

uniCATE: A Flexible Approach to Predictive Biomarker Discovery

Philippe Boileau – April 2022

Collaborators

Co-authors:

- Nina Qi
- Mark van der Laan
- Sandrine Dudoit
- Ning Leng

Additional support:

- Zoe June Assaf
- Romain Banchereau

Motivation

IMmotion 150/151 Trials

A need for biomarkers predictive of clinical benefit

IMmotion 150/151 Trials

A need for biomarkers predictive of clinical benefit

- Tyrosine kinase inhibitors (e.g. sunitinib, bevacizumab) were the standard of care for metastatic renal cell carcinoma patients (mRCC) until recently.

IMmotion 150/151 Trials

A need for biomarkers predictive of clinical benefit

- Tyrosine kinase inhibitors (e.g. sunitinib, bevacizumab) were the standard of care for metastatic renal cell carcinoma patients (mRCC) until recently.
- These treatments are ineffective for some patients, with many developing resistance over time.

IMmotion 150/151 Trials

A need for biomarkers predictive of clinical benefit

- Tyrosine kinase inhibitors (e.g. sunitinib, bevacizumab) were the standard of care for metastatic renal cell carcinoma patients (mRCC) until recently.
- These treatments are ineffective for some patients, with many developing resistance over time.
- IMmotion 150: Immune checkpoint inhibitors (e.g. atezolizumab) improved patient outcomes when used in combination with bevacizumab over sunitinib.

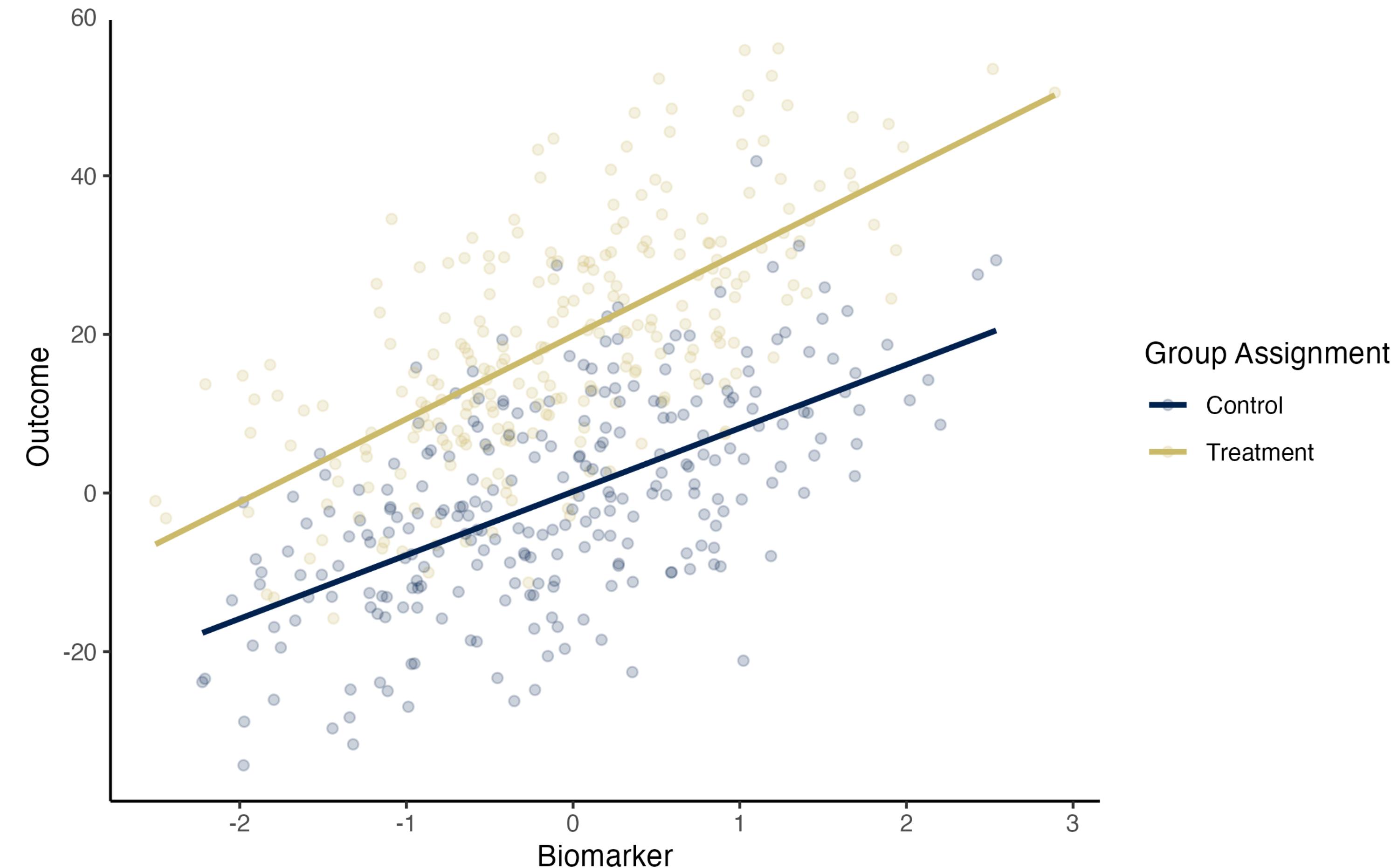
IMmotion 150/151 Trials

A need for biomarkers predictive of clinical benefit

- Tyrosine kinase inhibitors (e.g. sunitinib, bevacizumab) were the standard of care for metastatic renal cell carcinoma patients (mRCC) until recently.
- These treatments are ineffective for some patients, with many developing resistance over time.
- IMmotion 150: Immune checkpoint inhibitors (e.g. atezolizumab) improved patient outcomes when used in combination with bevacizumab over sunitinib.
- IMmotion 151: atezolizumab + bevacizumab improved PFS and OR over sunitinib in many patients with PD-L1 expressing mRCC, but not all.

Prognostic Biomarkers

Indicators of outcome, regardless of therapy

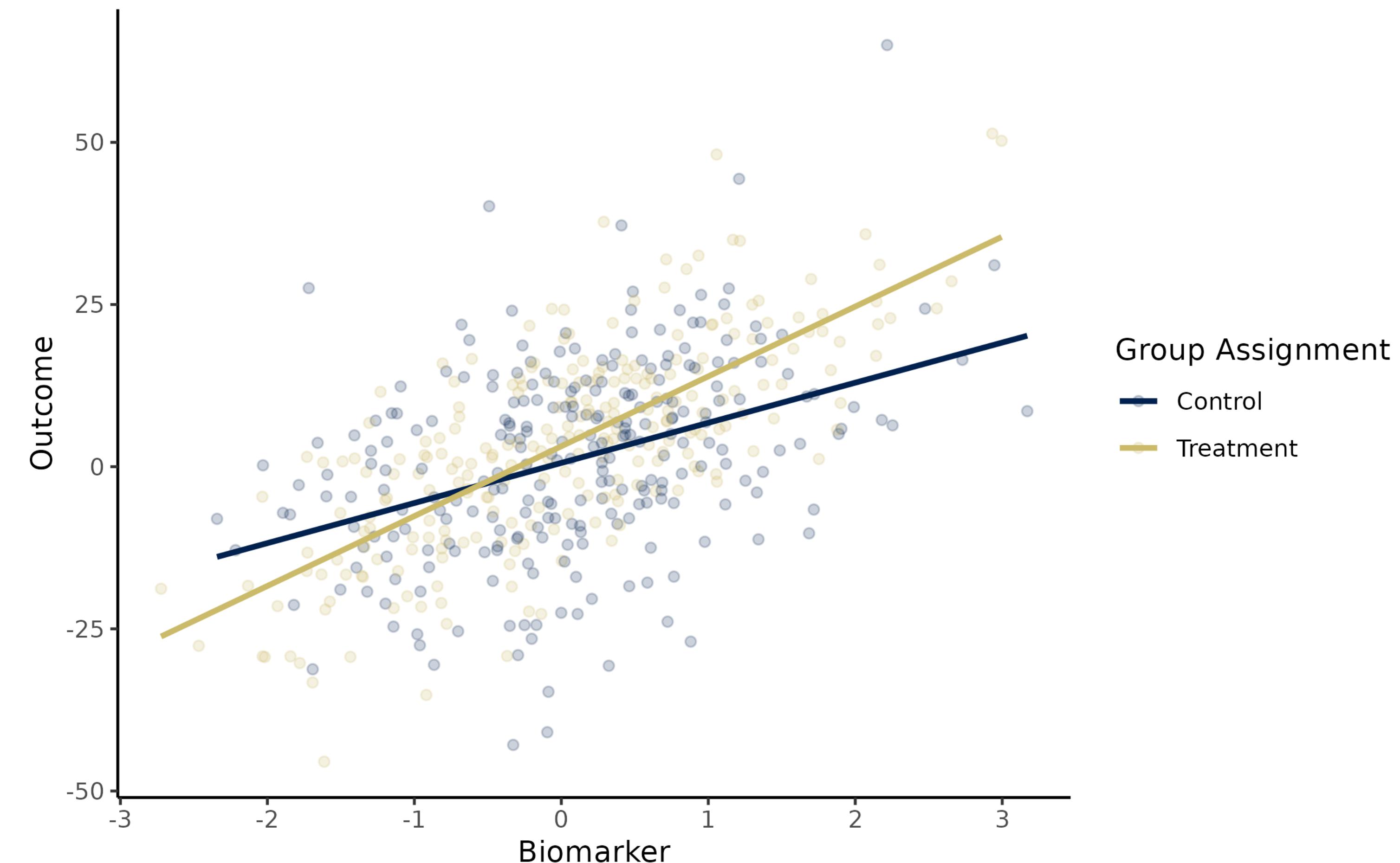


Predictive Biomarkers

Indicators of therapy effectiveness

Predictive Biomarkers

Indicators of therapy effectiveness



Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?
- **Targeted drug discovery:** What is the biological mechanism of a therapy?

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?
- **Targeted drug discovery:** What is the biological mechanism of a therapy?
- **Refined clinical trials:** Is a therapy more efficacious in a subset of the patient population?

Discovering Predictive Biomarkers

Identifying Predictive Biomarkers

A variable selection problem

Identifying Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:

Identifying Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms

Identifying Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Semiparametric methods targeting the conditional average treatment effect (CATE)

Identifying Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Semiparametric methods targeting the conditional average treatment effect (CATE)
- Harder when there are a large number of biomarkers: Penalized versions of the above methods are used.

Identifying Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Semiparametric methods targeting the conditional average treatment effect (CATE)
- Harder when there are a large number of biomarkers: Penalized versions of the above methods are used.
- **Bottom line:** Identification of predictive biomarkers is the byproduct of another inference procedure.

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

- Tian et al (2014) demonstrated that the treatment-biomarker interactions can be modeled directly through a minor transformation of the outcome.

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

- Tian et al (2014) demonstrated that the treatment-biomarker interactions can be modeled directly through a minor transformation of the outcome.
- In high-dimensions, the interaction coefficients of a linear model are estimated using penalized regression methods, like the LASSO.

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

- Tian et al (2014) demonstrated that the treatment-biomarker interactions can be modeled directly through a minor transformation of the outcome.
- In high-dimensions, the interaction coefficients of a linear model are estimated using penalized regression methods, like the LASSO.
- An “augmented” version of the methodology was developed, accounting for finite sample variation due to main effects. It is equivalent to fitting a traditional LASSO regression with treatment-biomarker interactions.

Issues with Penalized Regression Methods

Unreliable biomarker selection

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials
- Decreased signal to noise ratio in diagnostic assays

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

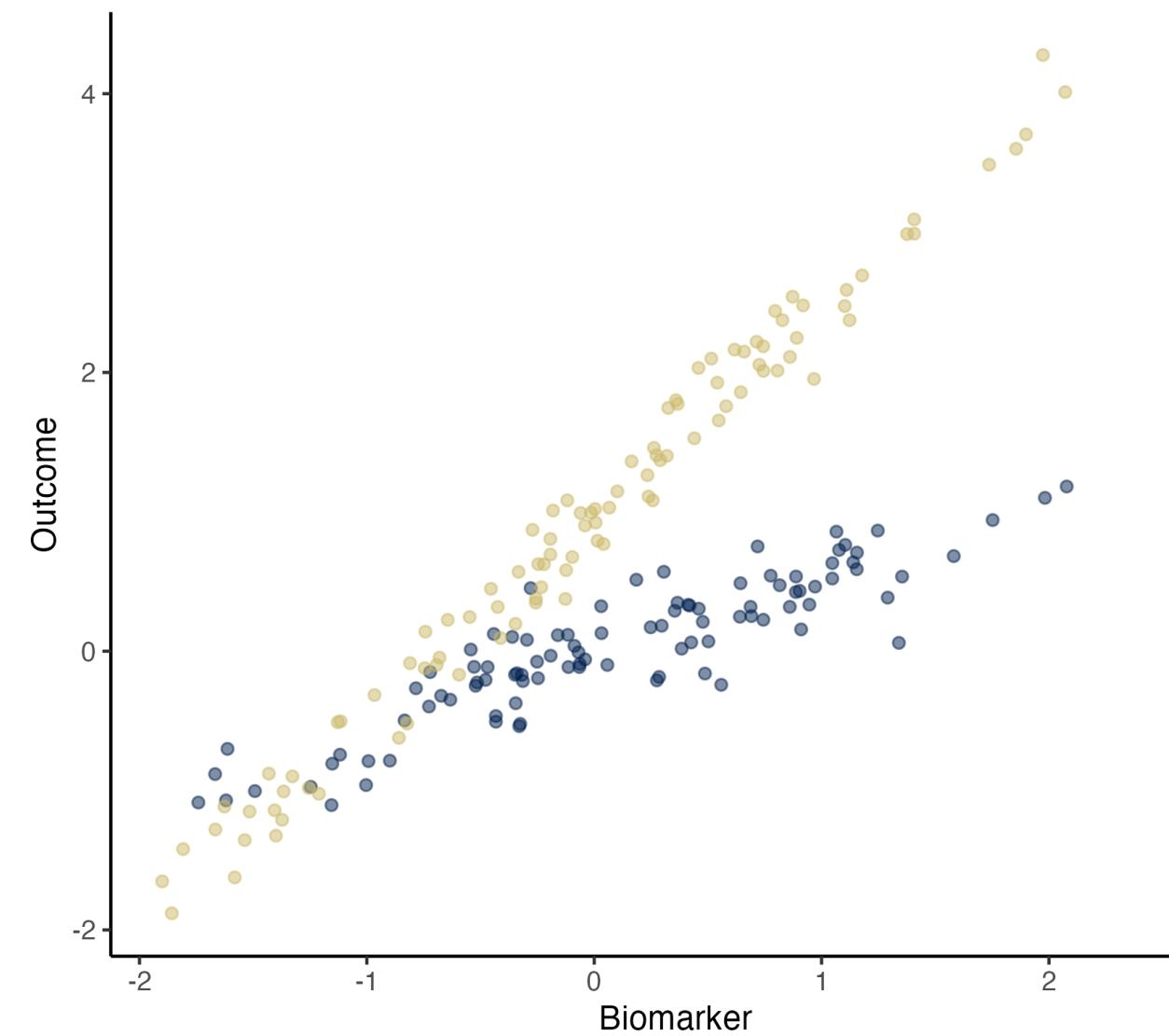
Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials
- Decreased signal to noise ratio in diagnostic assays

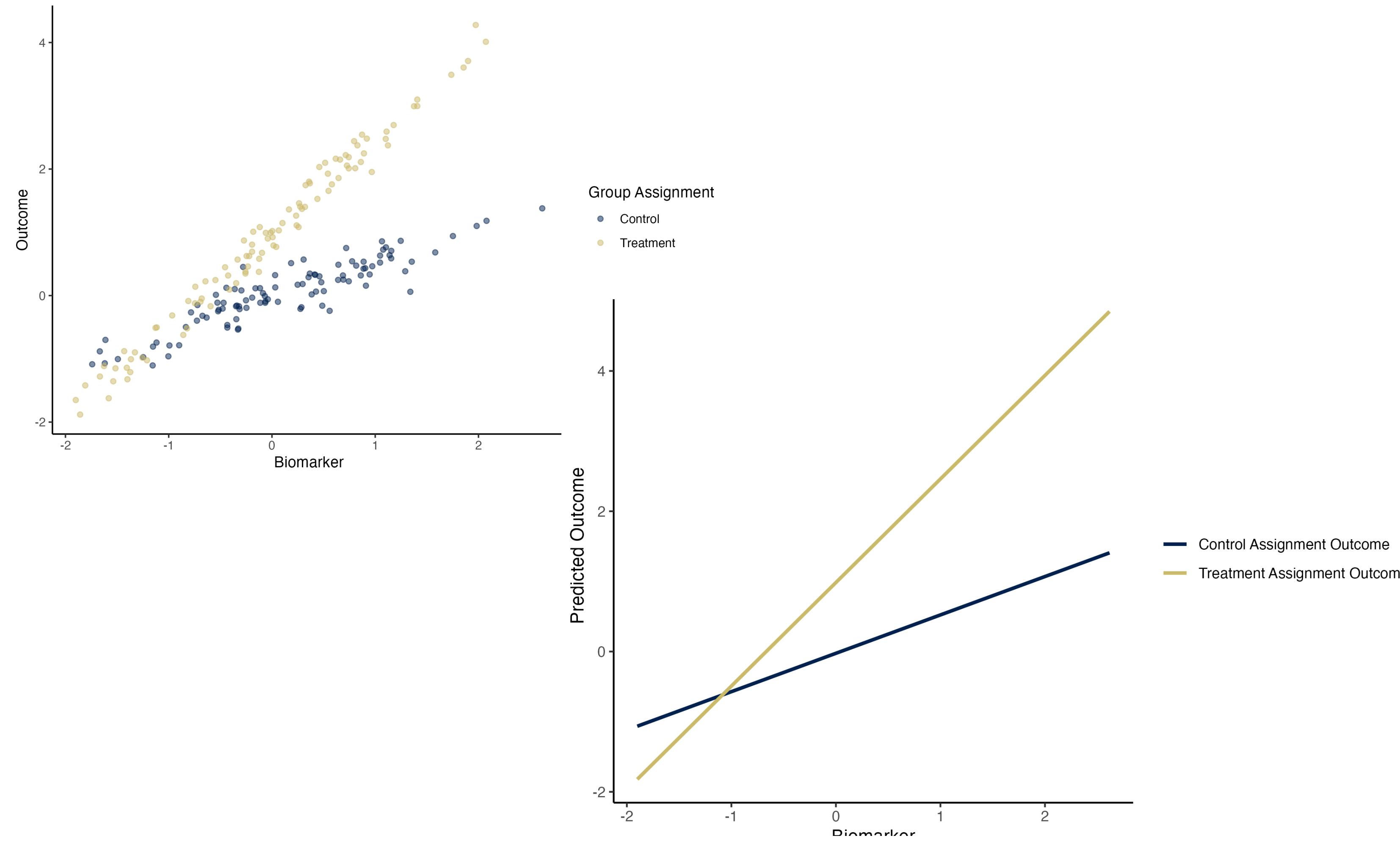
We need to consider alternative problem formulations.

A Dedicated (Variable Importance) Parameter

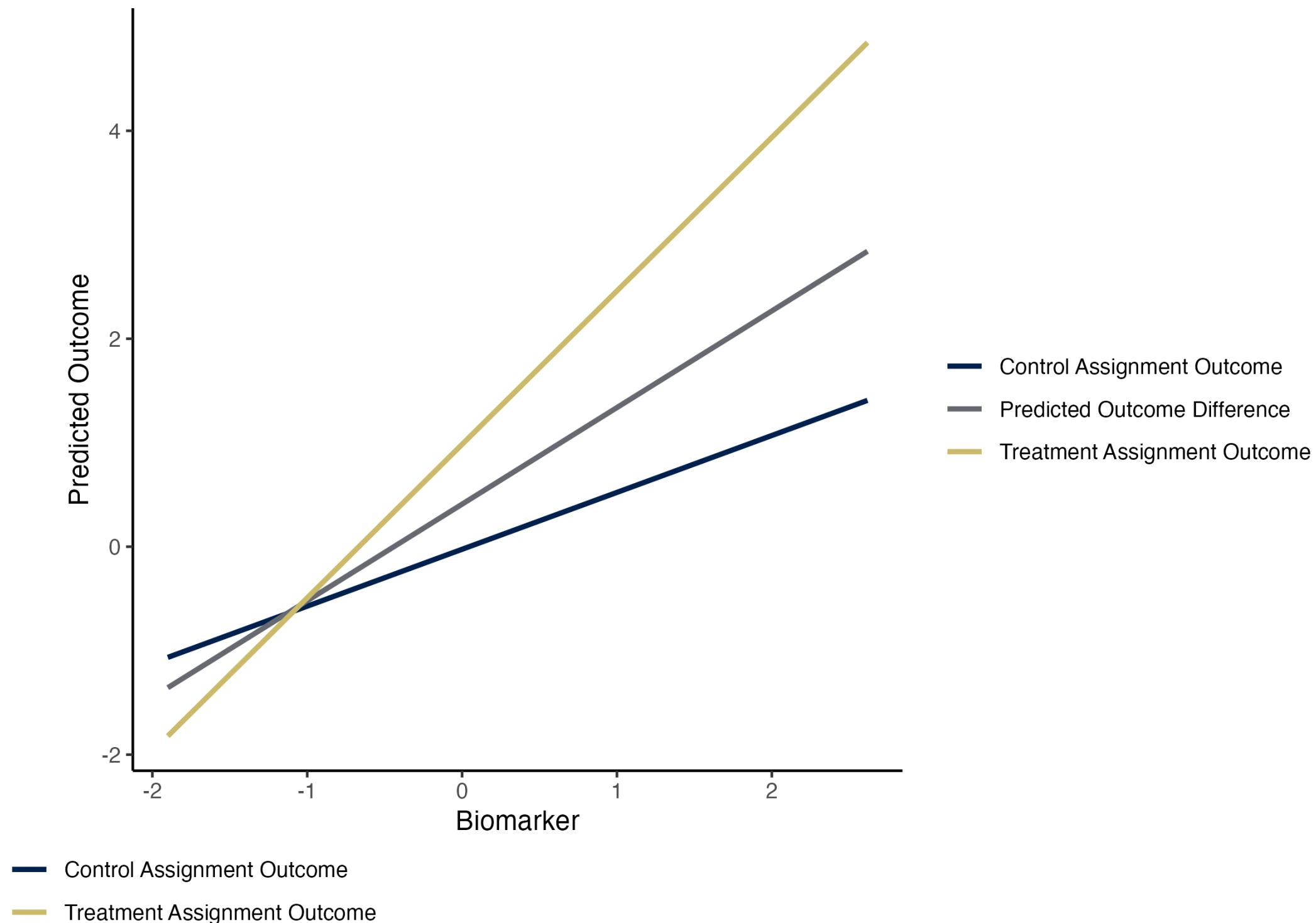
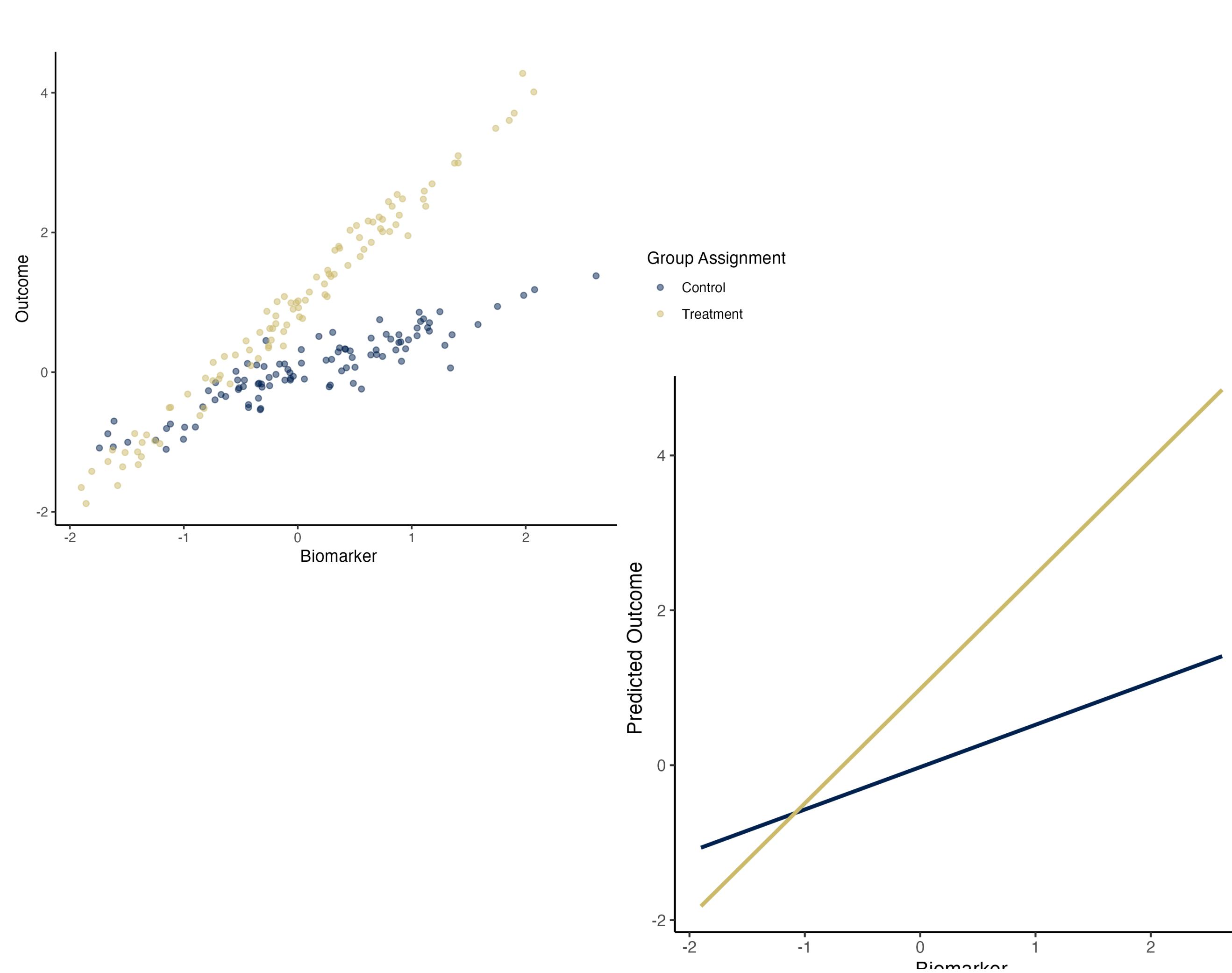
A Dedicated (Variable Importance) Parameter



A Dedicated (Variable Importance) Parameter



A Dedicated (Variable Importance) Parameter



uniCATE

Assumption-lean estimator of biomarker predictiveness

uniCATE

Assumption-lean estimator of biomarker predictiveness

Estimating this parameter for a centered biomarker is easy!

uniCATE

Assumption-lean estimator of biomarker predictiveness

Estimating this parameter for a centered biomarker is easy!

$$\frac{\frac{1}{n} \sum_{i=1}^n (\text{predicted outcome difference})_i (\text{biomarker})_i}{\frac{1}{n} \sum_{i=1}^n (\text{biomarker})_i^2}$$

uniCATE

Assumption-lean estimator of biomarker predictiveness

Estimating this parameter for a centered biomarker is easy!

$$\frac{\frac{1}{n} \sum_{i=1}^n (\text{predicted outcome difference})_i (\text{biomarker})_i}{\frac{1}{n} \sum_{i=1}^n (\text{biomarker})_i^2}$$

This estimator is **consistent** and **asymptotically Normal**. The only assumption: the biomarker has non-zero variance.

Ranking Biomarkers

uniCATE ranks biomarkers based on predictiveness

Ranking Biomarkers

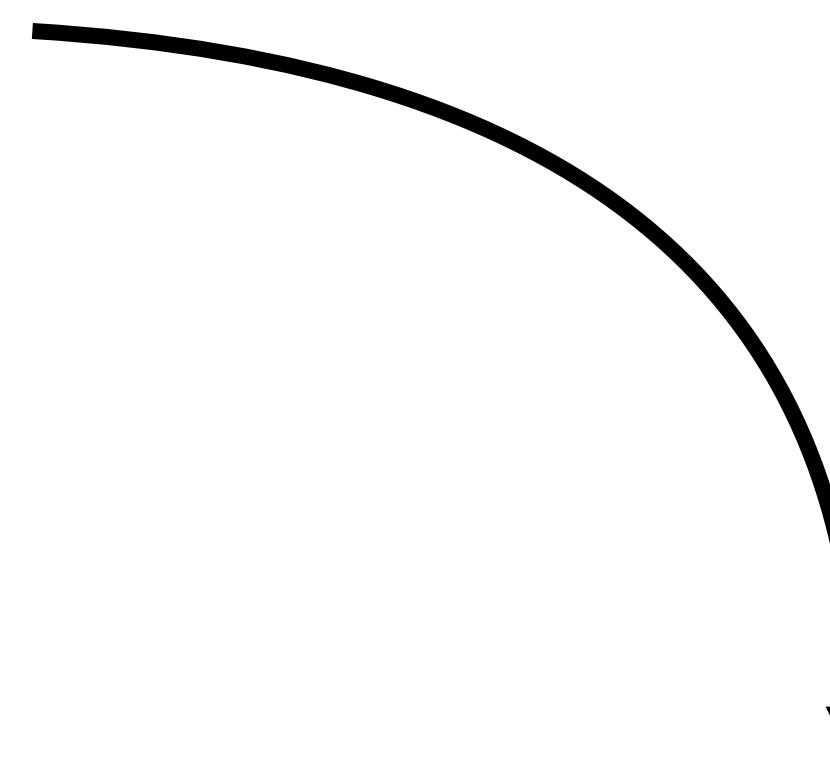
uniCATE ranks biomarkers based on predictiveness

```
n <- 200  
biomarker_1 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_2 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_3 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_4 <- rnorm(n, mean = 0, sd = 0.1)  
covariate <- 0.2 * rbinom(n, 1, 0.4)  
treatment <- rbinom(n, 1, 0.5)  
response <- covar + 1 * bio1 * treatment  
+ 2 * bio2 * treatment
```

Ranking Biomarkers

uniCATE ranks biomarkers based on predictiveness

```
n <- 200  
biomarker_1 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_2 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_3 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_4 <- rnorm(n, mean = 0, sd = 0.1)  
covariate <- 0.2 * rbinom(n, 1, 0.4)  
treatment <- rbinom(n, 1, 0.5)  
response <- covar + 1 * bio1 * treatment  
+ 2 * bio2 * treatment
```

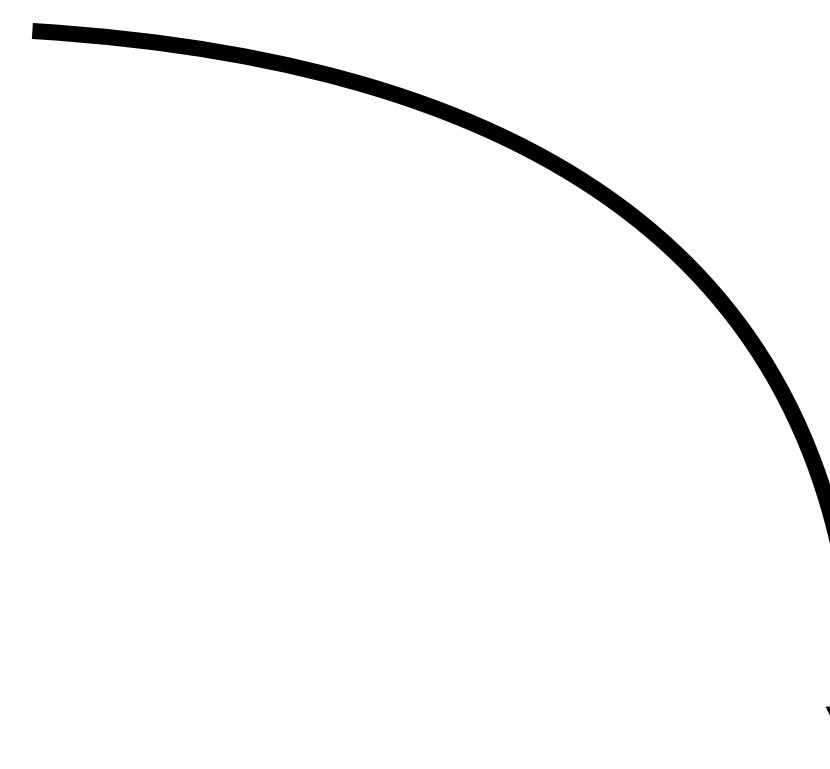


	Est.	SE	Z-score	P	P (BH)
biomarker_2	1.90	0.0768	24.7	3.58E-13	1.43E-13
biomarker_1	0.820	0.129	6.38	1.75E-10	3.51E-10
biomarker_3	0.0482	0.145	0.333	7.39E-01	9.03E-01
biomarker_4	-0.0202	0.166	-0.122	9.03E-01	9.03E-01

Ranking Biomarkers

uniCATE ranks biomarkers based on predictiveness

```
n <- 200  
biomarker_1 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_2 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_3 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_4 <- rnorm(n, mean = 0, sd = 0.1)  
covariate <- 0.2 * rbinom(n, 1, 0.4)  
treatment <- rbinom(n, 1, 0.5)  
response <- covar + 1 * bio1 * treatment  
+ 2 * bio2 * treatment
```

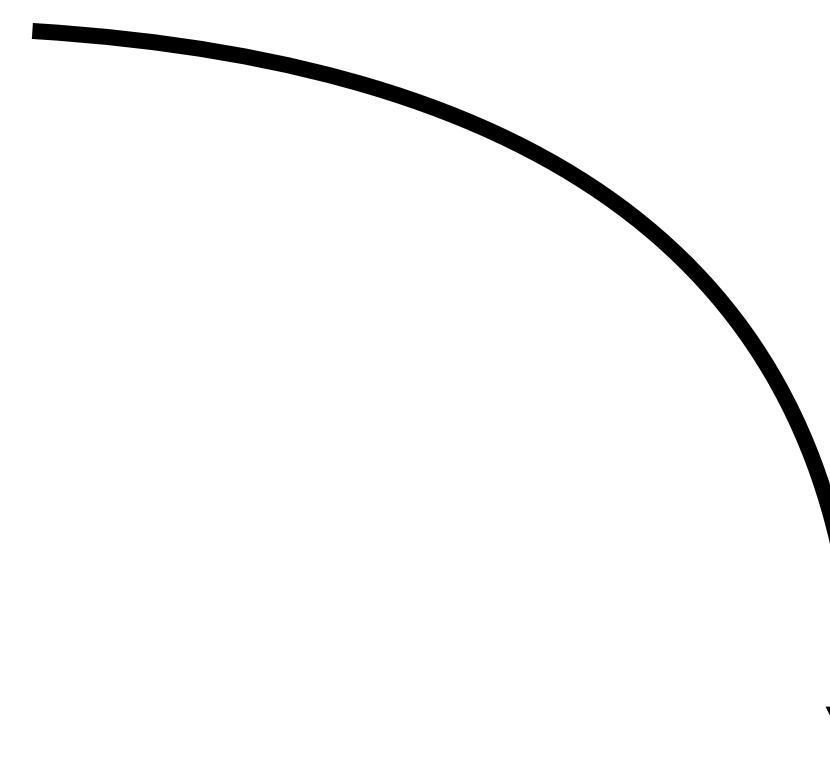


	Est.	SE	Z-score	P	P (BH)
biomarker_2	1.90	0.0768	24.7	3.58E-13	1.43E-13
biomarker_1	0.820	0.129	6.38	1.75E-10	3.51E-10
biomarker_3	0.0482	0.145	0.333	7.39E-01	9.03E-01
biomarker_4	-0.0202	0.166	-0.122	9.03E-01	9.03E-01

Ranking Biomarkers

uniCATE ranks biomarkers based on predictiveness

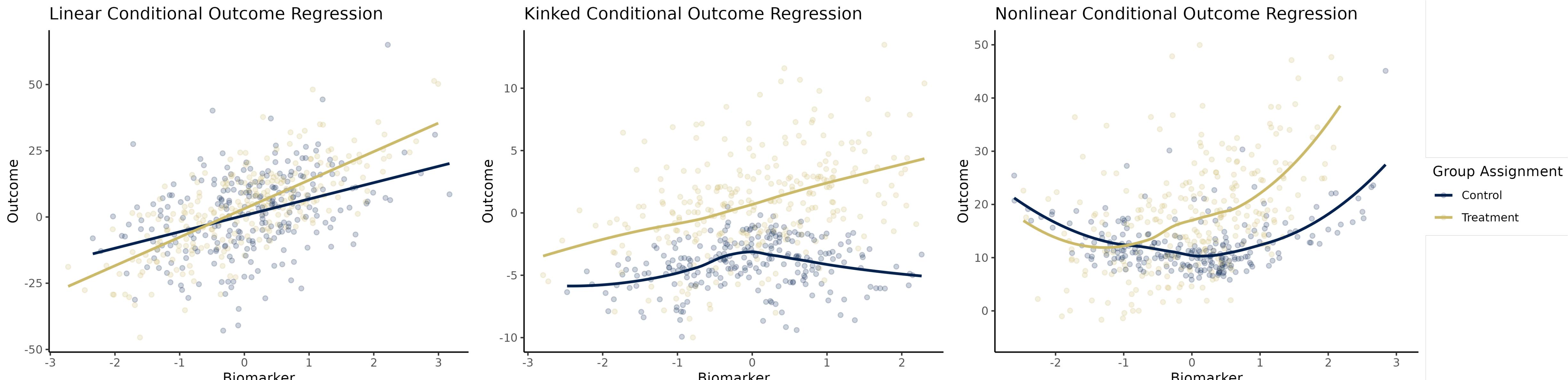
```
n <- 200  
biomarker_1 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_2 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_3 <- rnorm(n, mean = 0, sd = 0.1)  
biomarker_4 <- rnorm(n, mean = 0, sd = 0.1)  
covariate <- 0.2 * rbinom(n, 1, 0.4)  
treatment <- rbinom(n, 1, 0.5)  
response <- covar + 1 * bio1 * treatment  
+ 2 * bio2 * treatment
```



	Est.	SE	Z-score	P	P (BH)
biomarker_2	1.90	0.0768	24.7	3.58E-13	1.43E-13
biomarker_1	0.820	0.129	6.38	1.75E-10	3.51E-10
biomarker_3	0.0482	0.145	0.333	7.39E-01	9.03E-01
biomarker_4	-0.0202	0.166	-0.122	9.03E-01	9.03E-01

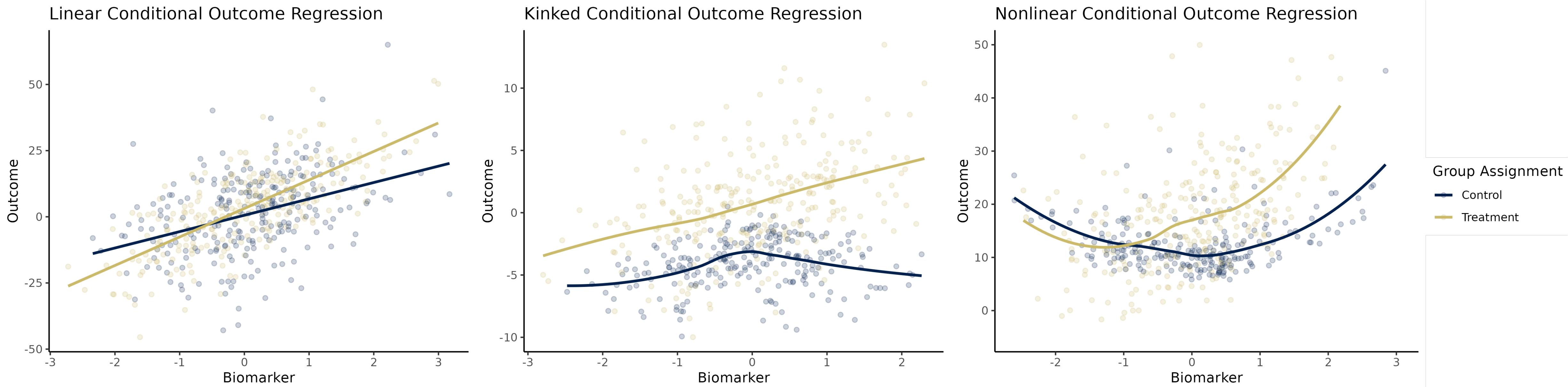
Confirmatory Simulations

Considered Biomarker-Outcome Relationships



Difficulty

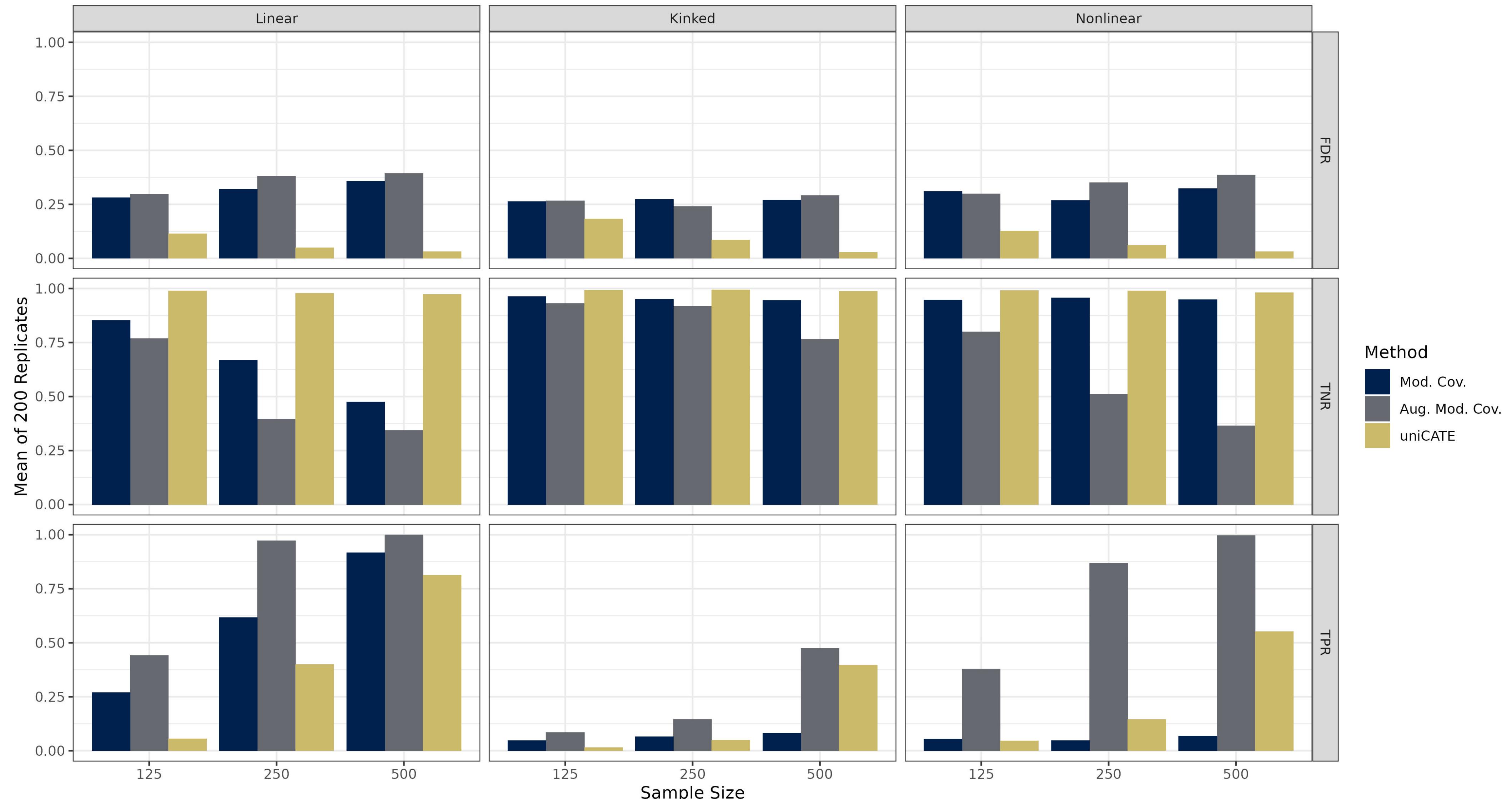
Considered Biomarker-Outcome Relationships



Difficulty

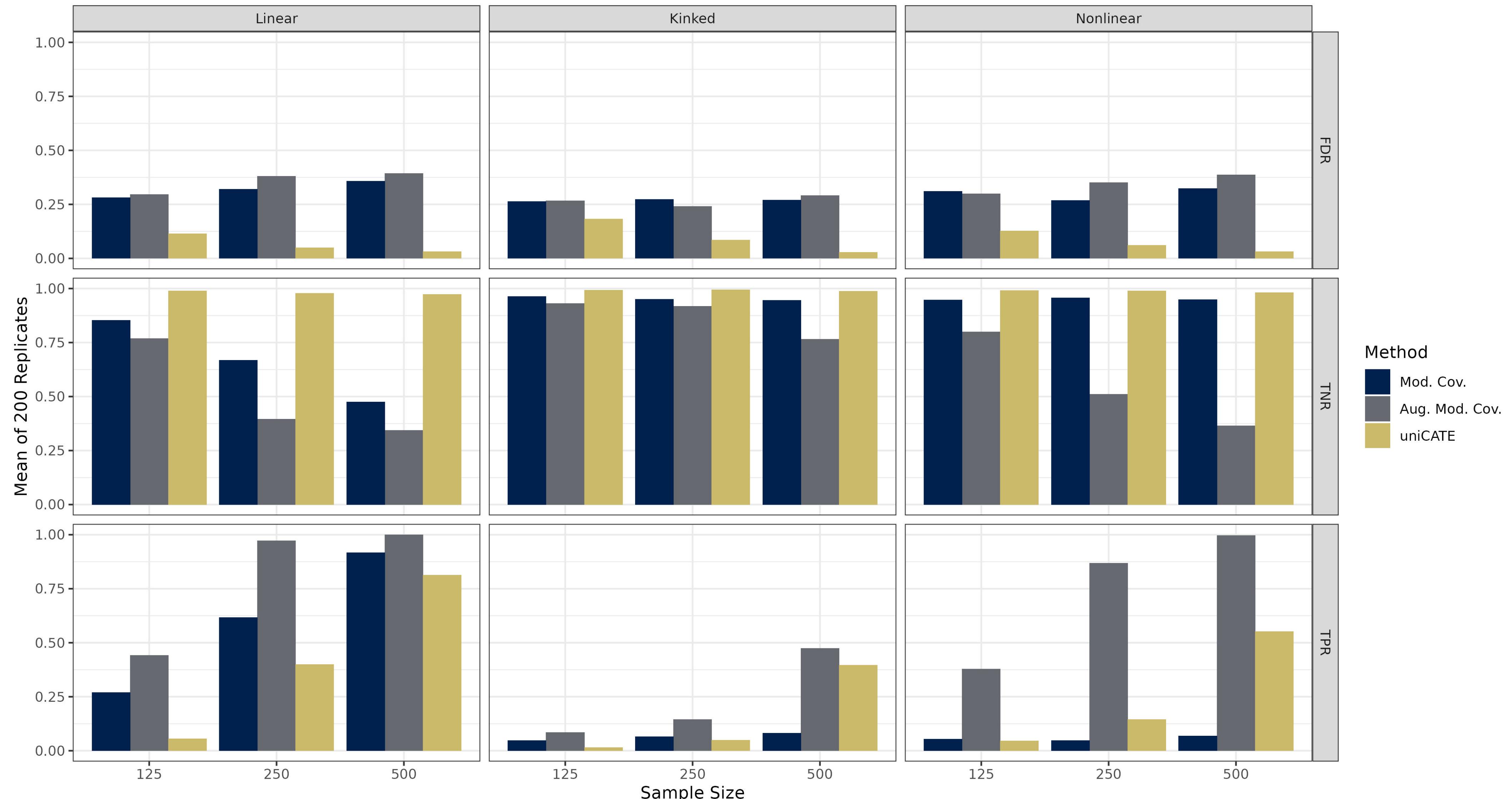
uniCATE Controls False Positive Rates

Classification of Non-Sparse, Moderate-Dimensional, and Uncorrelated Predictive Biomarkers



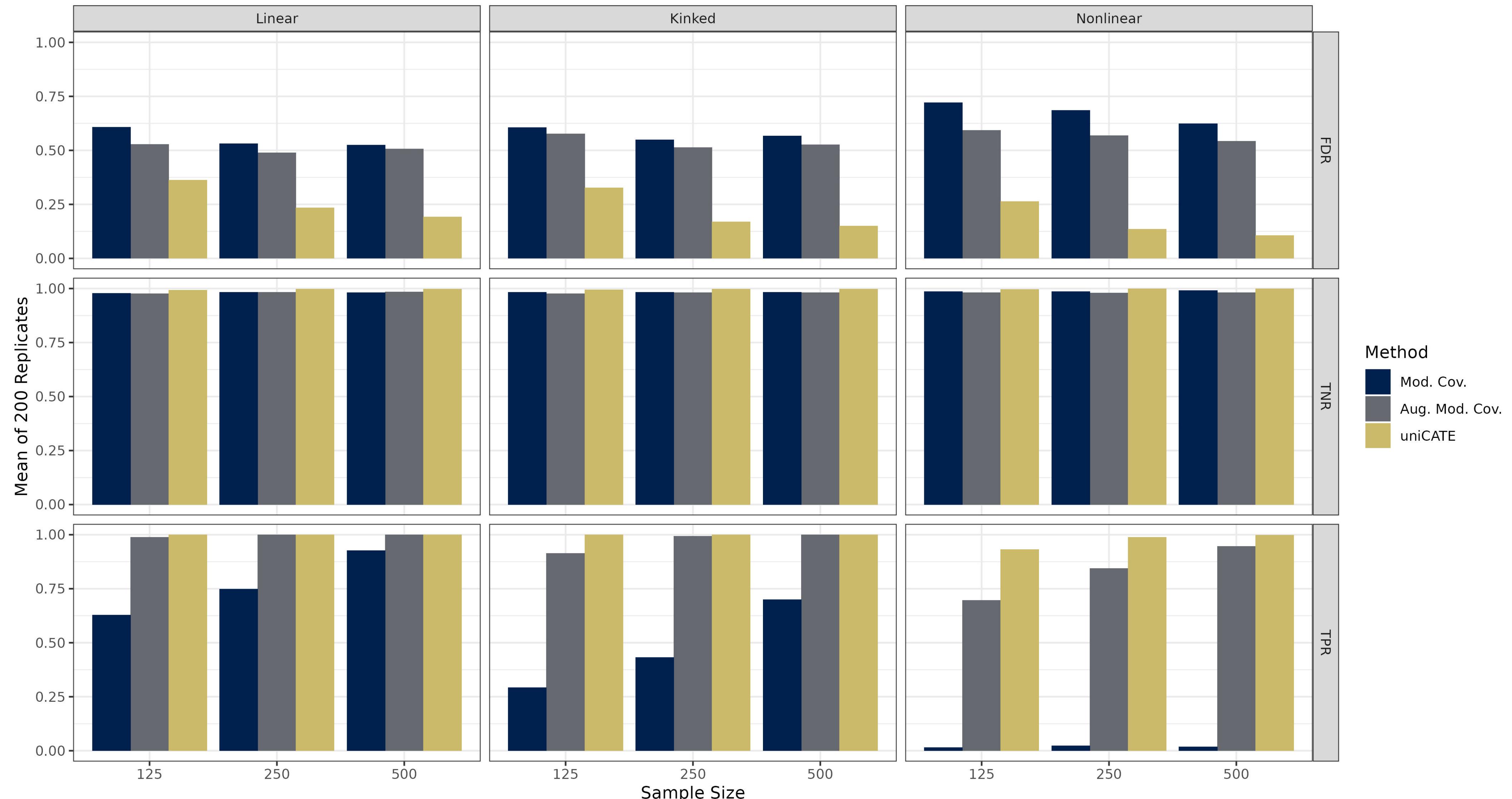
uniCATE Controls False Positive Rates

Classification of Non-Sparse, Moderate-Dimensional, and Uncorrelated Predictive Biomarkers



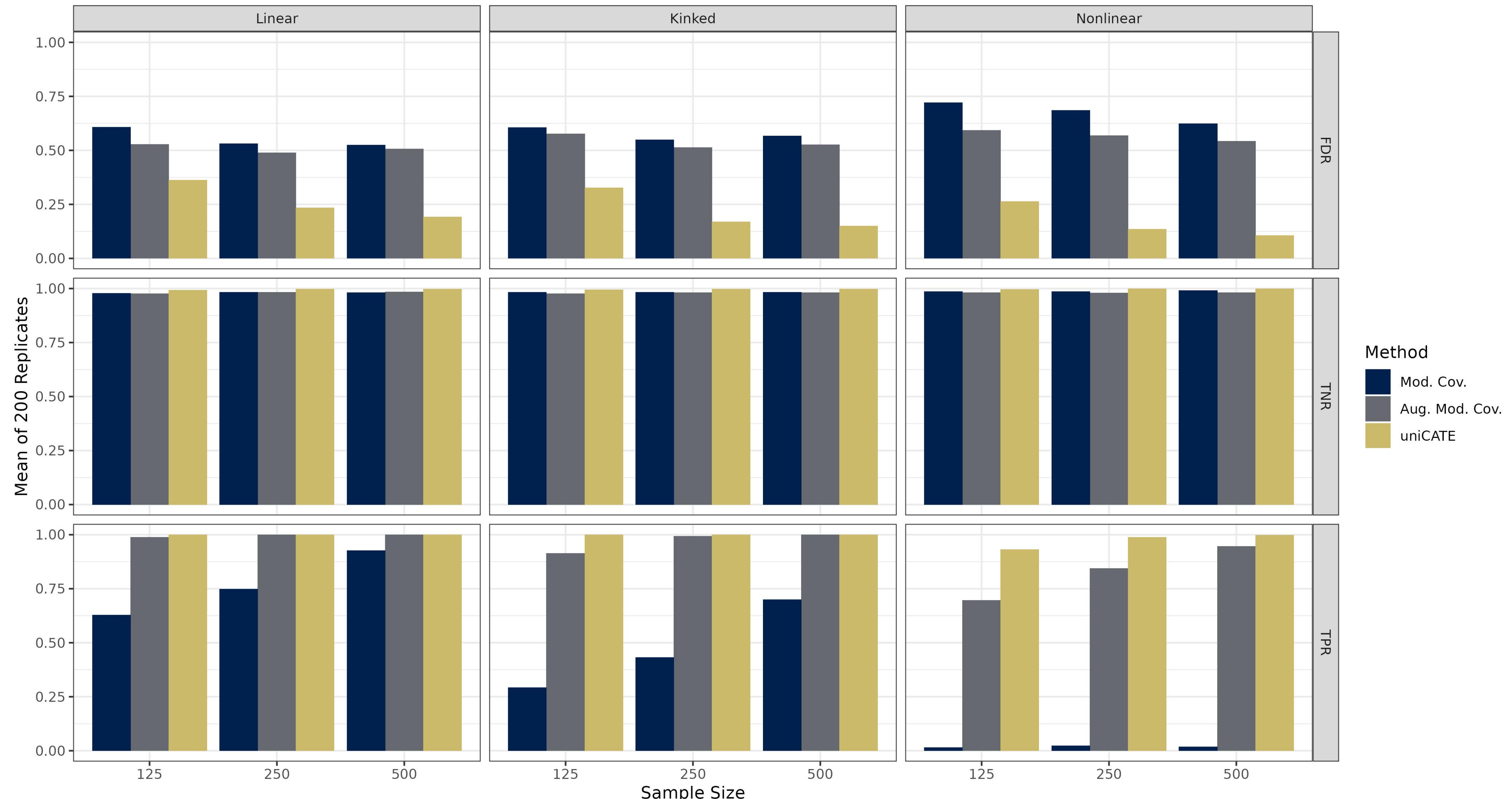
uniCATE Still Controls False Positive Rates

Classification of Sparse, High-Dimensional, and Correlated Predictive Biomarkers



uniCATE Still Controls False Positive Rates

Classification of Sparse, High-Dimensional, and Correlated Predictive Biomarkers



Application to IMmotion 150/151

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.
3. Objective response was used as the response variable.

Application to IMmotion 150

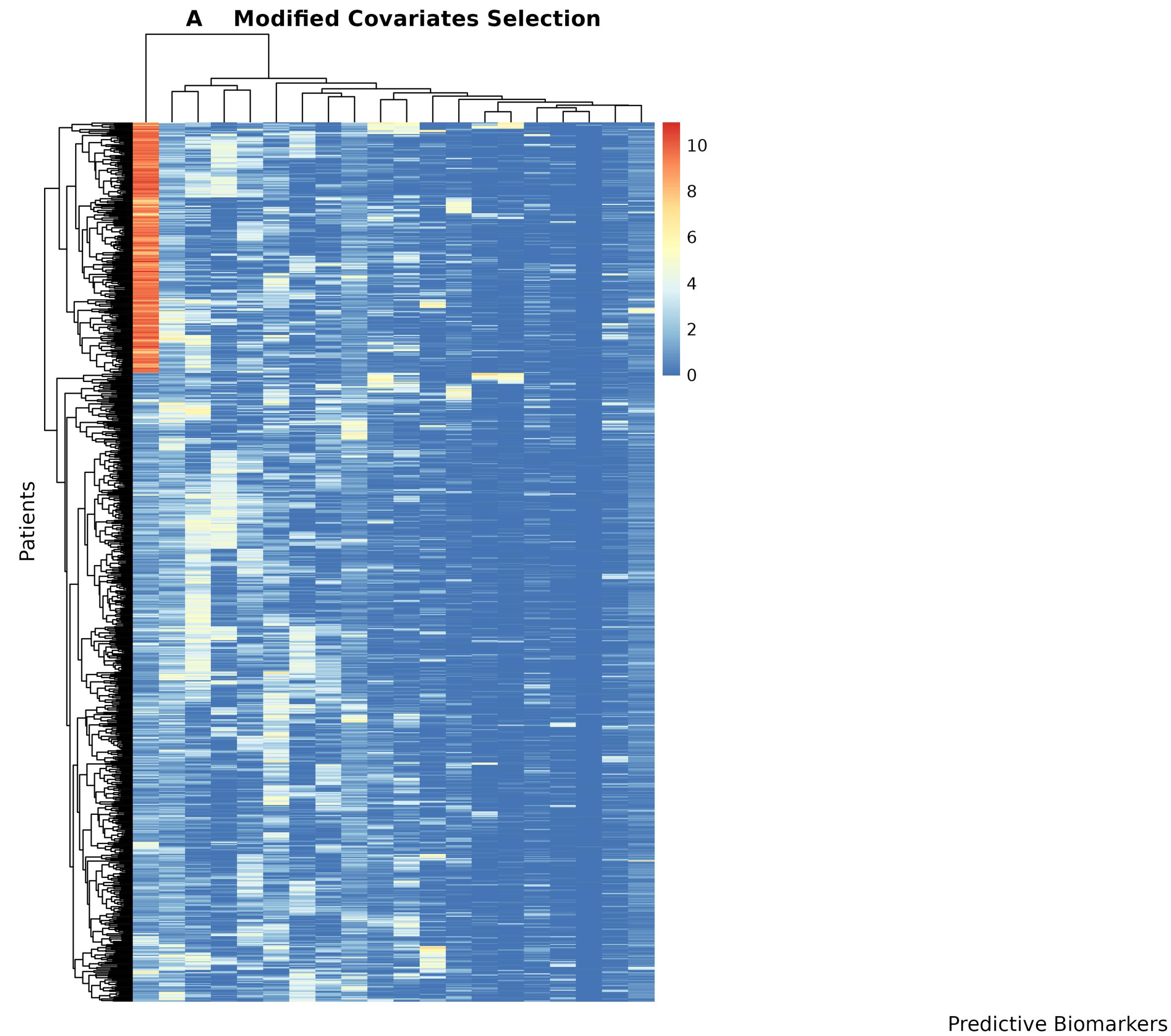
uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.
3. Objective response was used as the response variable.

92 genes were identified as predictive using a 5% FDR cutoff. They are associated with immune responses, including those mediated by B cells and lymphocytes.

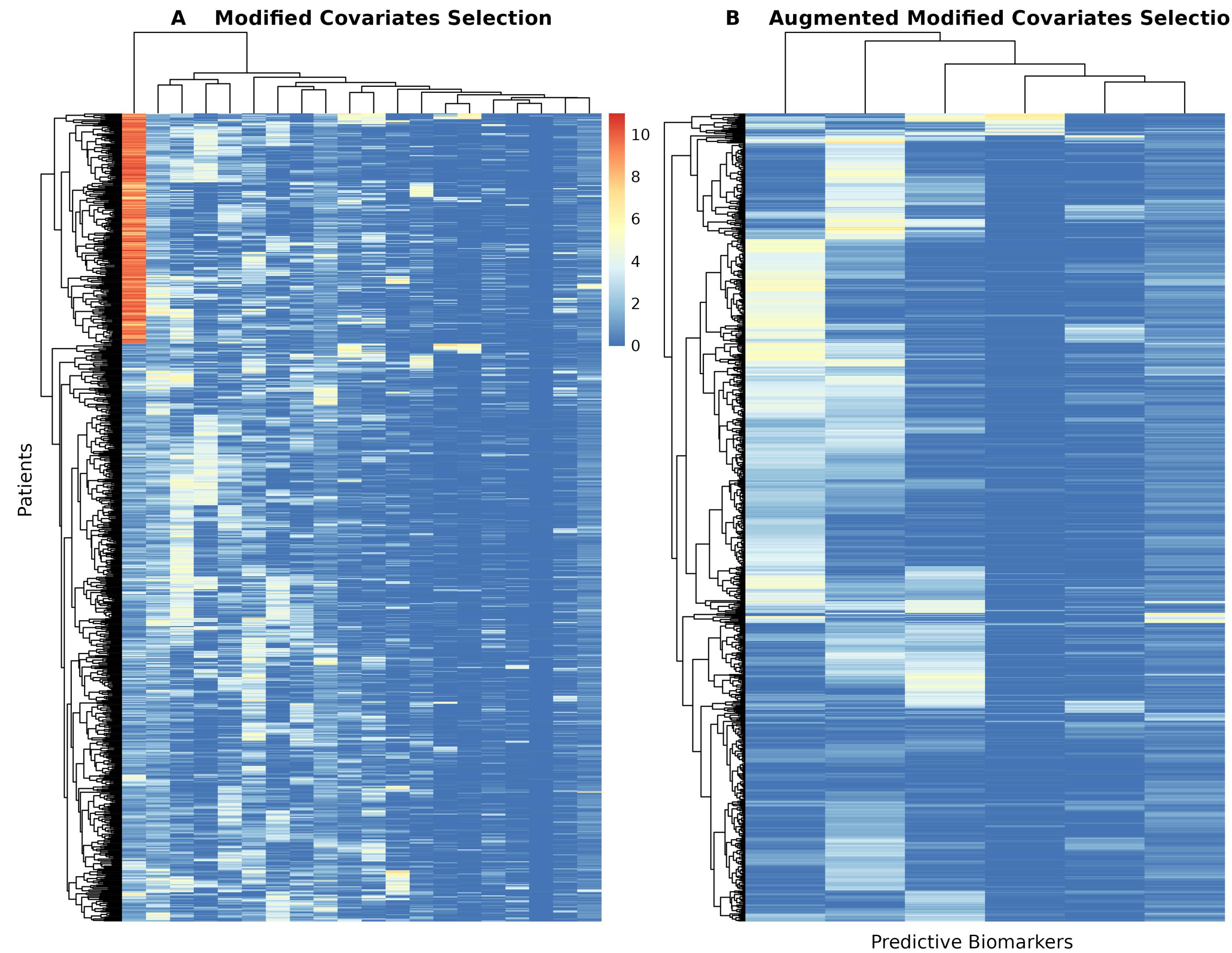
Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



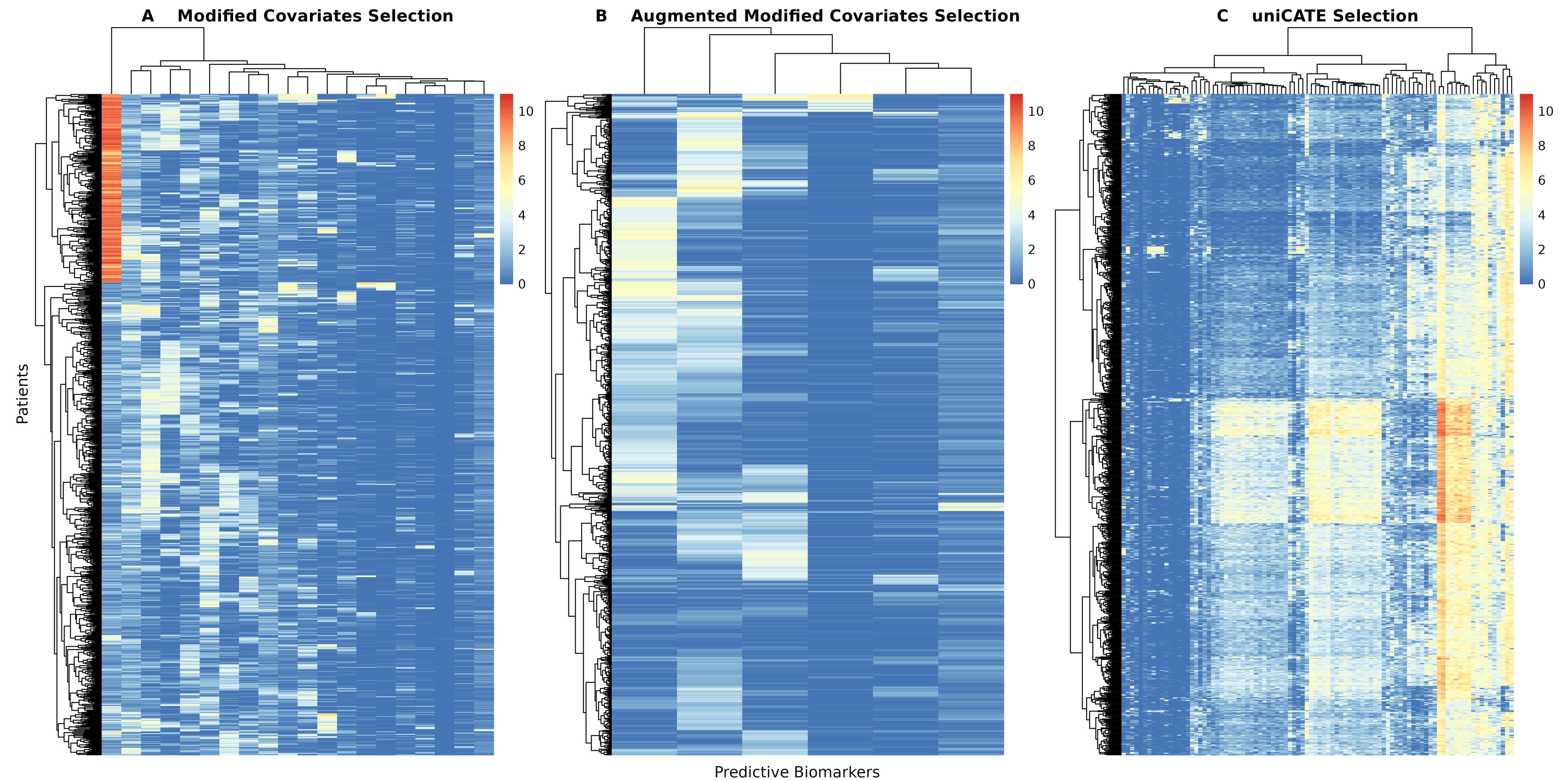
Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



Questions?

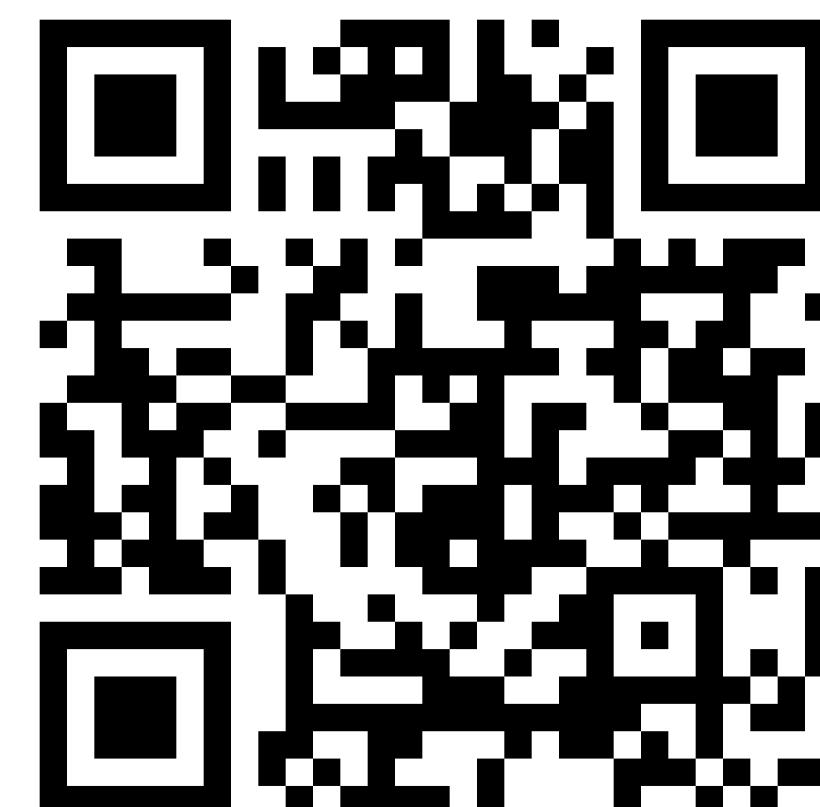
Next Steps

Extensions and follow-up projects

- uniCATE estimator was derived for continuous and binary outcomes. An analogous approach for time-to-event outcomes is available, too.
- uniCATE's target parameter is variable importance parameter on the absolute risk scale. Upcoming work will explore variable importance parameters on the relative risk scale.
- Current work is evaluating whether uniCATE's feature selections aid during treatment rule estimation, with positive results.

Conclusions

- uniCATE is an assumption-lean inference procedure that controls the rate of false positives in high dimensional randomized control trials.
- Check out uniCATE's implementation in the uniCATE R package, available at github.com/insightsengineering/uniCATE
- Manuscript is currently under review at Biostatistics. For the preprint:



Conclusions

- uniCATE is an assumption-lean inference procedure that controls the rate of false positives in high dimensional randomized control trials.
- Check out uniCATE's implementation in the uniCATE R package, available at github.com/insightsengineering/uniCATE
- Manuscript is currently under review at Biostatistics. For the preprint:

