Causal Inference:

Potential Outcomes and Confounders

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal

Determine whether a treatment has a causal effect

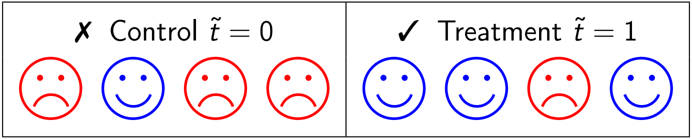
Does a new drug cure a disease?

Study to evaluate drug

Treatment $\tilde{t}$: If patient received drug $\tilde{t} = 1$, if not $\tilde{t} = 0$

Outcome $\tilde{y}$: If patient recovered $\tilde{y} = 1$, if not $\tilde{y} = 0$

Data:



$$0.25 = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) < p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) = 0.75$$

Is a new patient more likely to recover if treated?

Not necessarily!

Plan

Potential outcomes

Randomization

Confounding factors

Adjusting for confounders

Potential outcomes

Randomization

Confounding factors

Adjusting for confounders

Potential outcomes

Alternative scenarios defined for all patients

$\widetilde{\mathrm{po}}_0$: Outcome if patient is untreated
(*whether they receive the treatment or not!*)

$\widetilde{\mathrm{po}}_0 = 1$

$\widetilde{\mathrm{po}}_0 = 0$

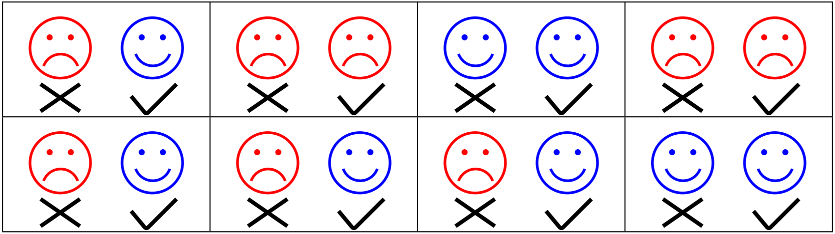
$\widetilde{\mathrm{po}}_1$: Outcome if patient is treated
(*whether they receive the treatment or not!*)

$\widetilde{\mathrm{po}}_1 = 1$

$\widetilde{\mathrm{po}}_1 = 0$

Causal effect

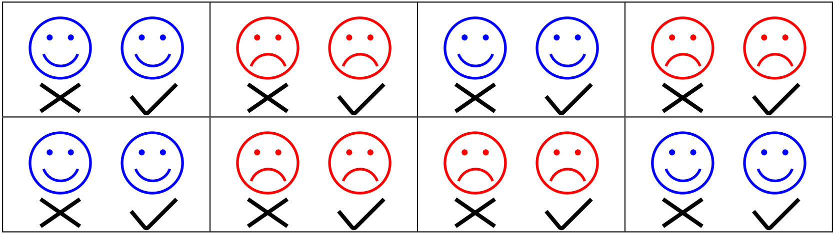
The drug has a causal effect if $\mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) < \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right)$



$$0.25 = \mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) < \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) = 0.75$$

No causal effect

The drug has a causal effect if $\mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = \mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right)$



$$0.5 = \mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = \mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right) = 0.5$$

Are we done here?

Fundamental problem of causal inference

Both potential outcomes are defined for every patient

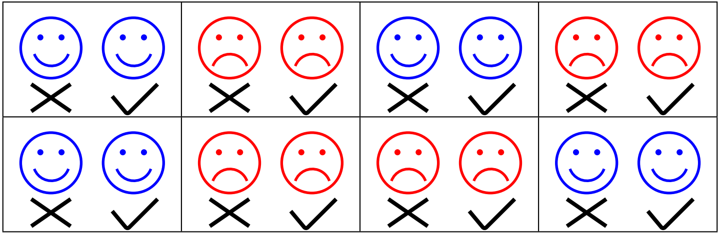
But we only see one of them at a time!

The treatment $\tilde{t}$ determines what outcome we observe

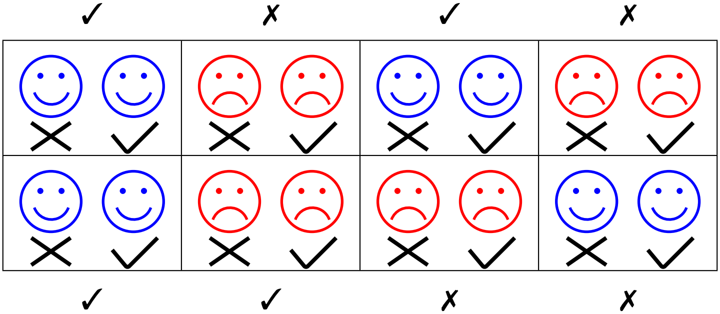
If $\tilde{t} = 0$, the observed outcome $\tilde{y} := \widetilde{\text{po}}_0$

If $\tilde{t} = 1$, the observed outcome $\tilde{y} := \widetilde{\text{po}}_1$

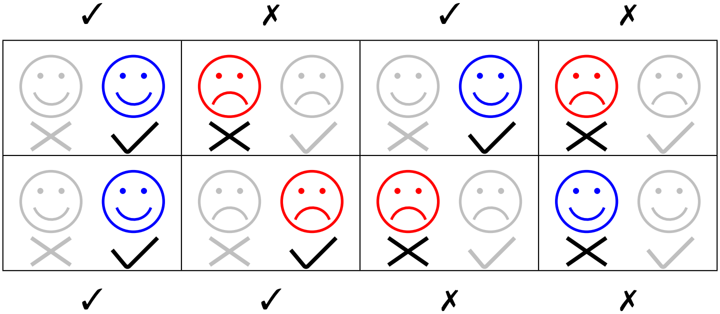
$$0.5 = \mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = \mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right) = 0.5$$



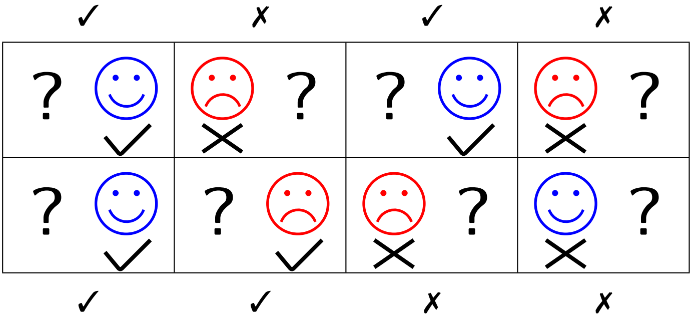
Treatment

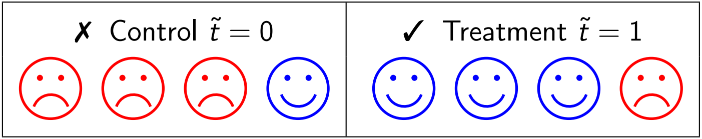


Unobserved potential outcomes are counterfactuals



Data





$$0.25 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0\right) < \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1\right) = 0.75$$

$$0.5 = \mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) = \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) = 0.5$$

Potential outcomes

Randomization

Confounding factors

Adjusting for confounders

Key question

When is the observed outcome representative of the potential outcomes?

$$p_{\widetilde{\text{po}}_0}(1) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) = p_{\widetilde{\text{po}}_0\,|\,\tilde{t}}(1\,|\,0)$$

$$p_{\widetilde{\text{po}}_1}(1) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) = p_{\widetilde{\text{po}}_1\,|\,\tilde{t}}(1\,|\,1)$$

When the treatment $\tilde{t}$ and the potential outcomes $\widetilde{\text{po}}_0$ / $\widetilde{\text{po}}_1$ are independent!

Wait a minute

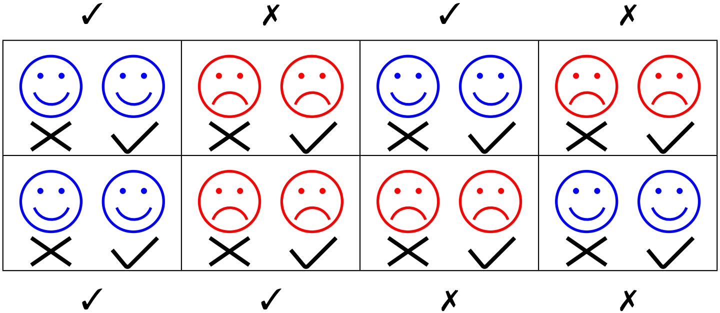
How can the outcome be independent from the treatment?

The potential outcomes need to be independent, not the observed outcome

Patients who are more likely to recover regardless of treatment, should not be more likely to receive treatment

The proportion of such patients in the control and treatment groups should be the same as in the whole population

Are the potential outcomes and the treatment independent?

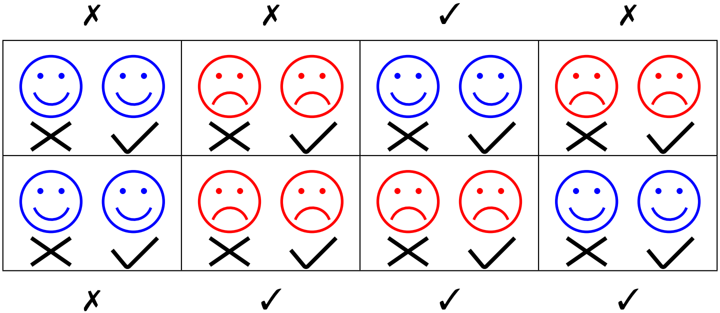


How can we guarantee independence?

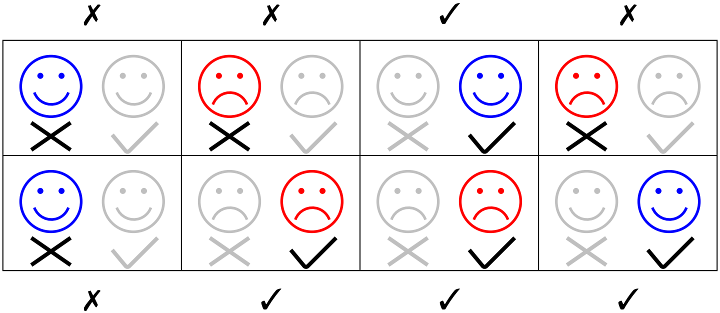
By assigning treatment at random

Randomization

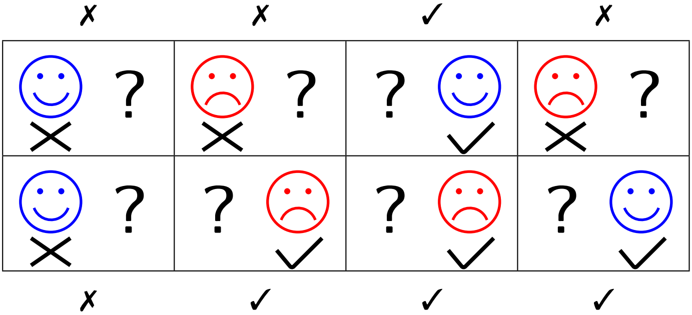
Flip a coin: tails, tails, heads, tails, tails, heads, heads, heads

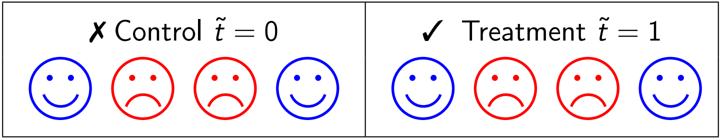


Observed potential outcomes



Data





$$0.5 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0\right) = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1\right) = 0.5$$

$$0.5 = \mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) = \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) = 0.5$$

Randomized controlled trial for COVID-19 vaccine

43,448 patients randomly divided into

- Treatment group of 21,720 patients: 8 cases (0.037%)
- Control group of 21,728 patients: 162 (0.746%)

Randomization guarantees that

$$\begin{aligned} 0.037\% &= \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1\right) \\ &= \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) \; < \mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) \\ &\qquad\qquad\qquad\quad = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0\right) = 0.746\% \end{aligned}$$

Potential outcomes

Randomization

Confounding factors

Adjusting for confounders

Confounding factor (a.k.a. confounder)

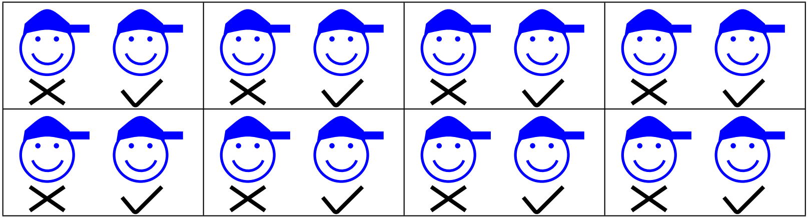
Governs dependence between treatment and potential outcomes

Drug example: Age $\tilde{a}$ is a confounder

Half the patients are young and half are old

$$p_{\tilde{a}}(\text{old}) = p_{\tilde{a}}(\text{young}) = 0.5$$

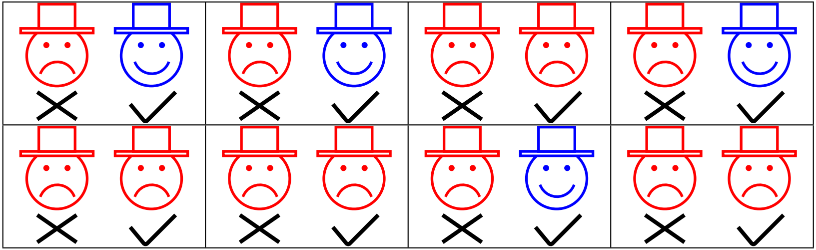
Young patients



$$1 = p_{\widetilde{\mathrm{po}}_0 \mid \tilde{a}}(1 \mid \text{young}) = p_{\widetilde{\mathrm{po}}_1 \mid \tilde{a}}(1 \mid \text{young}) = 1$$

No causal effect

Old patients



$$0 = p_{\widetilde{\text{po}_0} \mid \tilde{a}}(1 \mid \text{old}) < p_{\widetilde{\text{po}_1} \mid \tilde{a}}(1 \mid \text{old}) = 0.5$$

Causal effect

Overall causal effect

$$
\begin{aligned}
p_{\widetilde{\text{po}}_0}(1) &= p_{\tilde{a},\widetilde{\text{po}}_0}(\text{young}, 1) + p_{\tilde{a},\widetilde{\text{po}}_0}(\text{old}, 1) \\
&= p_{\tilde{a}}(\text{young}) p_{\widetilde{\text{po}}_0 \mid \tilde{a}}(1 \mid \text{young}) + p_{\tilde{a}}(\text{old}) p_{\widetilde{\text{po}}_0 \mid \tilde{a}}(1 \mid \text{old}) \\
&= 0.5 \cdot 1 + 0.5 \cdot 0 = 0.5
\end{aligned}
$$

$$
\begin{aligned}
p_{\widetilde{\text{po}}_1}(1) &= p_{\tilde{a}}(\text{young}) p_{\widetilde{\text{po}}_1 \mid \tilde{a}}(1 \mid \text{young}) + p_{\tilde{a}}(\text{old}) p_{\widetilde{\text{po}}_1 \mid \tilde{a}}(1 \mid \text{old}) \\
&= 0.5 \cdot 1 + 0.5 \cdot 0.5 = 0.75
\end{aligned}
$$

Causal effect: $0.5 = p_{\widetilde{\text{po}}_0}(1) < p_{\widetilde{\text{po}}_1}(1) = 0.75$

Treatment

The potential outcomes and age are dependent

If treatment and age are dependent

$$\begin{aligned}
\gamma_{\text{control}} &:= p_{\tilde{a}\,|\,\tilde{t}}(\text{young}\,|\,0) \neq p_{\tilde{a}}(\text{young}) = 0.5 \\
\gamma_{\text{treatment}} &:= p_{\tilde{a}\,|\,\tilde{t}}(\text{young}\,|\,1) \neq p_{\tilde{a}}(\text{young}) = 0.5
\end{aligned}$$

Then the treatment and potential outcomes are dependent!

Observed outcome

Assuming conditional independence between $\tilde{t}$ and $\widetilde{\text{po}}_0$ / $\widetilde{\text{po}}_1$ given age

$$
\begin{aligned}
& p_{\tilde{y} \mid \tilde{t}}(1 \mid 0) \\
& = p_{\widetilde{\text{po}}_{\mathbf{0}} \mid \tilde{t}}(1 \mid 0) \\
& = p_{\widetilde{\text{po}}_{\mathbf{0}}, \tilde{a} \mid \tilde{t}}(1, \text{young} \mid 0) + p_{\widetilde{\text{po}}_{\mathbf{0}}, \tilde{a} \mid \tilde{t}}(1, \text{old} \mid 0) \\
& = p_{\tilde{a} \mid \tilde{t}}(\text{young} \mid 0) p_{\widetilde{\text{po}}_{\mathbf{0}} \mid \tilde{a}, \tilde{t}}(1 \mid \text{young}, 0) + p_{\tilde{a} \mid \tilde{t}}(\text{old} \mid 0) p_{\widetilde{\text{po}}_{\mathbf{0}} \mid \tilde{a}, \tilde{t}}(1 \mid \text{old}, 0) \\
& = p_{\tilde{a} \mid \tilde{t}}(\text{young} \mid 0) p_{\widetilde{\text{po}}_{\mathbf{0}} \mid \tilde{a}}(1 \mid \text{young}) + p_{\tilde{a} \mid \tilde{t}}(\text{old} \mid 0) p_{\widetilde{\text{po}}_{\mathbf{0}} \mid \tilde{a}}(1 \mid \text{old}) \\
& = \gamma_{\text{control}}
\end{aligned}
$$

$$
\begin{aligned}
& p_{\tilde{y} \mid \tilde{t}}(1 \mid 1) \\
& = p_{\widetilde{\text{po}}_{\mathbf{1}} \mid \tilde{t}}(1 \mid 1) \\
& = p_{\tilde{a} \mid \tilde{t}}(\text{young} \mid 1) p_{\widetilde{\text{po}}_{\mathbf{1}} \mid \tilde{a}}(1 \mid \text{young}) + p_{\tilde{a} \mid \tilde{t}}(\text{old} \mid 1) p_{\widetilde{\text{po}}_{\mathbf{1}} \mid \tilde{a}}(1 \mid \text{old}) \\
& = \gamma_{\text{treatment}} + (1 - \gamma_{\text{treatment}}) 0.5
\end{aligned}
$$

Completely distorted by confounder

Observed outcome

Old patients don't want the drug:

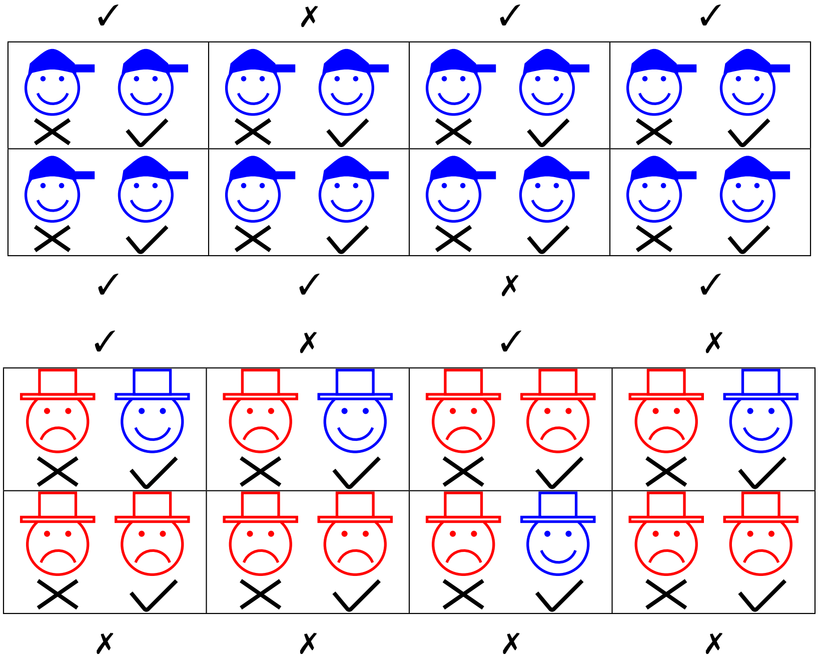
$\gamma_{\text{control}} := 0.25$, $\gamma_{\text{treatment}} = 0.75$

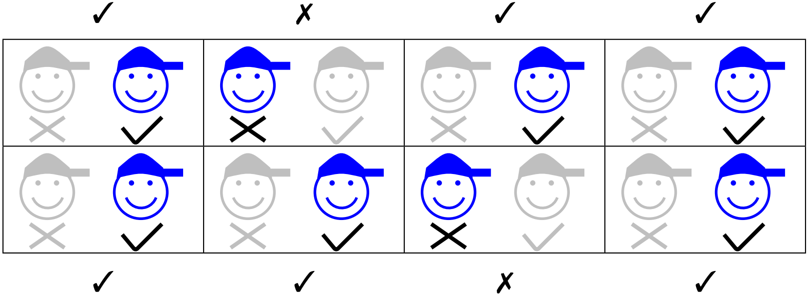
$$\begin{aligned} p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) &= \gamma_{\text{control}} \\ &= 0.25 \end{aligned}$$

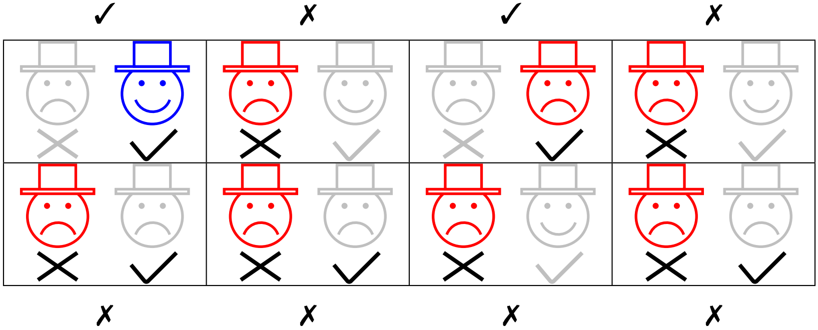
$$\begin{aligned} p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) &= \gamma_{\text{treatment}} + (1 - \gamma_{\text{treatment}})0.5 \\ &= 0.875 \end{aligned}$$

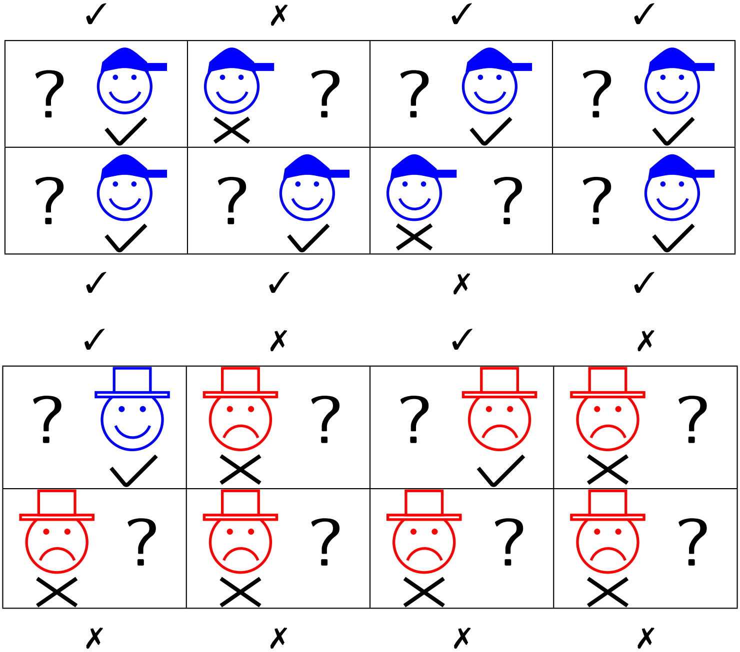
Observed effect: $0.25 = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) < p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) = 0.875$

Causal effect: $0.5 = p_{\widetilde{\text{po}}_0}(1) < p_{\widetilde{\text{po}}_1}(1) = 0.75$

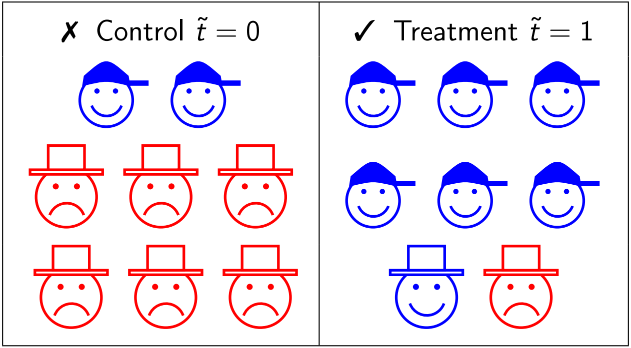








Data



$$0.25 = \mathrm{P}(\tilde{y} \mid \tilde{t} = 0) < \mathrm{P}(\tilde{y} \mid \tilde{t} = 1) = 0.875$$

Randomized controlled trial

Selection is random $\implies$ independent of age

$$\begin{aligned}\gamma_{\text{control}} &:= p_{\tilde{a}\,|\,\tilde{t}}(\text{young}\,|\,0) = p_{\tilde{a}}(\text{young}) = 0.5\\ \gamma_{\text{treatment}} &:= p_{\tilde{a}\,|\,\tilde{t}}(\text{young}\,|\,1) = p_{\tilde{a}}(\text{young}) = 0.5\end{aligned}$$

$$\begin{aligned}p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) &= \gamma_{\text{control}}\\ &= 0.5\end{aligned}$$

$$\begin{aligned}p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) &= \gamma_{\text{treatment}} + (1-\gamma_{\text{treatment}})0.5\\ &= 0.75\end{aligned}$$

Observed effect: $0.5 = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) < p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) = 0.75$

Causal effect: $0.5 = p_{\widetilde{\text{po}}_0}(1) < p_{\widetilde{\text{po}}_1}(1) = 0.75$

Potential outcomes

Randomization

Confounding factors

Adjusting for confounders

Problem solved?

Randomization neutralizes all confounders, even if they are unknown!

Problem: Randomization is very costly, and often *not possible*

Solution: Correct for known confounders

Adjusting for a confounder

If treatment is conditionally independent of the potential outcomes given the confounder, for each a

$$
\begin{aligned}
p_{\tilde{y}\,|\,\tilde{a},\tilde{t}}(1\,|\,a,0) &= p_{\widetilde{\text{po}}_0\,|\,\tilde{a},\tilde{t}}(1\,|\,a,0) \\
&= p_{\widetilde{\text{po}}_0\,|\,\tilde{a}}(1\,|\,a) \\
p_{\tilde{y}\,|\,\tilde{a},\tilde{t}}(1\,|\,a,1) &= p_{\widetilde{\text{po}}_1\,|\,\tilde{a}}(1\,|\,a)
\end{aligned}
$$

We observe true conditional causal effect!

Idea: Aggregate conditional causal effects

$$
\begin{aligned}
p_{\widetilde{\text{po}}_0}(1) &= \sum_{a\in\{\text{young},\text{old}\}} p_{\tilde{a}}(a)\, p_{\widetilde{\text{po}}_0\,|\,\tilde{a}}(1\,|\,a) \\
p_{\widetilde{\text{po}}_1}(1) &= \sum_{a\in\{\text{young},\text{old}\}} p_{\tilde{a}}(a)\, p_{\widetilde{\text{po}}_1\,|\,\tilde{a}}(1\,|\,a)
\end{aligned}
$$

Different to observed effect!

Compare

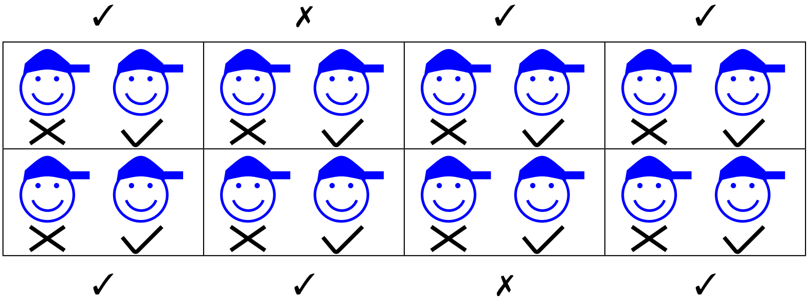
$$p_{\widetilde{\mathrm{po}}_0}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) p_{\widetilde{\mathrm{po}}_0 \,|\, \tilde{a}}(1 \,|\, a)$$

$$p_{\widetilde{\mathrm{po}}_1}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) p_{\widetilde{\mathrm{po}}_1 \,|\, \tilde{a}}(1 \,|\, a)$$

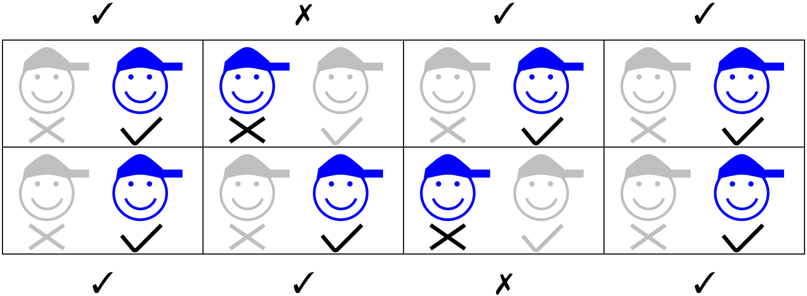
to

$$\begin{aligned} p_{\tilde{y} \,|\, \tilde{t}}(1 \,|\, 0) &= \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a} \,|\, \tilde{t}}(a \,|\, 0) p_{\tilde{y} \,|\, \tilde{a},\tilde{t}}(1 \,|\, a, 0) \\ &= \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a} \,|\, \tilde{t}}(a \,|\, 0) p_{\widetilde{\mathrm{po}}_0 \,|\, \tilde{a}}(1 \,|\, a) \\ p_{\tilde{y} \,|\, \tilde{t}}(1 \,|\, 1) &= \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a} \,|\, \tilde{t}}(a \,|\, 1) p_{\widetilde{\mathrm{po}}_1 \,|\, \tilde{a}}(1 \,|\, a) \end{aligned}$$

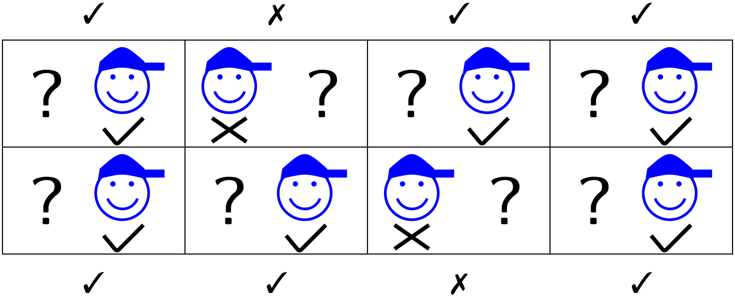
Young patients



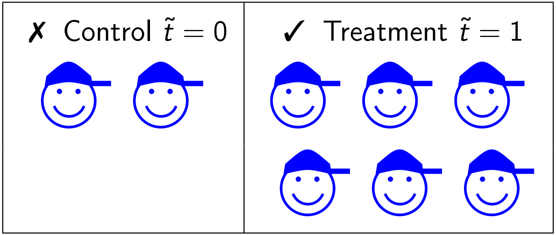
Young patients



Young patients



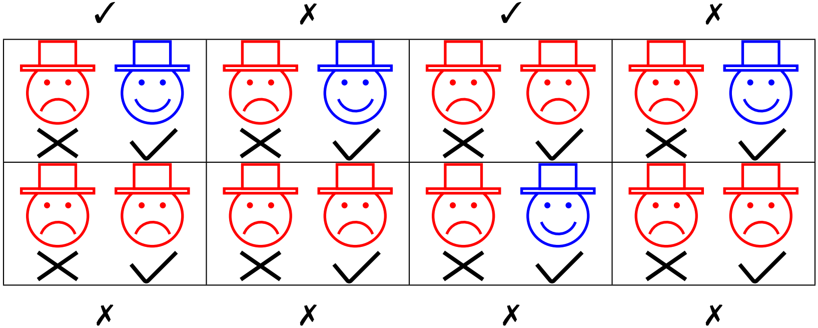
Stratified data



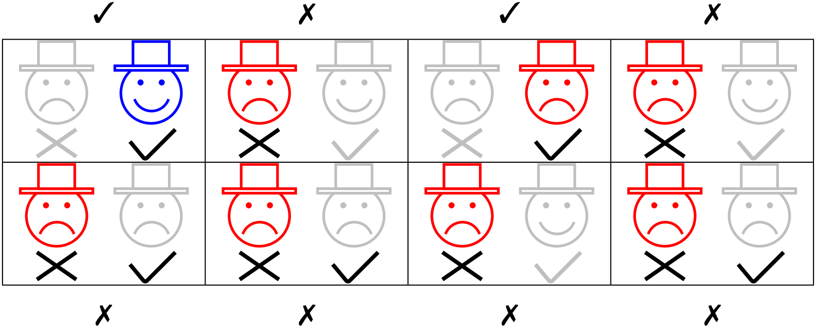
$$1 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = \text{young}\right) = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = \text{young}\right) = 1$$

$$1 = p_{\widetilde{\mathrm{po}}_0 \mid \tilde{a}}(1 \mid \text{young}) = p_{\widetilde{\mathrm{po}}_1 \mid \tilde{a}}(1 \mid \text{young}) = 1$$

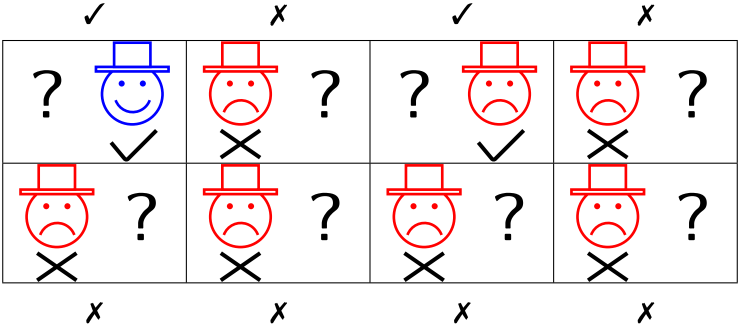
Old patients



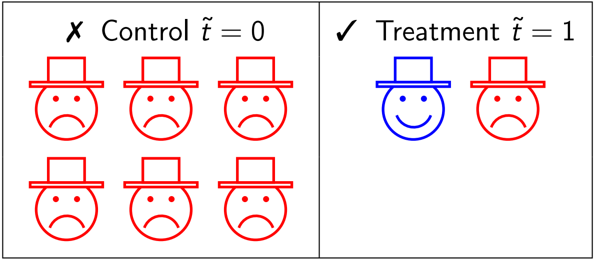
Old patients



Old patients



Stratified data



$$0 = \mathrm{P}\left(\tilde{y} \,|\, \tilde{t} = 0, \tilde{a} = \text{old}\right) < \mathrm{P}\left(\tilde{y} \,|\, \tilde{t} = 1, \tilde{a} = \text{old}\right) = 0.5$$

$$0 = p_{\widetilde{\text{po}}_0 \,|\, \tilde{a}}(1 \,|\, \text{old}) < p_{\widetilde{\text{po}}_1 \,|\, \tilde{a}}(1 \,|\, \text{old}) = 0.5$$

Adjusted causal effect

$$\sum_{a \in \{\text{young,old}\}} p_{\tilde{a}}(a) p_{\widetilde{\text{po}}_0 \,|\, \tilde{a}}(1 \,|\, a)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0 = 0.5 = p_{\widetilde{\text{po}}_0}(1)$$

$$\sum_{a \in \{\text{young,old}\}} p_{\tilde{a}}(a) p_{\widetilde{\text{po}}_1 \,|\, \tilde{a}}(1 \,|\, a)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0.5 = 0.75 = p_{\widetilde{\text{po}}_1}(1)$$

Compare to observed effect

$$p_{\tilde{y} \,|\, \tilde{t}}(1 \,|\, 0) = \sum_{a \in \{\text{young,old}\}} p_{\tilde{a} \,|\, \tilde{t}}(a \,|\, 0) p_{\widetilde{\text{po}}_0 \,|\, \tilde{a}}(1 \,|\, a)$$

$$= 0.25 \cdot 1 + 0.75 \cdot 0 = 0.25$$

$$p_{\tilde{y} \,|\, \tilde{t}}(1 \,|\, 1) = \sum_{a \in \{\text{young,old}\}} p_{\tilde{a} \,|\, \tilde{t}}(a \,|\, 1) p_{\widetilde{\text{po}}_1 \,|\, \tilde{a}}(1 \,|\, a)$$

$$= 0.75 \cdot 1 + 0.25 \cdot 5 = 0.875$$

Warning: Conditional independence assumption is crucial!

What have we learned?

Definition of potential outcomes

Why randomization allows us to perform causal inference

Confounders can completely distort observed effect

How to adjust for confounders (and when it works)