# Scenario 2: Presence of treatment effect heterogeneity (HTE)

Prepared by: Thomas Debray, Joana Caldas, and Stan Wijn

Generated: November 18, 2022

## Contents



# 1 Build Info

This report was generated using R version 4.2.2 (2022-10-31 ucrt). The following key packages were used:

- *pandoc* (version 2.19.2): Report output
- *MatchIt* (version 4.4.0): Nearest-neighbour matching
- *sandwich* (version 3.0.2): Estimation of (cluster-)robust standard errors
- *dplyr* (version 1.0.10): Facilitate data manipulation
- *tibble* (version 3.1.8): Facilitate data manipulation
- *ggplot2* (version 3.3.6): Visualization of results
- *mice* (version 3.14.0): Multivariate Imputation by Chained Equations
- *mvtnorm* (version 1.1.3): Draw x1 and 1x from bivaraite normal distribution
- *ggmice* (version 0.0.1): Visualize predictor matrix of imputations

# 2 Methods

## 2.1 Data-generating mechanism

We generated data for a super-population of $N = 100,000$ individuals treated with treatment A, B, or C. We generated two continuous variables $x_1$ and $x_2$ for each patient as baseline covariates, labeled as age (standardized) and disease severity (z-score). We generated a potential outcome, labeled as a test score for which higher values are better, under each of the three treatments. The test scores were generated from a linear model with age, disease severity, and the treatment indicators as main effects and an interaction term between the treatment indicators and age to introduce treatment effect heterogeneity. The outcome $y$ is continuous, and generated as follows:

$y_i \sim N(\mu_i, \sigma^2)$

with

$\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 I_{Bi} + \beta_4 I_{Ci} + (\beta_5 I_{Bi} + \beta_6 I_{Ci}) x_{1i}$

Using the following regression coefficients:

- Baseline outcome risk ($\beta_0$): **0**
- Confounder effect of $x_1$ ($\beta_1$): **1**
- Confounder effect of $x_2$ ($\beta_2$): **0.25**
- Treatment effect of B versus A ($\beta_3$): **-0.5**
- Treatment effect of C versus A ($\beta_4$): **-0.25**
- Standard deviation of the residual error ($\sigma$): **1**

Depending on the selected scenario, the values for $\beta_5$ and $\beta_6$ were modified.

**Scenario 2**: Presence of HTE

- Effect modification between B and x1 ($\beta_5$): **0.25**
- Effect modification between C and x1 ($\beta_6$): **0.125**

We assigned the treatment received as a function of the two baseline covariates $x_1$ and $x_2$. Thus, the observable dataset consisted of the two baseline covariates, the treatment received and one outcome – the one generated under the treatment that the patient received.

## 2.2 Sampling distributions and parameter information

The treatment groups were sampled in equal sizes. The two continuous variables $x_1$ and $x_2$ were drawn from a bivariate normal distribution using the *mvtnorm* package, using the following mean values:

Table 1: Sampling means

| Treatment | Mean age | Mean disease severity |
|---|---|---|
| A | 1 | -1 |
| B | 0 | 1 |
| C | -1 | -1 |

And the following covariance matrix:

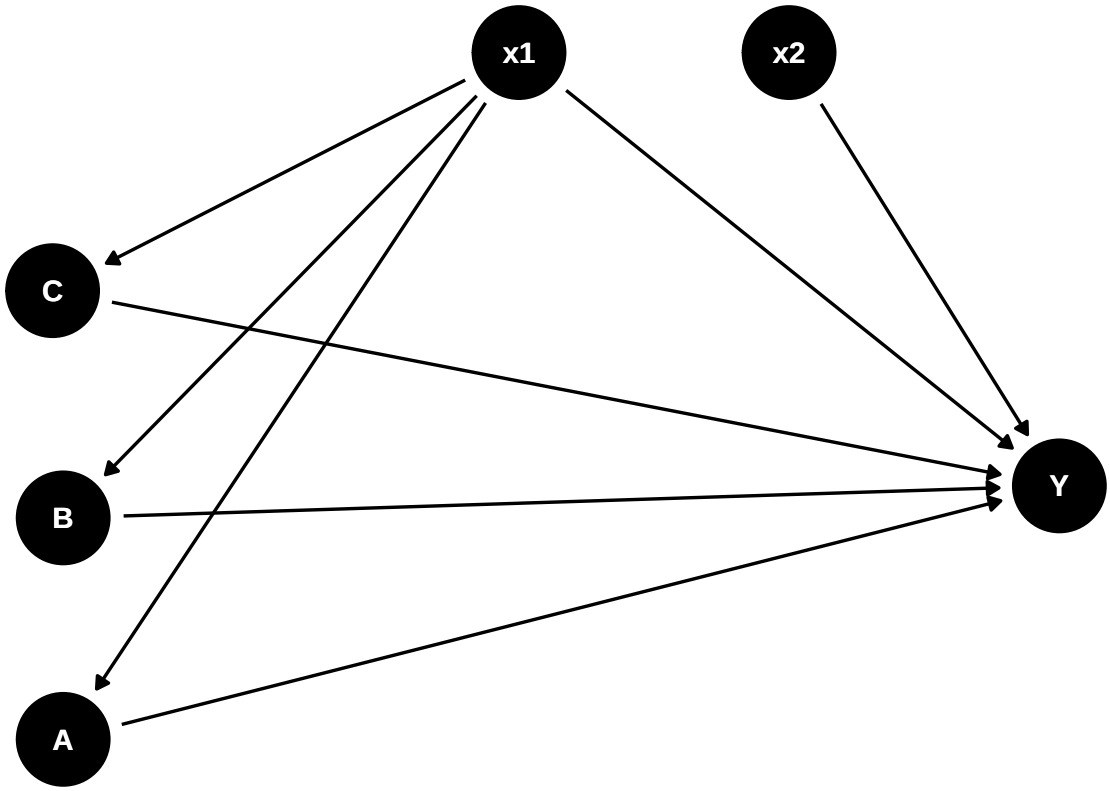
Table 2: Variance-covariance matrix of the baseline covariates

| | x1 | x2 |
|---|---|---|
| x1 | 2 | 0 |
| x2 | 0 | 2 |

## 2.3 Directed acyclic graph

To illustrate the relation between the treatment, the outcome and the treatment covariate interactions, a Directed acyclic graph is used. A, B, and C indicate the three treatment groups, Y is the outcome and x1 and x2 depict the covariates (age and disease severity). Both covariates influence the outcome $y_i$, but only in the presence of HTE do they also influence treatment B and C.

DAG in the presence of HTE




## 2.4 Implementation of pairwise matching

We applied 1:1 PS nearest-neighbor matching with a caliper and with replacement for each pairwise treatment comparison. We estimated the propensity score with a logistic regression model as a function of the two baseline covariates separately for each treatment comparison to mimic how pairwise matching would be repetitively applied in the context of three treatments being compared. We used a caliper of 0.05 standard deviation of the logit of the estimated PS. We assessed covariates balance pre- and post-matching with absolute standardized mean differences.

For each treatment comparison, we generated created two matched samples, separately targeting eachither of the ATTs. For example, for the comparison of treatment A vs B, we first match individuals treated with B to individuals treated with A, thus targeting the ATT-A. A first matched sample results from this matching. Second, starting from the original sample again, we now match individuals treated with A to individuals treated with B, targeting the ATT-B, which results in a second matched sample. For each treatment comparison and in each matched sample, we estimate the treatment effect (e.g., difference in means at a certain timepoint) using a linear regression model. We note that, in an application of pairwise matching to real data, researchers would typically generate only one matched sample, targeting one estimand or the other.

### 2.5 Evaluation and visualization

For the evaluation and visualization of covariate overlap, we compared the target populations before and after matching for all pairwise treatment comparisons with the bivariate ellipses. We also compared the estimated treatment effects in the matched sample.

## 3 Results

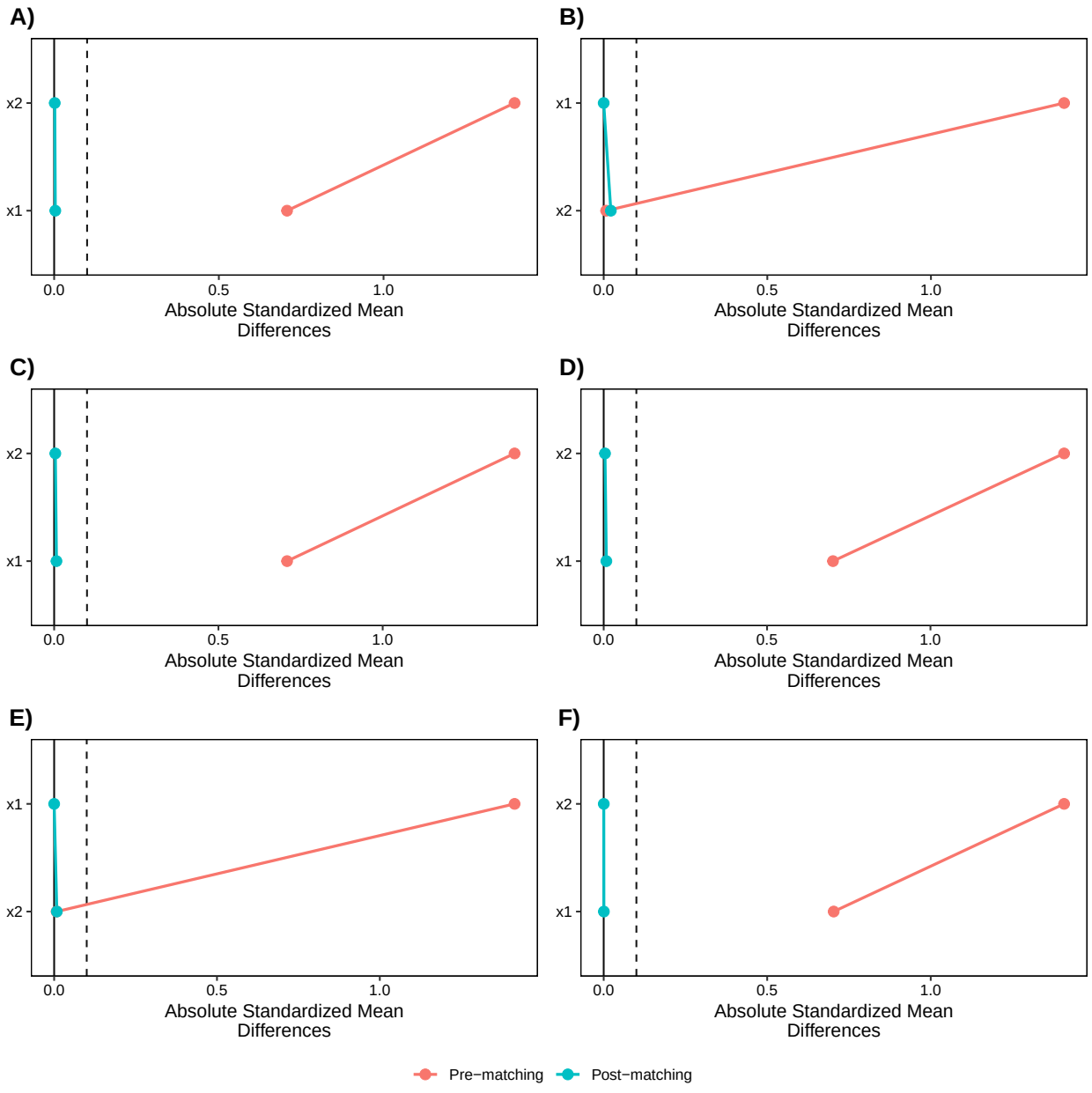
### 3.1 Scenario 2: Presence of HTE

There are three treatments A, B and C with a heterogeneous treatment effect. The treated populations are described by two continuous covariates $x_1$ and $x_2$. We generate a superpopulation of 120000 patients, in which there are **40000** observations with drug A, **40000** observations with drug B and **40000** observations with drug C. The continuous outcome $y$ is generated from the linear predictor:

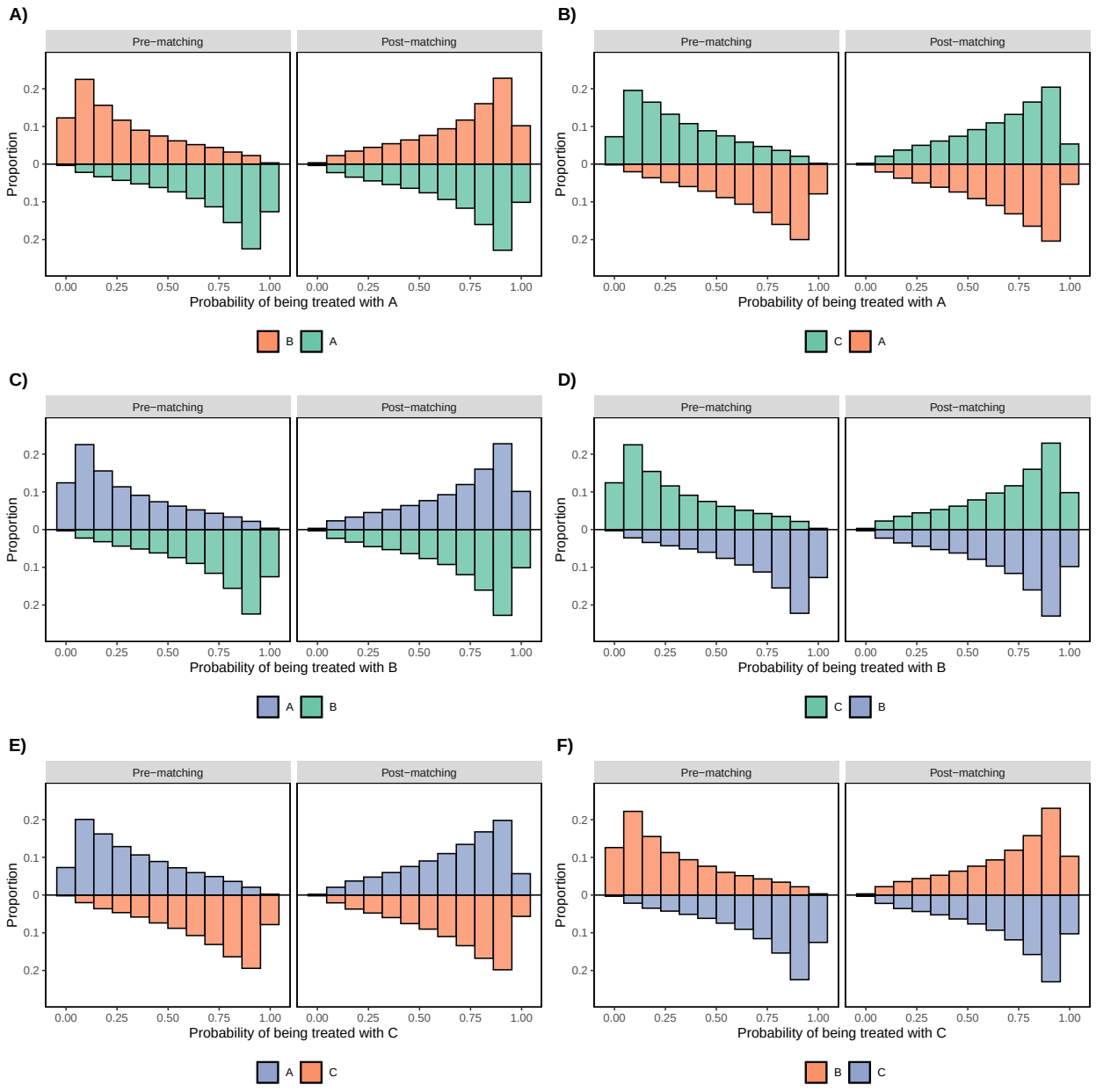
$\mu_i = 0 + 1\ x_{1i} + 0.25\ x_{2i} + \text{-}0.5\ I_{Bi} + \text{-}0.25\ I_{Ci} + (0.25\ I_{Bi} + 0.125\ I_{Ci})\ x_{1i}$

For all analyses, we apply nearest neighbour matching with a caliper width of 0.001 SD of the logit of the estimated PS. This results in a caliper of 0.002 for AvsB comparisons, a caliper of 0.002 for AvsC comparisons, and a caliper of 0.002 for BvsC comparisons. Afterwards, treatment effect estimates are derived using linear regression in the matched sample. A sandwich estimator is used to calculate robust standard errors.

## 3.2 Love plots

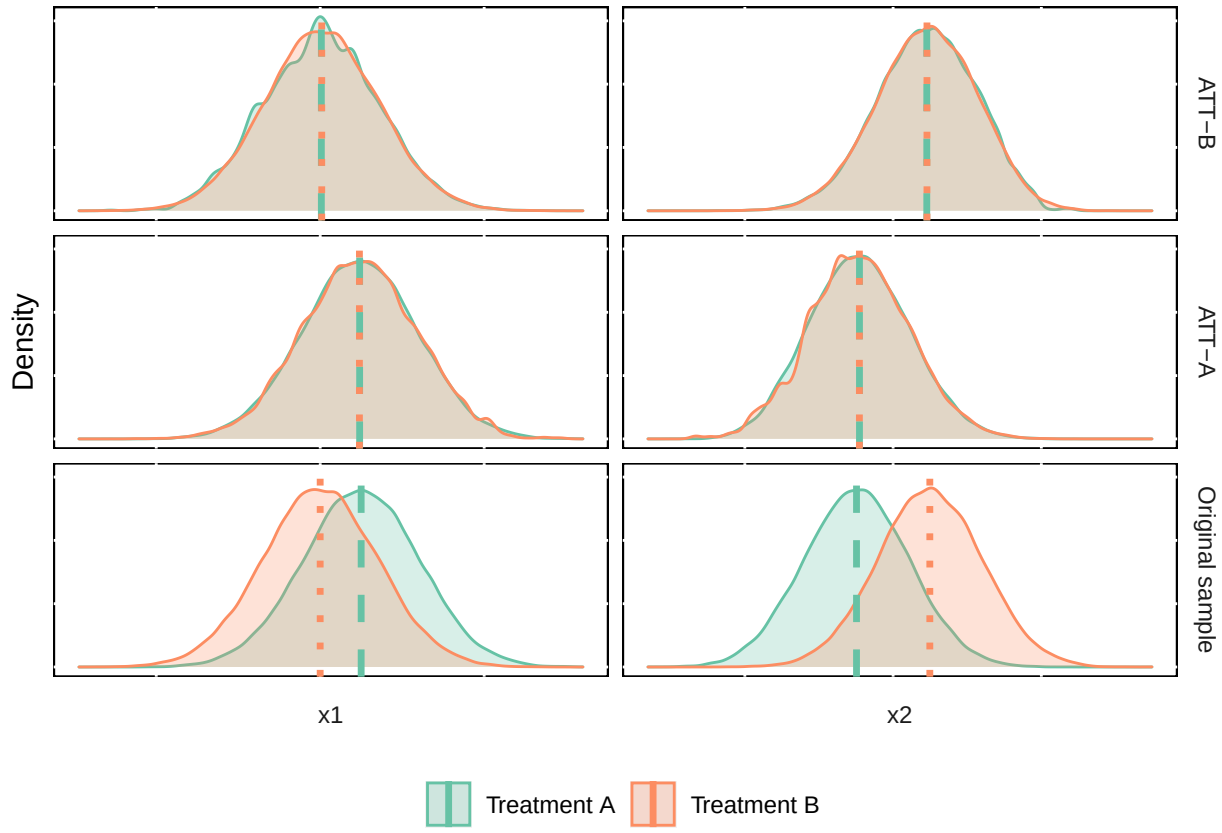


## 3.3 Balance plots



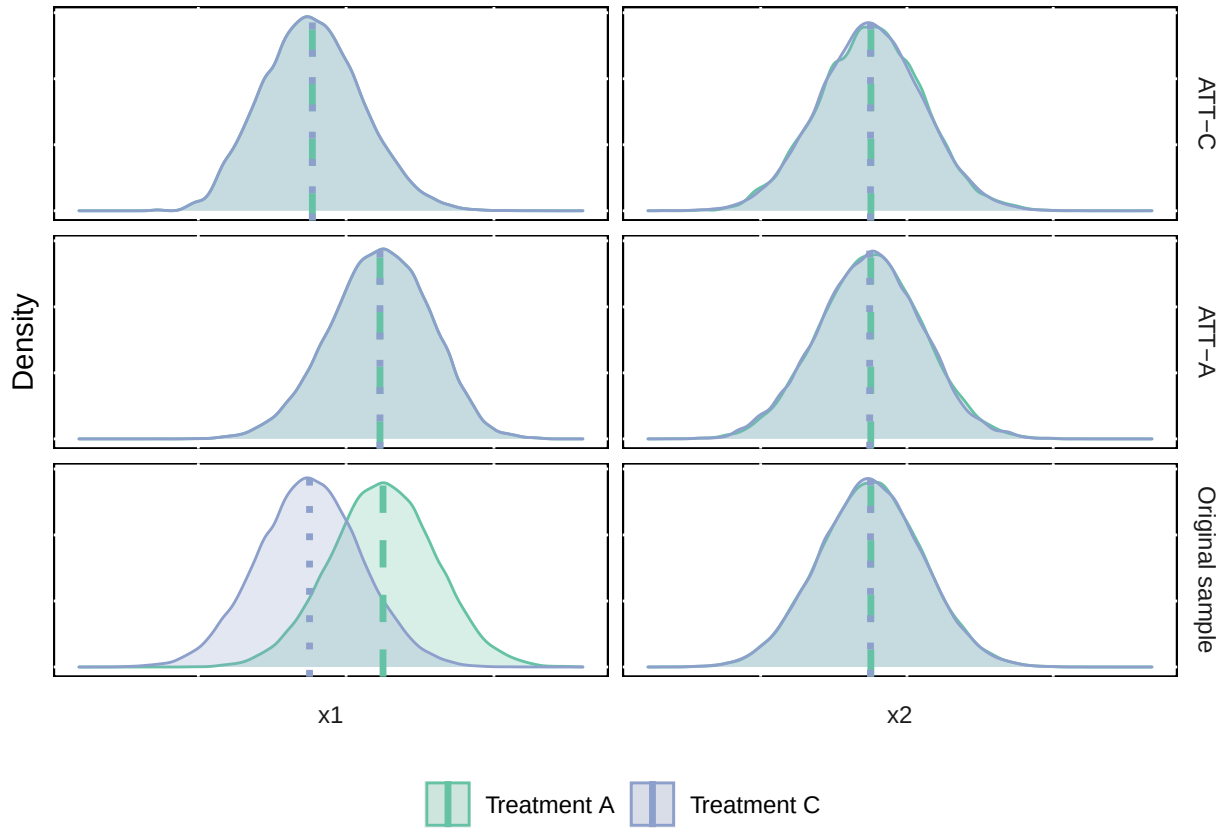
### 3.4 Difference between ATTA and ATTB

Joy plot displaying the differences between ATTA and ATTB



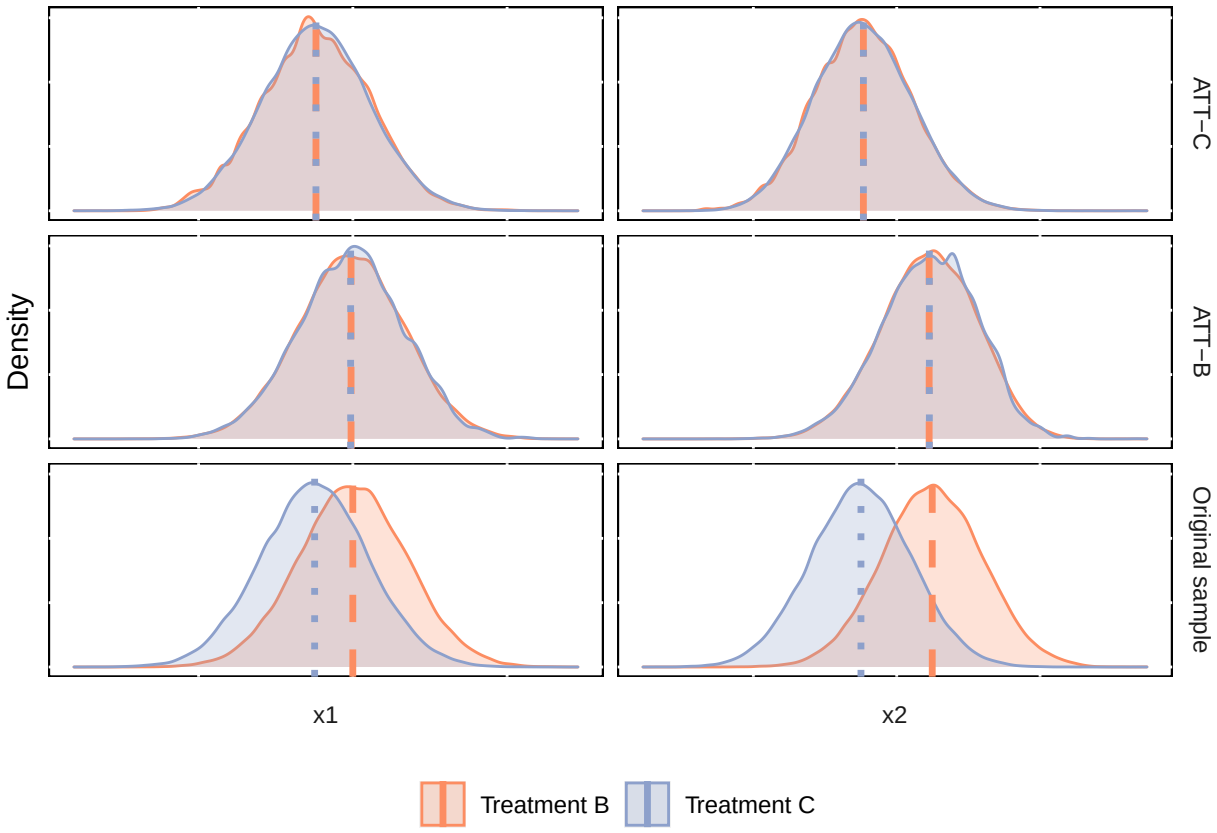
### 3.5 Difference between ATTA and ATTC

Joy plot displaying the differences between ATTA and ATTC

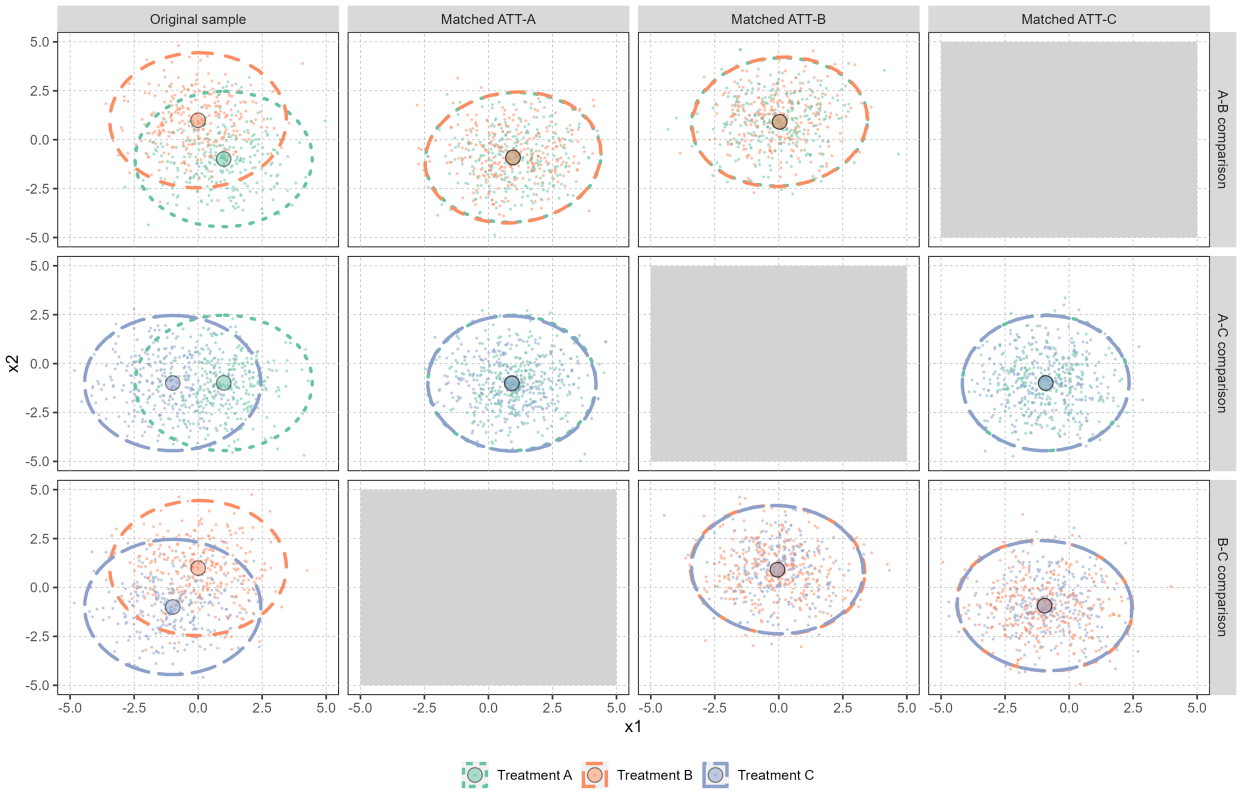


### 3.6 Difference between ATTB and ATTC

Joy plot displaying the differences between ATTB and ATTC



## 3.7 Overall distribution of covariates




## 3.8 Treatment effect estimates

Table 3: Treatment effect estimates for Scenario 2 when adopting matching with replacement

| | Comparison | Estimand | Replace | Reference | Estimate | SE | SS | ESS |
|---|---|---|---|---|---|---|---|---|
| x1 | A versus B | ATTA | TRUE | 0.250 | 0.2631928 | 0.0356748 | 51647 | 42111.28 |
| x11 | A versus B | ATTB | TRUE | 0.500 | 0.4771539 | 0.0321298 | 51782 | 42363.85 |
| x12 | A versus C | ATTA | TRUE | 0.125 | 0.1319471 | 0.0288779 | 53127 | 43732.07 |
| x13 | A versus C | ATTC | TRUE | 0.375 | 0.3568886 | 0.0272099 | 53234 | 43911.66 |
| x14 | B versus C | ATTB | TRUE | -0.250 | -0.2708809 | 0.0341529 | 51624 | 42238.68 |
| x15 | B versus C | ATTC | TRUE | -0.375 | -0.3718830 | 0.0368744 | 51958 | 42334.52 |