SPRING 2023

Spring Meeting Giovani Ricercatori

SID SIIA SIMI SIPREC SISA

Affrontiamo i fattori confondenti:

Propensity score e inferenza causale

Davide Bernasconi

Dipartimento di Medicina e Chirurgia
Università Milano-Bicocca, Monza

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

WORKSHOP: "Come approcciarsi ai test statistici: road to Propensity Score"

Outline

- Why propensity score (PS) methods?
- Key concepts in causal inference
- Quantify the causal effect in a simple observational study with one binary confounder
- Estimate the causal effect with multiple measured confounders: PS-based methods
- Conclusions
- Teamwork

Why propensity score (PS) methods?

Multivariate Behavioral Research • *Open Access* • Volume 46, Issue 3, Pages 399 - 424 • May 2011

An introduction to propensity score methods for reducing the effects of confounding in observational studies

Austin P.C.

[a] Institute for Clinical Evaluative Sciences, Department of Health Management, Policy and Evaluation, University of Toronto, Toronto, ON, M4N 3M5, G1 06, 2075 Bayview Avenue, Canada

About 6300 citations on Scopus (search done April 2023)

Key idea: mimic characteristics of an RCT

Notation

- A and Y are two binary random variables:
  - **A represents the Treatment/Exposure** (1=exposed,0=not exposed)
  - **Y represents the Outcome**

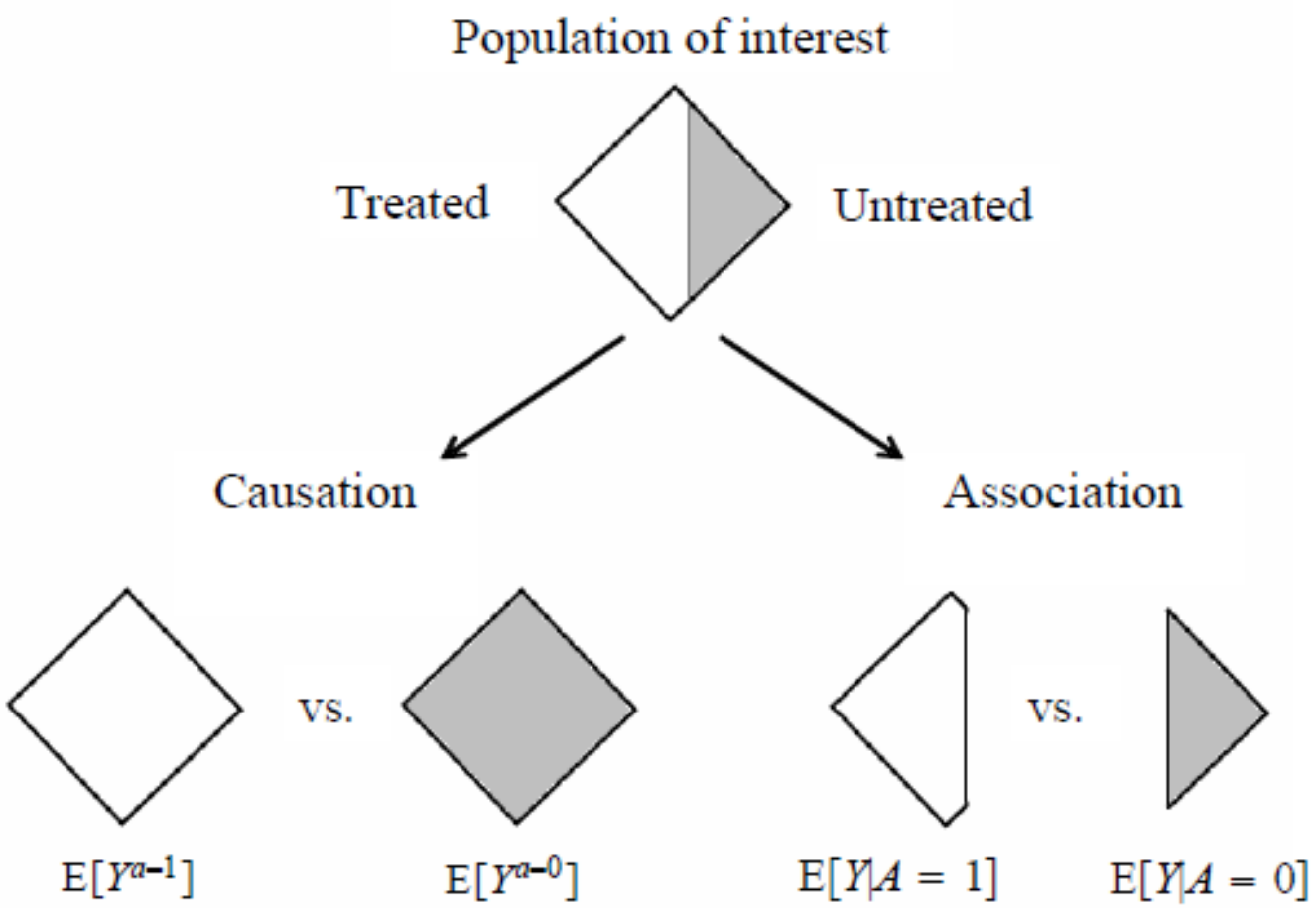
    (1=event 0=no event)

- We also define the **Counterfactual Outcome**:
  - $Y^{a=1}$ outcome Y we would observed under the exposure (a=1)
  - $Y^{a=0}$ outcome Y we would observe in the absence of exposure (a=0)

Definition of causal effect

Causal effect	
INDIVIDUAL 	POPULATION 
A has a causal effect on Y for the subject *i* if: $Y_i^{a=1} \neq Y_i^{a=0}$	A has a causal effect on Y in the population if: $P(Y^{a=1}=1) \neq P(Y^{a=0}=1)$ (or $E[Y^{a=1}] \neq E[Y^{a=0}]$)
$H_0: Y_i^{a=1} = Y_i^{a=0}$	$H_0: P(Y^{a=1}=1) = P(Y^{a=0}=1)$
Generally impossible to measure (exception: cross-over trials)	Under some conditions the **Average Causal Effect (ACE)** could be measured

Causation and association



Average causal effect, ACE
Comparison of **marginal** probabilities

Average effect in subgroups
Comparison of **conditional** probabilities

Association ≠ Causation (example)

Subjects	$Y^{a=0}$	$Y^{a=1}$
1	0	1
2	1	0
3	0	0
4	0	0
5	0	0
6	1	0
7	0	0
8	0	1
9	1	1
10	1	0
11	0	1
12	1	1
13	1	1
14	0	1
15	0	1
16	0	1
17	1	1
18	1	0
19	1	0
20	1	0

Average Causal Effect:

$$P(Y^{a=0}=1)=10/20=0.5$$

$$P(Y^{a=1}=1)=10/20=0.5$$

$$ACE=P(Y^{a=0}=1)-P(Y^{a=1}=1)=0$$

→ **No causal effect of A on Y**

Association ≠ Causation (example)

sub	$Y^{a=0}$	$Y^{a=1}$
1	0	
2	1	
3	0	
4	0	
5		0
6		0
7		0
8		1
9	1	
10	1	
11	0	
12		1
13		1
14		1
15		1
16		1
17		1
18		0
19		0
20		0

→

A	Y
0	0
0	1
0	0
0	0
1	0
1	0
1	0
1	1
0	1
0	1
0	0
1	1
1	1
1	1
1	1
1	1
1	1
1	0
1	0
1	0

Association:

$P(Y=1|A=0)=3/7=0.43$

$P(Y=1|A=1)=7/13=0.54$

$P(Y=1|A=0)-P(Y=1|A=1)\neq 0$

→ **A and Y are not independent**

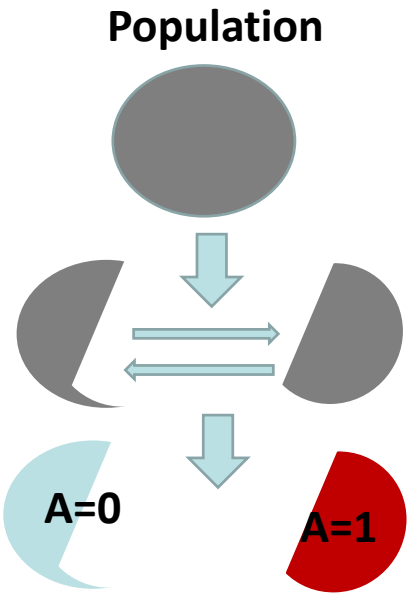
Randomized experiments

- In randomized experiments it is possible to estimate the average causal effect even if we observe only one outcome (either $Y^{a=0}$ or $Y^{a=1}$) for each subject
- Why? Because **exchangeability** holds:

$$P[Y^{a=1}=1|A=1]=P[Y^{a=1}=1|A=0]=P[Y^{a=1}]$$

Conditional probabilities = Marginal probability

$$\Rightarrow \mathbf{Y^a \perp A, \forall\ a}$$



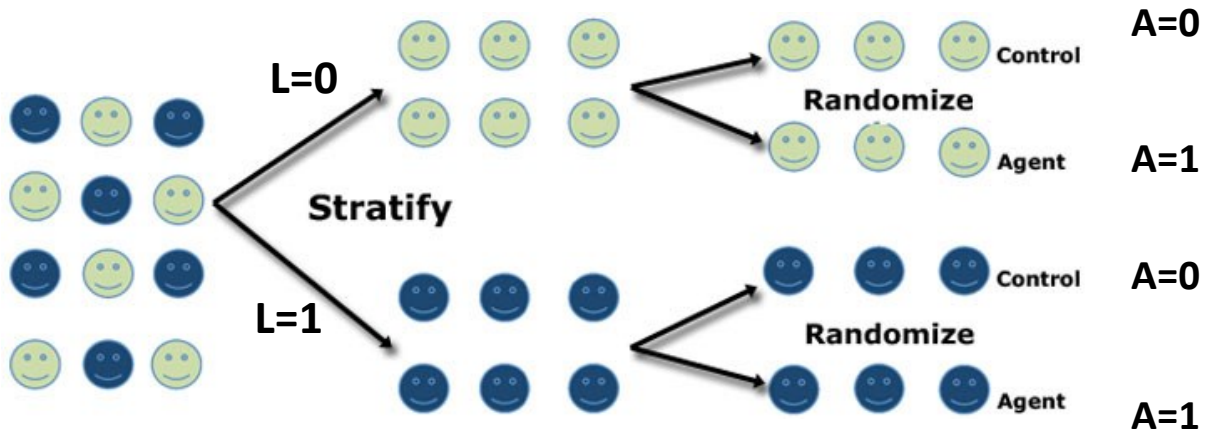
In an «ideal» randomized study:

Association = Causation

$$\mathbf{E[Y|A=1] = E[Y^{a=1}]}$$
$$\mathbf{E[Y|A=0] = E[Y^{a=0}]}$$

NB: In general is not possible to check validity of exchangeability from data

Randomization within strata



- Marginal exchangeability

- Conditional Exchangeability (on L)

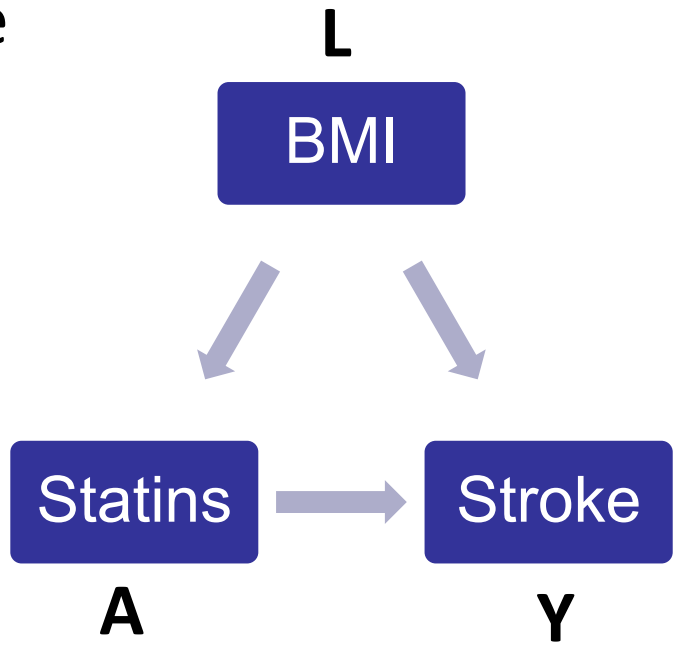
$$Y^a \perp A \mid L, \forall\, a \Leftrightarrow$$

$$E[Y^{a=1} \mid A=1, L=l] = E[Y^{a=1} \mid A=0, L=l] = E[Y^{a=1} \mid L=l]$$

Observational studies

Generally, in observational studies, subject exposed and not exposed are not exchangeable

$$Y^a \not\perp A$$



Common causes of exposure and outcome may exist and be measured (L)

In some situations, conditioned on these characteristics, exchangeability may hold

$$Y^a \perp A \mid L$$

Methods to estimate the causal effect

Context/assumption

- **Studies with randomization within strata**
- **Observational studies with conditional exchangeability**

Aim

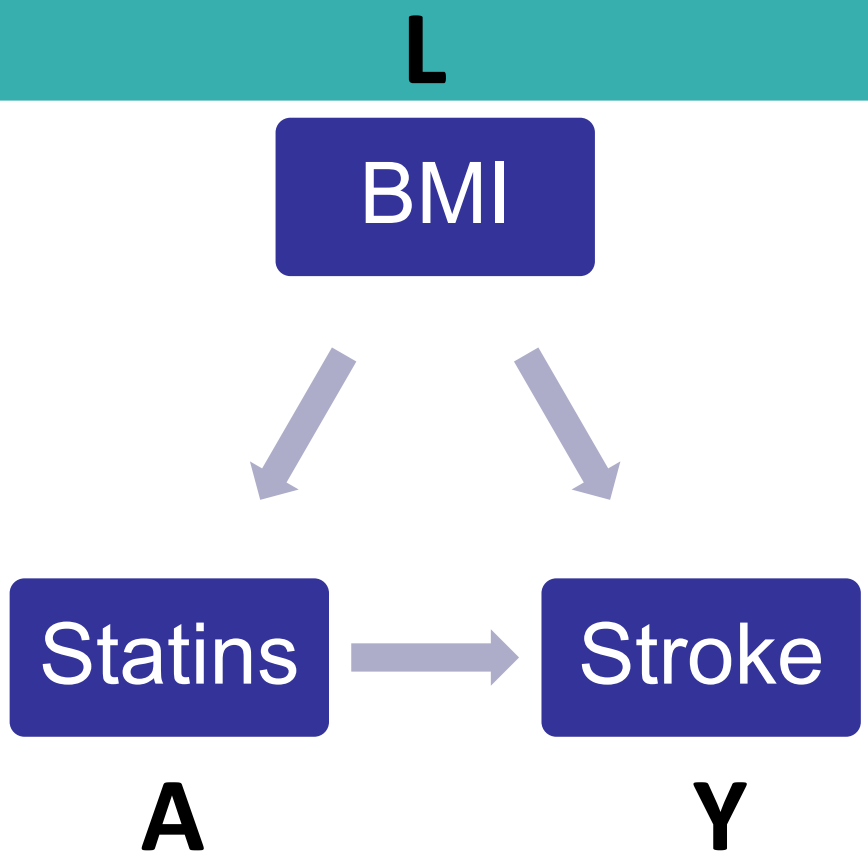
- **Estimate the Average Causal Effect (ACE)**

Methods

- **Stratification (effects within subgroups)**
- **Matching (ATT)**
- **IPW (ATE or ATT)**
- Standardization (aka G-computation)
- …

Toy example

subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0



Assumption:

- **No marginal exchangeability** ✗
- **Conditional Exchangeability (on L)** ✓

Stratification

Separately estimate the effect within the two strata L=0 and L=1

e.g. Relative Risk

$RR_{L=0}$ =

= P(Y=1|L=0,A=1)/P(Y=1|L=0,A=0)

= (2/4) / (1/4) = 2

subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Stratification

Separately estimate the effect within the two strata L=0 and L=1

e.g. Relative Risk

$RR_{L=0}$ =
= P(Y=1|L=0,A=1)/P(Y=1|L=0,A=0)

= (2/4) / (1/4) = 2

$RR_{L=1}$ = P(Y=1|L=1,A=1)/P(Y=1|L=1,A=0)

= (3/9) / (2/3) = 0.5

subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Matching

E.g. matching 1:1

- for each subject not exposed (A=0) in the stratum L=0 randomly match an exposed subject (A=1) in the same stratum L=0.
- Same for L=1.
- Exclude unmatched subjects

subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Matching

In the matched sample, L has the same distribution within exposed and not exposed groups.

Marginal exchangeability ✓

Estimate the average causal effect as in a randomized study

e.g. Relative Risk

$$RR = E[Y^{a=1}]/E[Y^{a=0}]$$

$$= P(Y=1|A=1)/P(Y=1|A=0)$$

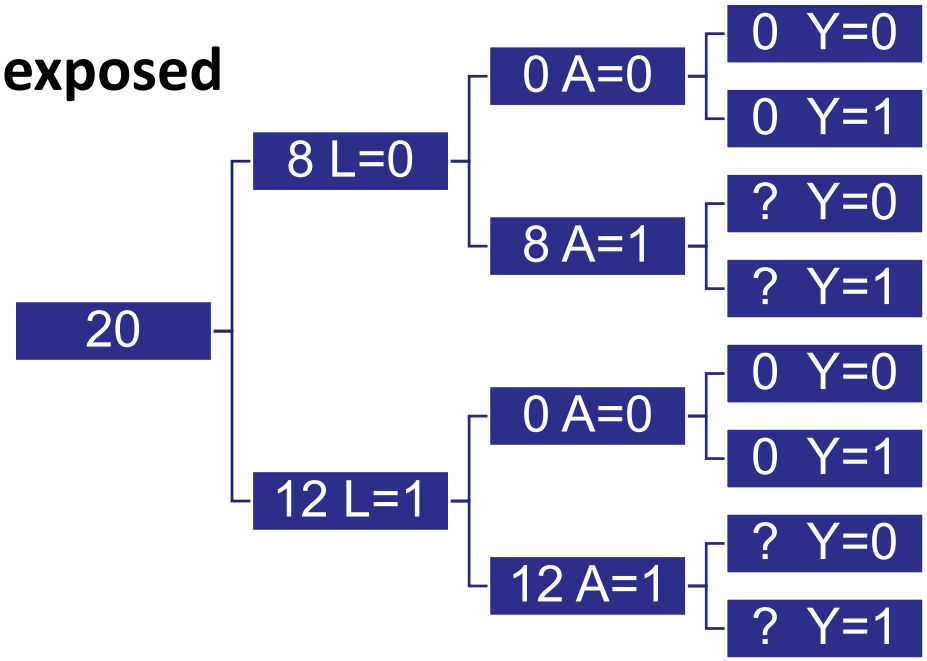
$$= (3/7) / (3/7) = 1$$

subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
14	1	1	1
15	1	1	0
16	1	1	0

Inverse Probabilty Weighting (IPW)

In each stratum, how many events would we expect if subjects are:

1. All exposed



subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Due to conditional exchangeability:

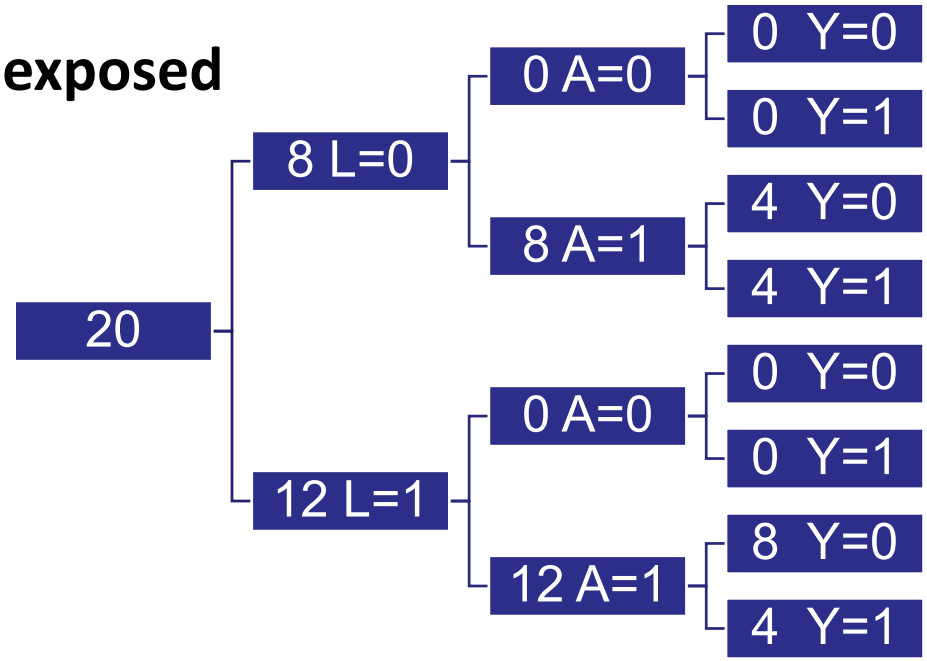
$E(Y^{a=1}|L=1)=P(Y=1|A=1,L=1)$

$E(Y^{a=1}|L=0)=P(Y=1|A=1,L=0)$

Inverse Probabilty Weighting (IPW)

In each stratum, how many events would we expect if subjects are:

1. All exposed



subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Due to conditional exchangeability:

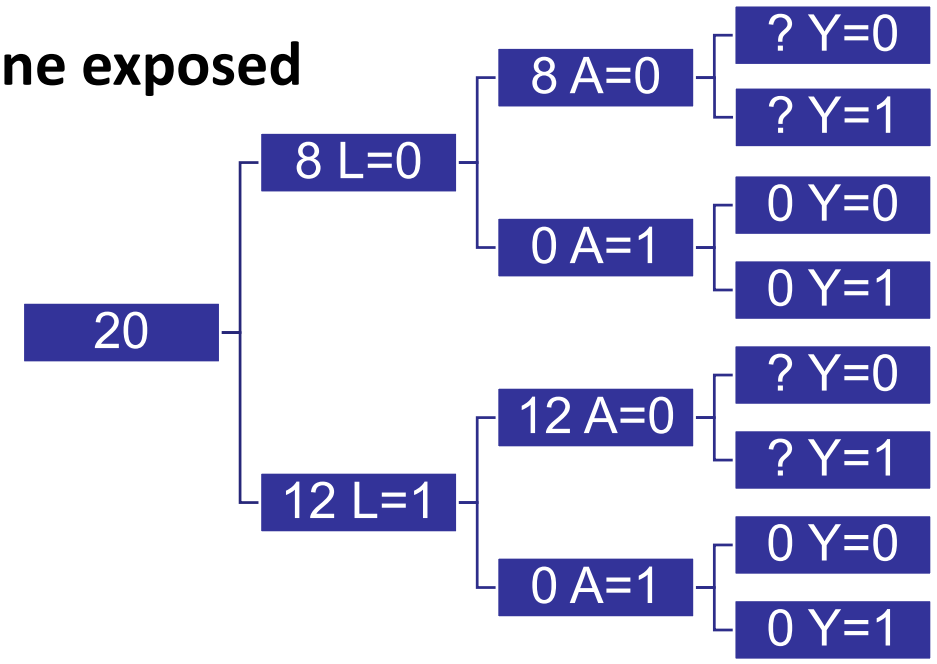
$E(Y^{a=1}|L=1)=P(Y=1|A=1,L=1)$

$E(Y^{a=1}|L=0)=P(Y=1|A=1,L=0)$

Inverse Probabilty Weighting (IPW)

In each stratum, how many events would we expect if subjects are:

2. None exposed



subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Due to conditional exchangeability:

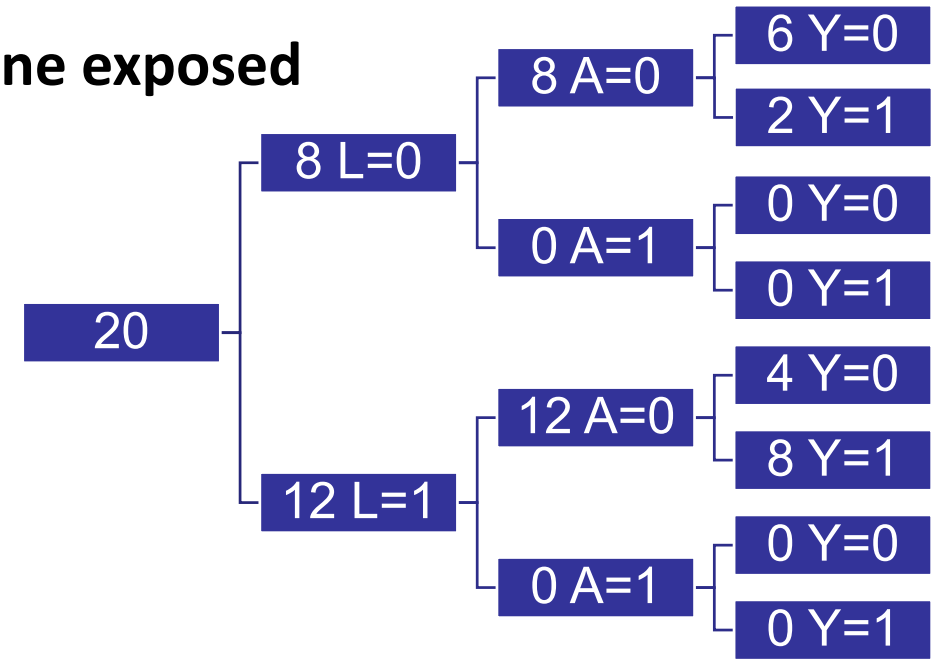
$E(Y^{a=0}|L=1)=P(Y=1|A=0,L=1)$

$E(Y^{a=0}|L=0)=P(Y=1|A=0,L=0)$

Inverse Probabilty Weighting (IPW)

In each stratum, how many events would we expect if subjects are:

2. None exposed



subject	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	1
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	0
16	1	1	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0

Due to conditional exchangeability:

$E(Y^{a=0}|L=1)=P(Y=1|A=0,L=1)$

$E(Y^{a=0}|L=0)=P(Y=1|A=0,L=0)$

Inverse Probabilty Weighting (IPW)

Lets pool together the two samples. In the new *pseudo*-population, L has the same distribution among exposed and non-exposed

Marginal exchangeability ✓

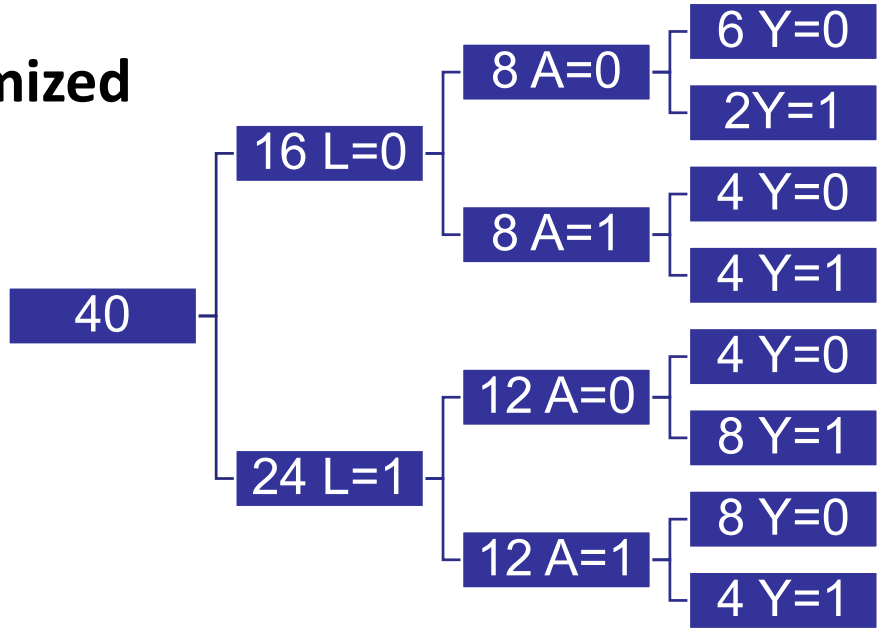
Estimate the ACE as in a randomized study

e.g. Relative Risk

RR = $E[Y^{a=1}]/E[Y^{a=0}]$

=P(Y=1|A=1)/P(Y=1|A=0)

= (8/20) / (10/20) = 0.4/ 0.5 = 0.8



Inverse Probabilty Weighting (IPW)

The *pseudo*-population (size: 2n), can also be created by weighting each individual by $w_A = 1/P(A=a|L=l)$

$P(A=0|L=0)=4/8$
$P(A=1|L=0)=4/8$
$P(A=0|L=1)=3/12$
$P(A=1|L=1)=9/12$

e.g. Relative Risk

$RR = E[Y^{a=1}]/E[Y^{a=0}]$

$=P(Y=1|A=1)/P(Y=1|A=0)$

$= (8/20) / (10/20) = 0.4/ 0.5$
$= 0.8$

subject	L	A	Y	Weights
1	0	0	0	2
2	0	0	1	2
3	0	0	0	2
4	0	0	0	2
5	0	1	0	2
6	0	1	0	2
7	0	1	1	2
8	0	1	1	2
9	1	0	1	4
10	1	0	1	4
11	1	0	0	4
12	1	1	1	1.33
13	1	1	1	1.33
14	1	1	1	1.33
15	1	1	0	1.33
16	1	1	0	1.33
17	1	1	0	1.33
18	1	1	0	1.33
19	1	1	0	1.33
20	1	1	0	1.33

How about dealing with multiple (measured) confounders?

Possible solution: **Propensity Score (PS)**

- For each subject *i*, **PS** is defined as **«the probability of treatment assignment conditional on observed baseline covariates»** (Rosenbaum & Rubin, Biometrika 1983)

$$\mathbf{PS_i = P(A_i = a \mid L_i)}$$

- PS is a measure of **balance**: conditional on PS, the **distribution of covariates** between treatment groups should be **similar**

- Typically estimated by logistic regression, e.g. *PS = P(A = 1 | L)*

$$\boldsymbol{logit\ (PS) = b_0 + b_1L_1 + b_2L_2 + \ldots}$$

- **PS matching**: match patients (e.g. 1:1) with similar PS

Inverse probaility of treatment weighting (IPTW)

- Individuals are weighted for the inverse of the probability of being treated with their actual treatment, given covariates:

$$W_i^{A=a} = \frac{1}{P(A_i = a \mid L_i)}$$

If subject *i* is treated with A=1:

$$W_i = \frac{1}{P(A_i = 1 \mid L_i)} = \frac{1}{PS_i}$$

If subject *i* is treated with A=0:

$$W_i = \frac{1}{P(A_i = 0 \mid L_i)} = \frac{1}{1\text{-}PS_i}$$

- In the weighted population, **marginal exchangeability** is achieved (provided there are no unmeasured confounders).

Conclusions

Issues to consider when using PS methods:

- Positivity assumption
- Absence of unmeasured confouders
- Check balance after PS matching or IPW
- Variable selection for PS model
- Not directly able to correct other type of bias in observational studies: (e.g. selection bias, ecc…)
- More complex settings:
  - Non-binary treatments
  - Time-dependent covariates
- Other causal methods not based on PS: standardization (a.k.a «G-computation»)

Conclusions

Some advantages of PS methods over outcome regression:

- Marginal vs conditional treatment effect
- Easier to estimate some effect measures (risk difference, RR or compare survival curves)
- Easier to check if balance is achieved with PS than to assess if outcome model is correct
- When outcome is rare and sample size is not big regression is limited but not PS

References

Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming.

Robins J, Hernan M, Brumback B (2000). Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology 11(5):550-60.

Ahern J, Hubbard A, Galea S. (2009) Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. Am J Epidemiol.;169(9):1140–1147.

Austin PC (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate behavioral research, 46(3), 399–424.

Edwards JK, Cole SR, Lesko CR, Mathews WC, Moore RD, Mugavero MJ, Westreich D. (2016) An illustration of inverse probability weighting to estimate policy-relevant causal effects. Am. J. Epidemiol; 184(4):336-344.

Bernasconi DP, Antolini L, Rossi E, Blanco-Lopez J, Andersen PK, Valsecchi MG (2022). A causal inference approach to compare treatment outcome in the absence of randomization and with informative censoring: an application to childhood leukemia in low-income countries. International Journal of Epidemiology 51(1):314-323.

Lavoro a gruppi

Leggere i due articoli ed identificare i seguenti aspetti:

- Tipo di studio (osservazionale/sperimentale?, prospettico retrospettivo?)
- Fattore di esposizione (binario/multicategorico?)
- Endpoint principale (continuo/binario/multicategorico/sopravvivenza?)
- Fattori confondenti (quali? quanti?)
- Metodo PS (matching/IPW?, come è stato stimato il PS?)
- Il bilanciamento dei confondenti è migliorato in seguito all'applicazione del metodo basato sul PS?
- Come è cambiata l'associazione stimata (quale misura di effetto è stata utilizzata?) tra il fattore di esposizione e l'outcome prima vs dopo l'applicazione del metodo basato sul PS?