

# Potential outcomes and counterfactuals

17.04.21

10:06

## Notation:

- treatment  $A$  (sometimes called "exposure") (e.g.  $A=1$  if receive active drug; if placebo  $A=0$ )
- outcome  $Y$  (e.g.  $Y=1$  if develop cardiov. disease within 2yrs; else  $Y=0$ )  
(e.g. time until death)

- potential outcomes  $Y^a$ : outcome would be observed if  $A=a$

NO DATA YET COLLECTED  $\rightarrow$  each person has 2 potential outcomes if  $|A|=2$ :  $Y^0, Y^1$

- counterfactuals: outcome that would have been observed, had the treatment been different

DATA ALREADY BEEN COLLECTED

$\rightarrow$  e.g. if in data we see that  $A=1$ ; then the observed outcome  $Y=Y^1$   
counterfactual outcome is  $Y^{A=0}=Y^0$ . &

if \_\_\_\_\_ " \_\_\_\_\_  $A=0$ ; then the observed outcome  $Y=Y^0$   
\_\_\_\_\_ " \_\_\_\_\_ is  $Y^{A=1}=Y^1$

! observed outcome vs. counterfactual outcome !

## Difference btw. outcome & potential outcome:

(1) Before the treatment is made, any/every outcome is a potential outcome:  $Y^0$  and  $Y^1$

(2) After the study, there is one observed (factual) outcome,  $Y=Y^A$ , and one counterfactual outcome,  $Y=Y^{1-A}$ .

$\Rightarrow$  <sup>1</sup> counterfactual outcomes  $Y^0, Y^1$  are typically assumed to be the same as <sup>2</sup> potential outcomes  $Y^0, Y^1$ . (these two terms are used interchangeably)