

MOTIVATION

- **Retrospective observational trials** can assess the impact of lifestyle factors on cancer incidence.
- **Causal inference** is the process of assessing when one thing causes another. However, in observational datasets treatment groups are not randomly selected, so assumptions for causal inference do not apply.
- **Confounding** is when a third variable influences both the covariate value and the outcome. Confounding means causes cannot be identified.
- **Matching Algorithms** subset data to create new control and treatment groups. This improves inference quality by reducing confounding. We compared performance of three matching algorithms in estimating a true effect.

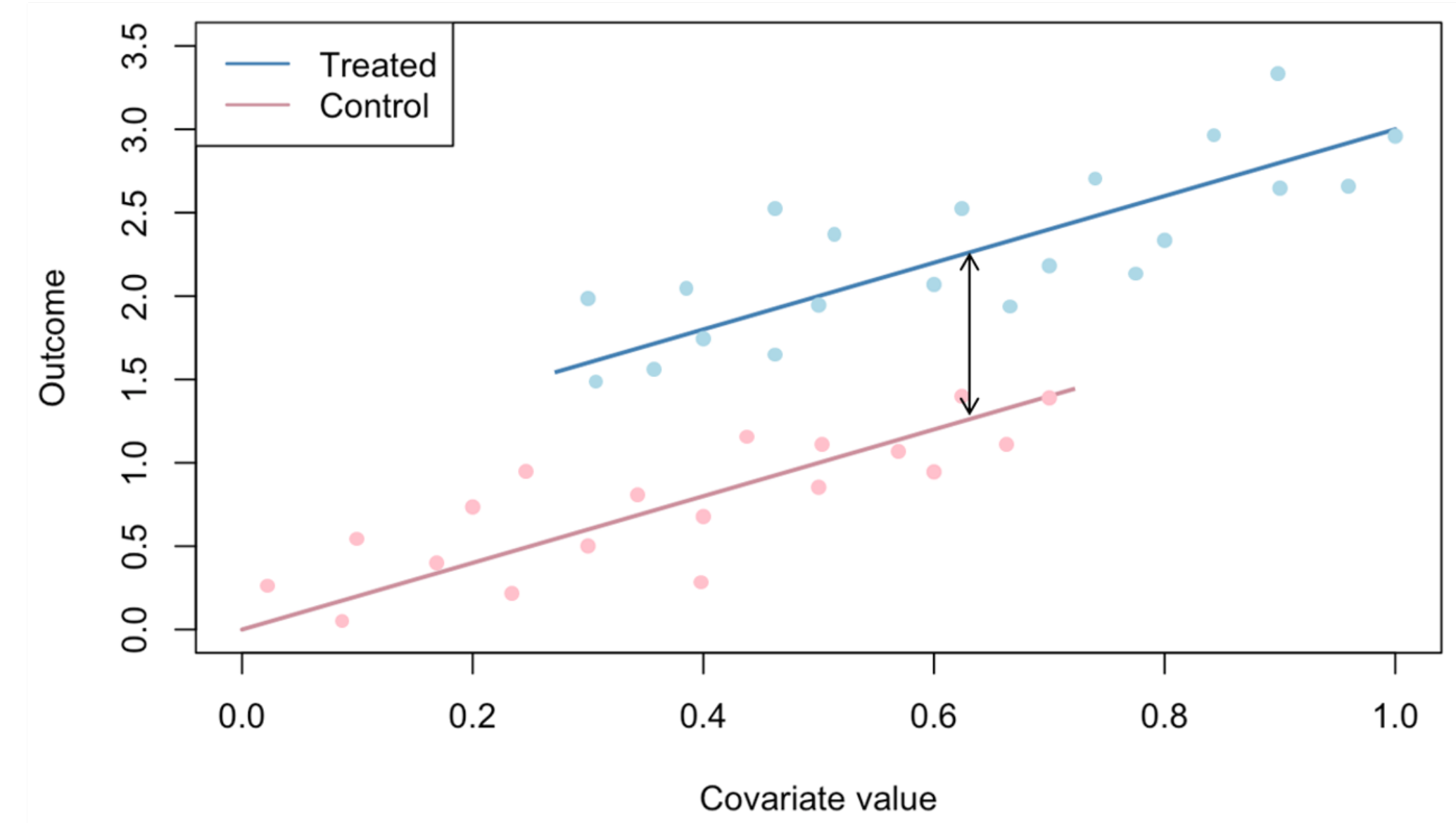


Figure 1: The **target estimand** is the **average effect of treatment on the treated (ATT)**, the true difference between groups. **Overlap** refers to the range of covariates with both treated and control data-points.

METHODS

- The performance of the algorithms was compared by performing 1000 Monte Carlo simulations for all algorithms and parameterizations (sample size, control-treated ratio, interaction effect, caliper size).
- We generated data with a **true** treatment effect $T = 1$. We measured bias, variance and mean squared error ($MSE = Bias + Variance^2$).

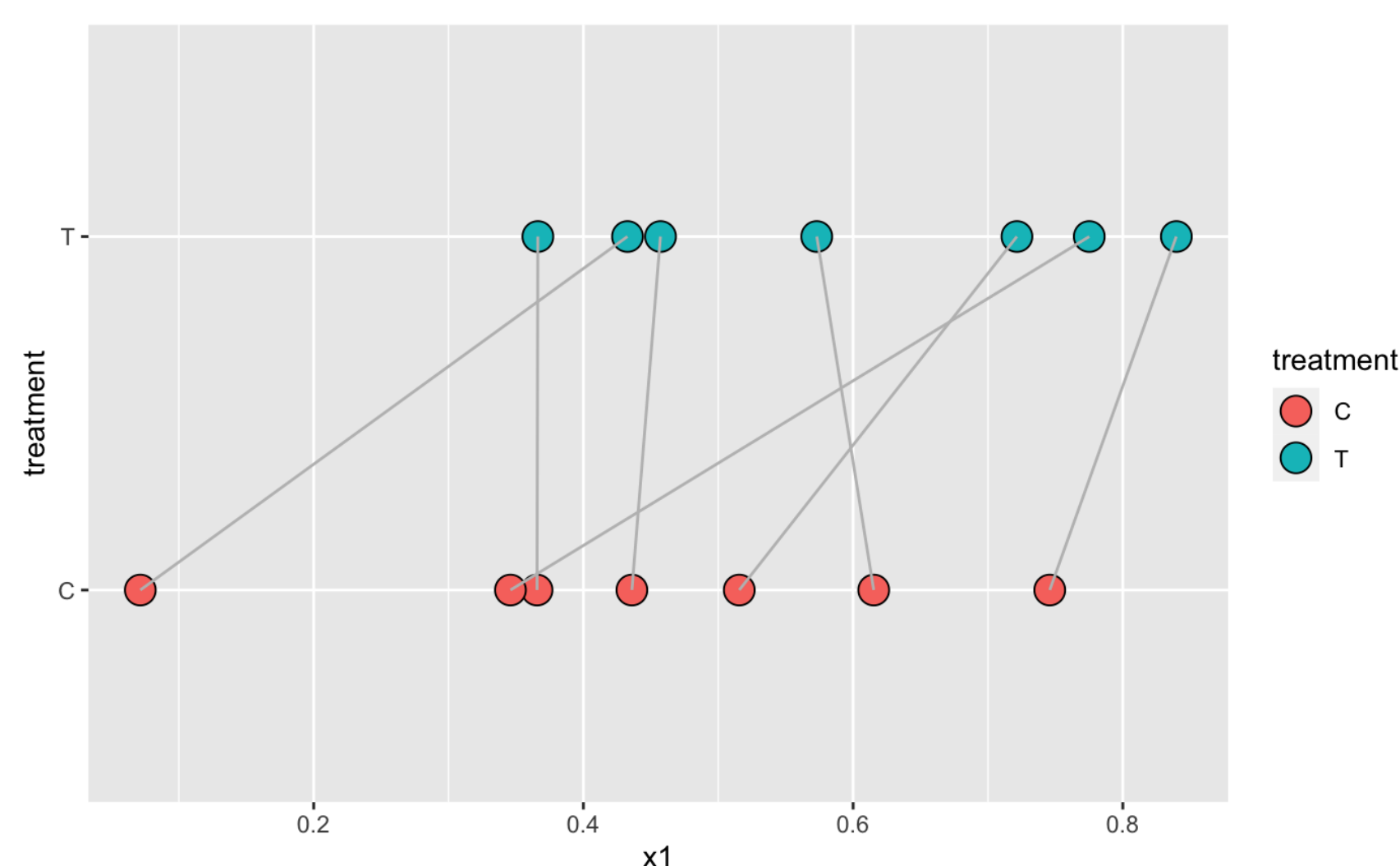


Figure 2: Matching without replacement: Treated units are matched to their nearest unmatched neighbour.

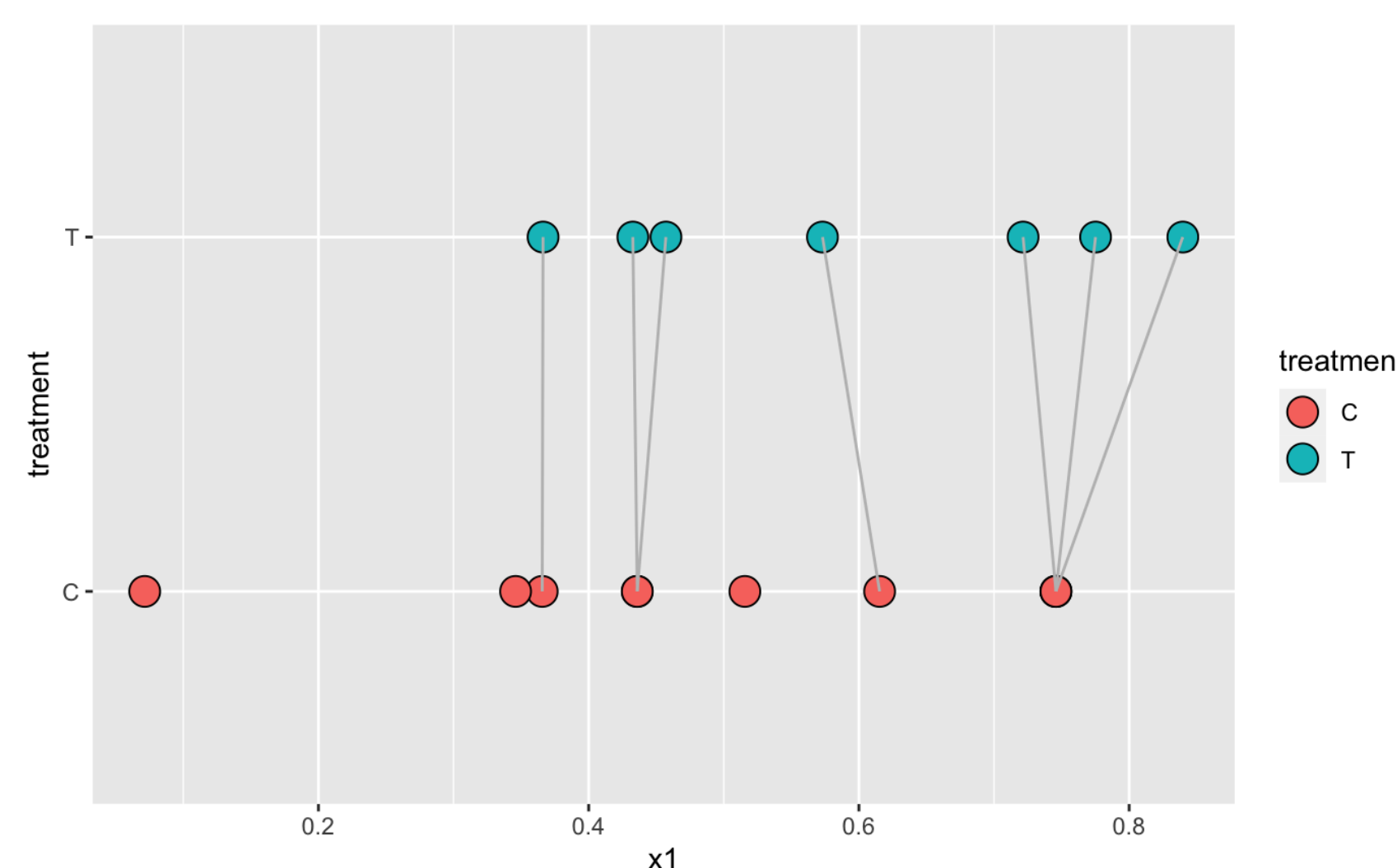


Figure 3: Matching with replacement: treated units are matched to the closest control unit, duplicates are possible.

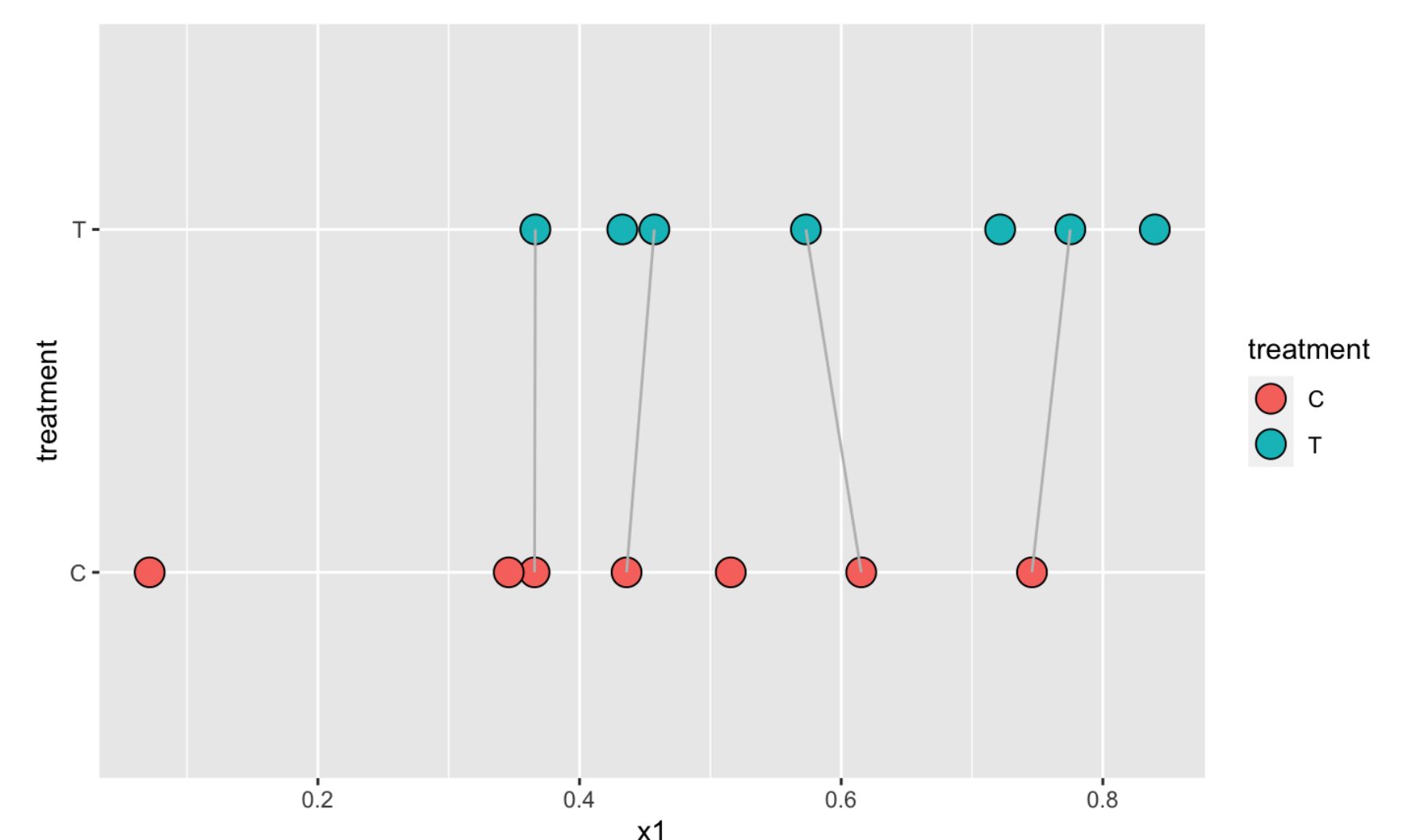


Figure 4: Caliper matching: matches can only occur within a set limit. Treated units may be discarded.

RESULTS

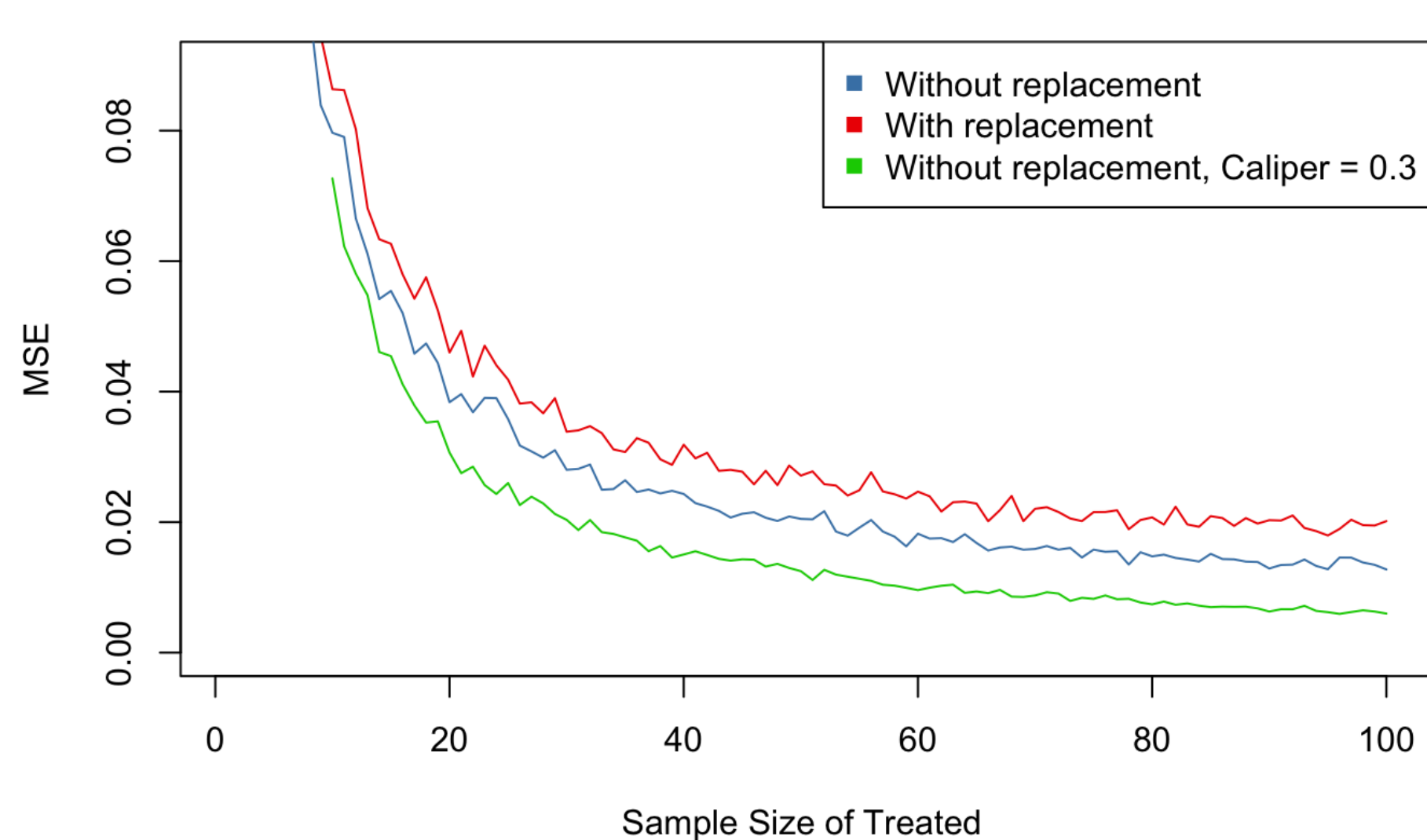


Figure 5: MSE reduces with sample size for all methods. It is lowest for caliper matching, due to reduced bias, and highest for matching with replacement due to higher variance.

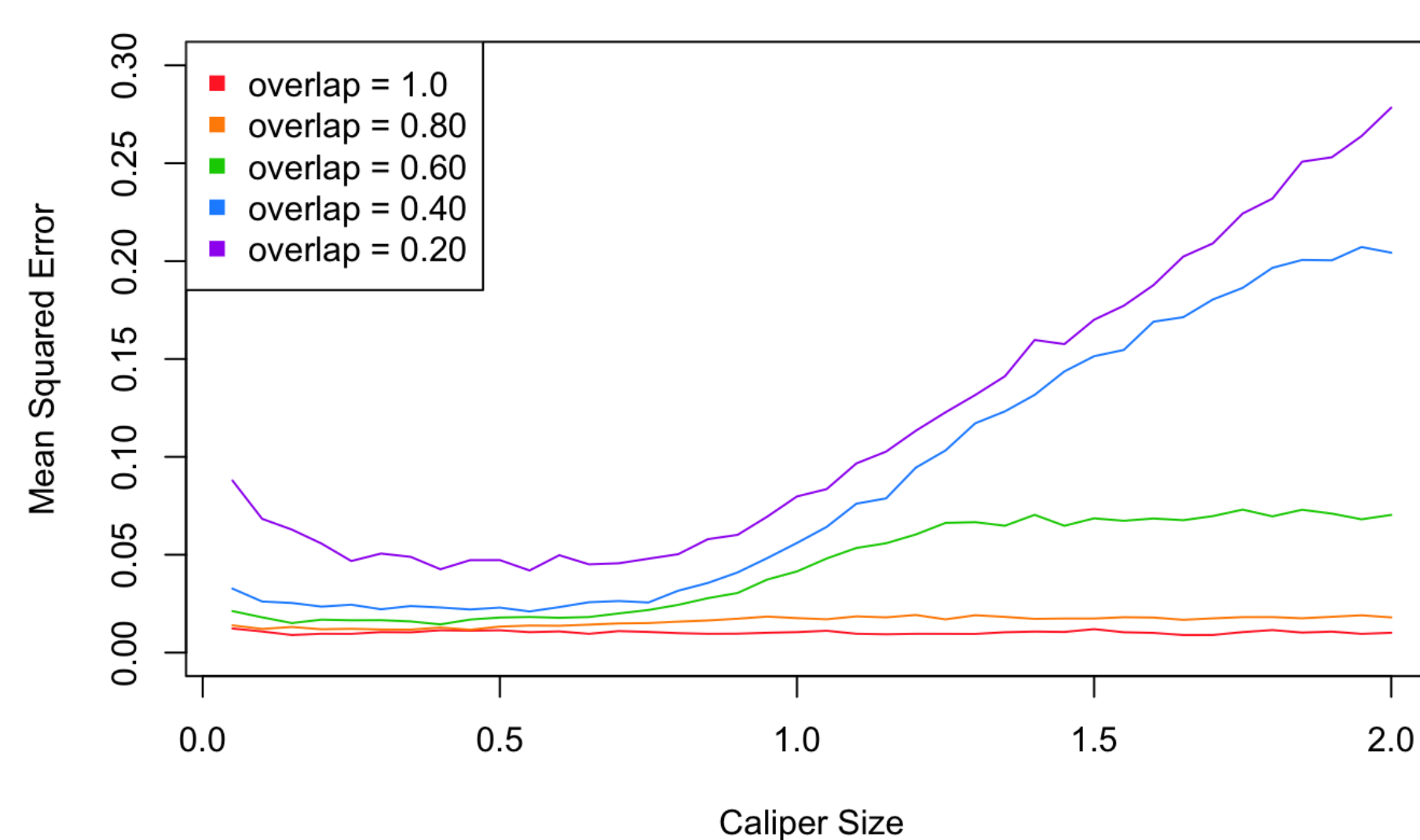


Figure 6: Larger calipers increase bias and smaller calipers increase variance. When overlap is low, the bias variance trade off is clear.

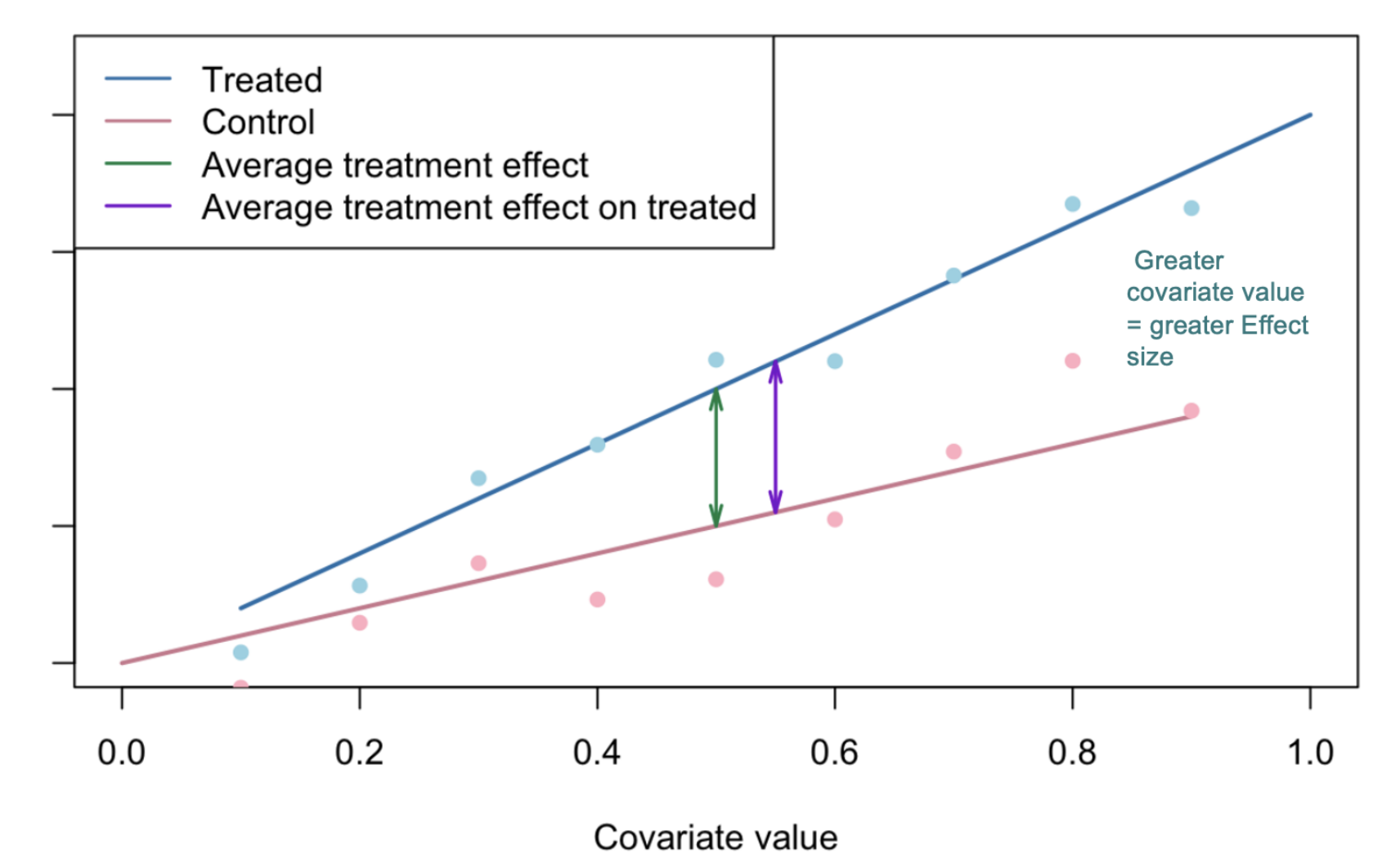


Figure 7: Interaction (where the covariate level influences the size of the effect) + overlap < 100% mean the target estimand is not measured correctly.

- In all simulations low overlap increases bias \rightarrow increased MSE.

CONCLUSIONS AND FUTURE RESEARCH

- Overlap < 100% introduces bias for all methods, even when there is theoretical positivity.
- Interaction effects also introduce bias. This adds to the bias from lack of overlap.
- The number of units discarded has a large effect on the MSE due to the Bias-Variance trade off.
- Future studies should extend this research to include multiple covariates. They should also test the conclusions using real data.

MORE INFORMATION AND CONTACT DETAILS

Master Thesis https://github.com/laura-ruth/LJS_thesis/blob/main/Thesis_ljs.pdf

Email L.R.Jansen-Storbacka@tudelft.nl