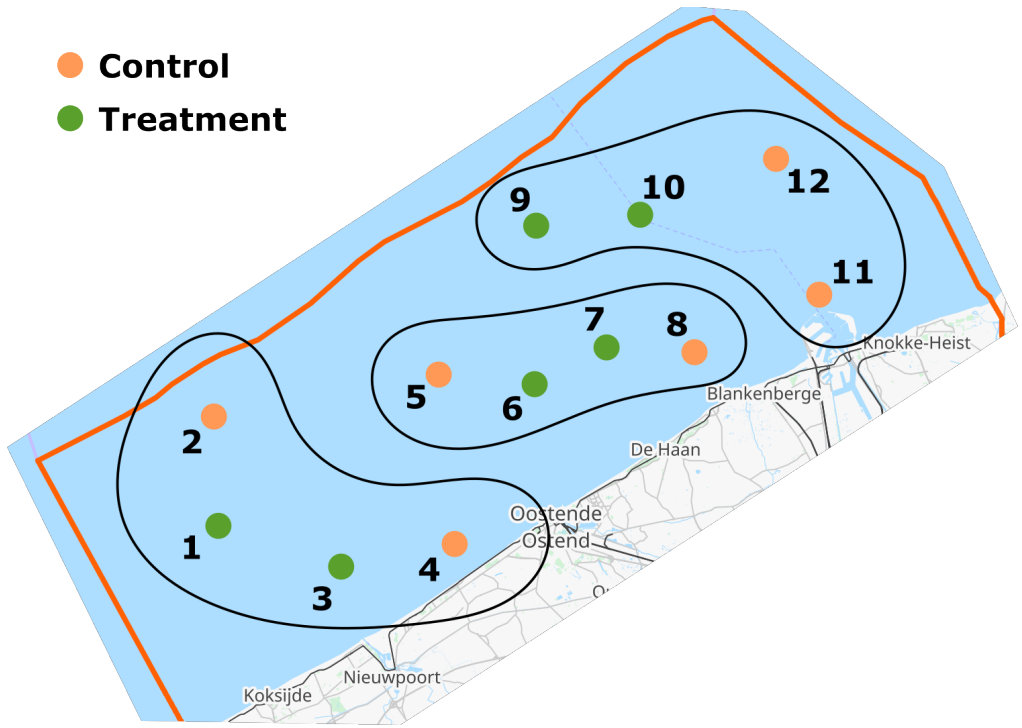
Toss	H		T		T		H		T		H	
Order	R	C	C	R	C	R	R	C	C	R	R	C
Area	1	2	3	4	5	6	7	8	9	10	11	12



Blocks of size 4

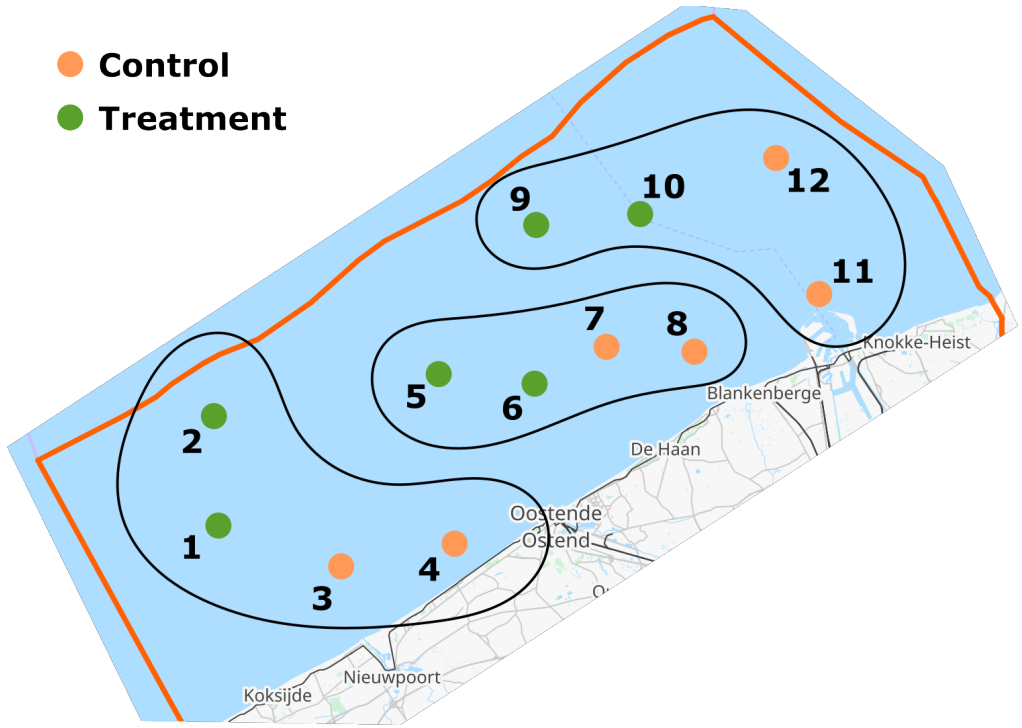
- Block size = 4: (1) RRCC, (2) RCRC, (3) RCCR, (4) CRCR, (5) CRRC, (6) CCRR
- Toss a die: 2, 5, 1

Toss	2				5				1			
Order	R	C	R	C	C	R	R	C	R	R	C	C
Area	1	2	3	4	5	6	7	8	9	10	11	12



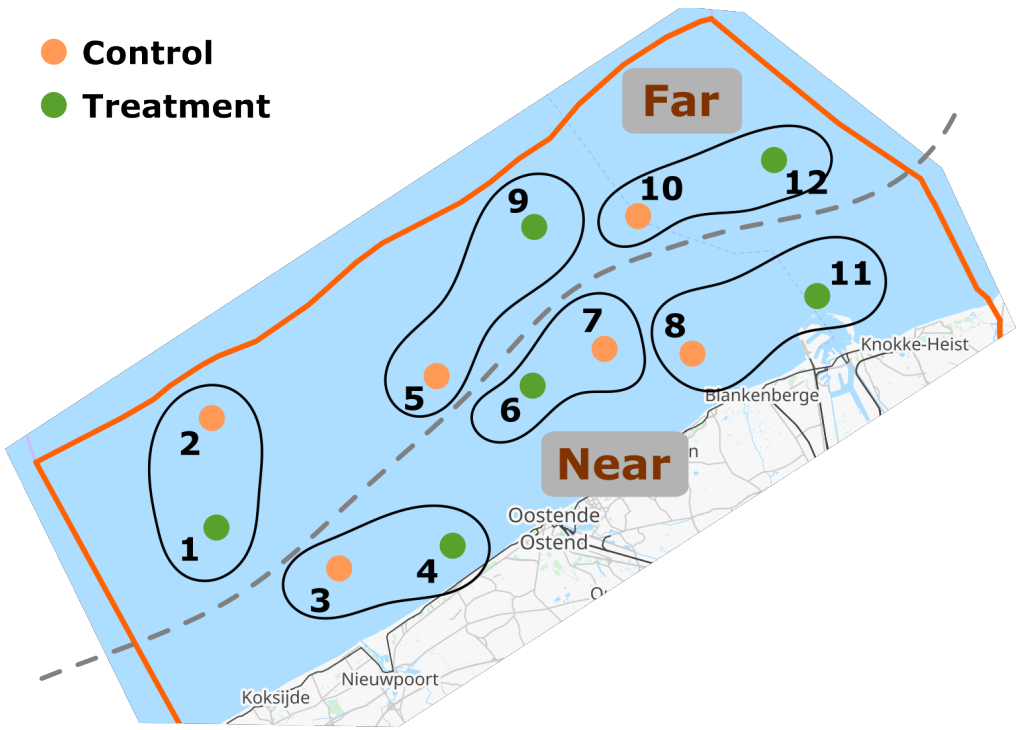
How comparable are the groups?

It is still possible that all control cages end up close to the shore.



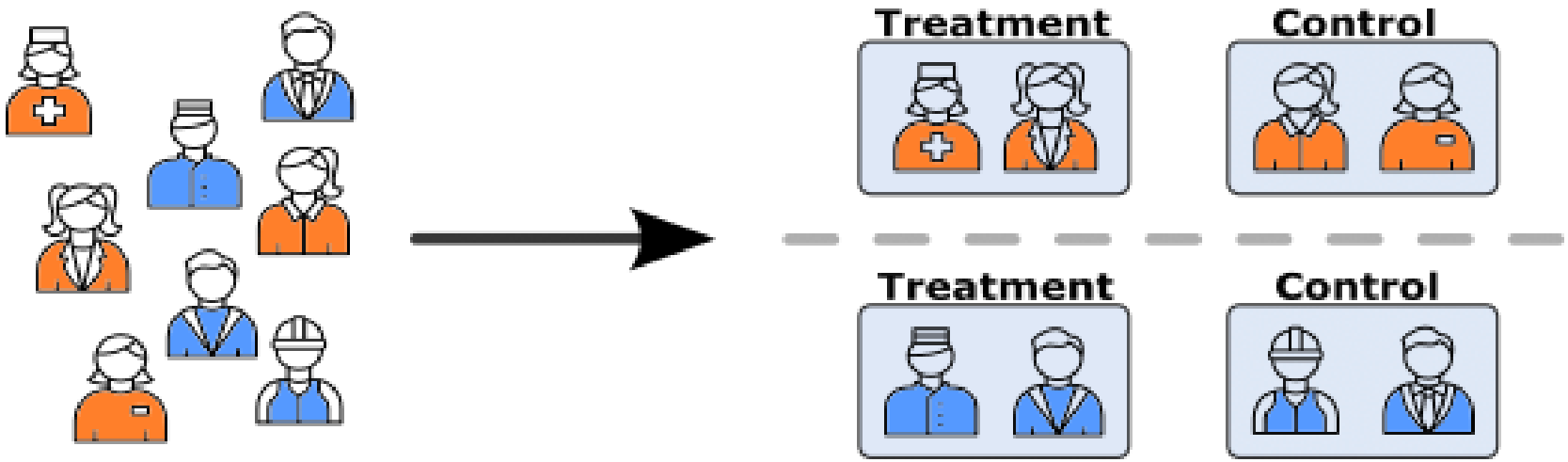
Stratified randomization

Stratified randomization (according to distance to shore: near or far) or **randomized complete block design** avoids the previous problem by separate balanced randomization per stratum.



Adjustment for confounding

- Suppose that the intervention groups we want to compare are inherently not comparable
- Then 'correction' through design is no longer possible
- Corrections will be made through **statistical analysis**:
  - via **stratum-specific analysis**: estimate intervention effect separately for areas close to shore versus further away, ...
  - via more advanced techniques (e.g., regression)



Section 3

Other experimental designs

Outline

1. The Salk vaccine field study
    1. Controlled randomized trials
2. Randomization techniques
    1. Simple randomization
    2. Block randomization
    3. Stratified randomization
3. **Other experimental designs**
    1. Pretest/posttest design
    2. Crossover design
    3. Factorial design
4. Observational studies
5. Simpson's paradox (optional)

Pre-test/post-test design

How?

- Measure characteristics (pre-test)
- Make intervention
- Measure characteristic again (post-test)

Note: no control group

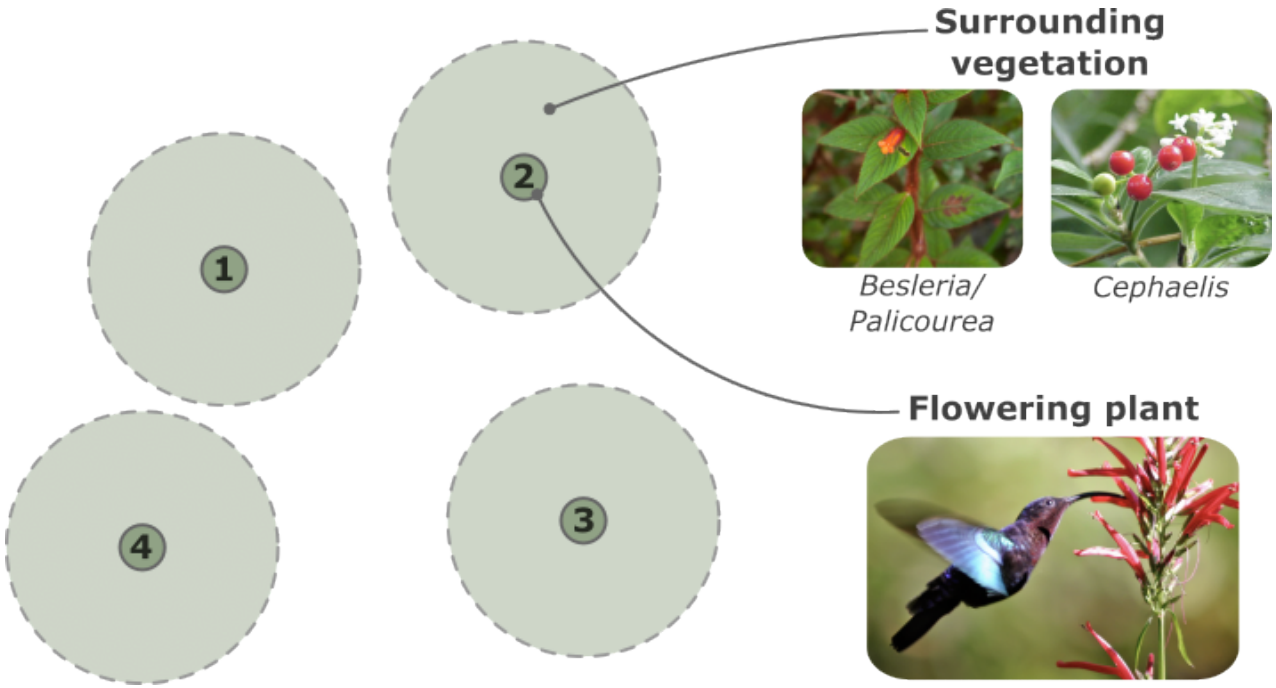
Example: social distancing

Change in Covid-19 case growth in the US before/after introduction of social distancing measures (Siedner et al, PLOS Medicine 17(10), 2020).

Possible confounders:

- Increases in Covid testing
- Spillover from states without social distancing

Crossover design



- Competition between 3 species with low vegetation in Central America
- **Interventions**: relative density Besleria/Palicourea and Cephaelia. A: 10:10, B: 90:10, C: 10:90, D: 50:50
- **Outcome**: number of times a flower is visited by hummingbirds

Crossover design

Period	Plant 1	Plant 2	Plant 3	Plant 4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

- Competition between 3 species with low vegetation in Central America
- **Interventions**: relative density Besleria and Cephaelia
  - A: 10:10, B: 90:10, C: 10:90, D: 50:50
- **Outcome**: number of times a flower is visited by hummingbirds
- **Characteristic**: each plant is subject to each intervention, but in different order

Crossover design

Period	Plant 1	Plant 2	Plant 3	Plant 4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Crossover designs give more information than

- **parallel designs** because every plant is evaluated for every intervention
- **pretest/posttest designs** because it allows to seperate time effect and intervention effect

Possible problem carry-over effect, interaction intervention–time

$\rightarrow$ Crossover design good for interventions with short term effects and stable response measurements over time

Factorial designs

Factorial designs

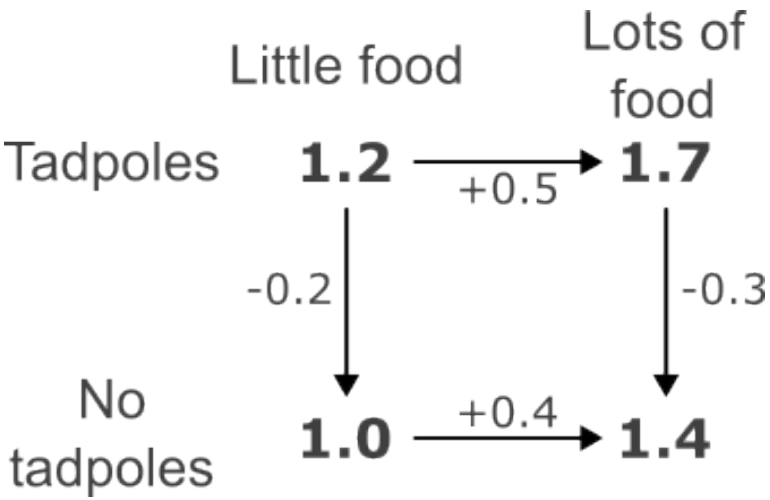
- Test more than 1 intervention at once
- Can detect interactions between interventions (whether 2 interventions strengthen or weaken each other)

Effect of tadpoles and food on salamander larvae length:

- Two factors with two levels each
- 4 groups total

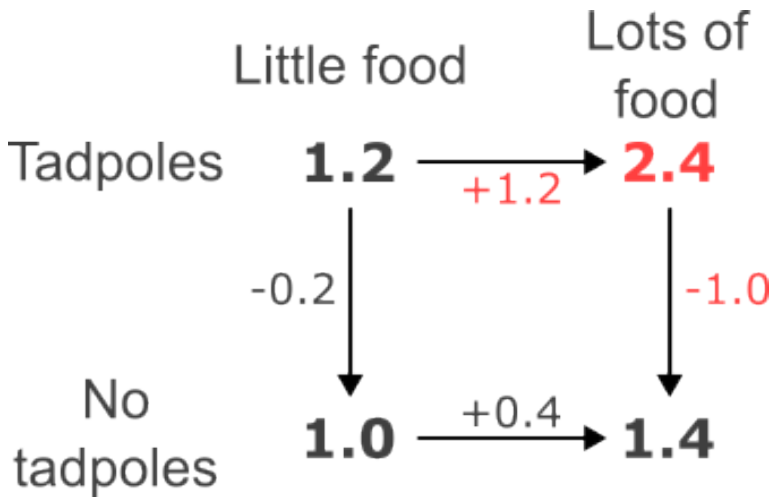
		Food	
		Little	Lots
Tadpoles	Y	A	B
	N	C	D

Factorial designs: interactions



Effect of food does not depend on presence of tadpoles.

- $F = (0.5 + 0.4)/2 = 0.45$
- $TP = (-0.2 - 0.3)/2 = -0.25$
- $I = (0.5 - 0.4)/2 = 0.05$



Effect of food gets a **boost** from the presence of tadpoles.

- $F = (0.5 + 0.4)/2 = 0.45$
- $TP = (-0.2 - 0.3)/2 = -0.25$
- $I = (0.5 - 0.4)/2 = 0.05$

F, TP: effect of food, tadpoles. I: interaction.

Section 4

Observational studies

Outline

1. The Salk vaccine field study
    1. Controlled randomized trials
2. Randomization techniques
    1. Simple randomization
    2. Block randomization
    3. Stratified randomization
3. Other experimental designs
    1. Pretest/posttest design
    2. Crossover design
    3. Factorial design
4. **Observational studies**
5. Simpson's paradox (optional)

Observational studies

- No experiments performed
- Observe which subject is exposed to which exposure
- Existence of confounders can thus not be excluded

Hydroxychloroquine in Covid-19 patients (NEJM 2020)

- 1376 Covid-19 patients in New York City
- 811 receive hydroxychloroquine, rest do not
- Time to intubation/death is compared between both groups

> **CONCLUSIONS**
>
> In this observational study involving patients with Covid-19 who had been admitted to the hospital, hydroxychloroquine administration was not associated with either a greatly lowered or an increased risk of the composite end point of intubation or death. Randomized, controlled trials of hydroxychloroquine in patients with Covid-19 are needed. (Funded by the National Institutes of Health.)

Example: vitamin E and risk for heart disease

Prospective observational study for the effect of vitamin E

After controlling for age and several coronary risk factors, we observed a ***lower risk of coronary disease among men with higher intakes of vitamin E*** *(P for trend = 0.003).*

For men consuming more than 60 IU per day of vitamin E, the multivariate relative risk was 0.64 *(95% confidence interval, 0.49 to 0.83)* ***as compared to those consuming less than 7.5 IU per day****.*

- Data suggest positive effect of vitamin E
- Based on this study, it was advised to administer vitamin E on large scale
- But are vitamin E users in fact **comparable** to nonusers?

Randomized studies

Randomized studies to verify: vitamin E vs. placebo

- Food intervention studies in Linxian (China): RR 0.90 (95% CI: 0.76–1.07)
- Alpha-Tocopherol, Beta Carotene Cancer Prevention Study: RR 1.18 (95% CI: 0.62–2.27)
- Gruppo Italiano per lo Studio della Sopravivenza nell' Infarto miocardico Prevenzione Study: RR 0.95 (95% CI: 0.86–1.05)
- Heart Outcomes Prevention Evaluation Study: RR 1.05 (95% CI: 0.95–1.16)
- Primary Prevention Project: RR 1.07 (95% CI: 0.74–1.56)
- Heart Projection Study: RR 1.06 (95% CI: 0.95–1.18)

{RR = relative risk, CI = confidence interval}

Observational studies versus experimental studies

- Observational studies typically face confounding
- Adjusting for confounding only possible w.r.t. factors that are measured
- Hence observational studies are more limited to deduct causal conclusions
- Randomized studies exclude 'in principle' the existence of confounding factors
- Therefore they are the **gold standard** for causal decision making, but are not always possible or feasible

Section 5

Simpson's paradox (optional)

Outline

1. The Salk vaccine field study
    1. Controlled randomized trials
2. Randomization techniques
    1. Simple randomization
    2. Block randomization
    3. Stratified randomization
3. Other experimental designs
    1. Pretest/posttest design
    2. Crossover design
    3. Factorial design
4. Observational studies
5. **Simpson's paradox (optional)**

Example

Should a doctor prescribe a drug, given the following recovery rates (in a trial of 700 people)?

Drug	No drug
273 out of 350 (78%)	289 out of 350 (**82%**)

More people are better off without it: don't prescribe the drug!

Example

Should a doctor prescribe a drug, given the following recovery rates (in a trial of 700 people)?

Drug	No drug
273 out of 350 (78%)	289 out of 350 (**82%**)

More people are better off without it: don't prescribe the drug!

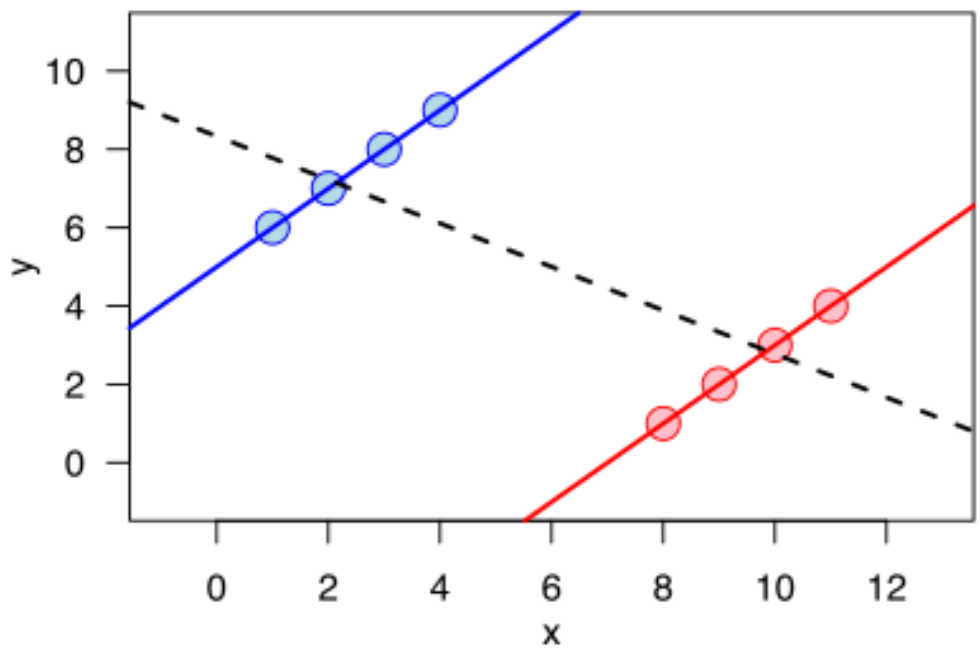
What if we know the gender of the patient?

	Drug	No drug
Men	81 out of 87 (**93%**)	234 out of 270 (87%)
Women	192 out of 263 (**73%**)	55 out of 80 (69%)
Total	273 out of 350 (78%)	289 out of 350 (**82%**)

Both men and women do better with the drug: prescribe it!

Simpson's paradox

- Trend **appears** in groups of data
- Trend **reverses** when groups are combined

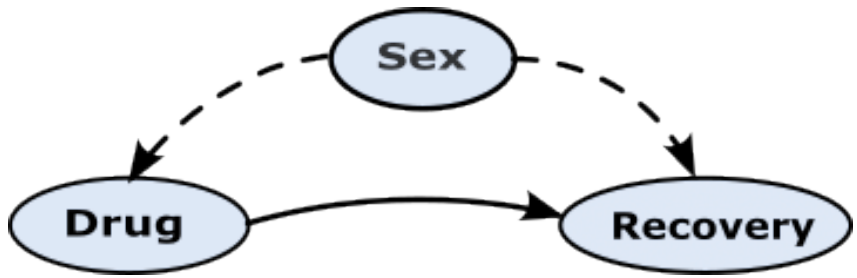


Simpson's paradox: resolution (in this case)

Suppose we know:

- Women are more likely to take the drug
- Estrogen has a negative effect on recovery

Then: sex is a confounder (common cause for drug/recovery). In this case, it makes sense to consider recovery rates separated by sex. This is referred to as **controlling** for sex.



Resolving Simpson's paradox requires knowledge about the mechanism that generated the data (cannot be done with just statistics alone)