Causal estimation

David Hajage david.hajage@aphp.fr

3 Juin 2025

SANTÉ SORBONNE UNIVERSITÉ

Sommaire

SANTÉ
SORBONNE
UNIVERSITÉ

1

Introduction

"Everyday" Causality

When I drop my pen, it falls.

- **Necessary** association: the pen always falls if I drop it.
- **Specific** association: the pen falls only if I drop it.
- **Immediate** association: the pen falls immediately if I drop it.

$\Rightarrow$ Causal relationship is simple to demonstrate.

Causality in Epidemiology

Now imagine:

- When I drop the pen, sometimes it falls, sometimes it doesn't. When it falls, it only does so 5 years after being dropped.
- When I don't drop the pen, it still sometimes falls (but less often).

$\Rightarrow$ Dropping the pen changes the **probability** that the pen will fall in 5 years.

Causality in Epidemiology

Now imagine:

- When I drop the pen, sometimes it falls, sometimes it doesn't. When it falls, it only does so 5 years after being dropped.
- When I don't drop the pen, it still sometimes falls (but less often).

$\Rightarrow$ Dropping the pen changes the **probability** that the pen will fall in 5 years.

Most causal relationships in epidemiology are neither necessary, specific, nor immediate.

Causal relationship? Yes, but more difficult to demonstrate.

SANTÉ SORBONNE UNIVERSITÉ

2

Potential counterfactual outcome, and causal inference

SANTÉ SORBONNE UNIVERSITÉ

2

Potential counterfactual outcome, and causal inference

Presentation of the sample

First name	A	Y
Michael	1	0
Danica	1	0
Mitchell	0	0
Joanna	0	0
Blanca	0	0
Simon	1	1
Rojesh	1	0
Hamdi	1	0
Rahma	0	1
Matthew	0	0
Safar	0	0
Samuel	1	1
Cheyenne	1	0
Joshua	1	0

Presentation of the sample

First name	A	Y
Michael	1	0
Danica	1	0
Mitchell	0	0
Joanna	0	0
Blanca	0	0
Simon	1	1
Rojesh	1	0
Hamdi	1	0
Rahma	0	1
Matthew	0	0
Safar	0	0
Samuel	1	1
Cheyenne	1	0
Joshua	1	0

A: medication taken ($A = 1$) or not taken ($A = 0$)

Y: disease cured ($Y = 1$) or not cured ($Y = 0$)

Simon

First name	A	Y
Michael	1	0
Danica	1	0
Mitchell	0	0
Joanna	0	0
Blanca	0	0
Simon	1	1
Rojesh	1	0
Hamdi	1	0
Rahma	0	1
Matthew	0	0
Safar	0	0
Samuel	1	1
Cheyenne	1	0
Joshua	1	0

- Simon took the medication and the disease was cured.
- Did the medication **cause** the cure? We don't know. Why?

Simon

First name	A	Y
Michael	1	0
Danica	1	0
Mitchell	0	0
Joanna	0	0
Blanca	0	0
Simon	1	1
Rojesh	1	0
Hamdi	1	0
Rahma	0	1
Matthew	0	0
Safar	0	0
Samuel	1	1
Cheyenne	1	0
Joshua	1	0

- Simon took the medication and the disease was cured.
- Did the medication **cause** the cure? We don't know. Why?

- Because we need to know what Simon would become **without** the medication.

Potential counterfactual outcome

- A denotes the observed treatment
- Y denotes the observed outcome
- $Y^{A=a}$ denotes the potential outcome with the therapeutic alternative $A = a$:
  - $Y^{A=0} = Y^0$ is the outcome we would observe if the medication is not taken
  - $Y^{A=1} = Y^1$ is the outcome we would observe if the medication is taken
- Only one of the values Y^0 and Y^1 is observed:
  - if $A = 0$, Y^0 is observed
  - if $A = 1$, Y^1 is observed
- The outcome we do not observe is called the counterfactual outcome

Potential counterfactual outcome

- A denotes the observed treatment
- Y denotes the observed outcome
- $Y^{A=a}$ denotes the potential outcome with the therapeutic alternative $A = a$:
  - $Y^{A=0} = Y^0$ is the outcome we would observe if the medication is not taken
  - $Y^{A=1} = Y^1$ is the outcome we would observe if the medication is taken
- Only one of the values Y^0 and Y^1 is observed:
  - if $A = 0$, Y^0 is observed
  - if $A = 1$, Y^1 is observed
- The outcome we do not observe is called the counterfactual outcome

- Suppose we could observe the unobservable...

Potential counterfactual outcome

First name	Y0	Y1
Michael	0	0
Danica	1	0
Mitchell	0	1
Joanna	0	0
Blanca	0	0
Simon	0	1
Rojesh	1	0
Hamdi	1	0
Rahma	1	1
Matthew	0	0
Safar	0	1
Samuel	1	1
Cheyenne	0	0
Joshua	0	0

- Simon took the medication and the disease was cured: $Y = Y^1 = 1$
- If Simon had not taken the medication, the disease would not have been cured: $Y^0 = 0$
- For Simon, the medication has a causal effect
- What about the others?

Potential counterfactual outcome

First name	Y0	Y1	Causal effect
Michael	0	0	No
Danica	1	0	Yes, harmful
Mitchell	0	1	Yes, favorable
Joanna	0	0	No
Blanca	0	0	No
Simon	0	1	Yes, favorable
Rojesh	1	0	Yes, harmful
Hamdi	1	0	Yes, harmful
Rahma	1	1	No
Matthew	0	0	No
Safar	0	1	Yes, favorable
Samuel	1	1	No
Cheyenne	0	0	No
Joshua	0	0	No

- The individual causal effect is defined by $Y^1 - Y^0$
- It is not necessarily the same for all individuals
  (because they are not pens being dropped)

Potential counterfactual outcome

First name	Y0	Y1	A	Y
Michael	0	0	1	0
Danica	1	0	1	0
Mitchell	0	1	0	0
Joanna	0	0	0	0
Blanca	0	0	0	0
Simon	0	1	1	1
Rojesh	1	0	1	0
Hamdi	1	0	1	0
Rahma	1	1	0	1
Matthew	0	0	0	0
Safar	0	1	0	0
Samuel	1	1	1	1
Cheyenne	0	0	1	0
Joshua	0	0	1	0

In reality, we never observe both Y^0 AND Y^1, but rather A and Y

$\Rightarrow$ The individual causal effect is unknown

Let's try to be less ambitious then

SANTÉ SORBONNE UNIVERSITÉ

2

Potential counterfactual outcome, and causal inference

Population causal effect

A 'less ambitious' causal effect is the population, or marginal, causal effect:

$$E(Y^1) - E(Y^0)$$

In the case of a binary outcome:

$$E(Y^1) - E(Y^0) = Pr(Y^1 = 1) - Pr(Y^0 = 1)$$

Potential counterfactual outcome

First name	Y0	Y1	Causal effect
Michael	0	0	No
Danica	1	0	Yes, harmful
Mitchell	0	1	Yes, favorable
Joanna	0	0	No
Blanca	0	0	No
Simon	0	1	Yes, favorable
Rojesh	1	0	Yes, harmful
Hamdi	1	0	Yes, harmful
Rahma	1	1	No
Matthew	0	0	No
Safar	0	1	Yes, favorable
Samuel	1	1	No
Cheyenne	0	0	No
Joshua	0	0	No

We have :

- $Pr(Y^1 = 1) = 5/14$
- $Pr(Y^0 = 1) = 5/14$
- $Pr(Y^1 = 1) - Pr(Y^0 = 1) = 0$

$\Rightarrow$ no population level causal effect

Population causal effect

- 'Less ambitious', perhaps, but...
- We cannot calculate $Pr(Y^1 = 1) - Pr(Y^0 = 1)$ for the same reasons we cannot calculate $Y^1 - Y^0$
- This is the fundamental problem of causal inference

'Doing' causal inference means seeking to estimate quantities from observed data, which serve as reasonable substitutes for unobservable quantities (such as $Pr(Y^1 = 1)$ and $Pr(Y^0 = 1)$, which involve counterfactual outcomes).

Crude estimation of the population causal effect

- What could be substitutes for $Pr(Y^1 = 1)$ and $Pr(Y^0 = 1)$?
- The most obvious substitutes are $Pr(Y = 1|A = 1)$ and $Pr(Y = 1|A = 0)$
- These quantities serve as good substitutes if those who took or did not take the medication are exchangeable $\Rightarrow$ Y^0 and Y^1 independent of A
- This would notably be the case if medication intake were random $\Rightarrow$ This is why randomized controlled trials are the gold standard for therapeutic evaluation

Crude estimation of the population causal effect

First name	Y0	Y1	A	Y
Michael	0	0	1	0
Danica	1	0	1	0
Mitchell	0	1	0	0
Joanna	0	0	0	0
Blanca	0	0	0	0
Simon	0	1	1	1
Rojesh	1	0	1	0
Hamdi	1	0	1	0
Rahma	1	1	0	1
Matthew	0	0	0	0
Safar	0	1	0	0
Samuel	1	1	1	1
Cheyenne	0	0	1	0
Joshua	0	0	1	0

8 individuals out of 14 took the medication.

- We 'know' that $Pr(Y^1 = 1) = 5/14$
- We estimate that $Pr(Y = 1|A = 1) = 2/8$

Likewise:

- We 'know' that $Pr(Y^0 = 1) = 5/14$
- We estimate that $Pr(Y = 1|A = 0) = 1/6$

Finally:

- We know that $Pr(Y^1 = 1) - Pr(Y^0 = 1) = 0$
- We estimate that $Pr(Y = 1|A = 1) - Pr(Y = 1|A = 0) = 1/12$

Crude estimation of the population causal effect

- $Pr(Y = 1|A = 1) - Pr(Y = 1|A = 0) = 1/12$
- If association implied causality, then we would conclude that the medication has a favorable effect.
- Besides chance (*i.e.*, sampling fluctuations), what could explain the observed association, other than a causal effect?

Crude estimation of the population causal effect

- $Pr(Y = 1|A = 1) - Pr(Y = 1|A = 0) = 1/12$
- If association implied causality, then we would conclude that the medication has a favorable effect.
- Besides chance (*i.e.*, sampling fluctuations), what could explain the observed association, other than a causal effect?
  lack of exchangeability

SANTÉ SORBONNE UNIVERSITÉ

2

Potential counterfactual outcome, and causal inference

Initial severity

First name	A	Y	L
Michael	1	0	1
Danica	1	0	0
Mitchell	0	0	0
Joanna	0	0	1
Blanca	0	0	0
Simon	1	1	1
Rojesh	1	0	1
Hamdi	1	0	0
Rahma	0	1	1
Matthew	0	0	1
Safar	0	0	0
Samuel	1	1	1
Cheyenne	1	0	1
Joshua	1	0	1

L : Initial severity of the disease ($L = 1$ in case of severity)

Exchangeability?

First name	A	Y	L
Michael	1	0	1
Danica	1	0	0
Mitchell	0	0	0
Joanna	0	0	1
Blanca	0	0	0
Simon	1	1	1
Rojesh	1	0	1
Hamdi	1	0	0
Rahma	0	1	1
Matthew	0	0	1
Safar	0	0	0
Samuel	1	1	1
Cheyenne	1	0	1
Joshua	1	0	1

- Among individuals $L = 1$:
  $Pr(A = 1|L = 1) = 6/9$
- Among individuals $L = 0$:
  $Pr(A = 1|L = 0) = 2/5$

Thus, individuals with a severe initial disease are more likely to take the medication.

Exchangeability?

First name	Y0	Y1	A	Y	L
Michael	0	0	1	0	1
Danica	1	0	1	0	0
Mitchell	0	1	0	0	0
Joanna	0	0	0	0	1
Blanca	0	0	0	0	0
Simon	0	1	1	1	1
Rojesh	1	0	1	0	1
Hamdi	1	0	1	0	0
Rahma	1	1	0	1	1
Matthew	0	0	0	0	1
Safar	0	1	0	0	0
Samuel	1	1	1	1	1
Cheyenne	0	0	1	0	1
Joshua	0	0	1	0	1

- Among individuals $L = 1$:
  $Pr(Y^0 = 1|L = 1) = 3/9$
- Among individuals $L = 0$:
  $Pr(Y^0 = 1|L = 0) = 2/5$

So, in the absence of medication, individuals with a severe initial disease are less likely to be cured.
(and the same applies with Y^1)

Initial severity

What does L represent in the relationship between A and Y?

Initial severity

What does L represent in the relationship between A and Y?

$\Rightarrow$ a confounding factor

Stratification on initial severity

First name	A	Y	L
Michael	1	0	1
Danica	1	0	0
Mitchell	0	0	0
Joanna	0	0	1
Blanca	0	0	0
Simon	1	1	1
Rojesh	1	0	1
Hamdi	1	0	0
Rahma	0	1	1
Matthew	0	0	1
Safar	0	0	0
Samuel	1	1	1
Cheyenne	1	0	1
Joshua	1	0	1

Among individuals $L = 1$:

- $Pr(Y = 1|A = 1 \cap L = 1) = 2/6$
- $Pr(Y = 1|A = 0 \cap L = 1) = 1/3$

Among individuals $L = 0$:

- $Pr(Y = 1|A = 1 \cap L = 0) = 0/2$
- $Pr(Y = 1|A = 0 \cap L = 0) = 0/3$

So, within each stratum of severity, we find no association between medication intake and favorable outcome

Stratification on initial severity

- We have seen that individuals with $A = 1$ and $A = 0$ are not exchangeable, due to a severity imbalance
- But, let's assume that within each severity stratum, individuals with $A = 1$ and $A = 0$ are exchangeable
- This is called the conditional exchangeability assumption (according to L)
- In this situation, the association between A and Y has a causal interpretation within each severity stratum L
- Under the conditional exchangeability assumption (conditionally on L), we find no causal effect of medication intake on recovery in each of the strata

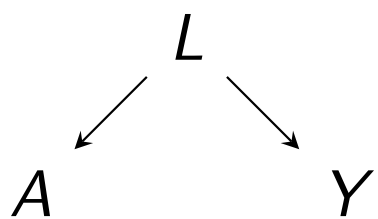
SANTÉ SORBONNE UNIVERSITÉ

2

Potential counterfactual outcome, and causal inference

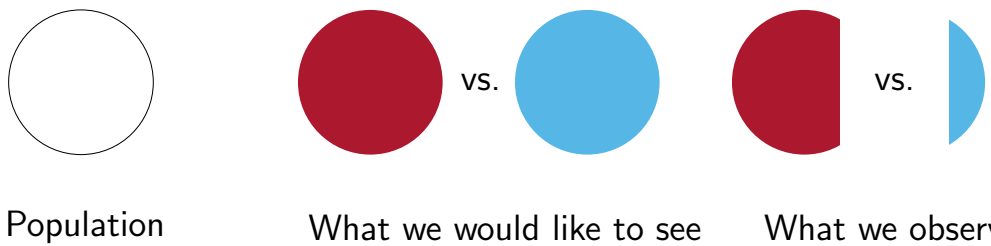
Summary I

- In epidemiology, cause-effect relationships are generally neither necessary, nor specific (nor immediate)
- Counterfactual reasoning allows defining what is called the causal effect
- At the individual level, the causal effect is unknown
- At the population level, the causal (marginal) effect is unknown, but can be estimated under certain assumptions
- In the absence of exchangeability (*i.e.*, outside of a randomized controlled trial), an association cannot be interpreted causally
- In our example, the apparent association was only related to the open path between A and Y passing through L



Summary II

- This leads us to define the causal effect in terms of **counterfactual interventions** For example: what would be the average difference in a population if we gave *everyone* intervention $A = 1$ versus *everyone* intervention $A = 0$?



- The connection with a randomized controlled trial is direct: an observational analysis should seek to estimate a quantity similar to what would have been estimated in a randomized trial (referred to as *trial emulation*)
- Counterfactual notations such as $Y^{A=a}$ are just a way of expressing what we seek to identify

Summary III

- The conditional exchangeability assumption is one of the fundamental assumptions for interpreting associations causally in observational settings
- In practice, there is rarely only one confounding factor
- Causal inference in observational settings therefore requires identifying a group of variables (L_1, L_2, L_3, etc.) for which there is conditional exchangeability
- This requires specialized knowledge about the diseases under study
- Causal diagrams (cf. Charles Assaad's course) are a valuable tool for identifying these variables (or questioning conditional exchangeability)

What's Next

- Stratification is a simple and effective method for accounting for a categorical confounding factor
- It quickly becomes impractical, if not unusable, when trying to account for many factors, or continuous factors
- In the following sections, we will see a method (standardization) that can be seen as a direct extension of the counterfactual reasoning
- We will then introduce methods based on propensity score

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

- Fundamental Assumptions
- Marginal effect
- From stratification to regression
- From regression to standardization
- Causal effect in treated individuals
- Standardization: general method
- Application with R
- Summary

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Fundamental assumptions of causal inference in observational settings

1) Absence of interference
2) Consistency
3) Conditional exchangeability
4) Positivity

Absence of interference

The exposure of one individual does not affect the potential outcome of another individual.

Example of situations where the assumption may be violated:
Situations where the exposure is ‘shared’

- Vaccination: the vaccination of one individual may influence the likelihood of infection in another individual

Consistency

The potential outcome of an individual, if they were to receive the observed exposure, is equal to the observed outcome[1]:

$$Y^{A=a} = Y \text{ for individuals where } A = a$$

Example of situations where the assumption may be violated:
Situations where the exposure has multiple 'versions'

- Obesity: obesity vs no obesity. But is the potential outcome the same if BMI = 40 vs BMI = 50?
- Surgery: surgery vs no surgery. But is the potential outcome the same if experienced surgeon vs beginner surgeon?

This highlights the importance of well defined interventions[2]

[1] Equivalent to accept that the data are generated by a structural causal model

[2] See "Does water kill?" https://doi.org/10.1016%2Fj.annepidem.2016.08.016

Conditional exchangeability

1) **Exchangeability** (in randomized trials)

Whether actually exposed or not, the potential outcome remains unchanged
$\rightarrow Y^0$ and Y^1 are independent of A: $Y^a \perp A \; \forall \; a \in (0,1)$
Then,

$$E(Y^a|A=1) = E(Y^a|A=0) \; \forall \; a \in (0,1)$$

2) **Conditional exchangeability** (or ignorability[3]) (in observational studies)

$\rightarrow Y^0$ and Y^1 are independent of A conditional on L:
$Y^a \perp A|L \; \forall \; a \in (0,1)$
Then,

$$E(Y^a|A=1, L=l) = E(Y^a|A=0, L=l) \; \forall \; a \in (0,1) \; \forall \; l$$

[3] Equivalent to the assumption that there exists a set of variables that satisfies the backdoor criterion

Positivity

All individuals in the population of interest have a non-zero probability of being exposed and unexposed.

$$Pr(A = a|L = l) > 0 \; \forall \; a \in (0, 1) \; \forall \; l$$

Will be further discussed later.

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Average Treatment Effect (ATE)

- We have already defined the population marginal effect by $E(Y^1) - E(Y^0)$. It is a difference of two means
- From now, we refer to it as *average treatment effect* or ATE
- In the case where Y is binary, $E(Y^a)$ is a probability, so ATE is a *risk difference*: $RD = Pr(Y^1 = 1) - Pr(Y^0 = 1)$
- In the case where Y is binary, other measures of association are frequently used, including:
  Risk ratio: $RR = E(Y^1)/E(Y^0) = Pr(Y^1 = 1)/Pr(Y^0 = 1)$
  Odds ratio: $OR = \frac{E(Y^1)}{1-E(Y^1)} / \frac{E(Y^0)}{1-E(Y^0)} = \frac{Pr(Y^1=1)}{1-Pr(Y^1=1)} / \frac{Pr(Y^0=1)}{1-Pr(Y^0=1)}$

In the following, we will mainly focus on the risk difference.

Identifiability of the marginal causal effect in a randomized trial

Why does randomization work?

$$ATE = E(Y^1) - E(Y^0)$$

1) Exchangeability
   $\Rightarrow E(Y^a) = E(Y^a|A = a) \; \forall \; a \in (0, 1)$
   Both groups ($A = 1$ and $A = 0$) are representative of the same (entire) population
2) Consistency
   $\Rightarrow E(Y^a|A = a) = E(Y|A = a) \; \forall \; a \in (0, 1)$

Thus:

$$ATE = E(Y|A = 1) - E(Y|A = 0)$$

Causal effect **in the entire** population, directly estimable by comparing the average outcome of each group.

Identifiability of the marginal causal effect in an observational study

1) Conditional Exchangeability
   $\Rightarrow E(Y^a|L=l) = E(Y^a|A=a, L=l)\ \forall\ a \in (0,1)\ \forall\ l$
   Both groups ($A=1$ and $A=0$) are representative of the same population within each stratum l
2) Consistency
   $\Rightarrow E(Y^a|A=a, L=l) = E(Y|A=a, L=l)\ \forall\ a \in (0,1)\ \forall\ l$

Thus:

$$ATE_l = E(Y|A=1, L=l) - E(Y|A=0, L=l)$$

Causal effect **in each stratum** defined by the value $L=l$

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

In a randomized trial

- We have already shown that in an experimental situation, thus in the absence of confounders:

$$ATE = E(Y|A = 1) - E(Y|A = 0) = \beta$$

- Rearranging terms:

$$E(Y|A = 1) = E(Y|A = 0) + \beta$$

- Noting $a \in (0, 1)$:

$$E(Y|A = a) = E(Y|A = 0) + \beta a$$
$$E(Y|A = a) = \alpha + \beta a$$

- We obtain a simple linear regression model on Y:
  β is the ATE, and $\widehat{\beta}$ is its estimate

In an observational study

- In an observational study, with a confounding factor L:

$$ATE_l = E(Y|A = 1, L = l) - E(Y|A = 0, L = l) = \beta_l$$

- Rearranging terms:

$$E(Y|A = a, L = l) = E(Y|A = 0, L = l) + \beta_l a$$
$$E(Y|A = a, L = l) = (\alpha + \gamma_l) + (\beta + \delta_l)a$$

$\alpha + \gamma_l$: stratum-specific intercept (γ_l depends on the value l)
$\beta + \delta_l$: ATE_l, stratum-specific effect (δ_l depends on the value l)

- We obtain a linear regression model adjusted on L, including an interaction term between A and L
- ATE_l is estimated by $\widehat{\beta} + \widehat{\delta_l}$
- If L is not an effect modifier (if the interaction $\delta_l = 0$), then $ATE_l = ATE = \beta$ (causal effect identical in all strata)

In an observational study

- The previous regression model allows estimating the stratum-specific effect ATE_l, which we call conditional effect of exposure A (conditional on the value l)
- To obtain it, we estimate a linear regression model including an interaction term between A and L:
  $E(Y|A, L) = \alpha + \gamma L + \beta A + \delta A \times L$
- This model generalizes to multiple confounding factors:
  $E(Y|A = a, L_1, L_2, \dots) =$
  $\alpha + \gamma_1 L_1 + \gamma_2 L_2 + \dots + \beta A + \delta_1 A \times L_1 + \delta_2 A \times L_2 + \dots$
- The validity of conditional estimates depends on:
  - the validity of the fundamental assumptions of causal inference
  - the proper specification of the model

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Marginal effect vs conditional Effect

So far we have seen:

- randomized trials, which allow estimating the population-level marginal causal effect ATE
- observational studies, which allow estimating *the* stratum-specific (or conditional) causal effects ATE_l

Marginal effect vs conditional Effect

So far we have seen:

- randomized trials, which allow estimating the population-level marginal causal effect ATE
- observational studies, which allow estimating *the* stratum-specific (or conditional) causal effects ATE_l

How can we estimate the population-level marginal effect in an observational study?

Standardization I

- The population-level marginal effect is the weighted average of the conditional effects:

$$ATE = \sum_l ATE_l \times Pr(L = l)$$

- The weights are $Pr(L = l)$, directly reflecting the distribution of L in the population
- Example:
  - effect in men: $ATE_h = 1$
  - effect in women: $ATE_f = 2$
  - if $2/3$ of the population are men:
    $ATE = 1 \times \frac{2}{3} + 2 \times \frac{1}{3} = \frac{4}{3} \approx 1.33$
- Again, if L is not an effect modifier, $ATE = ATE_l = \beta$

  Example:
  - $ATE_h = ATE_f = 2$, still $2/3$ of the population are men
  - $ATE = 2 \times \frac{2}{3} + 2 \times \frac{1}{3} = \frac{6}{3} = 2$

Standardization II

- The estimation of ATE is obtained by replacing all these quantities with their estimates in the analyzed sample:

$$\widehat{ATE} = \sum_l \widehat{ATE}_l \times \widehat{Pr}(L = l) = \sum_l (\widehat{\beta} + \widehat{\delta}_l) \times \widehat{Pr}(L = l)$$

- As $\widehat{ATE}_l = \hat{E}(Y^1) - \hat{E}(Y^0) = \hat{E}(Y|A = 0, L = l) - \hat{E}(Y|A = 1, L = l)$, the ATE could be obtained from the difference of average counterfactual outcome with either of the treatment alternatives:

$$\hat{E}(Y^a) = \sum_l \hat{E}(Y|A = a, L = l)\widehat{Pr}(L = l) \; \forall \; a \in (0, 1)$$

- This process is called standardization (or marginalization) (or G-computation)

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

ATE vs ATT

- ATE is the marginal causal effect in the entire population:
  $ATE = E(Y^1) - E(Y^0)$
- Literally, it represents the difference in average outcomes if the population were entirely exposed vs if the population were not exposed to the treatment of interest
- Another interesting question: what would be the difference in average outcomes of the exposed individuals if they hadn't been exposed?
- This is called the average treatment effect on the treated (ATT):

$$ATT = E(Y^1|A = 1) - E(Y^0|A = 1)$$

(in a randomized controlled trial, $ATE = ATT$)

- Literally, it represents the difference in average outcomes if the population of treated individuals had ultimately not been exposed to the treatment of interest
- Formally, it is a conditional average treatment effect (CATE), conditional on receiving the treatment

Identifiability of ATT in an observational study

- Still relies on the fundamental assumptions
  - Since the population of interest is the entire group of treated individuals, the positivity assumption is only required in this subpopulation: $Pr(A = a|L = l, A = 1) > 0 \, \forall \, a \in (0, 1) \, \forall \, l$
- Can be obtained by standardizing the conditional effects:

$$ATT = \sum_l ATE_l \times Pr(L = l, A = 1)$$

- The weights are $Pr(L = l, A = 1)$ here reflecting the distribution of L in the subpopulation of treated individuals

What about the untreated?

- Similarly, one might wonder what would be the difference in average outcomes of the non-exposed individuals if they had been exposed?
- This is called the average treatment effect on the controlled (ATC):

$$ATC = E(Y^1|A = 0) - E(Y^0|A = 0)$$

(in a randomized controlled trial, $ATE = ATT = ATC$)

- The identifiability conditions are symmetrical (see previous slide, with $A = 0$)

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Standardization in short I

- Standardization allows obtaining the average counterfactual outcome with either of the treatment alternatives:

$$E(Y^a) = \sum_l E(Y|A = a, L = l) \times Pr(L = l) \; \forall \; a \in (0, 1)$$

By replacing with estimates:

$$\widehat{E}(Y^a) = \sum_l \widehat{E}(Y|A = a, L = l)\widehat{Pr}(L = l) \; \forall \; a \in (0, 1)$$

- Then, we can estimate the marginal effect with different measures of association:

$$\widehat{DR} = \widehat{E}(Y^1) - \widehat{E}(Y^0) \quad (= \textstyle\sum_l(\widehat{\beta} + \widehat{\delta}_l) \times \widehat{Pr}(L = l))$$
$$\widehat{RR} = \widehat{E}(Y^1)/\widehat{E}(Y^0)$$
$$\widehat{OR} = \frac{\widehat{E}(Y^1)}{1-\widehat{E}(Y^1)} \Big/ \frac{\widehat{E}(Y^0)}{1-\widehat{E}(Y^0)}$$

Standardization in short II

OK, so practically, how do we do it?

Algorithm to estimate ATE

In a sample of size n:

1) Estimate a regression model on Y adjusted for confounders $L = L_1, L_2, \ldots$ separately by exposure
   - One model for individuals who received the treatment ($A = 1$):
     $E(Y|A = 1) \sim \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 + \ldots$
   - One model for individuals who did not receive the treatment ($A = 0$):
     $E(Y|A = 0) \sim \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \ldots$
   - (Equivalent to a single model including all interactions $A \times L_k$)

Algorithm to estimate ATE

In a sample of size n:

1) Estimate a regression model on Y adjusted for confounders $L = L_1, L_2, \ldots$ separately by exposure
   - One model for individuals who received the treatment ($A = 1$):
     $E(Y|A = 1) \sim \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 + \ldots$
   - One model for individuals who did not receive the treatment ($A = 0$):
     $E(Y|A = 0) \sim \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \ldots$
   - (Equivalent to a single model including all interactions $A \times L_k$)
2) Predict potential outcomes for each individual i
   - For each individual i, calculate $\widehat{Y}_{1,i}$ according to the model for $A = 1$
   - For each individual i, calculate $\widehat{Y}_{0,i}$ according to the model for $A = 0$

Algorithm to estimate ATE

In a sample of size n:

1) Estimate a regression model on Y adjusted for confounders $L = L_1, L_2, \ldots$ separately by exposure
   - One model for individuals who received the treatment ($A = 1$): $E(Y|A = 1) \sim \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 + \ldots$
   - One model for individuals who did not receive the treatment ($A = 0$): $E(Y|A = 0) \sim \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \ldots$
   - (Equivalent to a single model including all interactions $A \times L_k$)
2) Predict potential outcomes for each individual i
   - For each individual i, calculate $\widehat{Y}_{1,i}$ according to the model for $A = 1$
   - For each individual i, calculate $\widehat{Y}_{0,i}$ according to the model for $A = 0$
3) Calculate the average predictions $\widehat{E}(Y^1) = \sum_{i=1}^{n} \widehat{Y}_{1,i}/n$ and $\widehat{E}(Y^0) = \sum_{i=1}^{n} \widehat{Y}_{0,i}/n$

Algorithm to estimate ATE

In a sample of size n:

1) Estimate a regression model on Y adjusted for confounders $L = L_1, L_2, \ldots$ separately by exposure
   - One model for individuals who received the treatment ($A = 1$):
     $E(Y|A = 1) \sim \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 + \ldots$
   - One model for individuals who did not receive the treatment ($A = 0$):
     $E(Y|A = 0) \sim \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \ldots$
   - (Equivalent to a single model including all interactions $A \times L_k$)
2) Predict potential outcomes for each individual i
   - For each individual i, calculate $\widehat{Y}_{1,i}$ according to the model for $A = 1$
   - For each individual i, calculate $\widehat{Y}_{0,i}$ according to the model for $A = 0$
3) Calculate the average predictions
   $\widehat{E}(Y^1) = \sum_{i=1}^{n} \widehat{Y}_{1,i}/n$ and $\widehat{E}(Y^0) = \sum_{i=1}^{n} \widehat{Y}_{0,i}/n$
4) Estimate the desired measure of association
   $\widehat{DR} = \widehat{E}(Y^1) - \widehat{E}(Y^0)$ $\quad$ $\widehat{RR} = \widehat{E}(Y^1)/\widehat{E}(Y^0)$
   $\widehat{OR} = \frac{\widehat{E}(Y^1)}{1-\widehat{E}(Y^1)} / \frac{\widehat{E}(Y^0)}{1-\widehat{E}(Y^0)}$

Algorithm to estimate ATE

First name	Y0	Y1	A	Y	L	Predicted Y0	Predicted Y1
Michael	0	0	1	0	1	...	...
Danica	1	0	1	0	0	...	...
Mitchell	0	1	0	0	0	...	...
Joanna	0	0	0	0	1	...	...
Blanca	0	0	0	0	0	...	...
Simon	0	1	1	1	1	...	...
Rojesh	1	0	1	0	1	...	...
Hamdi	1	0	1	0	0	...	...
Rahma	1	1	0	1	1	...	...
Matthew	0	0	0	0	1	...	...
Safar	0	1	0	0	0	...	...
Samuel	1	1	1	1	1	...	...
Cheyenne	0	0	1	0	1	...	...
Joshua	0	0	1	0	1	...	...

$$\widehat{E}(Y^1) = \sum_{i=1}^{n} \widehat{Y}_{1,i}/n$$
$$\widehat{E}(Y^0) = \sum_{i=1}^{n} \widehat{Y}_{0,i}/n$$

Algorithm to estimate ATT

In a sample of size n:

1) Step 1 remains unchanged
2) Step 2 remains unchanged
3) Calculate the average predictions among exposed individuals

$$\widehat{E}(Y^1|A=1) = \frac{\sum_{i=1}^{n} \widehat{Y}_{1,i} \times A}{\sum_{i=1}^{n} A} \text{ and } \widehat{E}(Y^0|A=1) = \frac{\sum_{i=1}^{n} \widehat{Y}_{0,i} \times A}{\sum_{i=1}^{n} A}$$

4) Estimate the desired measure of association

$$\widehat{DR} = \widehat{E}(Y^1|A=1) - \widehat{E}(Y^0|A=1)$$
$$\widehat{RR} = \widehat{E}(Y^1|A=1) / \widehat{E}(Y^0|A=1)$$
$$\widehat{OR} = \frac{\widehat{E}(Y^1|A=1)}{1-\widehat{E}(Y^1|A=1)} \Big/ \frac{\widehat{E}(Y^0|A=1)}{1-\widehat{E}(Y^0|A=1)}$$

Algorithm to estimate ATT

First name	Y0	Y1	A	Y	L	Predicted Y0	Predicted Y1
Michael	0	0	1	0	1	...	...
Danica	1	0	1	0	0	...	...
Mitchell	0	1	0	0	0	...	...
Joanna	0	0	0	0	1	...	...
Blanca	0	0	0	0	0	...	...
Simon	0	1	1	1	1	...	...
Rojesh	1	0	1	0	1	...	...
Hamdi	1	0	1	0	0	...	...
Rahma	1	1	0	1	1	...	...
Matthew	0	0	0	0	1	...	...
Safar	0	1	0	0	0	...	...
Samuel	1	1	1	1	1	...	...
Cheyenne	0	0	1	0	1	...	...
Joshua	0	0	1	0	1	...	...

$$\widehat{E}(Y^1|A=1) = \frac{\sum_{i=1}^{n} \widehat{Y}_{1,i} \times A}{\sum_{i=1}^{n} A}$$

$$\widehat{E}(Y^0|A=1) = \frac{\sum_{i=1}^{n} \widehat{Y}_{0,i} \times A}{\sum_{i=1}^{n} A}$$

Variance estimation

- The previous algorithm allows for easily estimating the marginal effect 'by hand', as we will see in an example
- The variance of this estimation is less straightforward to obtain and must be estimated using an appropriate method
  - delta method (for example, using the R package `marginaleffects`)
  - bootstrap

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Database lalonde I

```r
library("cobalt")
data("lalonde", package = "cobalt")
str(lalonde)
```

```
## 'data.frame':    614 obs. of  9 variables:
##  $ treat   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ age     : int  37 22 30 27 33 22 23 32 22 33 ...
##  $ educ    : int  11 9 12 11 8 9 12 11 16 12 ...
##  $ race    : Factor w/ 3 levels "black","hispan",..: 1 2 1 1 1 1 1 1 1 3 ...
##  $ married : int  1 0 0 0 0 0 0 0 0 1 ...
##  $ nodegree: int  1 1 0 1 1 1 0 1 0 0 ...
##  $ re74    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ re75    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ re78    : num  9930 3596 24909 7506 290 ...
```

Database `lalonde` II

A data frame with 614 observations (185 treated, 429 control). There are 9 variables measured for each individual.

- 'treat' is the treatment assignment (`1=treated`, `0=control`).
- `age` is age in years.
- `educ` is education in number of years of schooling.
- `race` is the individual's race/ethnicity, (Black, Hispanic, or White).
- `married` is an indicator for married (1=married, 0=not married).
- `nodegree` is an indicator for whether the individual has a high school degree (`1=no degree`, `0=degree`).
- `re74` is income in 1974, in U.S. dollars.
- `re75` is income in 1975, in U.S. dollars.
- 're78' is income in 1978, in U.S. dollars.

'treat' is the treatment variable, 're78' is the outcome, and the others are pre-treatment covariates.

Y CONTINUOUS, 'BY HAND'

```
mod1 <- lm(re78 ~ age + educ + race + married + nodegree + re74,
           data = subset(lalonde, treat == 1)) ## model among treat = 1
mod0 <- lm(re78 ~ age + educ + race + married + nodegree + re74,
           data = subset(lalonde, treat == 0)) ## model among treat = 0
```

```
## income predicted if treat = 1
y1ate <- predict(mod1, newdata = lalonde)
## income predicted if treat = 0
y0ate <- predict(mod0, newdata = lalonde)
```

```
mean(y1ate) - mean(y0ate) ## ATE
```

```
## [1] 1155.916
```

```
## income predicted if treat = 1 among treat = 1
y1att <- predict(mod1, newdata = subset(lalonde, treat == 1))
## income predicted if treat = 0 among treat = 1
y0att <- predict(mod0, newdata = subset(lalonde, treat == 1))
```

```
mean(y1att) - mean(y0att) ## ATT
```

```
## [1] 1691.123
```

Y CONTINUOUS, WITH THE marginaleffects PACKAGE

```r
library(marginaleffects)
mod <- lm(re78 ~ treat * (age + educ + race + married + nodegree + re74),
          data = lalonde) # model with interactions
## ATE
avg_comparisons(mod, variables = "treat")
```

```
##
##  Estimate Std. Error    z Pr(>|z|)   S 2.5 % 97.5 %
##      1156       1023 1.13    0.258 2.0  -849   3161
##
## Term: treat
## Type: response
## Comparison: 1 - 0
```

```r
## ATT
avg_comparisons(mod, variables = "treat", newdata = subset(lalonde, treat == 1))
```

```
##
##  Estimate Std. Error    z Pr(>|z|)   S 2.5 % 97.5 %
##      1691        825 2.05   0.0403 4.6  75.1   3307
##
## Term: treat
## Type: response
```

Y BINARY, 'BY HAND' I

```
lalonde$re78bin <- lalonde$re78 > 0 ## income > 0 ?
table(lalonde$re78bin)
```

```
##
## FALSE  TRUE
##   143   471
```

```
## model among treat = 1
mod1 <- glm(re78bin ~ age + educ + race + married + nodegree + re74,
            data = subset(lalonde, treat == 1), family = binomial)
## model among treat = 0
mod0 <- glm(re78bin ~ age + educ + race + married + nodegree + re74,
            data = subset(lalonde, treat == 0), family = binomial)
```

Y BINARY, 'BY HAND' II

```r
## P(income > 0) predicted if treat = 1
y1ate <- predict(mod1, newdata = lalonde, type = "response")
## P(income > 0) predicted if treat = 0
y0ate <- predict(mod0, newdata = lalonde, type = "response")
```

```r
mean(y1ate) - mean(y0ate) ## Risk difference ATE
```

```
## [1] 0.09917026
```

```r
log(mean(y1ate) / mean(y0ate)) ## log(RR) ATE
```

```
## [1] 0.123929
```

```r
log((mean(y1ate)/(1-mean(y1ate))) / (mean(y0ate)/(1-mean(y0ate)))) ## log(OR) ATE
```

```
## [1] 0.6336351
```

Y BINARY, 'BY HAND' III

```r
## P(income > 0) predicted if treat = 1 among treat = 1
y1att <- predict(mod1, newdata = subset(lalonde, treat == 1), type = "response")
## P(income > 0) predicted if treat = 0 among treat = 1
y0att <- predict(mod0, newdata = subset(lalonde, treat == 1), type = "response")
```

```r
mean(y1att) - mean(y0att) ## Risk difference ATT
```

```
## [1] 0.05125712
```

```r
log(mean(y1att) / mean(y0att)) ## log(RR) ATT
```

```
## [1] 0.07013562
```

```r
log((mean(y1att)/(1-mean(y1att))) / (mean(y0att)/(1-mean(y0att)))) ## log(OR) ATT
```

```
## [1] 0.2613539
```

Y BINARY, WITH THE marginaleffects PACKAGE I

```r
library(marginaleffects)
mod <- glm(re78bin ~ treat * (age + educ + race + married + nodegree + re74),
           data = lalonde, family = binomial) ## model with interactions

## Risk difference ATE
avg_comparisons(mod, variables = "treat")
```

```
##
##  Estimate Std. Error    z Pr(>|z|)   S  2.5 % 97.5 %
##    0.0992     0.0405 2.45   0.0144 6.1 0.0197  0.179
##
## Term: treat
## Type: response
## Comparison: 1 - 0
```

Y BINARY, WITH THE marginaleffects PACKAGE II

```
## log(RR) ATE
avg_comparisons(mod, variables = "treat", comparison = "lnratioavg")
```

```
##
##  Estimate Std. Error    z Pr(>|z|)   S  2.5 % 97.5 %
##     0.124     0.0498 2.49   0.0129 6.3 0.0263  0.222
##
## Term: treat
## Type: response
## Comparison: ln(mean(1) / mean(0))
```

```
## log(OR) ATE
avg_comparisons(mod, variables = "treat", comparison = "lnoravg")
```

```
##
##  Estimate Std. Error    z Pr(>|z|)   S 2.5 % 97.5 %
##     0.634       0.29 2.19   0.0287 5.1 0.066    1.2
##
## Term: treat
## Type: response
## Comparison: ln(odds(1) / odds(0))
```

Y BINARY, WITH THE marginaleffects PACKAGE III

```r
## Risk difference ATT
avg_comparisons(mod, variables = "treat", newdata = subset(lalonde, treat == 1))
```

```
##
##  Estimate Std. Error     z Pr(>|z|)   S   2.5 % 97.5 %
##    0.0513     0.0531 0.965    0.334 1.6 -0.0528  0.155
##
## Term: treat
## Type: response
## Comparison: 1 - 0
```

```r
## log(RR) ATT
avg_comparisons(mod, variables = "treat", newdata = subset(lalonde, treat == 1),
                comparison = "lnratioavg")
```

Y BINARY, WITH THE marginaleffects PACKAGE IV

```
##
##  Estimate Std. Error     z Pr(>|z|)   S   2.5 % 97.5 %
##    0.0701     0.0736 0.953    0.341 1.6 -0.0741  0.214
##
## Term: treat
## Type: response
## Comparison: ln(mean(1) / mean(0))
```

Y BINARY, WITH THE marginaleffects PACKAGE V

```
## log(OR) ATT
avg_comparisons(mod, variables = "treat", newdata = subset(lalonde, treat == 1),
                comparison = "lnoravg")
```

```
##
##  Estimate Std. Error     z Pr(>|z|)   S  2.5 % 97.5 %
##     0.261      0.267 0.979    0.328 1.6 -0.262  0.785
##
## Term: treat
## Type: response
## Comparison: ln(odds(1) / odds(0))
```

SANTÉ SORBONNE UNIVERSITÉ

3

Standardization

Summary I

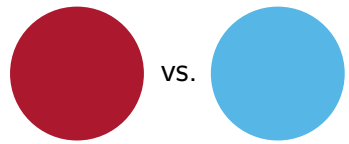
- Identifying a causal effect in an observational study relies on fundamental assumptions: no interference, coherence, conditional exchangeability, and positivity.
- Under these assumptions, regression models allow estimating conditional causal effects (stratum-specific), as well as potential outcomes for each individual.
- Standardization of conditional effects enables estimating a marginal causal effect, of the same nature as that estimated in a randomized controlled trial.

Summary II

- The causal effect can pertain to the entire population (ATE), or to the population of actually treated individuals (ATT), or even the untreated individuals (ATC).
- These causal effects depend on the distribution of confounding factors L acting as effect modifiers. Two samples from two populations with different distributions of these factors may lead to two different effect estimations (without any bias). (similar to how two randomized trials conducted in two different populations might yield different effect estimates)

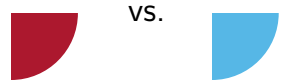
Note: Marginal vs. conditional effect

An effect can be marginal, like the average treatment effect (ATE):



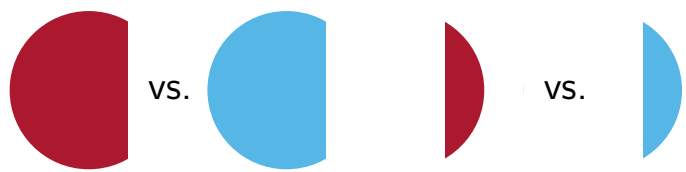
$$ATE$$

Or conditional like the conditional average treatment effect (CATE), effect conditional on certain levels of the variables L, for example:



$$CATE\ (ATE_l)$$

The average treatment effect on the treated (ATT) and on the controlled (ATC) are conditional CATEs on exposure A:



$$ATT\ (A = 1) \qquad ATC\ (A = 0)$$

Vocabulary note

Conditional *vs.* Adjusted?

- The terms 'conditional' and 'adjusted' are often used interchangeably.
- The terms 'marginal' and 'unadjusted' as well.
- However, we now know that a marginal effect can be estimated from an... adjusted model.

Vocabulary note

Conditional *vs.* Adjusted?

- The terms 'conditional' and 'adjusted' are often used interchangeably.
- The terms 'marginal' and 'unadjusted' as well.
- However, we now know that a marginal effect can be estimated from an... adjusted model.

We should differentiate:

- marginal/conditional: type of estimand targeted
- unadjusted/adjusted: type of analysis performed

SANTÉ SORBONNE UNIVERSITÉ

4

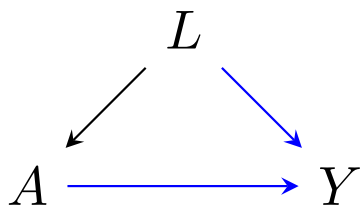
Propensity score

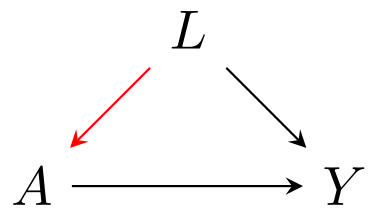
- Why?
- Definition, properties, estimation
- The steps
- Estimation Variance
- Summary
- Application with R

Change in Approach

The method we are going to see now still relies on the fundamental assumptions.

But instead of modeling $E(Y|A, L)$, we will model $E(A|L)$.





SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

- Number of individuals in the intervention group: 20
- Number of individuals in the control group: 20
- Number of individuals in the intervention group in stratum A: 10
- Number of individuals in the intervention group in stratum B: 10
- Number of individuals in the control group in stratum A: 10
- Number of individuals in the control group in stratum B: 10

$\Rightarrow$ Balance

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

- Risk difference in stratum A: $9/10 - 5/10 = 4/10$
- Risk difference in stratum B: $5/10 - 1/10 = 4/10$

$\Rightarrow$ The stratum is not an effect modifier

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

- Risk difference in stratum A: $9/10 - 5/10 = 4/10$
- Risk difference in stratum B: $5/10 - 1/10 = 4/10$

$\Rightarrow$ The stratum is not an effect modifier

- Overall risk difference: $14/20 - 6/20 = 8/20 = 4/10$

$\Rightarrow$ Marginal effect $=$ Conditional effect

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

Starting again with odds ratio:

- Odds ratio in stratum A: $\frac{9 \times 5}{5 \times 1} = 9$
- Odds ratio in stratum B: $\frac{5 \times 9}{1 \times 5} = 9$

$\Rightarrow$ The stratum is not an effect modifier

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

Starting again with odds ratio:

- Odds ratio in stratum A: $\frac{9\times5}{5\times1} = 9$
- Odds ratio in stratum B: $\frac{5\times9}{1\times5} = 9$

$\Rightarrow$ The stratum is not an effect modifier

- Overall odds ratio: $\frac{14\times14}{6\times6} \approx 5,44$

$\Rightarrow$ Marginal effect $\neq$ Conditional effect ??

Collapsibility

Here are the results of a randomized trial, reported based on an initial characteristic (stratum A or B).

	Stratum A		Stratum B		Together	
Allocation	Dead	Alive	Dead	Alive	Dead	Alive
Intervention	9	1	5	5	14	6
Control	5	5	1	9	6	14

$$E(Y|A,L) = \alpha + \gamma L + \beta A$$

Linear model and L is not an effect modifier $\Rightarrow ATE = ATE_l = \beta$

Collapsibility

- Indeed, the model used to estimate the odds ratio is not a linear model

$$E(Y|A,L) = \alpha + \gamma L + \beta A$$

with β = mean difference
but a logistic model[4]

$$logit\{E(Y|A,L)\} = \alpha + \gamma L + \beta A$$

with $\beta = log(OR)$

- However, the expectation of a logit is not equal to the logit of the expectation
- This phenomenon illustrates the non-collapsibility of the odds ratio

[4] $logit(x) = log\left(\frac{x}{1-x}\right)$

Collapsibility

- Neither of these two types of effects is biased here:
  - the conditional effect quantified by the odds ratio is 9 (instead of analyzing each stratum separately, one can obviously estimate the conditional OR associated with A globally with a logistic regression adjusted on L)
  - the marginal effect quantified by the odds ratio is 5.44
  - for these two estimations, there is no confounding bias
- We are talking about two different estimands.

Collapsibility

- To estimate the marginal odds ratio from the logistic model $logit\{E(Y|A,L)\} = \alpha + \gamma L + \beta A$, one can use standardization (which we did at the end of the previous section).
- This property (collapsibility) sometimes favors certain measures of association over others.

Collapsibility

Outcome type	Measure of Association	Collapsibility *
Continuous	Mean Difference	Yes
Binary	Odds Ratio	No
	Risk Difference	Yes
	Risk Ratio	Yes
Censored	Hazard Ratio	No
	Risk Difference at a Given Time	Yes
	Restricted Mean Survival Time Difference	Yes

* Do conditional and marginal measures coincide?

Difficulties in modeling Y

- Correct model specification is challenging when the number of variables L for adjustment is large.
- Verifying this correct model specification from the data is not straightforward.
  - Extrapolation phenomenon: multiple different specifications tailored to the observed data.
- Logistic and Cox models are biased in the case of rare events.
  - Finite sample bias: the smaller the number of events per variable, the more biased these models will be.

Which leads us to the following question...

In an observational cohort study, can we estimate the marginal effect more directly, i.e., without modeling Y and then standardizing?

SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

Definition

The propensity score (e) is the individual probability of being exposed to the treatment ($A = 1$), given the subject's characteristics L before exposure.

$$e = P(A = 1|L = l)$$

(where L represents all of the subject's characteristics).

Properties

Conditional Exchangeability

If Y^0 and Y^1 are independent of A conditional on L:

$$Y^a \perp A|L \; \forall \; a \in (0,1)$$

then Y^0 and Y^1 are independent of A conditional on e:

$$Y^a \perp A|e \; \forall \; a \in (0,1)$$

$\Rightarrow$ Reduces the problem of accounting for multiple confounding factors to a single dimension.

Estimation

In an observational study, the probability of being treated is unknown and **must be estimated**.

Most often, this estimation relies on a logistic regression model, where the exposure to the treatment A is the dependent variable, and the k observed covariates L are the explanatory variables:

$$P(A = 1|L = l) = \frac{\exp(\beta_0 + \beta_1 L_1 + \beta_2 L_2 + \cdots + \beta_k L_k)}{1 + \exp(\beta_0 + \beta_1 L_1 + \beta_2 L_2 + \cdots + \beta_k L_k)}$$

Once the parameters (the βs) of the model are estimated, the propensity score for each subject is calculated based on their individual characteristics.

The validity of methods based on the propensity score relies on the correct specification of this model.

The idea

- Identify and compare treated and untreated subjects who have the same initial probability of being exposed to the treatment, similar to what occurs in a randomized controlled trial
- Subjects who are treated and untreated but have a similar initial probability of being treated will also tend to have similar initial characteristics
- This approach aims to replicate conditions of *quasi-randomization*
- The propensity score is estimated for each study, and within the same study, each time a different treatment is analyzed

SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not
2) Estimate the propensity score (see previous slides)

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not
2) Estimate the propensity score (see previous slides)
3) Use the propensity score to balance the characteristics of subjects between the exposed and unexposed groups

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not
2) Estimate the propensity score (see previous slides)
3) Use the propensity score to balance the characteristics of subjects between the exposed and unexposed groups
4) Assess the balance before and after using the propensity score, and return to step 2 until the balance is satisfactory

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not
2) Estimate the propensity score (see previous slides)
3) Use the propensity score to balance the characteristics of subjects between the exposed and unexposed groups
4) Assess the balance before and after using the propensity score, and return to step 2 until the balance is satisfactory
5) Estimate the (marginal) effect of the exposure

The Different Steps

1) List *a priori* all variables to be considered, whether measured or not
2) Estimate the propensity score (see previous slides)
3) Use the propensity score to balance the characteristics of subjects between the exposed and unexposed groups
4) Assess the balance before and after using the propensity score, and return to step 2 until the balance is satisfactory
5) Estimate the (marginal) effect of the exposure

At each step, there are one or more questions to consider...

Which variables?

Step 1: Which variables?

- Non-parsimonious model: there is no limit (other than "technical" constraints) on the number of variables to include
- Ideally, the model linking exposure to characteristics includes:
  1) All confounding factors
  2) All variables related solely to Y (*a.k.a.* prognostic factors)
  3) **No** variables related solely to A (*a.k.a.* instrumental variables)

Step 1: Which variables?

- Non-parsimonious model: there is no limit (other than "technical" constraints) on the number of variables to include
- Ideally, the model linking exposure to characteristics includes:
  1) All confounding factors
  2) All variables related solely to Y (*a.k.a.* prognostic factors)
  3) **No** variables related solely to A (*a.k.a.* instrumental variables)
- Including prognostic factors improves the precision of the estimation, while including instrumental variables decreases precision and may even increase bias

Step 1: Which variables?

- Non-parsimonious model: there is no limit (other than "technical" constraints) on the number of variables to include
- Ideally, the model linking exposure to characteristics includes:
  1) All confounding factors
  2) All variables related solely to Y (*a.k.a.* prognostic factors)
  3) **No** variables related solely to A (*a.k.a.* instrumental variables)
- Including prognostic factors improves the precision of the estimation, while including instrumental variables decreases precision and may even increase bias
- Estimating the propensity score model is, therefore, quite counter-intuitive: we are not looking for variables that best predict A!
- In fact, you should choose the same variables as when modeling $E(Y|A, L)$ (where this choice is less counter-intuitive).

Step 1: Which variables?

- Selection criteria: literature review, expert opinion, ... Produce a DAG!
- **Do not** rely only on the "p-value" of the included variables or on stepwise selection processes
  - Omitting a variable, even one weakly associated, biases the estimation of the treatment effect
  - And "p-value > 0.05" does not mean "no association," but rather "no statistically significant association"
- **Do not** use global statistics (such as AUC for "area under the ROC curve") to evaluate the "quality" of the propensity model
  - On the contrary, an AUC close to 1 indicates that the treatment effect cannot be evaluated in the study

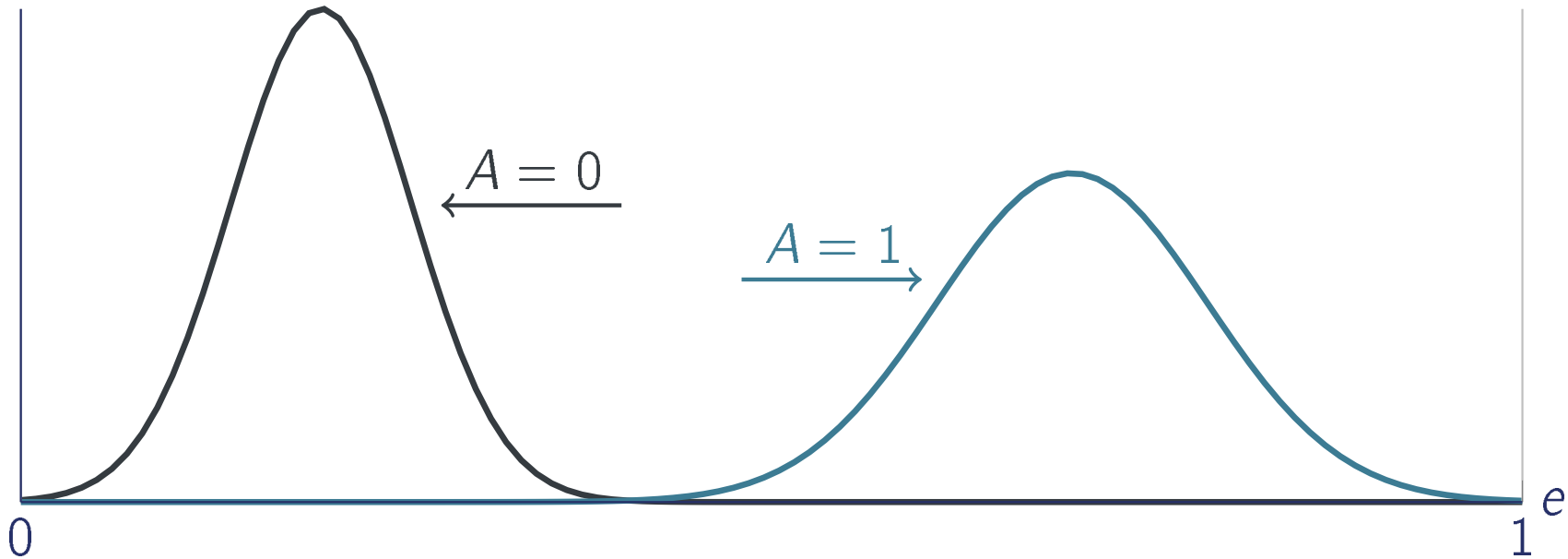
Step 2

How to evaluate the quality of the model?

Step 2: Evaluate the quality of the model

Once the propensity model and the propensity score (PS) for each patient are estimated, we can plot the distribution (e.g., histogram) of the PS in each treatment group.
When developing a logistic model, it is common practice to evaluate the quality of predictions with the AUC. The larger the AUC, the less overlap there is in the distribution of the PS between treated and untreated groups.
$\Rightarrow$ Let's examine the different possible scenarios.

Step 2: Distribution of the PS by treatment group

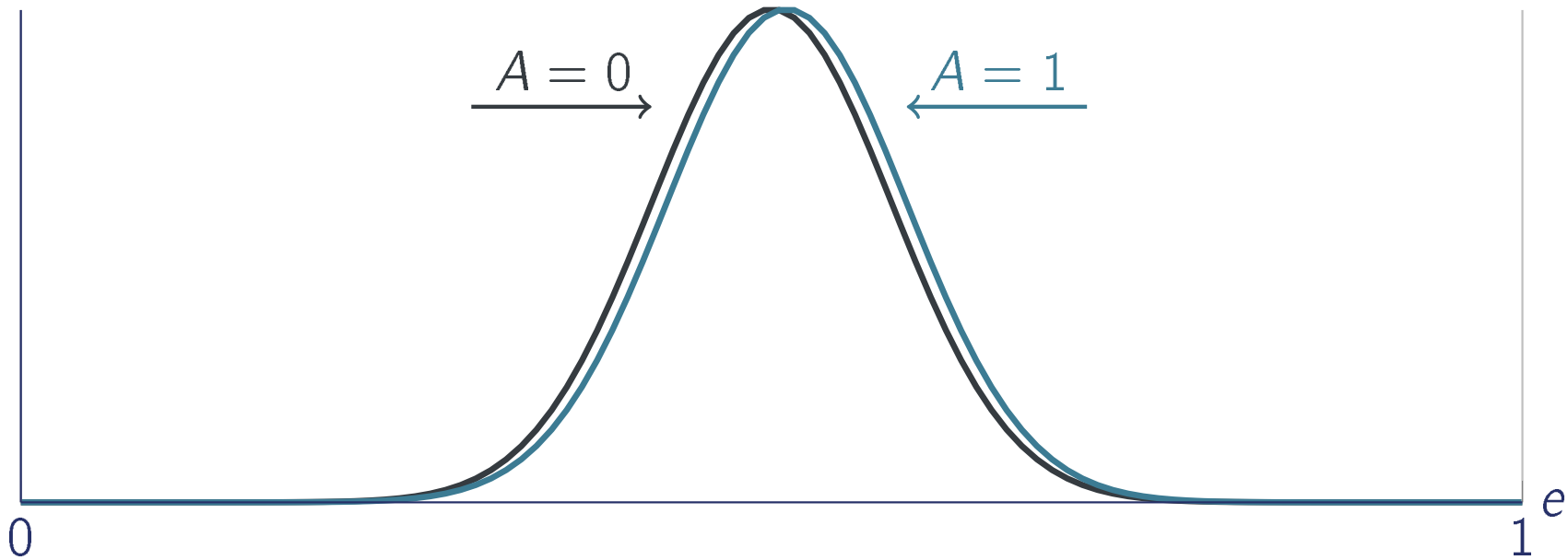


Perfectly predictable treatment exposure ($AUC = 1$)

$\longrightarrow$ two groups of patients: the **"always"** and the **"never"**

$\longrightarrow$ situation where the treatment effect cannot be estimated

Step 2: Distribution of the PS by treatment group

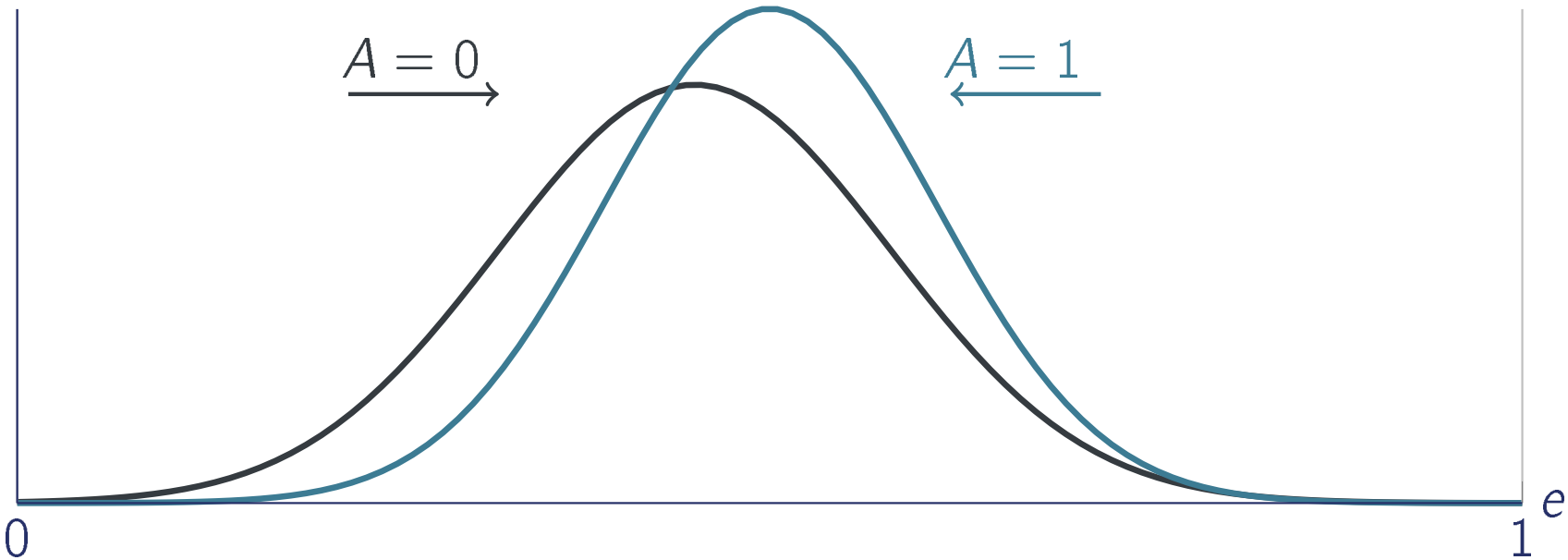


Non-predictable treatment exposure ($AUC = 0.5$)

$\longrightarrow$ each patient in the treated group corresponds to a patient in the untreated group with the same propensity to receive the treatment

$\longrightarrow$ situation so ideal that the use of the propensity score does not change the estimation of the treatment effect (because it is close to what would be observed in a randomized trial)

Step 2: Distribution of the PS by treatment group



Partially predictable exposure to treatment ($0.5 < AUC < 1$)

$\longrightarrow$ the two distributions overlap

$\longrightarrow$ "ideal" situation for using the propensity score

Step 2: Distribution of the PS by treatment group

- A certain degree of randomness in the assignment of the treatment is necessary, even in an observational study!
  Which fundamental assumption is being referred to here?

Step 2: Distribution of the PS by treatment group

- A certain degree of randomness in the assignment of the treatment is necessary, even in an observational study!
  Which fundamental assumption is being referred to here? The positivity assumption: there are both treated and untreated subjects for all levels of all considered variables

Step 2: Distribution of the PS by treatment group

- A certain degree of randomness in the assignment of the treatment is necessary, even in an observational study!
  Which fundamental assumption is being referred to here? The positivity assumption: there are both treated and untreated subjects for all levels of all considered variables
- Consequence: if it is known (*a priori* or during the exploration of the database) that some individuals are always or never treated, they should be removed from the analysis
- Therefore, the AUC is not a good measure of the quality of the propensity model: one should neither seek to maximize nor minimize it (again: we are not looking for variables that best predict A)

Reminder: Fundamental Assumptions of Causal Inference in Observational Studies

1) No interference
2) Consistency
3) Conditional exchangeability
4) Positivity

What methods of utilization?

Step 3: Methods for using the estimated propensity score

Four methods are available:

- Adjustment
- Stratification
- Matching
- Weighting (Inverse Probability of Treatment Weighting, IPTW)

Step 3: Methods for using the estimated propensity score

Four methods are available:

- Adjustment
- Stratification
- Matching
- Weighting (Inverse Probability of Treatment Weighting, IPTW)

Weighting is a particularly elegant, flexible method, with good performance in most situations.

Principle of Weighting

Each subject is weighted, with the weight $\hat{w}$ directly derived from the estimated propensity score (PS) of each individual.

- $\hat{w} = 1/\hat{e}$ for treated individuals
- $\hat{w} = 1/(1 - \hat{e})$ for untreated individuals

Intuition behind Weighting

- If a treated individual has a low estimated probability of receiving treatment ($A = 1$ and small $\hat{e}$), it means there are many similar individuals to them who did not receive treatment ($A = 0$ and small $\hat{e}$).
- After weighting, this individual will be overrepresented in the analysis (if $\hat{e}$ is small, $1/\hat{e}$ is large), to "compensate" for the rarity of their characteristics in the treated group compared to the untreated group.

Intuition behind Weighting

- If a treated individual has a low estimated probability of receiving treatment ($A = 1$ and small $\hat{e}$), it means there are many similar individuals to them who did not receive treatment ($A = 0$ and small $\hat{e}$).
- After weighting, this individual will be overrepresented in the analysis (if $\hat{e}$ is small, $1/\hat{e}$ is large), to "compensate" for the rarity of their characteristics in the treated group compared to the untreated group.
- Conversely, if a treated individual has a high estimated probability of receiving treatment ($A = 1$ and large $\hat{e}$), they will be underrepresented after weighting ($1/\hat{e}$ small), to "compensate" for the frequency of their characteristics in the treated group compared to the untreated group.

Intuition behind Weighting

- If a treated individual has a low estimated probability of receiving treatment ($A = 1$ and small $\hat{e}$), it means there are many similar individuals to them who did not receive treatment ($A = 0$ and small $\hat{e}$).
- After weighting, this individual will be overrepresented in the analysis (if $\hat{e}$ is small, $1/\hat{e}$ is large), to "compensate" for the rarity of their characteristics in the treated group compared to the untreated group.
- Conversely, if a treated individual has a high estimated probability of receiving treatment ($A = 1$ and large $\hat{e}$), they will be underrepresented after weighting ($1/\hat{e}$ small), to "compensate" for the frequency of their characteristics in the treated group compared to the untreated group.

- And vice versa

Illustration

In the weighted pseudo-population, treated and untreated groups will have balanced characteristics on average, thus restoring *exchangeability*.
(**only on the variables included in the propensity score model**)

Illustration

Initial population

Characteristics	**Total**, N = 10000	**Treated** N = 3,489	**Untreated** N = 6,5
Dosage, mean (SD)	423 (50)	438 (48)	415 (49)
Male, n (%)	4,099 (41%)	1,613 (46%)	2,486 (38%)

Illustration

Initial population

Characteristics	**Total**, N = 10000	**Treated** N = 3,489	**Untreated** N = 6,5
Dosage, mean (SD)	423 (50)	438 (48)	415 (49)
Male, n (%)	4,099 (41%)	1,613 (46%)	2,486 (38%)

Weighted population

Characteristics	**Total**, N = 19989	**Treated** N = 9,985	**Untreated** N = 10,
Dosage, mean (SD)	423 (50)	423 (49)	423 (50)
Male, n (%)	8,185 (41%)	4,084 (41%)	4,101 (41%)

How to check balance?

Step 4 : Checking balance

- The goal is to achieve balance in terms of characteristics between treated and untreated patients
- **No need to perform usual group comparison tests**: just because a difference is not statistically significant doesn't mean it doesn't introduce bias in the estimation of treatment effect

Step 4: Checking Balance

- Plot the distribution of the propensity score by treatment group before and after weighting $\rightarrow$ overlapping distributions
- Plot the distribution of each initial characteristic by treatment group before and after weighting $\rightarrow$ overlapping distributions
- Calculate the standardized difference between the two groups before and after weighting for each characteristic $\rightarrow |d| < 10\%$.

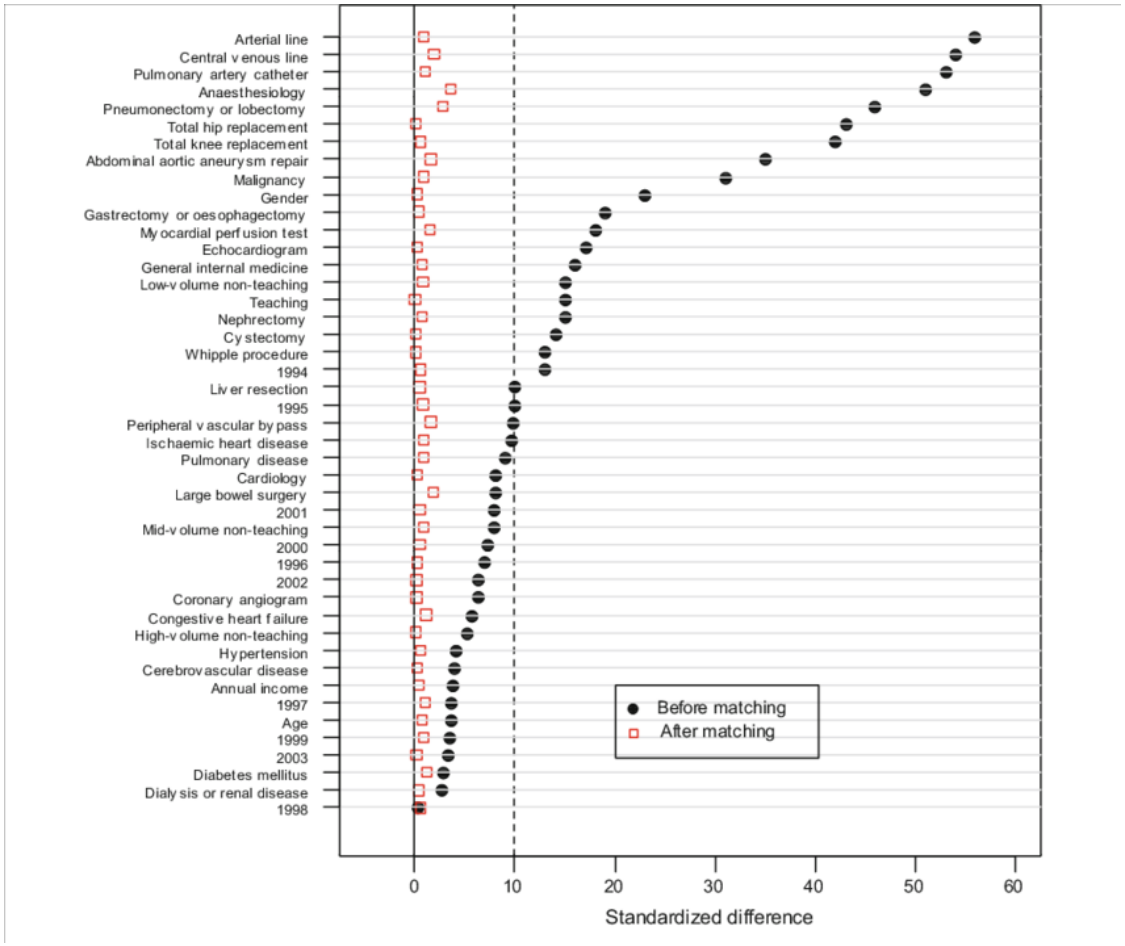
Quantitative characteristic

$$d = 100 \frac{\hat{\mu}_{A=1} - \hat{\mu}_{A=0}}{\sqrt{\frac{\hat{\sigma}^2_{A=1} + \hat{\sigma}^2_{A=0}}{2}}}$$

Qualitative characteristic

$$d = 100 \frac{\hat{p}_{A=1} - \hat{p}_{A=0}}{\sqrt{\frac{\hat{p}_{A=1}(1-\hat{p}_{A=1}) + \hat{p}_{A=0}(1-\hat{p}_{A=0})}{2}}}$$

Example



Gayat et al. Intensive Care Medicine, 2010

Step 5

How to estimate the effect?

Estimating the marginal treatment effect

- In the weighted sample, once balance is achieved
- Direct comparison of treated and untreated on Y: once the propensity score has restored balance, the analysis is “univariable”!

Weighting on the propensity score

- Calculating the treatment effect in the weighted sample
- Estimating the ATE

$$ATE = \frac{\sum_{i=1}^{n} Y_i A_i w_i}{A_i w_i} - \frac{\sum_{i=1}^{n} Y_i (1 - A_i) w_i}{(1 - A_i) w_i}$$

$$= \text{"Weighted mean of Y among treated"}$$

$$\text{minus "Weighted mean of Y among untreated"}$$

Weighting on the propensity score

Characteristic	Total, N = 19989	Treated N = 9,985	Untreated N = 10,
Dosage, mean (SD)	423 (50)	423 (49)	423 (50)
Male, n (%)	8,185 (41%)	4,084 (41%)	4,101 (41%)
Dead, n (%)	13,384 (67%)	6,317 (63%)	7,067 (71%)

$$\widehat{ATE} = 63 - 71 = -8\%$$

Weighting on the propensity score

SCOOP! Estimating the ATT is possible with weighting!

- $\hat{w}' = 1$ for treated individuals
- $\hat{w}' = \hat{e}/(1 - \hat{e})$ for untreated individuals

(Other types of weights exist, depending on the target population)

Weighting on the propensity score

Characteristic	Total, N = 6982	Treated N = 3,489	Untreated N = 3,49
Dosage, mean (SD)	438 (48)	438 (48)	438 (49)
Male, n (%)	3,228 (46%)	1,613 (46%)	1,615 (46%)
Dead, n (%)	4,943 (71%)	2,320 (66%)	2,623 (75%)

$$\widehat{ATT} = 66 - 75 = -9\%$$

SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

Estimation Variance

Attention must be paid to estimating the variance of the effect!

1) Take into account how the analyzed sample was constituted (correlation between individuals in the same stratum or matched, weighting of individuals who then contribute more than once to the analysis): calculate **robust variances**
   - Otherwise, underestimation of the variance is possible! $\Rightarrow$ increased risk of obtaining a significant effect mistakenly
   - In R: `survey` package and `svyglm` function, `cluster` option in `coxph` function...
2) Take into account the propensity score estimation step
   - Otherwise, over or underestimation of the variance is possible!
   - More challenging

Regardless of the method, the simplest way is to estimate the variance of the effect by bootstrap, by re-estimating the propensity model in each bootstrap sample

SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

Summary

- Reduces the problem of considering multiple confounding factors to a single dimension
- Estimation of ATE or ATT (and others exist!)
  - Choice depending on the research objective!
- Particularly useful:
  - if exposure is not too rare, many factors to consider, and few events of interest
  - if interested in multiple different outcomes Y
  - if interested in the marginal treatment effect
- But only takes into account **measured** factors: does not solve the problem of unmeasured confounding any better!

SANTÉ SORBONNE UNIVERSITÉ

4

Propensity score

Database

- Outcome: Time to death
- Exposure Trt (1 ou 0)
- Confounding factors : one binary (B), one continuous (C)

Exposure effect : $\log(HR) = \log(1/2) = -0.69$

Database

```
head(df)
```

```
##   id       time status      age smoking Trt
## 1  1  1.8100826      0 45.19058       1   0
## 2  2  0.1849345      1 54.77195       1   1
## 3  3 18.8250355      0 47.05518       1   0
## 4  4 14.2559600      0 47.18936       0   0
## 5  5  6.1861925      0 43.76213       1   0
## 6  6  6.0899116      1 49.11622       1   0
```

Naive estimation

```r
library(survival)
mod <- coxph(Surv(time, status) ~ Trt, data = df)
summary(mod)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ Trt, data = df)
##
##   n= 5000, number of events= 3023
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## Trt -0.25037   0.77852  0.04533 -5.523 3.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## Trt    0.7785      1.284    0.7123    0.8508
##
## Concordance= 0.513  (se = 0.004 )
## Likelihood ratio test= 32  on 1 df,   p=2e-08
## Wald test            = 30.51  on 1 df,   p=3e-08
## Score (logrank) test = 30.66  on 1 df,   p=3e-08
```

Conditional model

```r
mod <- coxph(Surv(time, status) ~ Trt + age + smoking, data = df)
summary(mod)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ Trt + age + smoking, data = df)
##
##   n= 5000, number of events= 3023
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## Trt     -0.832446  0.434984  0.048588 -17.13   <2e-16 ***
## age      0.140381  1.150712  0.006372  22.03   <2e-16 ***
## smoking  1.491806  4.445118  0.042980  34.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## Trt         0.435      2.299    0.3955    0.4784
## age         1.151      0.869    1.1364    1.1652
## smoking     4.445      0.225    4.0860    4.8358
##
## Concordance= 0.735  (se = 0.004 )
## Likelihood ratio test= 1683  on 3 df,   p=<2e-16
## Wald test            = 1532  on 3 df,   p=<2e-16
```

PS ESTIMATION

```r
## Régression logistique
modT <- glm(Trt ~ age + smoking, data = df, family = "binomial")
summary(modT)
```

```
##
## Call:
## glm(formula = Trt ~ age + smoking, family = "binomial", data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -15.88242    0.66486 -23.888   <2e-16 ***
## age           0.28002    0.01282  21.848   <2e-16 ***
## smoking       0.66992    0.07798   8.591   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5309.3  on 4999  degrees of freedom
## Residual deviance: 4667.5  on 4997  degrees of freedom
## AIC: 4673.5
##
## Number of Fisher Scoring iterations: 5
```

Weights estimation

```r
## Weights (ATE)
df$denom <- ifelse(df$Trt == 1, df$probT, 1 - df$probT)
df$wATE <- 1/df$denom
## Weights (ATT)
df$denom <- ifelse(df$Trt == 1, 1, (1 - df$probT)/df$probT)
df$wATT <- 1/df$denom
```

Weights estimation

```
head(df)
```

```
##   id       time status      age smoking Trt      probT     denom     wATE
## 1  1  1.8100826      0 45.19058       1   0 0.07187021 12.913972 1.077436
## 2  2  0.1849345      1 54.77195       1   1 0.53112187  1.000000 1.882807
## 3  3 18.8250355      0 47.05518       1   0 0.11545547  7.661348 1.130525
## 4  4 14.2559600      0 47.18936       0   0 0.06485559 14.418870 1.069354
## 5  5  6.1861925      0 43.76213       1   0 0.04934562 19.265223 1.051907
## 6  6  6.0899116      1 49.11622       1   0 0.18861141  4.301906 1.232455
##          wATT
## 1 0.07743551
## 2 1.00000000
## 3 0.13052533
## 4 0.06935356
## 5 0.05190700
## 6 0.23245510
```

Treatment effect

```r
## ATE
mod <- coxph(Surv(time, status) ~ Trt + cluster(id), data = df, weight = wATE)
summary(mod)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ Trt, data = df, weights = wATE,
##     cluster = id)
##
##   n= 5000, number of events= 3023
##
##         coef exp(coef) se(coef) robust se      z Pr(>|z|)
## Trt -0.68569   0.50374  0.02771   0.06237 -10.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## Trt    0.5037      1.985    0.4458    0.5692
##
## Concordance= 0.581  (se = 0.008 )
## Likelihood ratio test= 625.1  on 1 df,   p=<2e-16
## Wald test            = 120.9  on 1 df,   p=<2e-16
## Score (logrank) test = 635.3  on 1 df,   p=<2e-16,   Robust = 147.9  p=<2e-16
##
```

Treatment effect

```r
## ATT
mod <- coxph(Surv(time, status) ~ Trt + cluster(id), data = df, weight = wATT)
summary(mod)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ Trt, data = df, weights = wATT,
##     cluster = id)
##
##   n= 5000, number of events= 3023
##
##         coef exp(coef) se(coef) robust se      z Pr(>|z|)
## Trt -0.64851   0.52283  0.05414   0.04755 -13.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## Trt    0.5228      1.913    0.4763    0.5739
##
## Concordance= 0.579  (se = 0.007 )
## Likelihood ratio test= 145.4  on 1 df,   p=<2e-16
## Wald test            = 186  on 1 df,   p=<2e-16
## Score (logrank) test = 148.2  on 1 df,   p=<2e-16,   Robust = 200.6  p=<2e-16
##
```

SANTÉ SORBONNE UNIVERSITÉ

5

Going further

- Double Robustness
- Longitudinal data
- IPCW
- Trial emulation

SANTÉ SORBONNE UNIVERSITÉ

5

Going further

Double Robustness

Longitudinal data

IPCW

Trial emulation

Doubly robust estimators I

- Standardization on one hand, and methods based on the propensity score on the other hand, rely on models that need to be correctly specified.
- Doubly robust methods provide two chances to "be right": they work if **at least** one of the models $A \sim f(L)$ or $Y \sim g(A, L)$ is correctly specified.

Doubly robust estimators II

In fact... we've already seen it!

- After weighting on the propensity score, estimating the treatment effect relies on a univariable model, including only $A \Rightarrow$ the outcome model $Y \sim g(A, L)$ is misspecified (as it does not include L)
- The models used for standardization are estimated using the same weight for all individuals $\Rightarrow$ the propensity model $A \sim f(L)$ is misspecified (as it does not include L)

Doubly robust estimators III

We can “deduce” that standardizing from models weighted on the propensity score is a doubly robust method

Weighting and standardizing : example I

```r
set.seed(123)
## Simulation of 1000 individuals
n <- 1000
## Two independent variables N(0,1)
L1 <- rnorm(n)
L2 <- rnorm(n)
## A ~ L1 + L2
A <- rbinom(n, 1, prob = plogis(L1 + L2))
## Y ~ A + L1 + L2
Y <- rnorm(n, mean = A + L1 + L2)
df <- data.frame(L1 = L1, L2 = L2, A = A, Y = Y)
df[1:3, ]
```

```
##           L1          L2 A          Y
## 1 -0.5604756 -0.99579872 0 -2.2350820
## 2 -0.2301775 -1.03995504 1  0.3041802
## 3  1.5587083 -0.01798024 1  1.8362135
```

Theoretical effet of A on Y : 1

Weighting and standardizing : example II

Correctly specified models:

```
Agood <- glm(A ~ L1 + L2, data = df, family = binomial) ## Propensity model
df$pgood <- predict(Agood, type = "response") ## Propensity score
df$wgood <- ifelse(df$A == 1, 1/df$pgood, 1/(1-df$pgood)) ## Weights
Y1good <- glm(Y ~ L1 + L2, data = df[df$A == 1, ]) ## Prediction model for A = 1
Y0good <- glm(Y ~ L1 + L2, data = df[df$A == 0, ]) ## Prediction model for A = 0
```

Weighting and standardizing : example III

Uncorrectly specified models: missing L2

```
Abad <- glm(A ~ L1, data = df, family = binomial) ## Propensity model
df$pbad <- predict(Abad, type = "response") ## Propensity score
df$wbad <- ifelse(df$A == 1, 1/df$pbad, 1/(1-df$pbad)) ## Weights
Y1bad <- glm(Y ~ L1, data = df[df$A == 1, ]) ## Prediction model for A = 1
Y0bad <- glm(Y ~ L1, data = df[df$A == 0, ]) ## Prediction model for A = 0
```

Weighting and standardizing : example IV

Standardization

- With **correctly** specified models

```
mean(predict(Y1good, newdata = df)) - mean(predict(Y0good, newdata = df))
```

```
## [1] 1.015952
```

- With **uncorrectly** specified models

```
mean(predict(Y1bad, newdata = df)) - mean(predict(Y0bad, newdata = df))
```

```
## [1] 1.911068
```

Weighting and standardizing : example V

Pondération

- With **correctly** specified models

```
glm(Y ~ A, data = df, weight = wgood)$coefficients["A"]
```

```
##        A
## 1.05794
```

- With **uncorrectly** specified models

```
glm(Y ~ A, data = df, weight = wbad)$coefficients["A"]
```

```
##        A
## 1.936891
```

Weighting and standardizing : example VI

Weighting + standardization

- Model for A **correctly** specified, models for Y **uncorrectly** specified

```
Y1badwgood <- glm(Y ~ L1, data = df[df$A == 1, ], weight = wgood)
Y0badwgood <- glm(Y ~ L1, data = df[df$A == 0, ], weight = wgood)
mean(predict(Y1badwgood, newdata = df)) - mean(predict(Y0badwgood, newdata = df))
```

```
## [1] 1.034155
```

- Model for A **uncorrectly** specified, models for Y **correctly** specified

```
Y1goodwbad <- glm(Y ~ L1 + L2, data = df[df$A == 1, ], weight = wbad)
Y0goodwbad <- glm(Y ~ L1 + L2, data = df[df$A == 0, ], weight = wbad)
mean(predict(Y1goodwbad, newdata = df)) - mean(predict(Y0goodwbad, newdata = df))
```

```
## [1] 1.032516
```

Doubly Robust Estimators

- There are other doubly robust estimators.
- Some may also incorporate machine learning methods, which allow for great flexibility regarding the 'models' involved.

American Journal of Epidemiology (2011)

Doubly Robust Estimation of Causal Effects

Funk, Michele Jonsson; Westreich, Daniel; Wiesen, Chris; Stürmer, Til; Brookhart, M. Alan; Davidian, Marie

Statistics in Medicine (2018)

Targeted maximum likelihood estimation for a binary treatment: A tutorial

Luque-Fernandez, Miguel Angel; Schomaker, Michael; Rachet, Bernard; Schnitzer, Mireille E.

SANTÉ SORBONNE UNIVERSITÉ

5

Going further

Double Robustness

Longitudinal data

IPCW

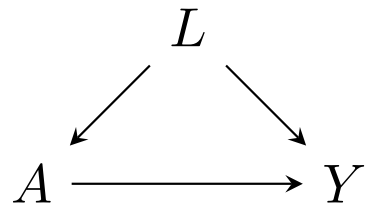
Trial emulation

G-METHODS

Jamie Robins and colleagues introduced three methods (the “g-methods”) to estimate a causal effect in the presence of time-dependent exposure and confounding factors:

- Inverse probability weighting of marginal structural models
  (Robins et al 2000, Epidemiology)
  $\longrightarrow$ Generalization of propensity score weighting
- G-computation formula
  (Robins 1986, Mathematical Modelling)
  $\longrightarrow$ Generalization of standardization
- G-estimation of structural nested models
  (Robins et al 1992, Epidemiology)
  $\longrightarrow$ ‘Generalization’ of propensity score adjustment

Inverse probability weighting of marginal structural models



$$E(Y^a) = \alpha + \beta a$$

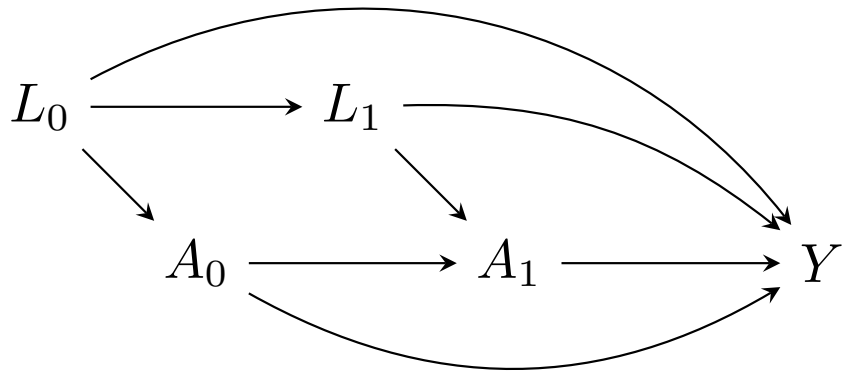
is a **marginal structural model**:

- "marginal" (as opposed to conditional): the model is not conditioned on L
- "structural": the counterfactual outcome is modeled

Propensity score weighting:

- Estimate $\widehat{\beta}$ using a regression model weighted on: $w_i = \frac{1}{P(A=a_i \mid L=l_i)}$
- Using these weights $\Rightarrow$ balance of L at baseline

Inverse probability weighting of marginal structural models

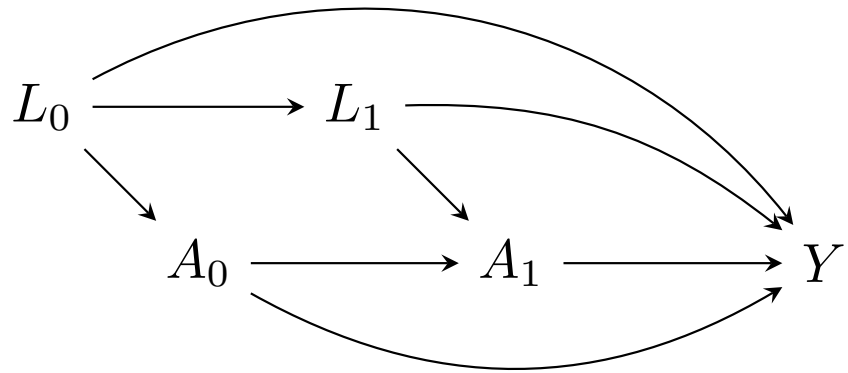


The marginal structural model of interest is now:

$$E(Y^{\bar{a}}) = \alpha + \gamma_0 a_0 + \gamma_1 a_1 + \gamma_{01} a_0 a_1$$

We note that there are 4 possible treatment trajectories: $a_0 = a_1 = 0$, $a_0 = a_1 = 1$, $a_0 = 0, a_1 = 1$ and $a_0 = 1, a_1 = 0$

Inverse probability weighting of marginal structural models



- $w_{0i} = \frac{1}{P(A_0 = a_{0i} | L_0 = l_{0i})} \Rightarrow$ balance of L_0 between $A_0 = 1$ and $A_0 = 0$
- $w_{1i} = \frac{1}{P(A_1 = a_{1i} | L_0 = l_{0i}, A_0 = a_{0i}, L_1 = l_{1i})} \Rightarrow$ balance of $\{L_0, A_0, L_1\}$ between $A_1 = 1$ and $A_1 = 0$
- $w_i = w_{0i} \times w_{1i} \Rightarrow$ balance of $\{L_0, A_0, L_1\}$ between all treatment trajectories

Inverse probability weighting of marginal structural models

General formula

$$w_i = \frac{1}{\prod_{k=0}^{K} Pr(A_k | \bar{A}_{k-1}, \bar{L}_k)}$$

- $\prod\_t = 0^K$: product of the terms from time 0 to time K
- $Pr(A_k \mid \bar{A}_{k-1}, \bar{L}_k)$: probability of being exposed at time k, conditional on
  - the exposure history up to the previous time point $k-1$
  - the history of confounders up to time k
- The notation $\bar{L}_k$ includes both time-varying and baseline (time-independent) variables such as age, sex, etc.
- Overall, the weight for an individual i is the inverse of the probability of receiving their own exposure trajectory, conditional on the history of confounders.

SANTÉ SORBONNE UNIVERSITÉ

5

Going further

Double Robustness

Longitudinal data

IPCW

Trial emulation

Informative missing data and censoring

Missing data (with censoring being a particular case) can be "handled" as we did for an intervention: through counterfactual reasoning!

What would have been the average outcome
if the entire population had no missing data?

$$\Leftrightarrow$$

$$E(Y^{C=0})^{5}$$

[5] $C = 0$: non-missing data

Informative missing data and censoring

Missing data (with censoring being a particular case) can be "handled" as we did for an intervention: through counterfactual reasoning!

What would have been the average outcome
if the entire population had no missing data?

$$\Leftrightarrow$$

$$E(Y^{C=0})^{5}$$

Inverse Probability of Censoring Weighting (IPCW): weight the population (containing missing data) to make it similar to what it would have been without missing data

[5] $C = 0$: non-missing data

SANTÉ SORBONNE UNIVERSITÉ

5

Going further

Double Robustness

Longitudinal data

IPCW

Trial emulation

Emulation of a target clinical trial

(Good) RCTs do not suffer (or suffer less) from:

- confounding bias due to randomization
- selection bias and measurement error because their protocols plan *a priori* selection criteria, intervention, outcome, and follow-up

Hernán and Robins (2016) proposed a formal approach to adopt the same design principles in observational data-based research

Journal of Clinical Epidemiology (2016)

Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses

Miguel A. Hernán; Brian C. Sauer; Sonia Hernández-Díaz; Robert Platt; Ian Shrier

American Journal of Epidemiology (2016)

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Hernán, Miguel A.; Robins, James M.

THANKS FOR WATCHING!

SANTÉ
SORBONNE
UNIVERSITÉ