SANOFI-B&P-Oncology

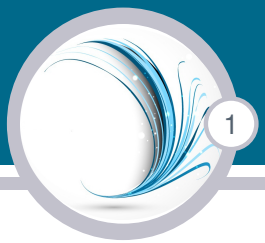
Matching Methodologies for Historical Data

Comparison Study for Matching Methodologies for Historical Data Borrowing

Alice Gosselin
alice.gosselin@sanofi.com

May, 2019

Content



Introduction
Historical data

- Regulatory and pharmaceutical industry requirements
  - Ethical concerns;
  - Large unmet needs;
  - Reduction of costs;
  - Reduction of timelines;
  - Reduction in sample sizes;
  - Gold standard: randomized clinical trial (RCT).
- Data environment
  - Increase in data availability;
  - Increase in amount of data.

$\Rightarrow$ Try to be as close as possible to a randomized clinical trial using historical data

- Key variables?
- Can balance be achieved on key covariates?
- Any condition for treatment assignment?

Introduction
Historical data

- Issue: treatment assignment may depend on covariates
  - If no dependency, one could approximate the treatment effect by difference in means without control on the covariates;
  - But in historical data framework, it is not possible → need to find a way to reduce this dependency.
- Goal: preprocess data to obtain a dataset with treated and control units with similar observed baseline covariates, allowing unbiased estimate of any quantity of interest such as the **Average Treatment effect on the Treated** (ATT)
- Adjusted data could then also be integrated into an informative Bayesian prior

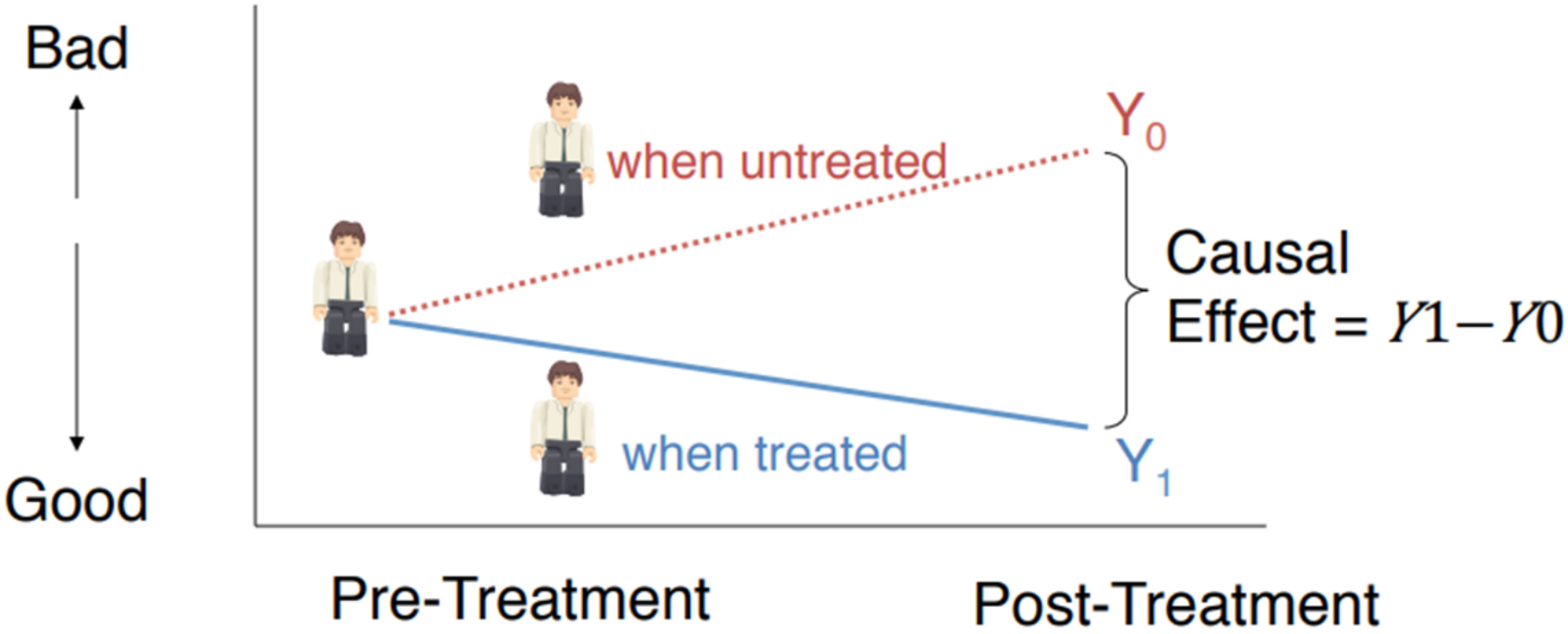
Notations

- Population $N = N_0 + N_1$
  - random sample $n = n_0 + n_1$
- Binary treatment

$$T_i = \begin{cases} 1 & \text{if treated} \\ 0 & \text{otherwise} \end{cases}$$

- X_{ij} value of j^{th} pretreatment characteristic of unit i
  - $[X_{i1}, X_{i2}, ..., X_{iJ}]$ vector of characteristics for unit i
- $f_{X|T=0}$ densities for covariates in the control group
- $f_{X|T=1}$ densities for covariates in the treatment group
- Observed outcome for unit i: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$
- Treatment effect $\tau_i = Y_i(1) - Y_i(0)$

Introduction
Quantity of Interest



$$ATT = E[Y(1) - Y(0) | T = 1]$$
$$= \underbrace{E[Y(1) | T = 1]}_{\mathbf{A}} - \underbrace{E[Y(0) | T = 1]}_{\mathbf{B}}$$

- **A**: estimated by the sample mean of the outcome among the treated units
- **B**: counterfactual mean - not observed, mean of the outcome of the treated units, had they, contrary to the fact, been in the control group
- **B** $\not\approx E[Y(0) | T = 0]$: if there exists covariates leading to treatment assignment, there is a risk that the outcome may be explained by covariates instead of treatment administration

Selection Bias due to observed and unobserved confounders

Strong ignorability of treatment assignment

- No unmeasured confounder: $Y(1), Y(0) \perp\!\!\!\perp T|X$
- Common support: $0 < P(T = 1|X) < 1$

- Given a set of covariates:
  - Treatment assignment and outcome are independent
  - Every patient has a non-zero chance of receiving the treatment

Propensity Score
Framework

- Balancing score $b(X_i)$, function of the observed covariates, such that:

$$X_i \perp\!\!\!\perp T_i | b(X_i)$$

- This balancing score can be taken as the propensity score:

$$ps(X) = P(T = 1|X)$$

→ single composite score of all observed, measured , potential confounders of the association between the treatment and the outcome

- $X \perp\!\!\!\perp T|ps(X) \Rightarrow$ observations with the same propensity score must have the same distribution of observed baseline characteristics independently of the treatment assignment

Propensity Score
Framework

- In case of randomized clinical trial, the propensity score is known
- When no randomization is possible, the propensity score needs to be estimated
  - Logistic regression model - more commonly used: treatment status ~ observed baseline characteristics
  - Generalized boosting method
  - Generalized method of moments

2 Models when using propensity score

- Selection model
  - always includes variables believed to have an impact on the selection process
  - misspecification can have high impact on the control of selection bias
- Outcome model

Propensity Score
Different methods

4 Methods to create a weighted/matched dataset using propensity score

- Matching
- Stratification
- Covariate adjustment
- Weighting

Propensity Score Weighting

- Each unit receives a weight:
  - For treated unit: $w_i = \frac{1}{ps_i}$
  - For control unit: $w_i = \frac{1}{1-ps_i}$

  where ps_i is the propensity score estimated for unit i
- Control influence of patients by weighting their responses with their propensity score

Propensity Score
Covariates choice - Different Approaches

- All measures baseline covariates
- All baseline covariates associated with the treatment assignment variable T
- All covariates that affect the outcome (*potential confounders*)
- All covariates affecting both treatment assignment T and outcome Y (*true confounders*)

Propensity Score
Balance checking

- Is the propensity score model well specified?
- Need to check the balance between treated and control units in terms of baseline covariates **before and after** matching
- First option: look at standardized difference, meaning a sum up difference between means and prevalences
- Second (and *preferred*) option: look at the entire distribution of the covariates

Entropy Balancing
Framework

- Obtain treated and control groups with similar moments of covariate distributions

- No need to verify covariate balance
- Not susceptible to model misspecification
- As for propensity score weighting, there is no need to drop any patients

Entropy Balancing
Balance constraints

- Recall: $ATT = E[Y(1)|T=1] - E[Y(0)|T=1]$
- Estimation of the counterfactual mean:

$$\widehat{E[Y(0)|T=1]} = \frac{\sum_{\{i|T=0\}} Y_i w_i}{\sum_{\{i|T=0\}} w_i}$$

Optimization Problem

- $w_i \mid \min_{w_i} H(w) = \sum_{\{i|T=0\}} h(w_i)$

R moment constraints:
$$\sum_{\{i|T=0\}} w_i c_{ri}(X_i) = m_r \quad r \in 1, ..., R \qquad (1)$$

Normalization:
$$\sum_{\{i|T=0\}} w_i = 1 \qquad (2)$$

Normalization:
$$w_i \geq 0 \quad \forall i|T=0 \qquad (3)$$

Entropy Balancing
Balance constraints

- $h(.)$ is a distance metric such as Kullback divergence

$$h(w_i) = w_i log(w_i/q_i)$$

- q_i are base weights, usual choice is uniform $q_i = 1/n_0$
- $h(.) \rightarrow$ loss function measuring the distance between the distribution of estimated control weights and the distribution of the base weights

- m_r: rth order moment of a given variable X_j from the treated group
- $c_{ri}(X_{ij})$: moment function for control group for a given variable X_j

$$c_{ri}(X_{ij}) = X_{ij}^r \quad \text{or} \quad (X_{ij} - \mu_j)^r$$

- Entropy balancing: no need to check for balance because the weights are estimated **directly using balance constraints**

Simulations
Introduction

- Approaches: **propensity score weighting** / **entropy balancing**
- Models: mis-specified model or specified model
- Measurement: mean Overall Response Rate

	Active Treatment	Control/SoC
Current Study	Group A	Group B
Historical Study	Group C	Group D

- Group A: patient-level data
- Group B: no data
- Group C: no data available
- Group D: patient-level or aggregated data

- Utlimate goal: compare Group A and Group B
- Simulations goal: preprocess the data through simulation of Group B and D

- Simulated variables:
  - trt: treatment indicator
    - 1: if patients are from *control* group of the **current clinical trial**
    - 0: if patients are from *control* group of the **historical clinical trial**
  - X1, X2: covariates
  - Y: outcome
- Two simulated datasets will be used: one for propensity score, one for entropy balancing (based on the previous one)

Propensity score requires patient-level data, while entropy balancing works also if historical data are only aggregated data

Simulations
Introduction

- $ORR_{histocontrol}$ = mean in response rate in the historical control group
- $ORR_{currentcontrol}$ = mean in response rate in the current control group
- Check imbalance between both groups in terms of X_1 and X_2

X1		X2	
historical	current	historical	current
0.2446	0.757	0.2072	0.1536

- Notice a high difference in mean of $X1$ between the two groups, this difference is smaller but still visible for mean of X_2

$\Rightarrow$ Use propensity score or entropy balancing in order to correct this imbalance

Simulations
Propensity Score

- Propensity scores are calculated based on a logistic regression
  - Model 1 - mis-specified: $trt \sim X_1 + X_2$
  - Model 2 - specified model: $trt \sim X_1 + X_2 + X_1 * X_2$
- Weight for each unit
  - For patients in current control group: $w_i = \frac{1}{ps_i}$
  - For patients in historical control group: $w_i = \frac{1}{1-ps_i}$
- Weighted outcome $Y_{w,i} = Y_i * w_i$

Propensity Score

- Weighted mean of Y_w in historical control group

$$ORR_{histocontrol-ps} = \frac{\sum_{i \in gp\ 0} Y_{w,i}}{\sum_{i \in gp\ 0} w_i}$$

- Weighted mean of Y_w in current control group

$$ORR_{currentcontrol-ps} = \frac{\sum_{i \in gp\ 1} Y_{w,i}}{\sum_{i \in gp\ 1} w_i}$$

Entropy Balancing

- Weights are created for each patient from current control group
- $ORR_{histocontrol-eb} = ORR_{histocontrol}$
- $ORR_{currentcontrol-eb}$ = weighted mean of Y_{eb} in current control group

Differences of interest

$$diff_{ps} = ORR_{histocontrol-ps} - ORR_{currentcontrol-ps}$$

$$diff_{eb} = ORR_{histocontrol-eb} - ORR_{currentcontrol-eb}$$

Simulations

Comparison - 10 000 simulations - Mis-specified model

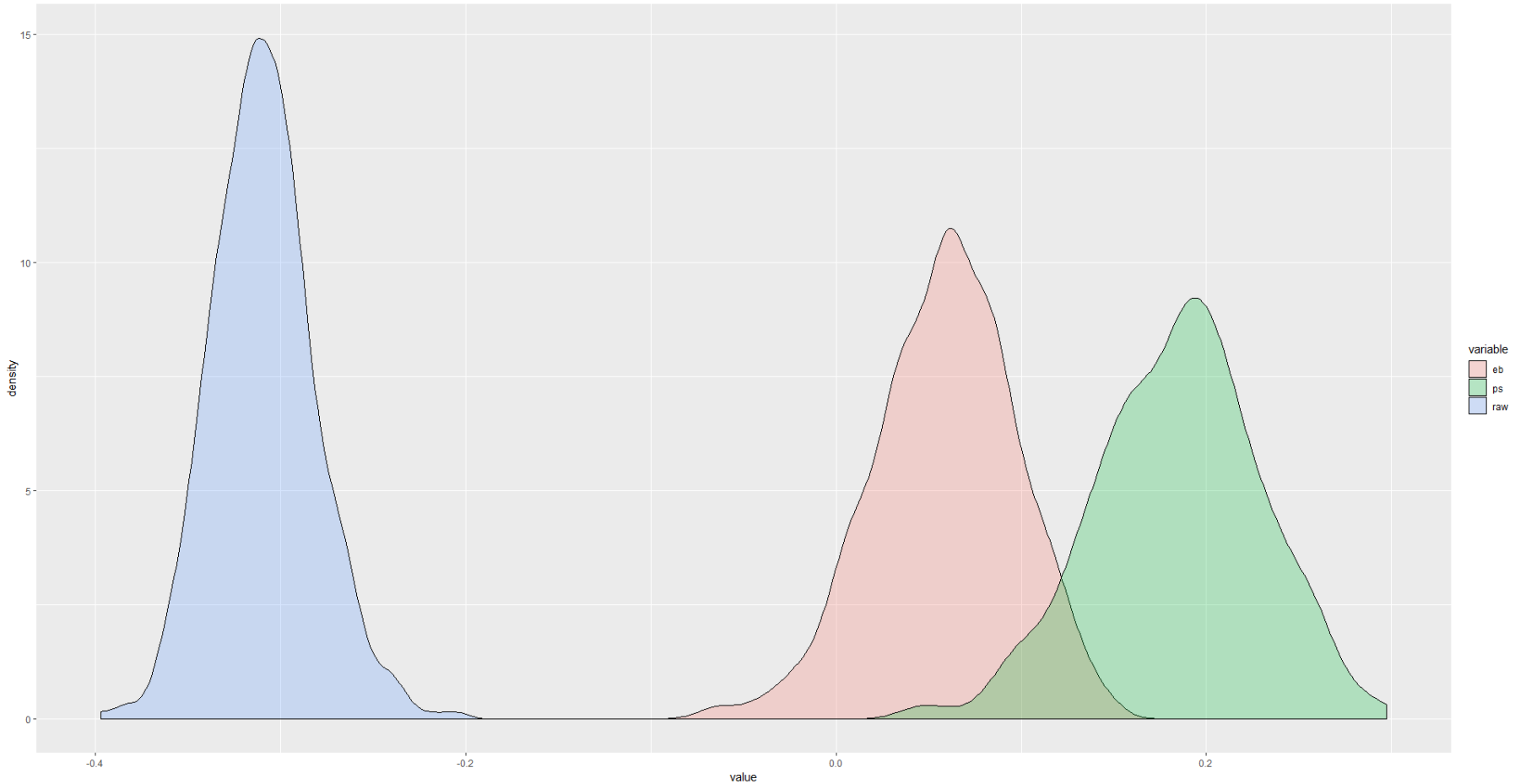


Figure: Distributions of difference in ORR between historical and current control groups, for raw data, propensity score and entropy balancing

Simulations

Comparison - 10 000 simulations - Specified model

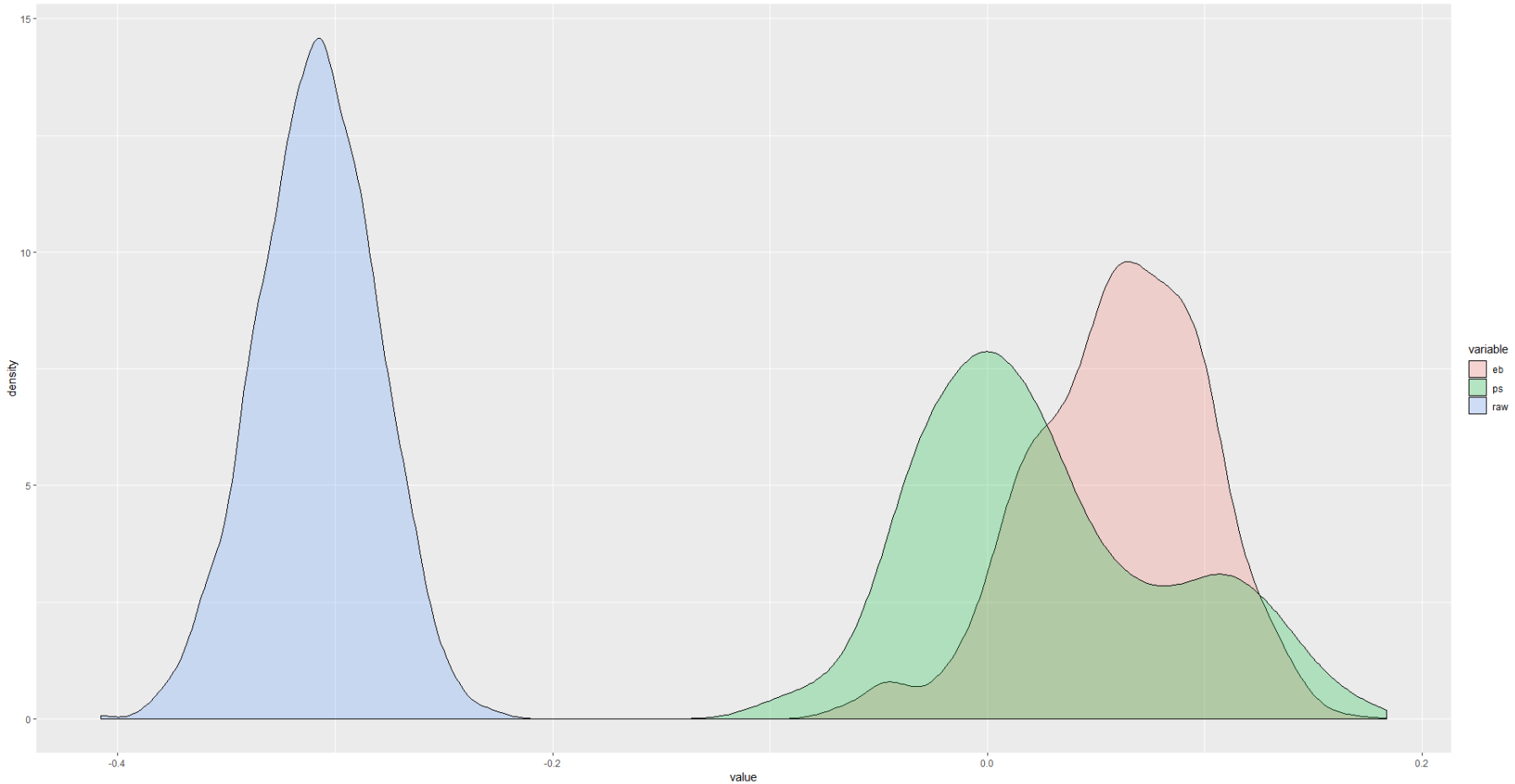


Figure: Distributions of difference in ORR between historical and current control groups, for raw data, propensity score and entropy balancing

Simulations

Comparison - 10 000 simulations - Specified model - Different set of parameters

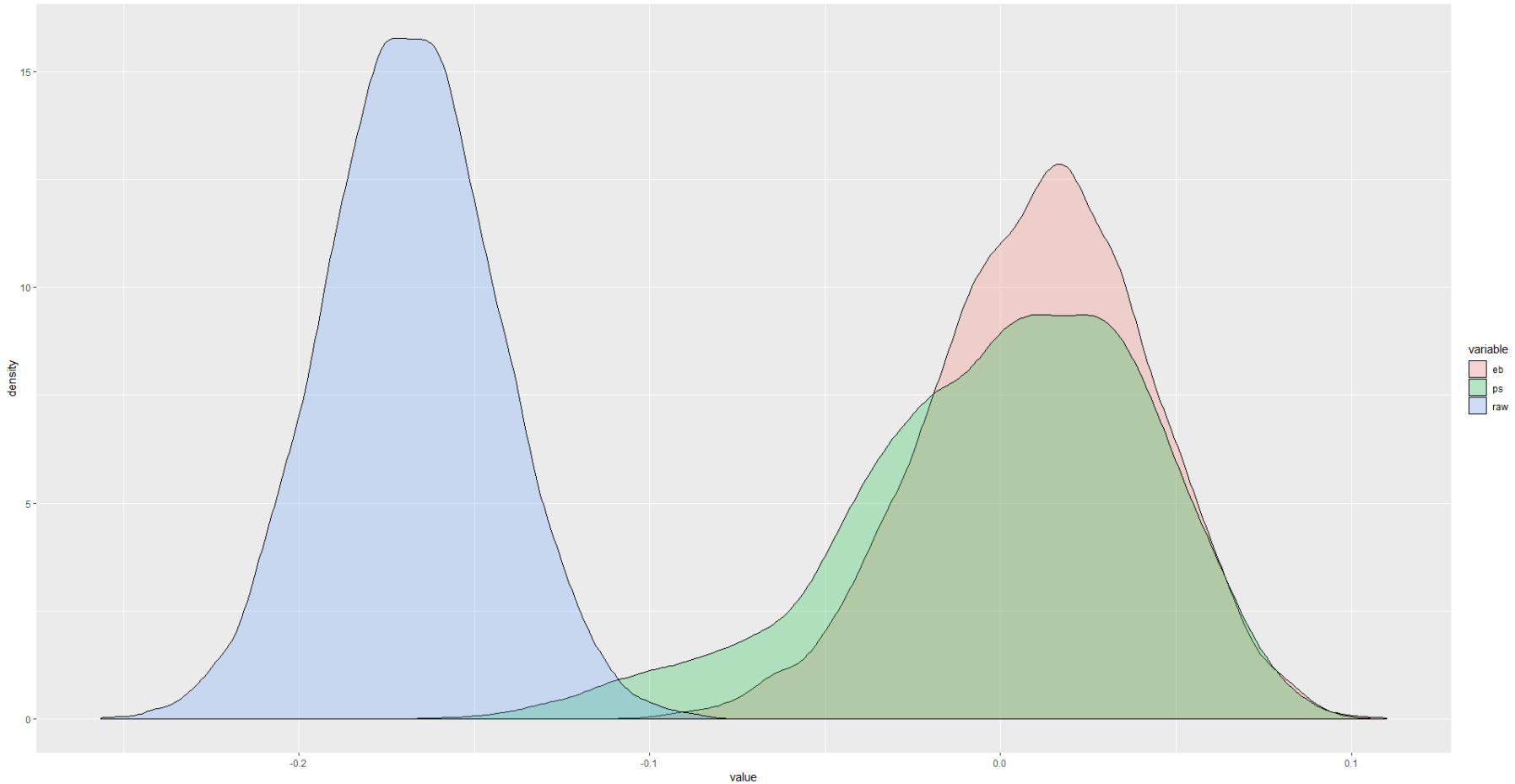


Figure: Distributions of difference in ORR between historical and current control groups, for raw data, propensity score and entropy balancing

Conclusion - Advantages and Potential Limitations

- For PS weighting and EB, all patients are used rather than only matched patients as for other PS methods
- Weights can be easily integrated to any further statistical analysis
- For EB
  - No balance checking needed
  - Not sensitive when model is mis-specified
  - Works with aggregated historical data

- For PS
  - Sensitive to mis-specification of the selection model
  - Unmatched treated units disregarded
  - Requires patient-level data for both historical and current studies
- For EB: optimization problem with inconsistent balance constraints or no solution because of limited data
- If limited overlap treated/control groups $\Rightarrow$ control units with very high weights $\Rightarrow$ increased variance of the analysis

Conclusion - Advantages

Key messages

- Patient-level data available for historical study and enough confidence in the specification of the model ⇒ Propensity score
- Patient-level data available for historical study but not enough confidence in the specification of the model ⇒ Entropy balancing
- Only aggregated data available for historical study ⇒ Entropy balancing

Thank you for your attention !

Example
Selection Bias

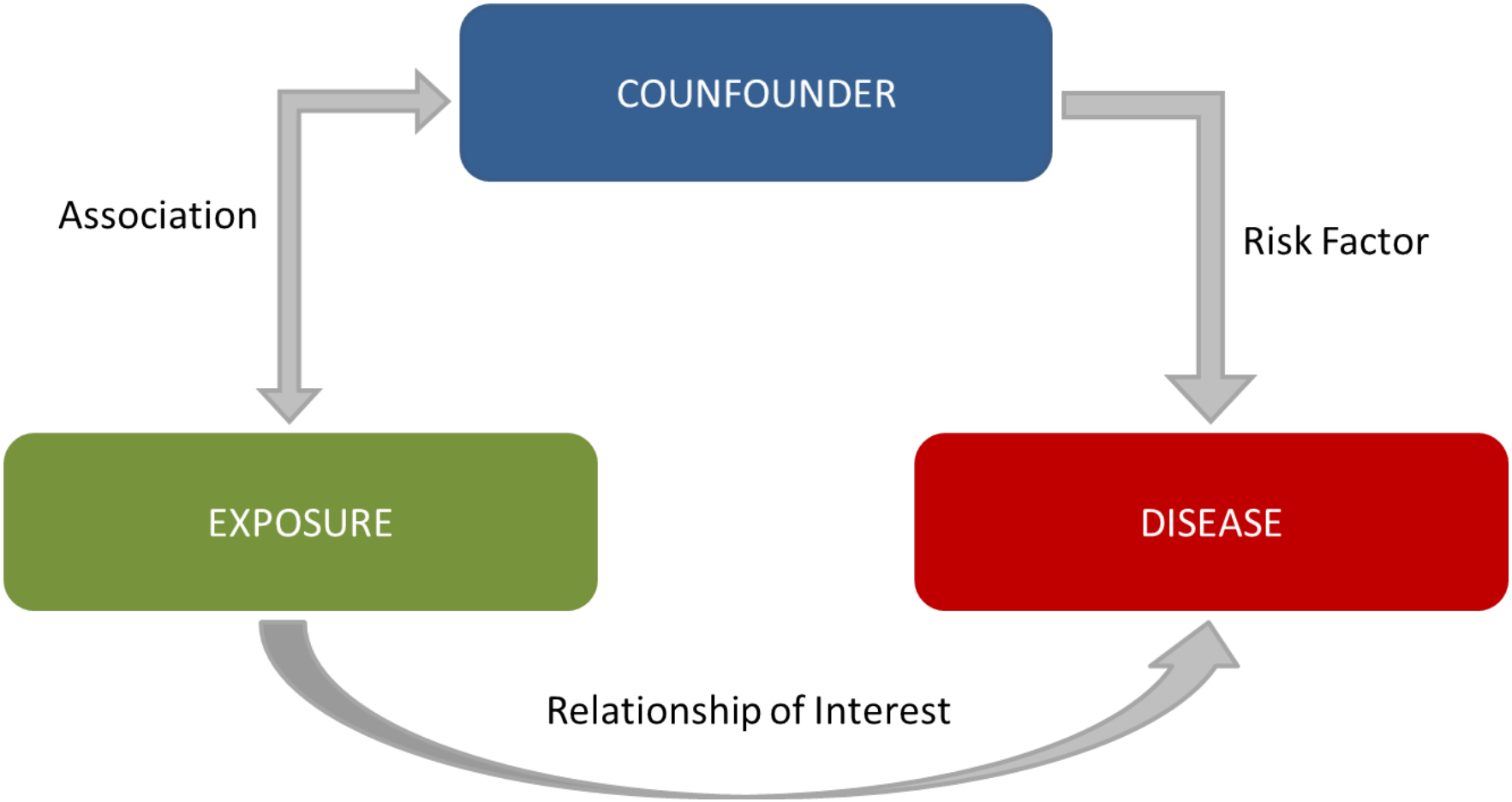
Context

- Car company A
- Objective: assess the impact of advertisement on new car sales
- Two groups of users:
  - Some users have looked for cars on-line
  - Other users have not

- Group 1 is more likely to be exposed to an ad from A
- Group 1 is more likely to buy a car regardless of their exposure to the ad, because we already know that users are interested in buying a car due to their on-line research

⇒ **Group 1 has a higher baseline likelihood of buying a car**
⇒ A comparison of exposed/unexposed groups, without consideration of this known difference, would produce an overly optimistic measurement of the effect of the advertisement. (Back)

Introduction
Quantity of Interest



Back

Example
Strong ignorability

- T: exposure to an advertisement
- Y: buy a car from car company A
- X, the vector of covariates, should include an indicator for *search for cars on-line*
- If no such indicator $\Rightarrow$ violation of strong ignorability assumption

Back

Propensity Score
Different Methods

Matching

- How many matches? 1:1 or M:1
- Replacement allowed or not
- Greedy or optimal algorithm?
- How close a match is acceptable?

Stratification

- Create strata based on propensity score with a threshold
- Compare outcome within each stratum
- Control unbalanced sample sizes between strata using weights

Covariate adjustment

- The design and the analysis of the study are not separate
- Requires the outcome variable
- Model: $Y \sim T + ps(X)$

Propensity Score
Different Methods - Matching

- Goal: create sample of matched treated and control units

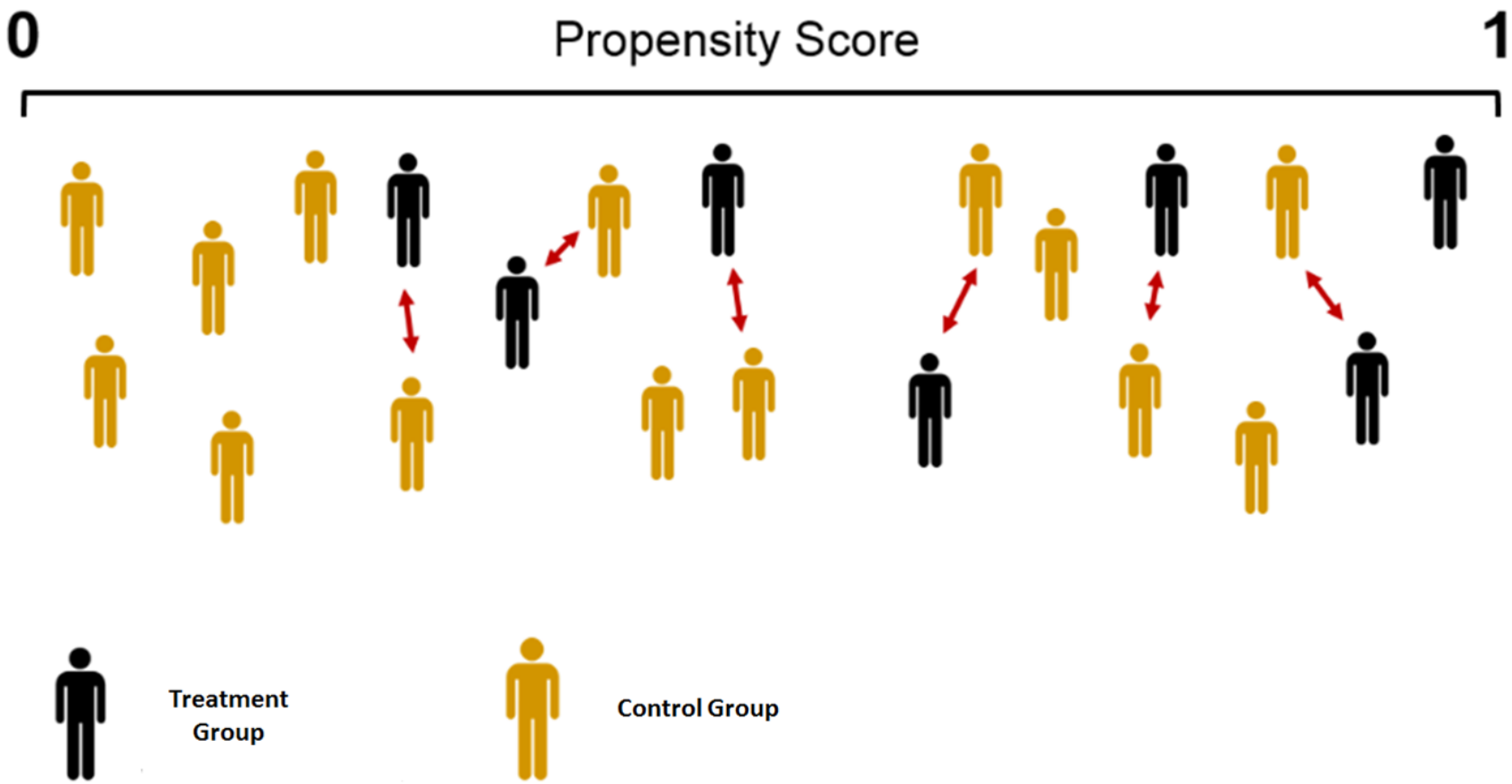
How many matches should be chosen?

- 1:1 - most common: 1 treated unit matched with 1 control unit with the closest propensity score
- M:1 - full matching:
  - 1 treated unit matched with *M* control units or
  - *M* treated units matched with 1 control unit

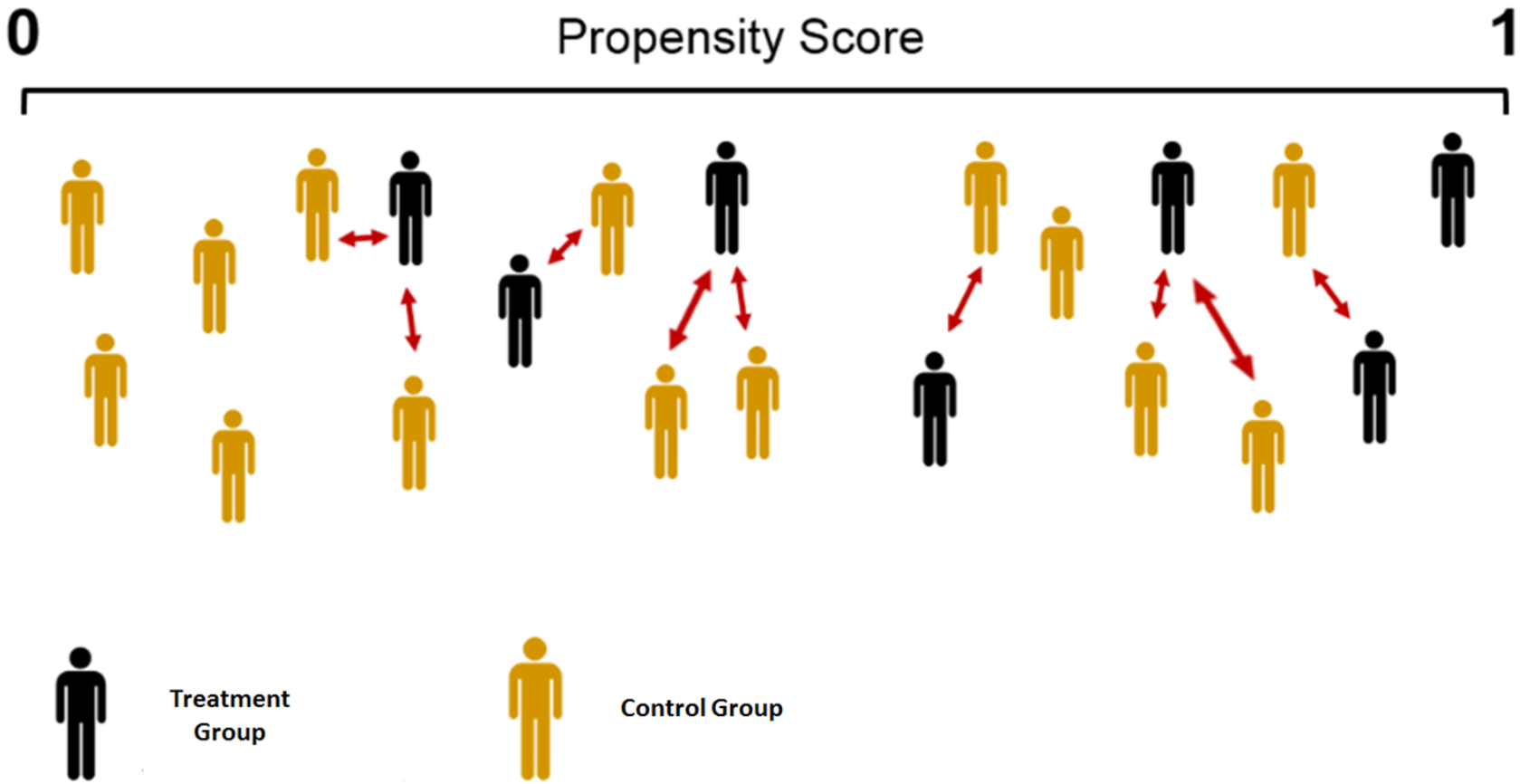
Replacement

- With replacement: one control unit can be matched several times
- Without replacement: once a control unit has been matched to a given treated unit, this control unit cannot be matched again with another treated unit

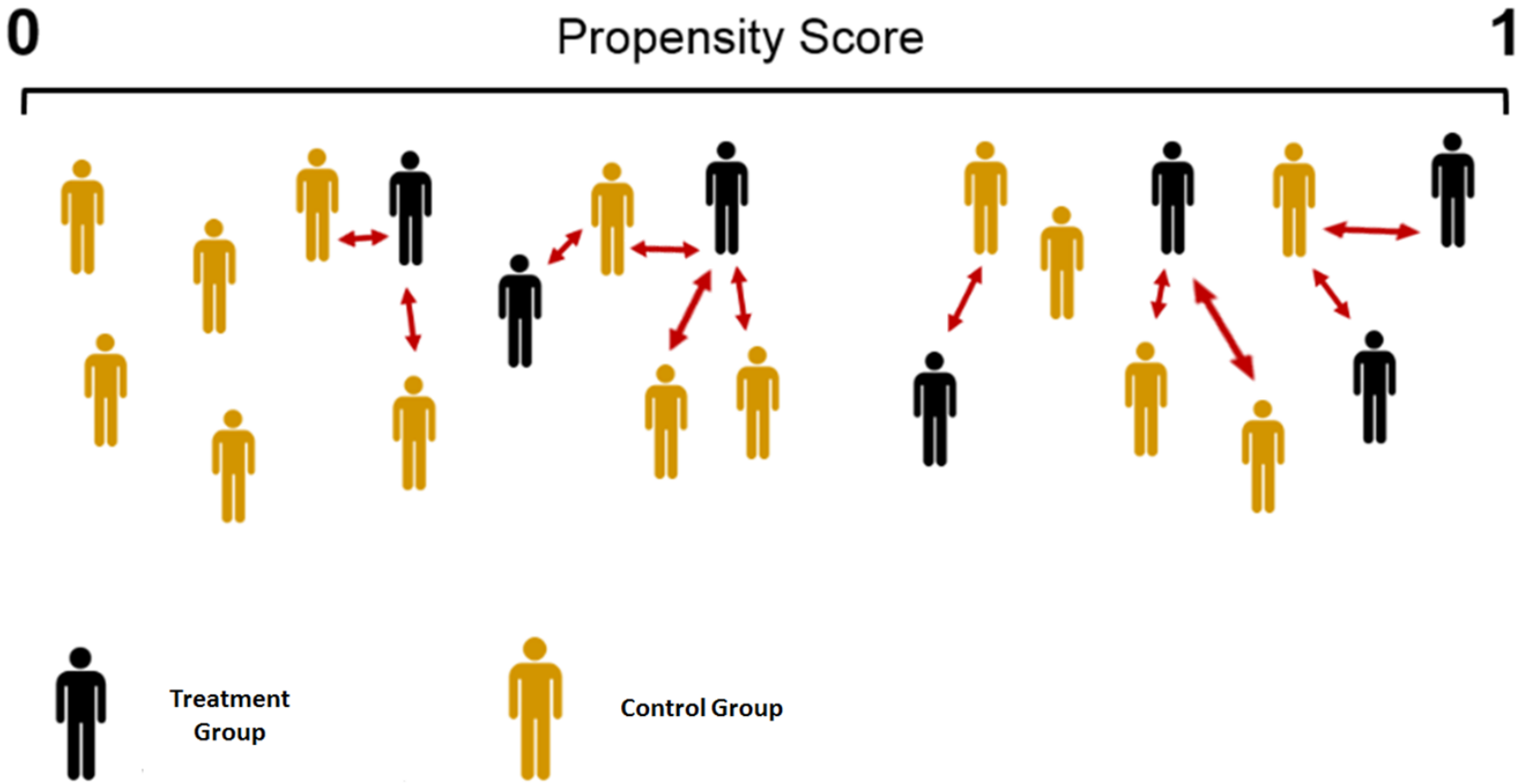
Propensity Score
Matching - 1:1

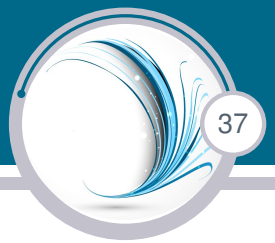


Propensity Score
Matching - M:1



Propensity Score
Matching - M:1 with replacement

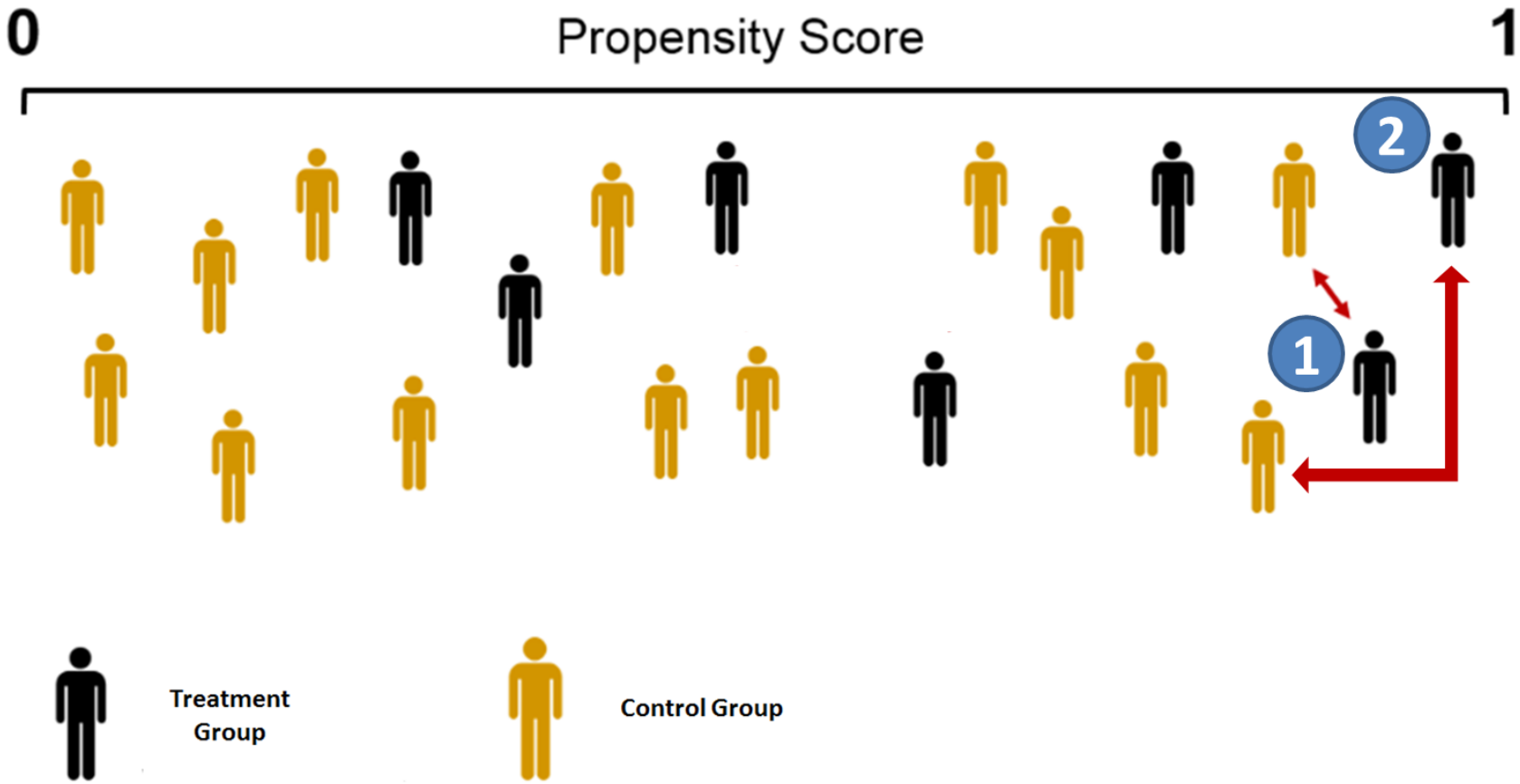




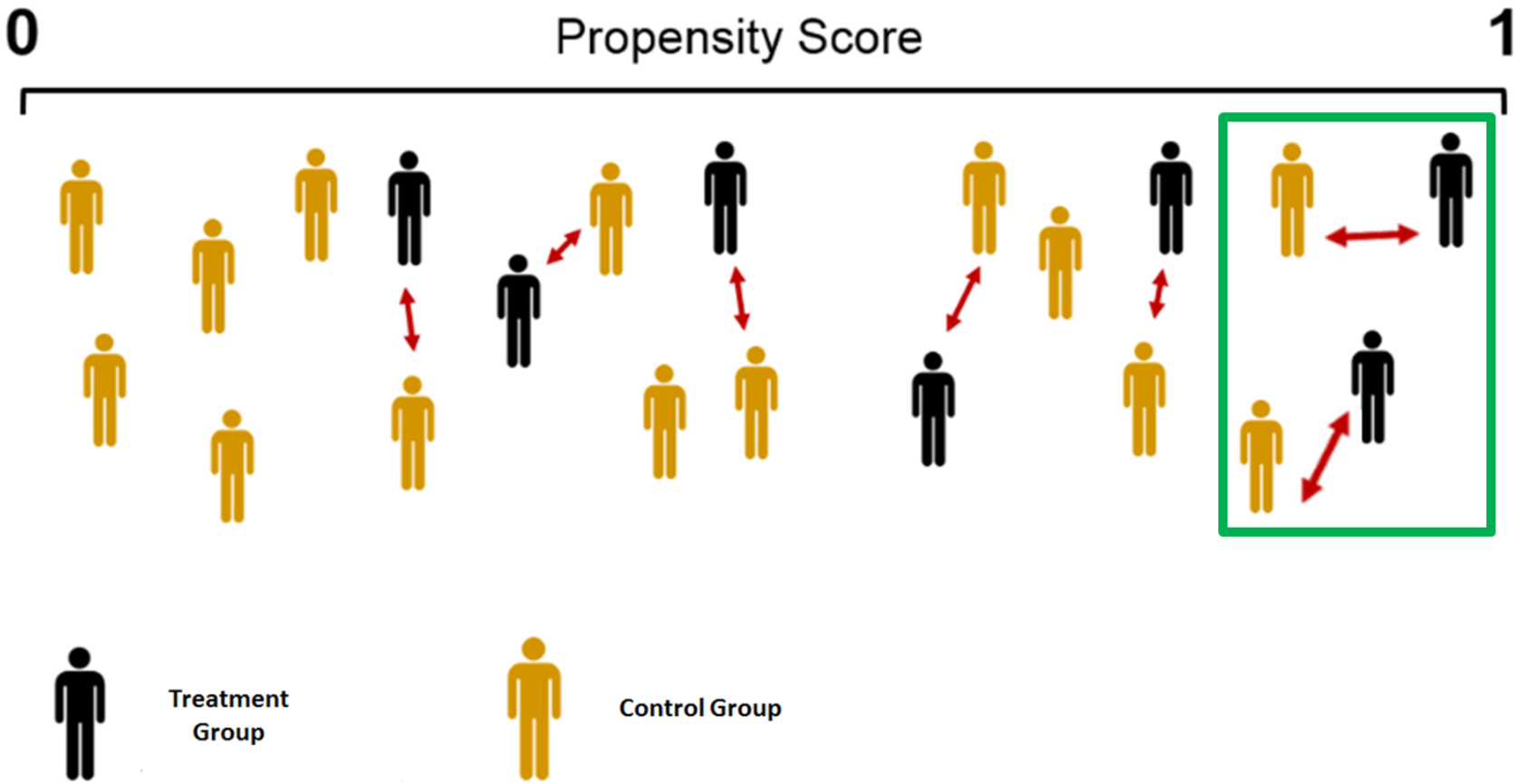
What type of algorithm should be chosen?

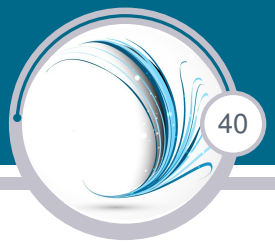
- Greedy: a treated unit is selected randomly, then the control unit whose propensity score is the closest to the treated unit is chosen as a match; without checking that this control unit might be better matched to another treated unit
- Optimal: matches formed so as to minimize the total within-pair difference of the propensity score

Propensity Score
Matching -Greedy



Propensity Score
Matching - Optimal



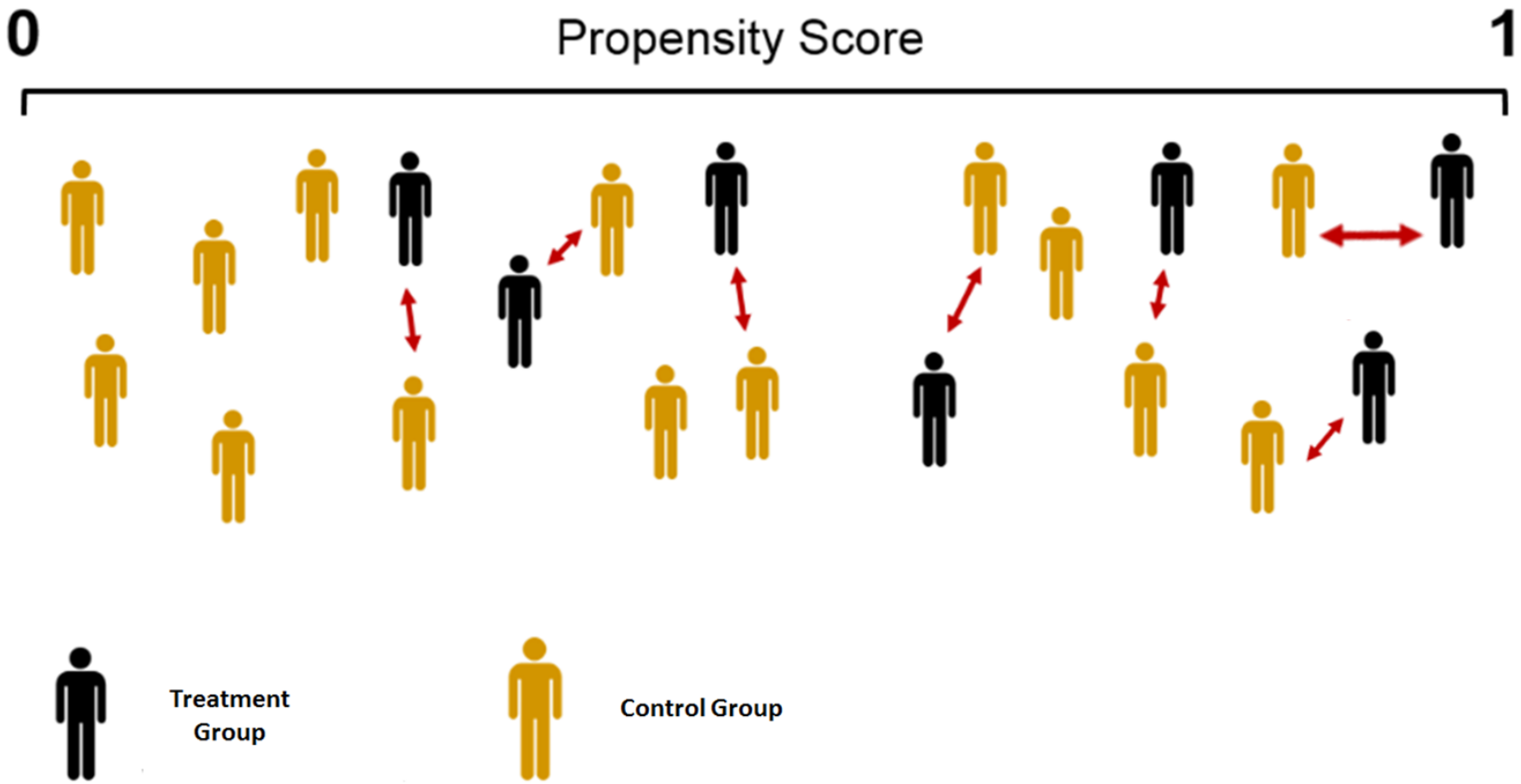


How close a match is acceptable?

- Nearest neighbor: consider the control unit with the propensity score closest to the one of the given treated unit
- Nearest neighbor in a specified caliper distance: absolute difference in the propensity scores of matched units must be below a prespecified threshold; if no control unit appears in this area, then the given treated unit will not be matched
- Radius
- Kernel...

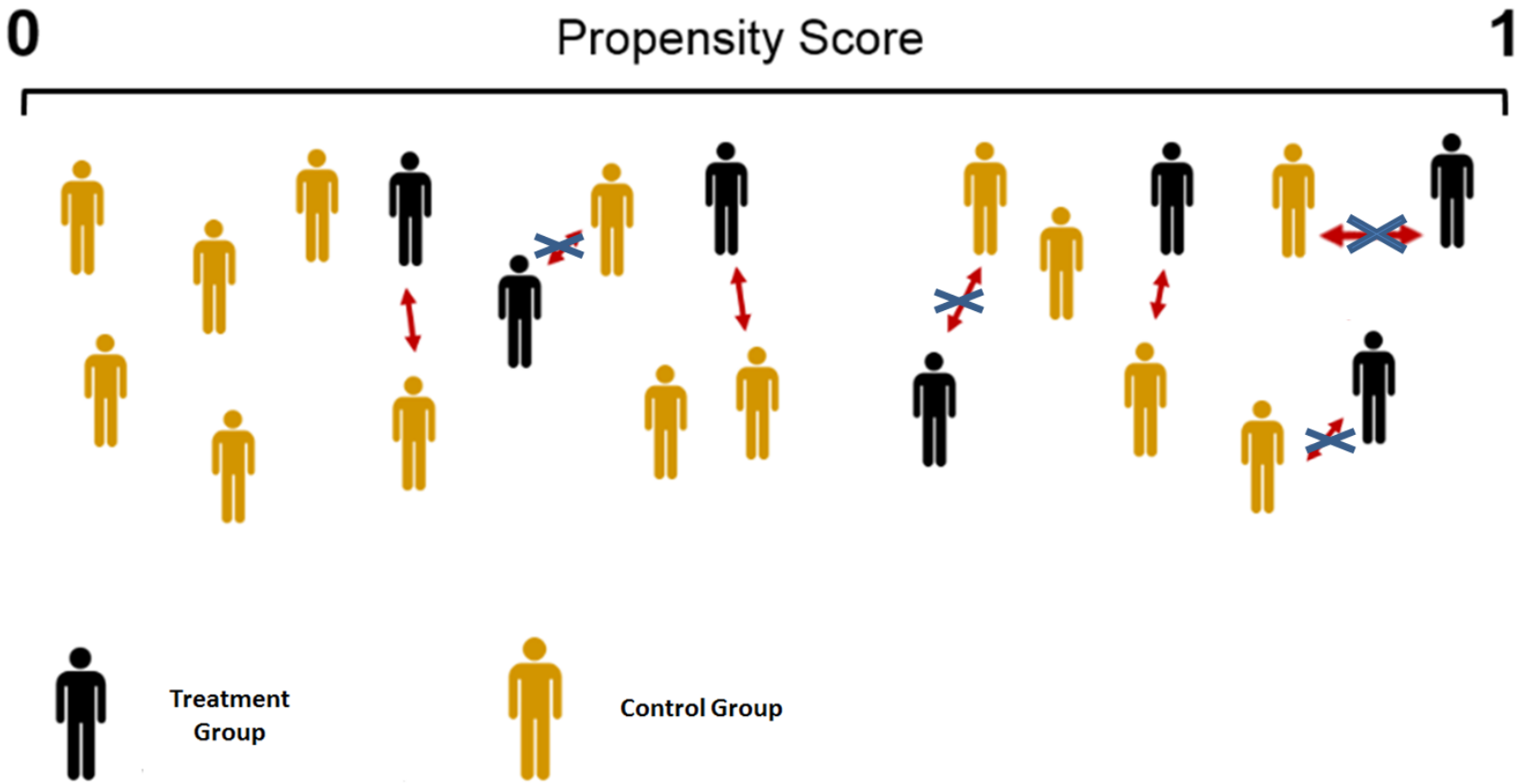
Once a matched dataset is obtained, one can directly compare outcomes between treated and control units

Propensity Score
Matching - Nearest Neighbor



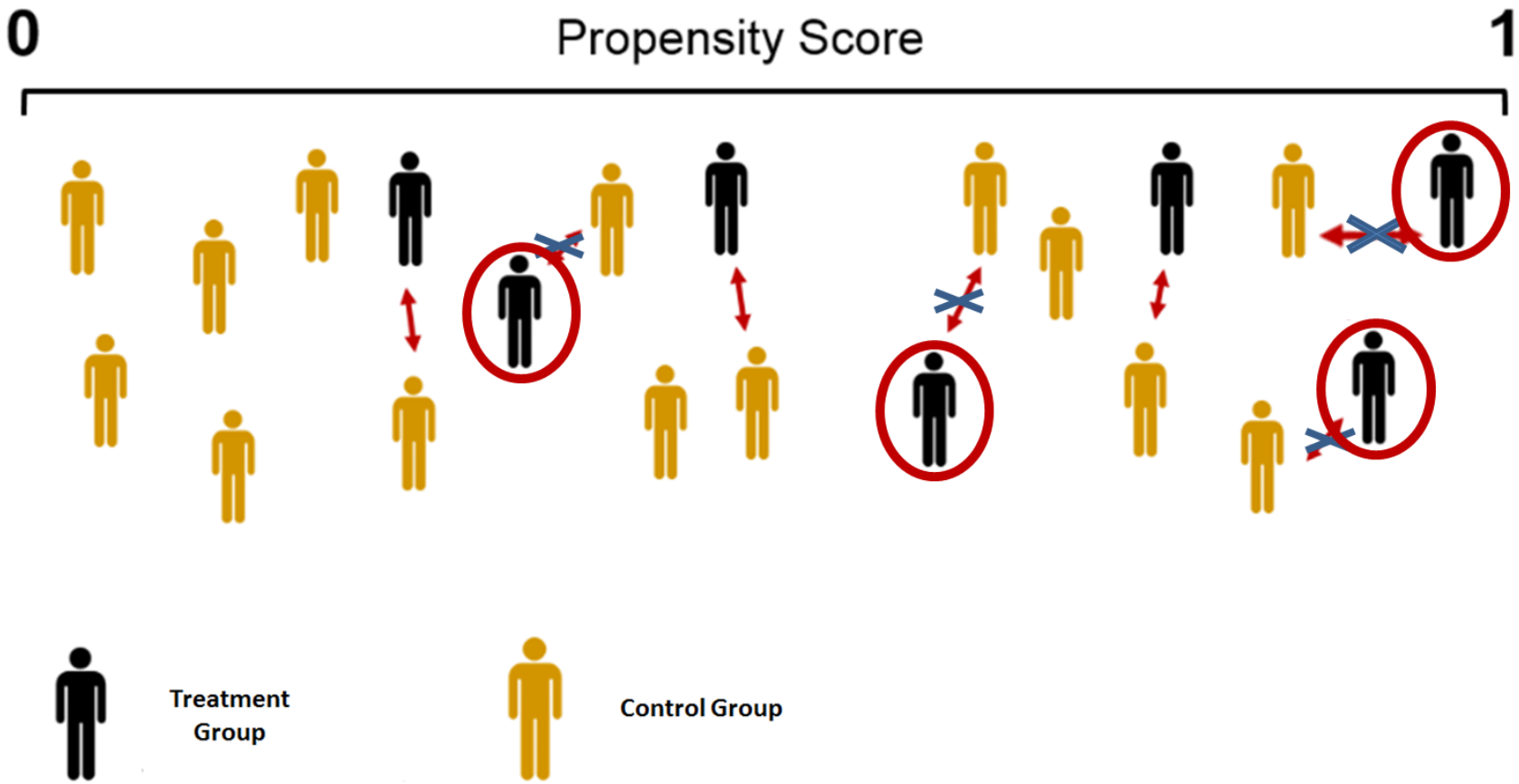
Propensity Score

Matching - Nearest Neighbor with a prespecified caliper



Propensity Score

Matching - Nearest Neighbor with a prespecified caliper



Back

Stratification

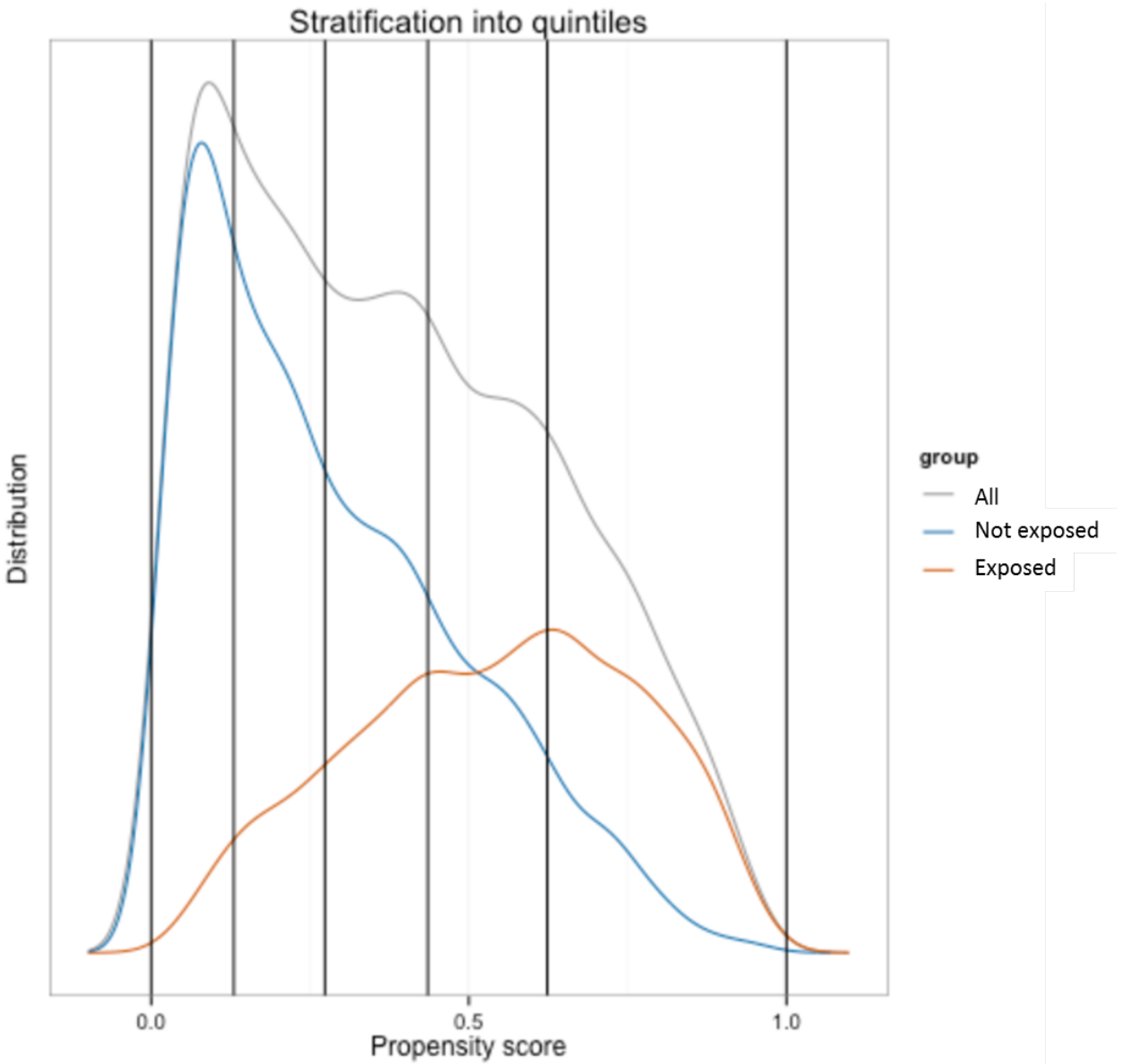
- Create strata based on propensity score with a prespecified threshold
  - Example: create 5 strata based on the 5 quintiles of the propensity score distribution
- 1 stratum = 1 matched sample
- Compare outcome within each stratum
- Control unbalanced sample sizes between strata using weights (for ATT): $\frac{1}{\#\text{ treated units in stratum } k}$

Covariate adjustment

- the design and the analysis of the study are not separate
- requires the outcome variable
- Model: $Y \sim T + ps(X)$

Example

Propensity score with stratification on quintiles (Back)



Variables (Back)

- *n.ss*: number of patients in historical control group + number of patients in current control group
- X_1 and X_2: covariates $\sim Bin(p_1)$ and $\sim Bin(p_2)$
- $logit.p = d_0 + d_1 * X_1 + d_2 * X_2 + d_3 * X_1 * X_2$
- trt: treatment indicator $\sim Bin(p_3 = \frac{exp(logit.p)}{1+exp(logit.p)})$
- $logit.p.Y = c_0 + c_1 * X_1 + c_2 * X_2 + c_3 * X_1 * X_2$
- Y: outcome $\sim Bin(p_4 = \frac{exp(logit.p.Y)}{1+exp(logit.p.Y)})$

trt	X1	X2	Y
1	0	0	0
0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	1
0	1	0	1

Simulations

Introduction - Parameters

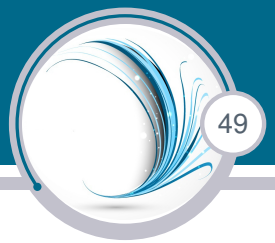
$$
\begin{aligned}
n.ss &= 4000 \\
p_1 &= 0.3 \\
p_2 &= 0.2 \\
d_0 &= -4 \\
d_1 &= 2 \\
d_2 &= -2 \\
d_3 &= 5 \\
c_0 &= -2 \\
c_1 &= 2 \\
c_2 &= 5 \\
c_3 &= -3
\end{aligned}
$$

Back

Simulations
Propensity Scores

trt	X1	X2	Y	psvalue	weight
1	0	0	0	0.04296074	23.277065
0	0	0	0	0.04296074	1.044889
0	0	0	0	0.04296074	1.044889
0	0	1	0	0.02958366	1.030486
0	0	0	1	0.04296074	1.044889
0	1	0	1	0.28482369	1.398257

Back



Variables

- Framework: for historical patients, we only have aggregated data
- Fro patients from current control group: data unchanged
- For patients from historical control group:
  - X_1 = mean of X_1 calculated previously for propensity score
  - X_2 = mean of X_2 calculated previously for propensity score
  - Y = mean of Y calculated previously for propensity score

Simulations
Entropy Balancing

trt	X1	X2	Y
1	1.0000000	0.0000000	0.0000000
1	1.0000000	1.0000000	1.0000000
1	0.0000000	0.0000000	0.0000000
1	0.0000000	0.0000000	0.0000000
1	1.0000000	0.0000000	0.0000000
1	1.0000000	0.0000000	1.0000000
1	1.0000000	1.0000000	0.0000000
1	1.0000000	1.0000000	1.0000000
1	1.0000000	0.0000000	1.0000000
0	0.2493816	0.2040702	0.2374635
0	0.2493816	0.2040702	0.2374635
0	0.2493816	0.2040702	0.2374635
0	0.2493816	0.2040702	0.2374635
0	0.2493816	0.2040702	0.2374635
0	0.2493816	0.2040702	0.2374635

Patients from current control group

Patients from historical control group

Back

Simulations
Comparison in means of covariates

Mean over all simulations of
mean(Xj in historical control group) - mean(Xj in current control group)

- No misspecification

	raw	after ps	after eb
X1	-0.6254	-0.0127	-0.0004070
X2	-0.3627	0.03388	-0.0001948

- Mis-specified model

	raw	after ps	after eb
X1	-0.6267	0.1619	-0.0004234
X2	-0.3630	0.1529	-0.0001939