Heva
A DOCAPOSTE company

Bonnes pratiques pour les appariements de populations dans les études comparatives

Nada Assi, Eléonore Herquelot, Arnaud Panès, Fanny Raguideau
Heva, Lyon

Congrès EMOIS, 4 et 5 avril 2024, Lille

Matching can be used in comparative observational studies to make two populations comparable

Here are the questions that then arise

- When to use matching?
- Which variables to include in the matching?
- Who are the eligible controls and how many to use?
- What are the different types of matching?
- When to start and end follow-up for the controls?
- How to assess the quality of the matching?
- Issues related to unmeasured confounders

Causality versus association

Clinical study

Randomization = exchangeable patients (comparable)

Association ~ causation

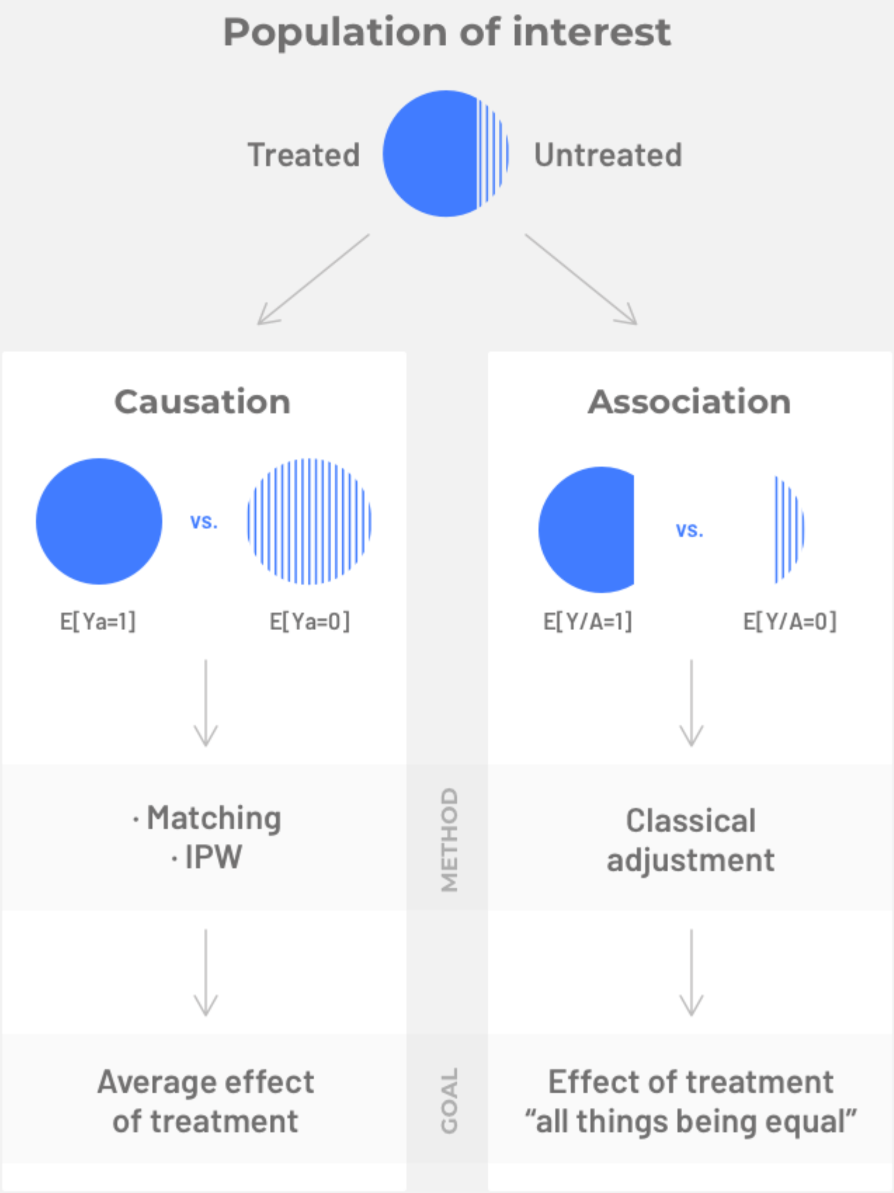
Real-Word study

No choice of treatment

Association ≠ causation because of confounding factors

Several methods to remove bias

→ **Matching reduces the effects of confounding factors (CF) between cases and controls in comparative observational studies**



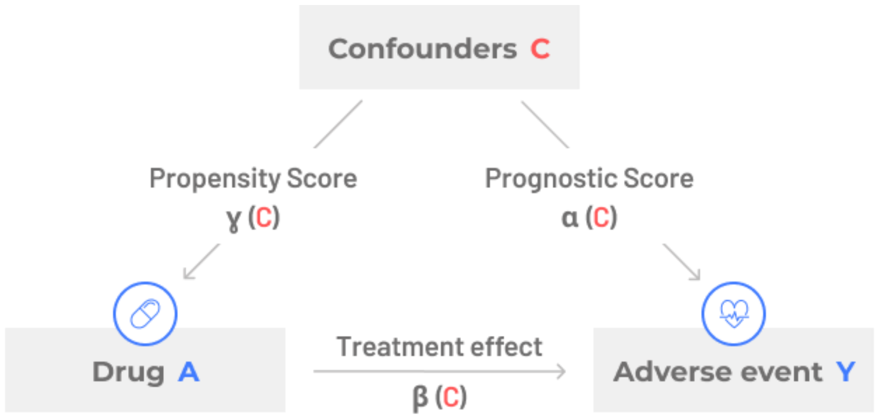
Adjustment methods

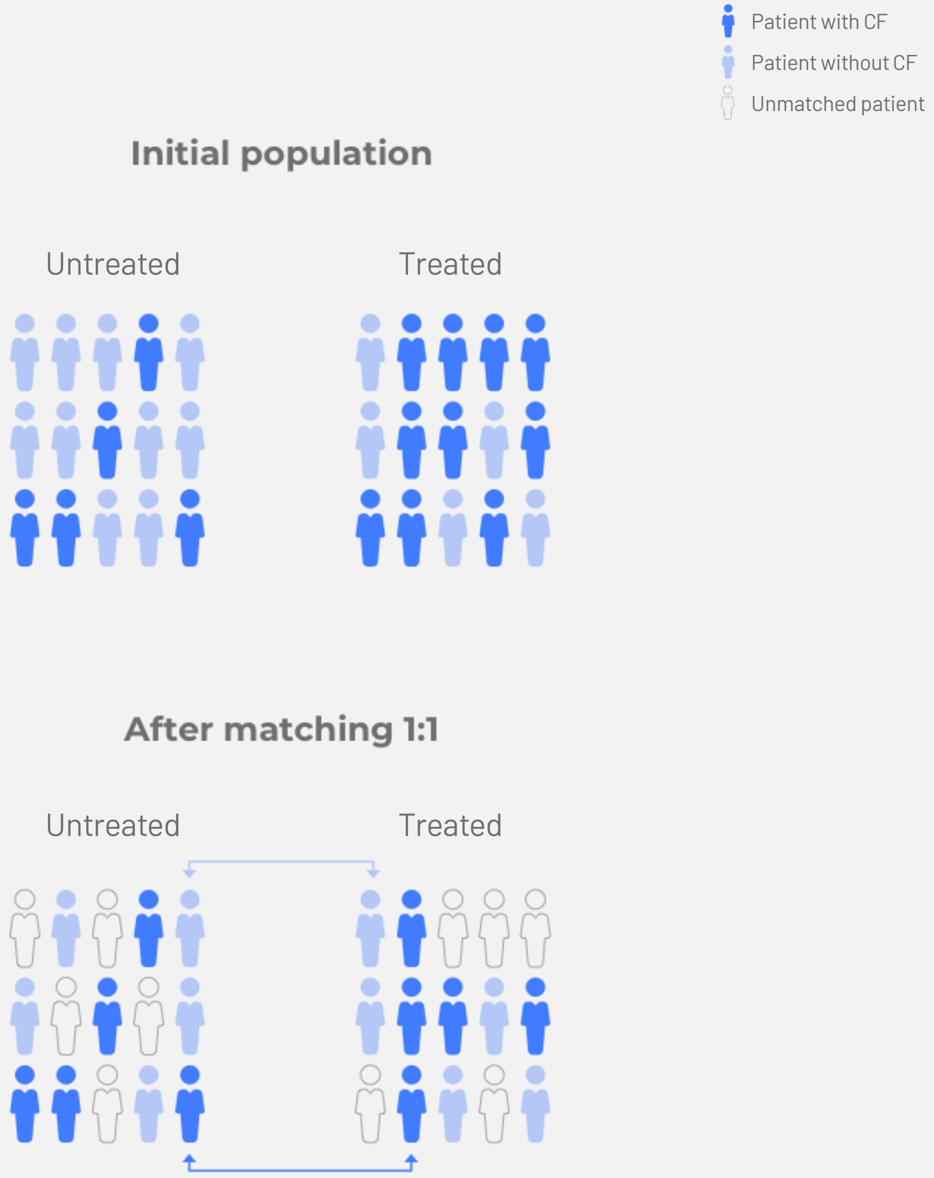
Simple cases

Constant exposure or treatment of interest constant

Absence of time-varying confounding factors

→ **All types of adjustment are possible (adjustment by multivariable models, matching, weighting)**





Examples of confounders in SNDS studies

Age & Gender

Socio-demographics characteristics

Socio-economic characteristics

CMUc, ACS,C2S, social disadvantage index based on ZIP code

Socio-demographics characteristics

Care offers

Departmental density of physicians of a given specialty per 100k inhabitants...

Socio-demographics characteristics

CNAM Cartography

algorithms specific for SNDS datasets to identify the main chronic pathology in France

Medical history

Heva designed algorithms specific to each studies with Scientific Committee

using diagnosis of hospital stays, treatments, medical acts, laboratory exams

Medical history

Other algorithms found in the literature

for example, more than 100 algorithm are available to detect asthmatic patients using Electronic Health Records

Medical history

Treatment or exposure

Characteristics at inclusion

Prescriber or nature of hospital stay

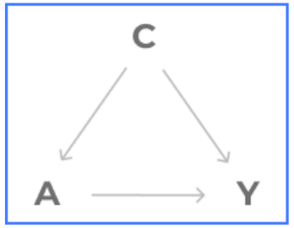
emergency, duration, diagnosis

Characteristics at inclusion

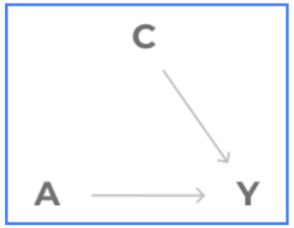
Which variables should be included?

It is tempting to include variables « just in case », however if these are not confounding factors then they can bias the results.

Include

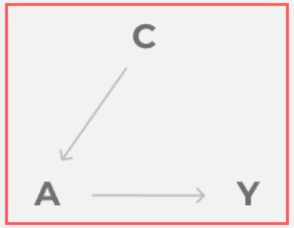


Confounding factor
→ Bias reduction

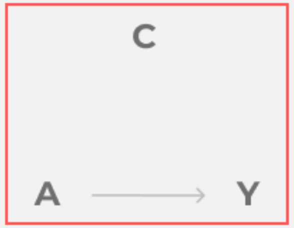


Predictive factor
→ Variance reduction
(not necessary for bias reduction)

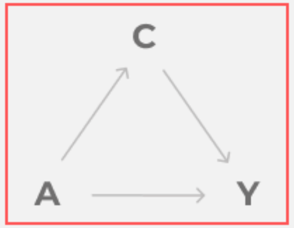
Do not include



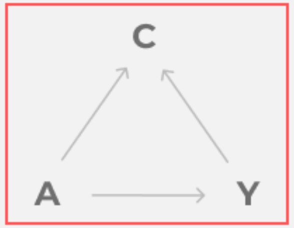
→ Bias increase



C is irrelevant here



C is a mediator
→ Bias increase



C is a collider
→ Bias increase

Who are the controls and how many do we need?

Someone like me overall but not me

Do we need controls?

Subjects can serve as their own control

Subjects present in the cohort that do not develop the outcome of interest or exposure

Use of another reference population (external or random sample)

Number of controls

Ratio 1:1[1]

For rare diseases, the number of controls can be increased to increase power

Increasing the number of controls increases precision but does not correct for selection bias[2]

Little gain beyond 2 controls[3] and rarely useful to go beyond 3[4]

More than 1 control

Fixed matching ratio: keep only those matched with the required number and exclude the other casesets

Variable matching ratio: Keep all casesets despite design imbalance – advised[5] because it removes more bias, outperforms 1:1 matching for treatment effect estimation but this difference in performance decreases with increase of available matches

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1706071/
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2962254/
3. https://pubmed.ncbi.nlm.nih.gov/3961313/
4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2962254
5. Practical Propensity Score Methods Using R, Walter Leite - University of Florida, USA – 2016 SAGE Publications (ISBN: 978-1-4522-8888-8)

Types of matching

Select a population with nearly perfect covariance balance

Direct/Exact Matching (EM)

1-3 CF*

→ Limit to 3 otherwise the number of combinations is too great and limiting

→ Include variables that are related to data structure

Propensity Score Matching

CF* >3

Calculation of a propensity score with SNDS variables using logistic regression for a given ratio

Matching with a greedy algorithm with caliper without replacement

Caliper value usually equal to 20% of the variability of the logarithm of the PS values

Mixed Approach (Exact + PSM)

CF* >3 for PSM / 1-3 CF* for EM

Combines the PSM with exact constraints on key confounders (related to structure or epidemic circulation e.g.)

→ More precise

*CF = Confounding Factor

Critical matching time points

Choice of the time of matching

« Organic» index date for controls:

- e.g., hospitalisation of interest in intervention groups

Absence of index date for controls:

- Fictitious date set at the 1st of January of the year of the case

  \+ easier for covariable computations
  \- but issues related to strange censoring patterns, with peaks at the beginning of the year and for comparability with respect to the pathology under scrutiny)

- Inclusion taking all case index dates for the controls \- can be computationally expensive
- Tip: Inclusion of controls on comparable events or similar pathology histories

Choice of end of follow-up time for controls

- Discontinuation of follow-up is independent between cases and controls
- Discontinuation of the case-control pair at the same time (e.g., if the control dies after 6 months, the case is discontinued at 6 months of follow-up, and vice versa) – e.g., in case of telesurveillance on a determined period, or when cases and controls have pathologies with diverging prognosis, other criteria

Validation of matching

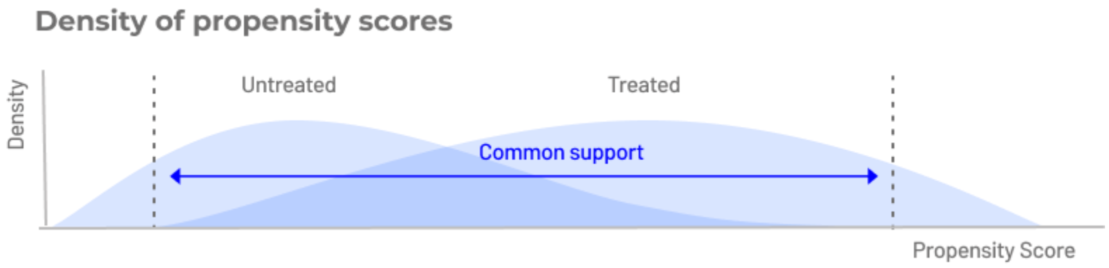
Investigate the PS common support

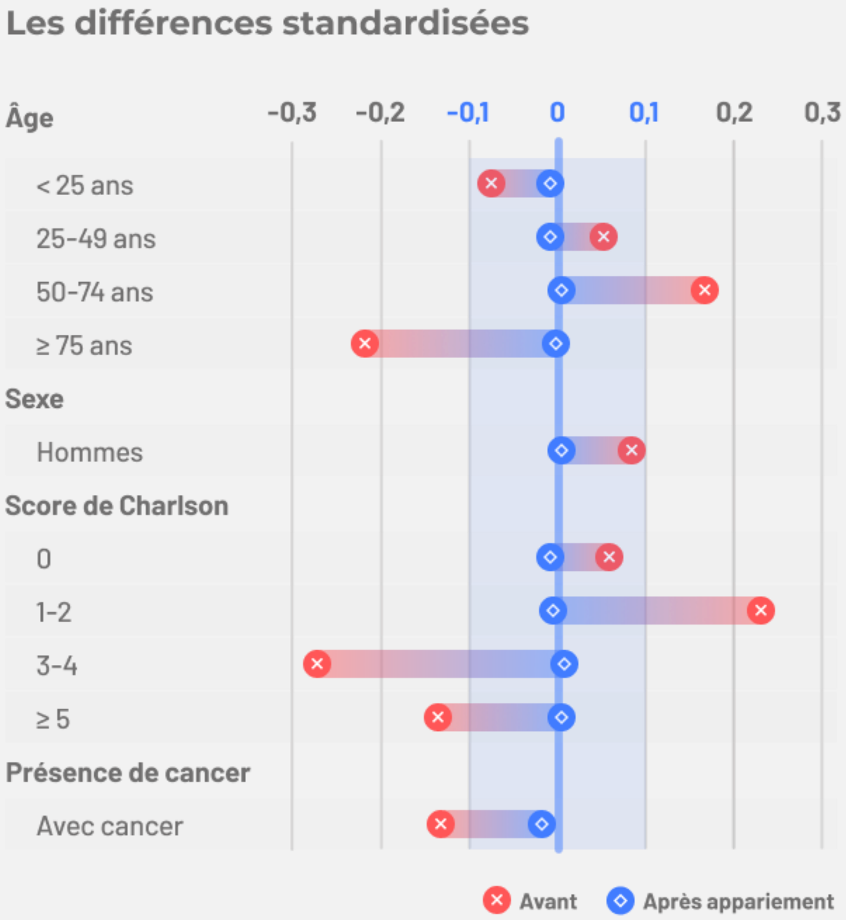
Exclusions may be warranted for extreme PS values (no overlap)

Compute the standardized differences

Good balance obtained if StdDiff < |0.1|

Otherwise consider double robust methods





Beyond Matching

Complex cases

Presence of at least one time-varying confounding factor that could influence the treatment/exposure of interest

Complex interactions between confounding factors

→ **Marginal structural model (MSM) or IPW model**

High dimensional propensity score

Use of all available codes (diagnoses, acts, lab tests,...) to create a propensity score

Automatic screening of variables

→ **HdPS = optimizes the chance to consider all available information in the database (even indirect variables)**

Presence of residual confounding

Prior Event Rate Ratio (PERR) method:

- Addresses confounding due to unmeasured variables
- Ratio between a risk ratio (RR) or hazard ratio (HR) computed during follow-up divided by the RR or HR 'prior' to start of follow-up that constitutes a control or baseline period

Take home message for matching best practices

What to look out for when matching

Design

Highly prevent or rare disease ?

What outcomes will be examined?

What are the potential confounders?

Define the time periods of the study

Controls

How to identify them and how many are needed?

What is their index date and when to stop their follow-up?

Choice of the matching type according to controls and number of confounders

Quality

Compute the standardized differences

Weight them if variable ratio matching

Double robust adjustment if covariate imbalance

Consider sensitivity methods if missing confounders are suspected

→ Remember = the missing confounding factors have a minimal impact on studied associations

→ Matching allows comparing groups of patients at the population level according to global distributions and is not a one-by-one comparison (no need for caseset stratification or conditional modeling unless complex interaction phenomena are suspected
→ Importance of carefully planned matching and potential sensitivity analysis steps with alternative adjustment methods (e.g., IPTW)
→ Choice of appropriate modeling after matching is crucial

Merci pour votre attention

Heva
A DOCAPOSTE company