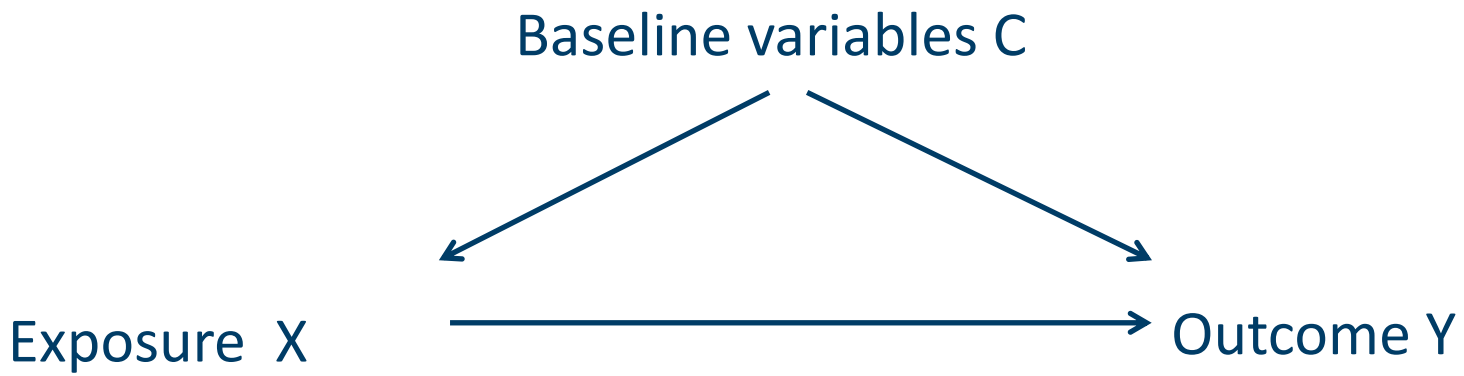
Leids Universitair Medisch Centrum

Causal inference I

Week 5: adjusting for confounding: propensity score methods

Confounding in observational studies

exposure groups are not comparable at baseline (start of study)



- Confounding may introduce spurious associations between exposure and observed outcome.
- Backdoor path in DAG
- No exchangeability: *Y(1), Y(0)* $\not\perp\!\!\!\perp$ *X*
  - X is not independent of $Y(x)$, for all x

Challenge: how to handle confounding

How to estimate causal effects from observational data?

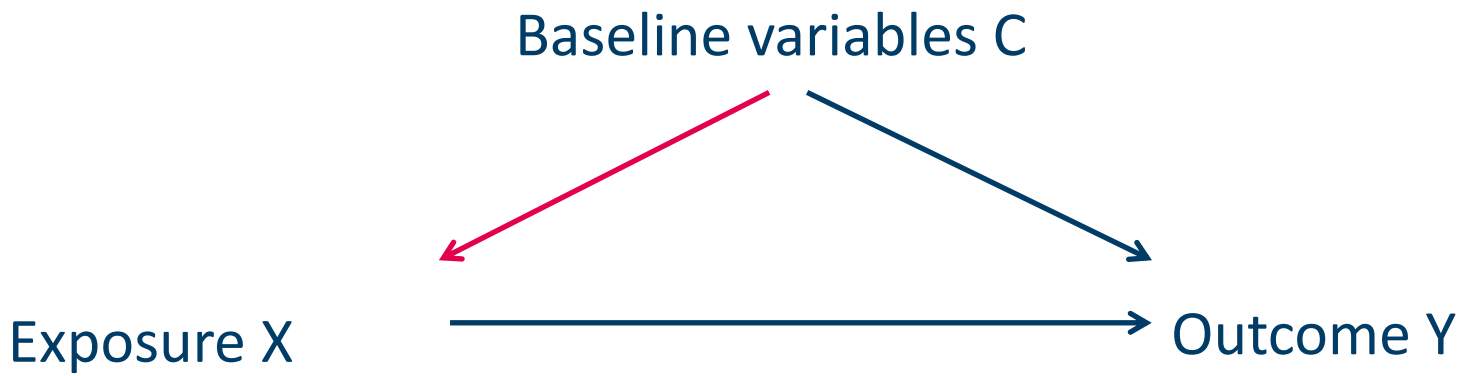
1. Try to identify all confounders and adjust for them

- Stratification
- Traditional regression methods.
- G-computation (Outcome regression + standardization)
- Propensity score methods (inverse probability weighing, matching, stratification) Today
- Double robust methods

2. Pseudo-randomization/ natural experiments

- Instrumental variable analysis
- Regression discontinuity design

Controlling for confounding using propensity score methods



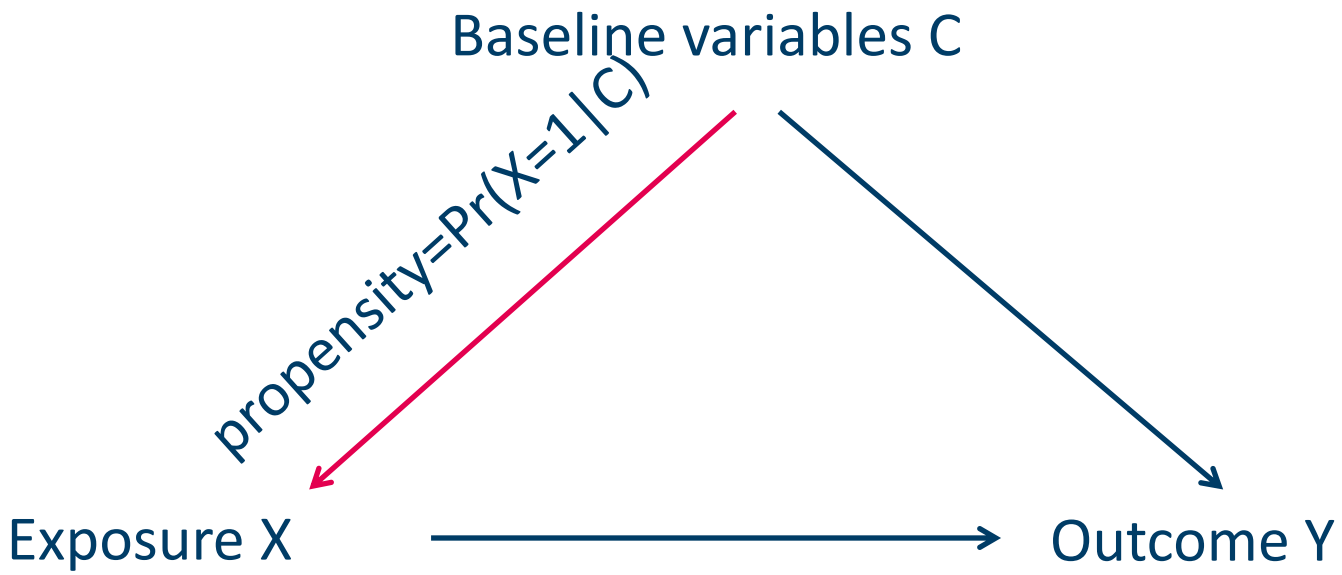
Model the relation between the baseline variables (C) and the exposure (X)

Definition of propensity scores

- Rosenbaum and Rubin (1983)
- Some patients have a higher probability to receive a certain exposure (a higher “propensity”)
- Summarize the information of the confounders as a propensity score
- This is the probability to receive exposure X=1, given baseline variables C
- Propensity score: $Pr(X=1 \mid C=c)$

Propensity score

Propensity score = Probability to receive X=1, given baseline variables C



Properties of propensity score (PS)

- PS summarizes the information on the baseline variables that is relevant for the exposure
- It is a scalar (number between 0 and 1), independent of the dimension of C
- Exposure assignment is independent of confounders (ignorable) for any given value of PS (as in a randomized study)
- Balancing score: conditioning on the PS value achieves covariate balance
- Note: balance is for measured confounders

→ The propensity score is sufficient to adjust for measured confounding

How to use a propensity score?

Compare treated and untreated individuals with the same propensity score

This can be done in different ways

1. Weighting
2. Matching

3. Stratification
   - basic comparisons within subclasses of patients with `similar' PS
4. Propensity score as covariate in outcome regression model

Weighting

Propensity score weighting

Idea: construct a pseudo-sample in which there are no imbalances on measured covariates between the exposure groups.

- Many people with a given characteristic who are exposed, and few unexposed, then $\Pr(X_i = 1 | C_i)$ will be large.
- These individuals will be downweighted in the exposed group and upweighted in the non exposed groups.
- Idem, individuals with low $\Pr(X_i = 1 | C_i)$ will upweighted in the exposed group and downweighted in the unexposed group

Weights

Use weights based on propensity score

- $1/\Pr(X_i = 1 \mid C_i)$ if exposed
- $1/(1 - \Pr(X_i = 1 \mid C_i)) = 1/(\Pr(X_i = 0 \mid C_i)$ if unexposed

Inverse probability weights (IPW)

Example: propensity of exposure is 0.2

- Exposed, weight is 1/0.2 = 5
- Not exposed, weight is 1/(1-0.2) = 1/0.8=1.25

Example

Binary confounder, binary exposure and binary outcome

C	0				1			
(n)	300				200			
X	0		1		0		1	
(n)	200		100		100		100	
Y	0	1	0	1	0	1	0	1
(n)	180	20	80	20	50	50	40	60

Receiving exposure X depends strongly on confounder C.

Outcome Y depends on X and C

Table

C	X	Y	n
0	0	0	180
0	0	1	20
0	1	0	80
0	1	1	20
1	0	0	50
1	0	1	50
1	1	0	40
1	1	1	60

Relation between exposure and confounder

Original data

	C=0	C=1	
X=0	200	100	300
X=1	100	100	200
	300	200	500

Exposure probabilities

$Pr(X = 1) =$

$Pr(X = 1|C = 1) =$

$Pr(X = 1|C = 0) =$

Relation between exposure and confounder

Original data

	C=0	C=1	
X=0	200	100	300
X=1	100	100	200
	300	200	500

exposure probabilities

$Pr(X = 1)$ = 200/500 =0.4

$Pr(X = 1|C = 1)$ = 100/200=0.5

$Pr(X = 1|C = 0)$ = 100/300=1/3

Weights

X=1, C=1:

X=0, C=1:

X=1, C=0:

X=0, C=0:

Relation between exposure and confounder

Original data

	C=0	C=1	
X=0	200	100	300
X=1	100	100	200
	300	200	500

exposure probabilities

$Pr(X = 1)$ = 200/500 =0.4

$Pr(X = 1|C = 1)$ = 100/200=0.5

$Pr(X = 1|C = 0)$ = 100/300=1/3

Weights

X=1, C=1: 1/0.5=2

X=0, C=1: 1/0.5=2

X=1, C=0: 1/(1/3) =3

X=0, C=0: 1/(2/3) = 3/2=1.5

Relation between exposure and confounder

Original data

	C=0	C=1	
X=0	200	100	300
X=1	100	100	200
	300	200	500

exposure probabilities

$Pr(X = 1)$ = 200/500 =0.4

$Pr(X = 1|C = 1)$ = 100/200=0.5

$Pr(X = 1|C = 0)$ = 100/300=1/3

Weights

X=1, C=1: 1/0.5=2

X=0, C=1: 1/0.5=2

X=1, C=0: 1/(1/3) =3

X=0, C=0: 1/(2/3) = 3/2=1.5

Reweighted data

	C=0	C=1	
X=0			
X=1			

Relation between exposure and confounder

Original data

	C=0	C=1	
X=0	200	100	300
X=1	100	100	200
	300	200	500

exposure probabilities

$Pr(X = 1)$ = 200/500 =0.4

$Pr(X = 1|C = 1)$ = 100/200=0.5

$Pr(X = 1|C = 0)$ = 100/300=1/3

Weights

X=1, C=1: 1/0.5=2

X=0, C=1: 1/0.5=2

X=1, C=0: 1/(1/3) =3

X=0, C=0: 1/(2/3) = 3/2=1.5

Reweighted data

	C=0	C=1	
X=0	300	200	500
X=1	300	200	500
	600	400	1000

Exchangeability in the reweighted data

Reweighted analyses

Estimated average potential outcome under no exposure

$\hat{E}[Y(0)] =$

$$\frac{0*180*1.5+1*20*1.5+0*50*2+1*50*2}{180*1.5+20*1.5+50*2+50*2} = \frac{130}{500}=0.26$$

$\hat{E}[Y(1)] =0.36$ (check it yourself)

Estimated ATE (average treatment effect) = 0.10

C	X	Y	n	weight
0	0	0	180	1.5
0	0	1	20	1.5
0	1	0	80	3
0	1	1	20	3
1	0	0	50	2
1	0	1	50	2
1	1	0	40	2
1	1	1	60	2

Estimation of propensity score

How to estimate the propensity score if there are several confounders?

How to estimate $\Pr(X = 1 | \mathrm{C})$?

- Usually estimated by logistic regression
  - Outcome is exposure X
  - Covariates: C
- Alternative: machine learning techniques

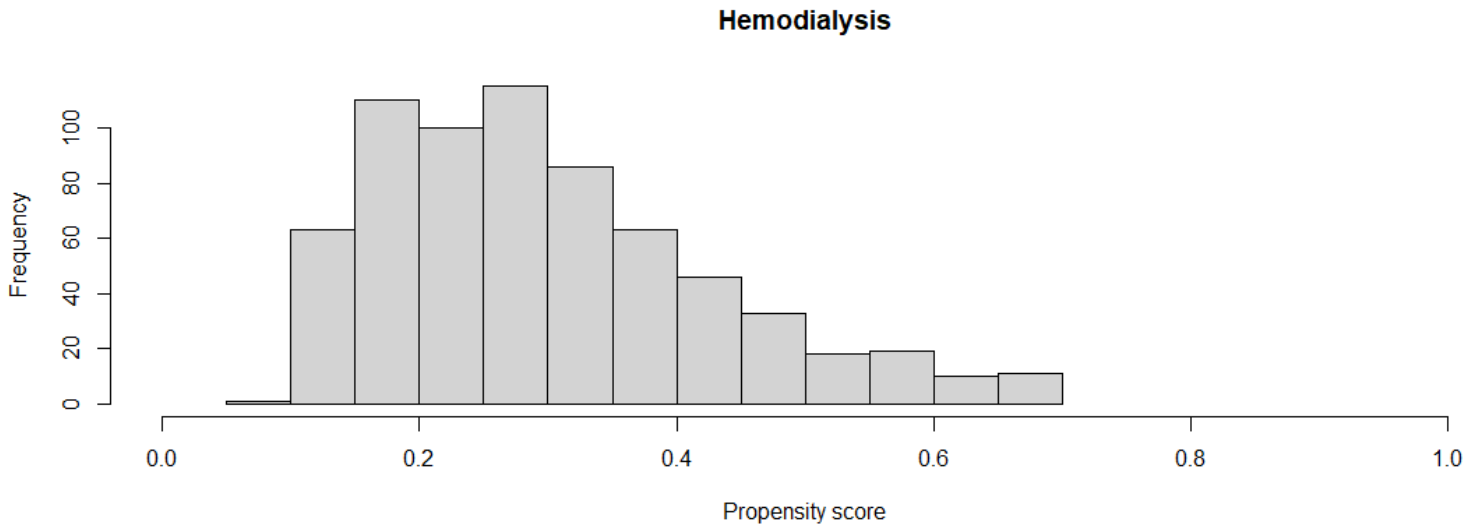
The kidney disease example

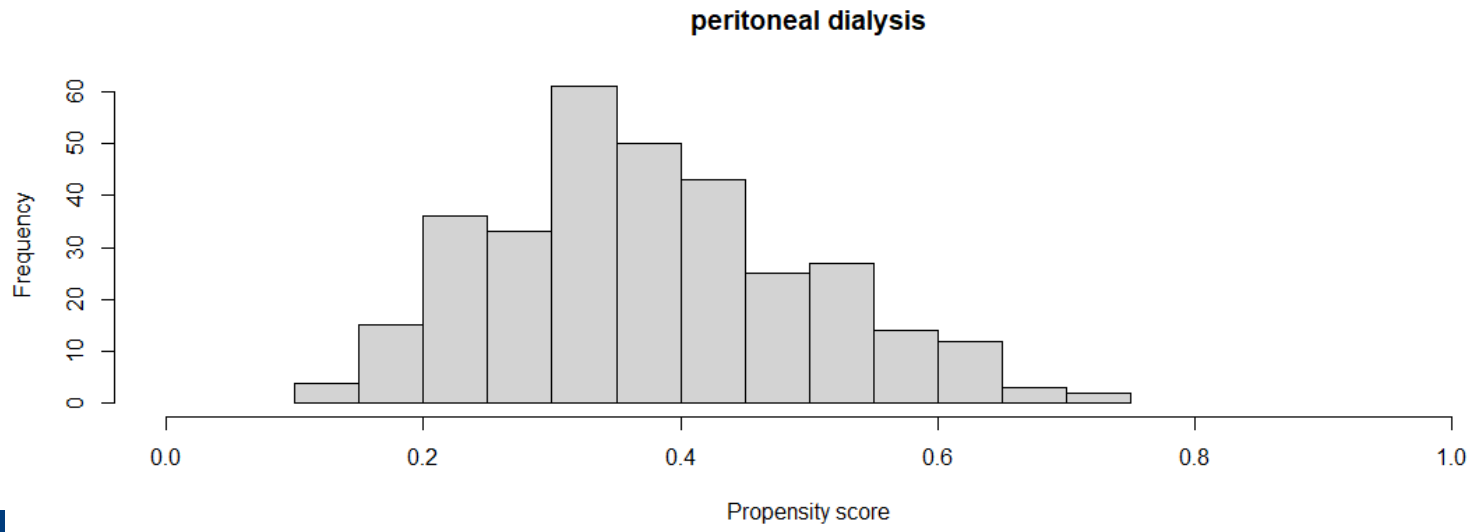
Propensity score:

Logistic regression model for

Pr (peritoneal| age, gender, physical function at baseline)

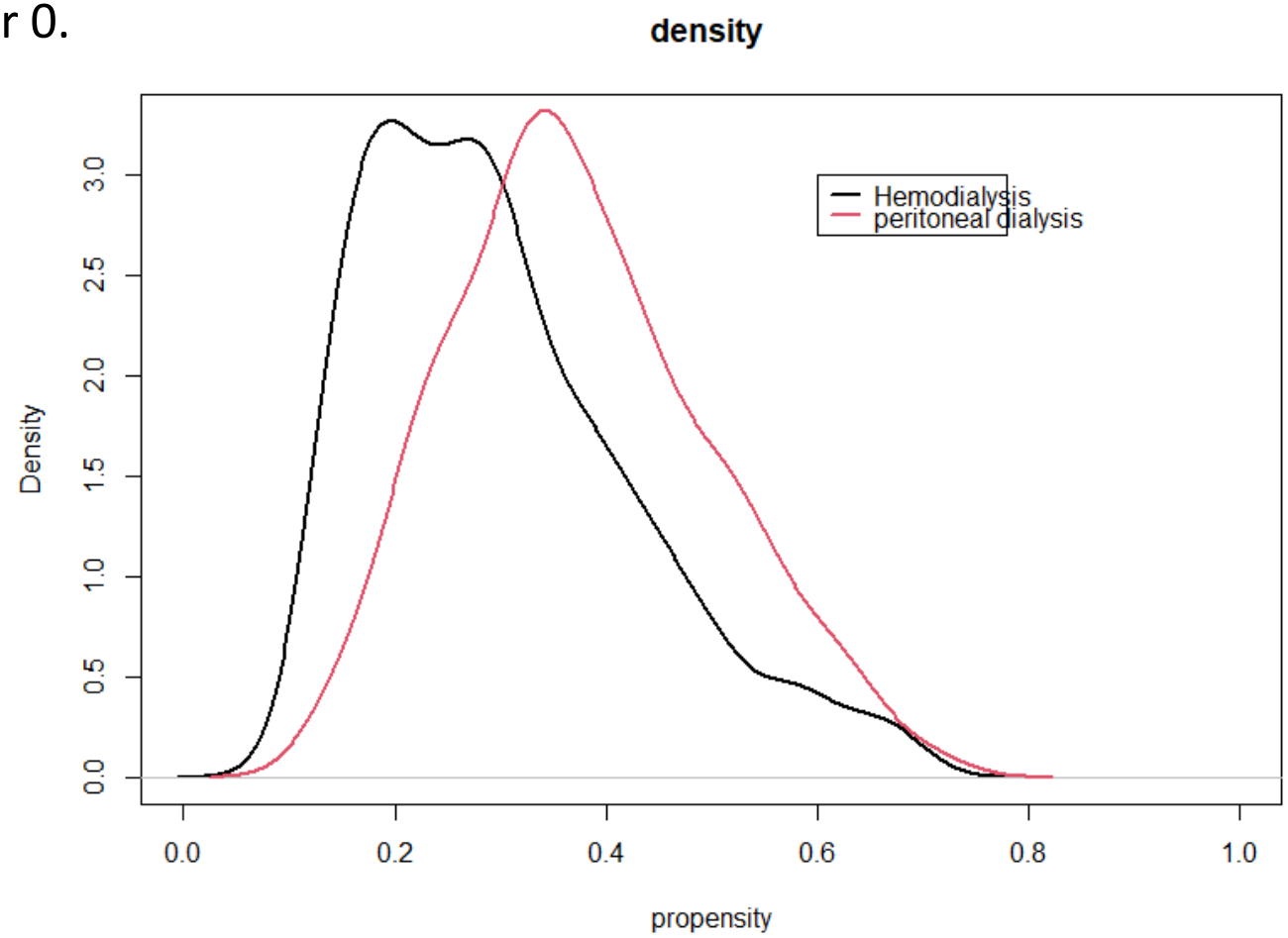
Histograms of propensity scores





Checking the positivity assumption

Distribution of PS in exposed and unexposed should overlap. No PS values close to 1 or 0.

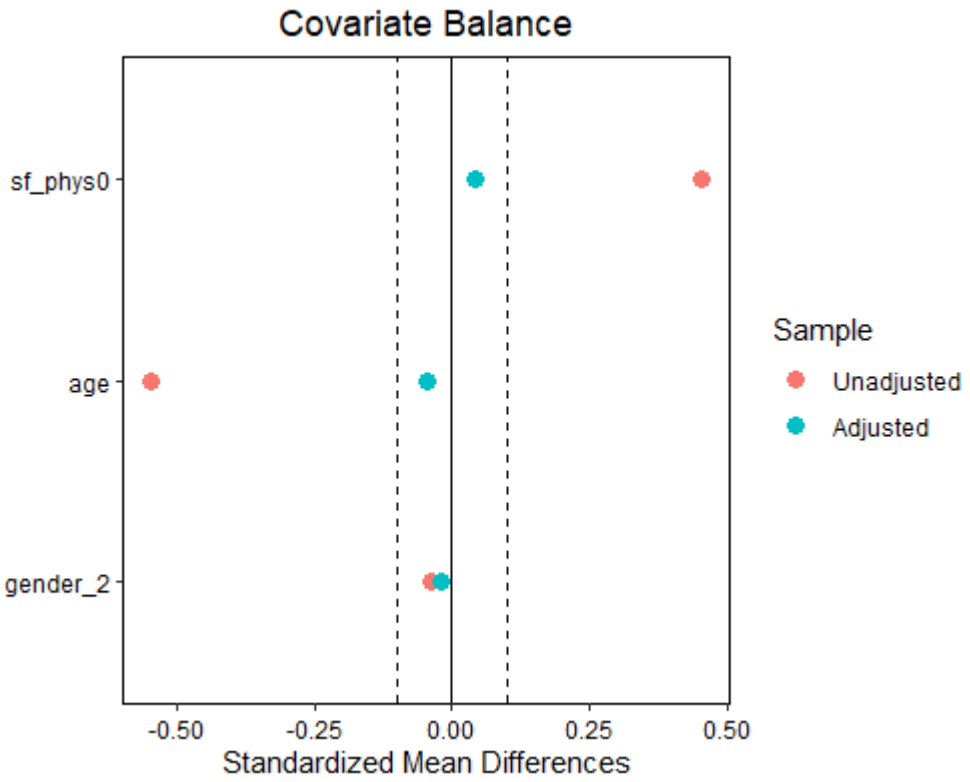


There should be balance in the covariables after reweighting

Often checked by looking at Standardized Mean Difference (SMD)

$$\text{SMD} = \frac{\bar{C}_{x=1} - \bar{C}_{x=0}}{SD_{pooled}}$$

Rule of thumb: SMD should be < 0.1 after reweighting



Propensity score estimation with reweighting (IPW)

1. Fit propensity models: fit a (e.g., logistic regression) model for the probability of being exposed.
2. Determine the weights:
   a. Use models in (1) to predict the probability that a person received the exposure he did in fact receive.
   b. Set each individual's weight to one over the probability computed in (2a). Optionally: truncate weights.
   c. Check for balance in the weighted sample.
      No balance: improve propensity model (go back to 1)
3. Weighting each individual by the weights computed in (2b), then perform standard analysis to relate exposure to outcome
   (e.g., calculating difference in mean outcome or fit regression model for the response given exposure).
4. Calculate se's and 95% CIs accounting for weighting (bootstrap or analytically).

Notes on propensity weighting

- Relations with survey sampling, where some groups are oversampled

- Observations with very high weight are very influential → Often large weights (e.g.>10, or > 99th quantile) are truncated
- Standard errors can be calculated analytically ("robust"methods, Sandwich methods), or by bootstrap (see r-causal.org 12.3)

- Standard weighting will yield an estimate of the ATE, the difference in the population.

- By changing the weights, other estimands, like treatment effect in the treated (ATT) can be estimated.

Nice to knows

- Methods can be extended to treatment regimes over time (marginal structural models
- Weights may be stabilized, by accounting for Pr(X=1) in numerator of weights.

Matching

Matching

Matching:

- Pair each individual in one exposure group to an individual with the most similar covariates from the opposite (unexposed) group

- One option is to match on PS values (but it is not the only way).

- The goal is to make X and C independent in the matched sample.

There are many different ways to perform matching

Matching to obtain ATT with PS

Assume we have n_1 exposed individuals and n_0 unexposed individuals

- Pair the n_1 exposed individuals to n_1 selected unexposed individuals with the closest value of the propensity scores;

- (For 1-1 matching, without replacement) The matched sample consists of $2n_1$ individuals, n_1 exposed and n_1 unexposed.

Matching methods, further specifications

- Common support restrictions. Discard individuals with very large or very small propensity.

- Use a caliper. Match only if the difference in (logit) propensity score is below a threshold (the caliper).

Note: these methods may reduce residual confounding, but the estimand no longer corresponds to the ATT (or ATE), but to the effect of the exposure in the subgroup of matchable individuals. (see r-causal.org 11.2.4 and 11.2.5)

Steps in matching using the Propensity score

1. Estimate a PS-model.
2. Evaluate the overlap (positivity assumption). If needed, truncate the study population accordingly (and report the truncation).
3. Implement a matching method .
4. Evaluate the quality of the matched sample (balance checks, empirical distribution plots).
5. If balance is not satisfactory, go back to step 1.

The matched subsample can be analysed as an RCT

Remarks on matching

Many choices to be made:

- 1-1 matching or 1-more matching?
- With of without replacement (can observation be a match more than once?)
- Choice of matching criterion
- Choice of matching method
- Maximum allowed distance in (logit) propensity score (the “caliper”)
- Trade off between residual confounding and variance due to removing unmatched individuals

General advice: try out and compare balance for different methods (without looking at effect on outcome)

Matching

Advantages

- Free from modelling of X -Y relationship.
- Poor overlap evident by lack of matching
- Easily described to non-technical audiences

Disadvantages of matching:

- Residual confounding (if imperfect matching).
- Subjectivity of choices.
- Inefficient (1:1 matching).

How to build a propensity score?

Variables to include in propensity model

- Variables which affect both exposure and outcome (confounders) should be included in propensity score
- Using variables only related to exposure, gives less precise results
- Using variables only related to outcome can increase precision of results

What is a good propensity score?

The propensity score should balance the data

Aim is not to build a good prediction model for the exposure

- Do not use automatic selection procedures

AUC (area under the curve) of the propensity model is not a measure of quality of propensity score

- Overlap between exposure groups is desirable
- In a randomized trial, the AUC of propensity model will be close to 0.50

When is a propensity score useful ?

In follow-up study with:

- Rare outcome(s)
- Non rare exposure levels
- Multiple confounders

  In this situation: limited possibility to adjust for confounders in outcome model

For matching in follow-up study when only few individuals can be followed and with one treatment/exposure of interest

- Matching on propensity scores gives better balance than matching on a few individual covariates

What to prefer, outcome regression/G-computation or propensity score methods?

Propensity score methods:

- It may be easier to specify a “good" exposure model than a good outcome model.
- Easier to check if there is positivity
- Easy to check if there is balance in baseline variables in the reweighted data
- More flexibility in modelling if outcome is rare
- Matching and weighting do not assume a model for the relation between exposure, confounders and outcome

What to prefer, outcome regression or exposure modelling?

But

- Propensity score methods are statistically inefficient compared to outcome regression.
- For weighting: a small number of large weights can be highly influential
- There are solutions for large weights (e.g. truncation) but they are not statistically principled.
- Matching: residual confounding (if imperfect matching) .
- Matching: Subjectivity of choices.

Of note

Assumptions

- Outcome regression/G computation: the outcome model should be correctly specified
- Propensity score methods: the propensity model needs to be correctly specified

There exist ‘double robust methods’ : they work if at least one of the two models is correct → in course Causal Inference 2