A comparison of subgroup identification methods in clinical drug development

Cynthia Huber [1], Norbert Benda [2,1] and Tim Friede [1]

[1] Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[2] Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany

24 October 2019

Acknowledgement: This work was co-funded by the BfArM and the University Medical Center Göttingen

UNIVERSITÄTSMEDIZIN GÖTTINGEN : UMG

Subgroups in drug development (1/2)

- Precision medicine aims at identifying (biomarker-defined) subpopulations which differ in the efficacy (or safety) of a specific treatment
  $\rightarrow$ Identifying treatment-by-biomarker or treatment-by-subgroup interactions
  $\rightarrow$ **Predictive biomarkers**
- In contrast, prognostic biomarkers predict the natural course of a disease
- Biomarker may refer to genetic markers or other baseline characteristics including demographic and clinical characteristics

Subgroups in drug development (2/2)

- The drug's mechanism of action often generates hypotheses whether a biomarker is predictive or not
- Without prior hypotheses about treatment-by-subgroup interactions methods identifying subgroups are needed
- Numerous available subgroup identification methods (Lipkovich et al., 2017; Ondra et al., 2016)
- Systematic and independent comparison studies are lacking
- Existing comparison studies do not always consider criteria relevant in regulatory settings

Subgroup identification methods

Methods for identifying subgroups with enhanced treatment effect and selecting cut-off values in case of continuous biomarkers include

- IT: **I**nteraction **T**rees (Su et al., 2009)
- MOB: **Mo**del-**b**ased Recursive Partitioning (Seibold et al., 2016)
- STIMA: **S**imultaneous **T**hreshold **I**nteraction **M**odeling **A**lgorithm (Dusseldorp et al., 2010)
- SIDES: **S**ubgroup **I**dentification based on **D**ifferential **E**ffect **S**earch (Lipkovich et al., 2011)
- ARDP: **A**daptive **R**efinement by **D**irected **P**eeling (Patel et al., 2016; LeBlanc et al., 2005)

Application to Amyotrophic Lateral Sclerosis (ALS) data (1/2)

- ALS is a rare disease with an annual incidence of 2/100 000 (Atassi et al., 2014)
- ALS affects the nervous system
- ALS leads to loss of muscle function and paralysis
- Median survival time is about 2-3 years (EMA, 2016)
- Two approved drugs in the EU: riluzole and edaravone
- Both treatments do not achieve substantial benefit for ALS patients (FDA, 2009, 2017)
- PRO-ACT (Pooled Resource Open Access ALS Clinical Trials) database

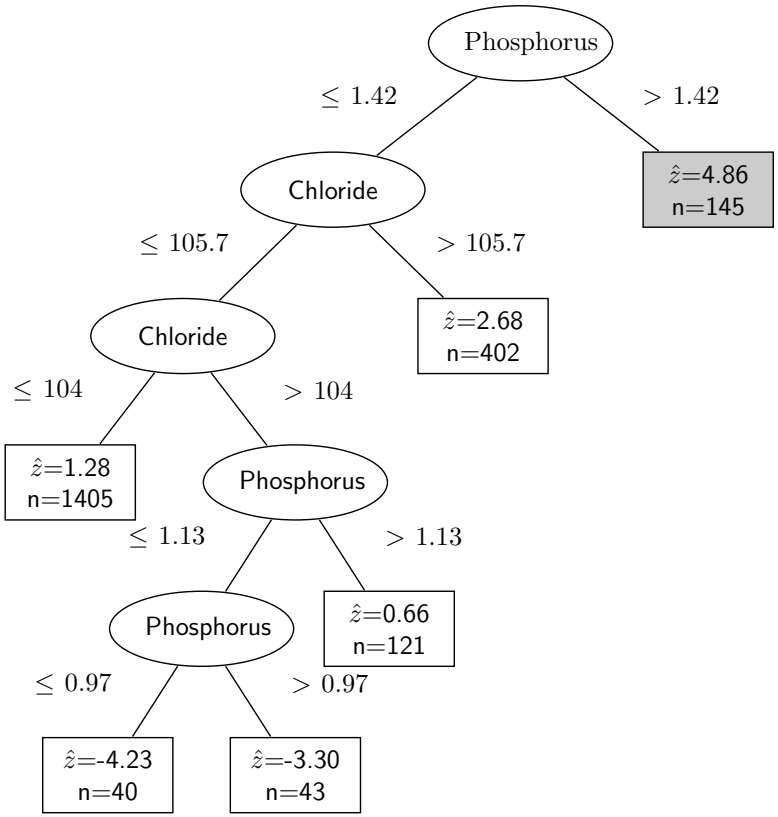
Application to ALS data (2/2)

- Outcome: ALSFRS (ALS Functional Rating Scale) after 6 month
- ALSFRS is the sum of 10 items regarding motor function
- Each item is rated on a scale from 0 to 4 with 0 indicating no function
- For illustrative purpose, preselection of two covariates based on p-values of treatment-by-biomarker interactions in a linear model: Phosphorus and chloride
- Data include 2156 observations
- Treatment effect in an identified subgroup $\hat{S}$:

$$z(\hat{S}) = E(Y|T = 1, \mathbf{X} \in \hat{S}) - E(Y|T = 0, \mathbf{X} \in \hat{S}),$$

with Y denoting the outcome, T the treatment indicator and $\mathbf{X}$ denoting the covariates

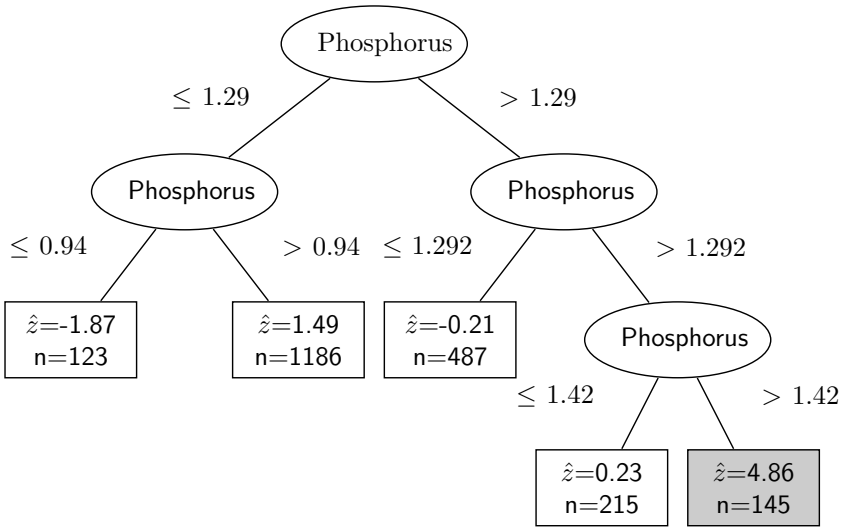
Interaction Tree (IT)



IT without pruning applied to ALS data

- Each leaf is associated with a linear model
- Model: $E(Y|\mathbf{X}) = \alpha + \beta_0 \cdot T + \beta_1 \cdot T \cdot I(X_j \leq c) + \gamma_1 \cdot I(X_j \leq c)$
- Selection of **split** is based on testing $H_0 : \beta_1 = 0$
- Post-pruning based on an interaction-complexity criterion

Model-based recursive partitioning (MOB)

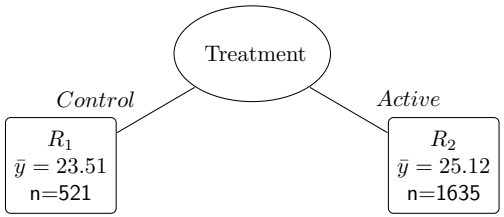


MOB applied to ALS data

- Each leaf is associated with a linear model
- Model: $E(Y|\mathbf{X}) = \alpha + \beta \cdot T$
- Selection of **splitting variable** is based on testing for parameter instabilities in α and β
- Partition if p-value of test is smaller than a nominal level

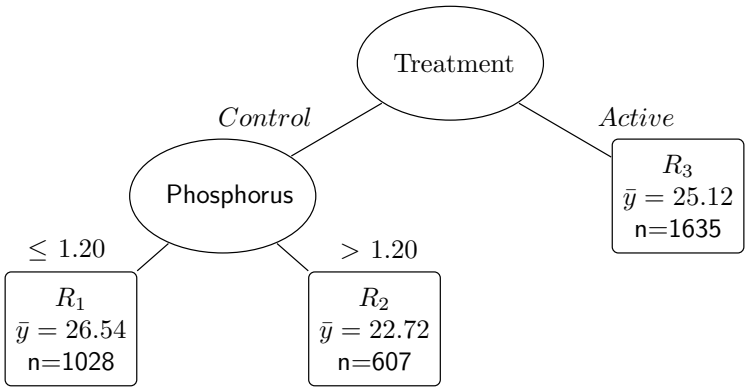
Simultaneous Threshold Interaction Modeling Algorithm (STIMA)

- STIMA uses a linear regression model for modelling main effects and a tree for modelling interactions with the treatment indicator



$$E(Y|\mathbf{X}) = \alpha + \beta_0 \cdot I(T = 1) + \sum_{j=1}^{p} \gamma_j \cdot X_j$$

- **Best split**: Highest increase in variance accounted for by an expanded model



Expanded model:

$$E(Y|\mathbf{X}) = \alpha + \beta_0 I(T = 1) + \beta_1 I(T = 0) I(X_{Phosphorus} > 1.2) + \sum_{j=1}^{p} \gamma_j X_j$$

Selection of a target population

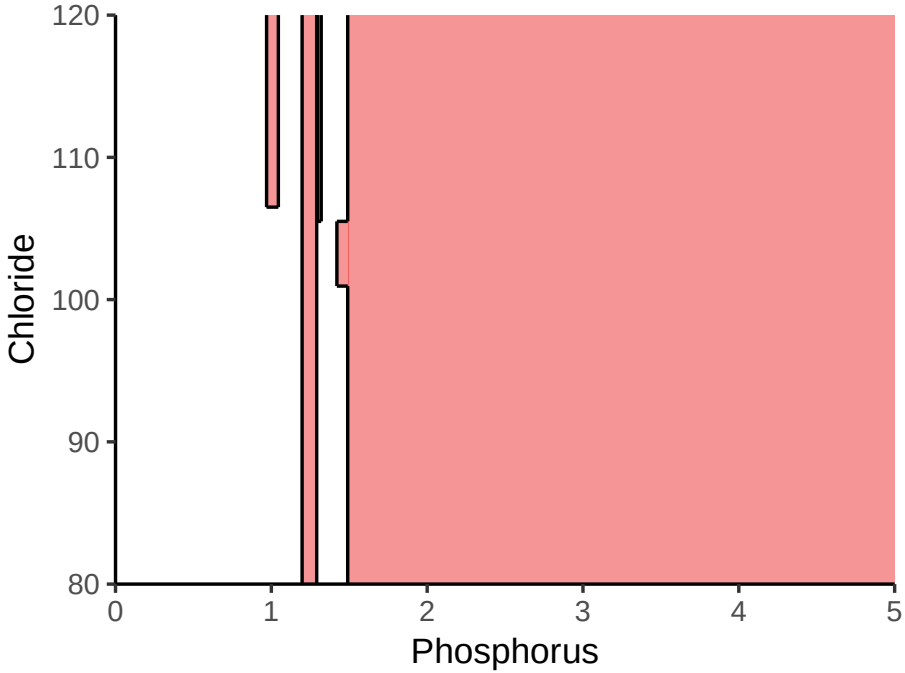
- Subgroups resulting from the methods are defined by
  - IT and MOB: Terminal nodes
  - STIMA: Combining terminal nodes of the $T = 1$ and $T = 0$ branches
- Identified subgroups $\hat{S}$ meeting the criteria

$$z(\hat{S}) > \text{mintrt}$$

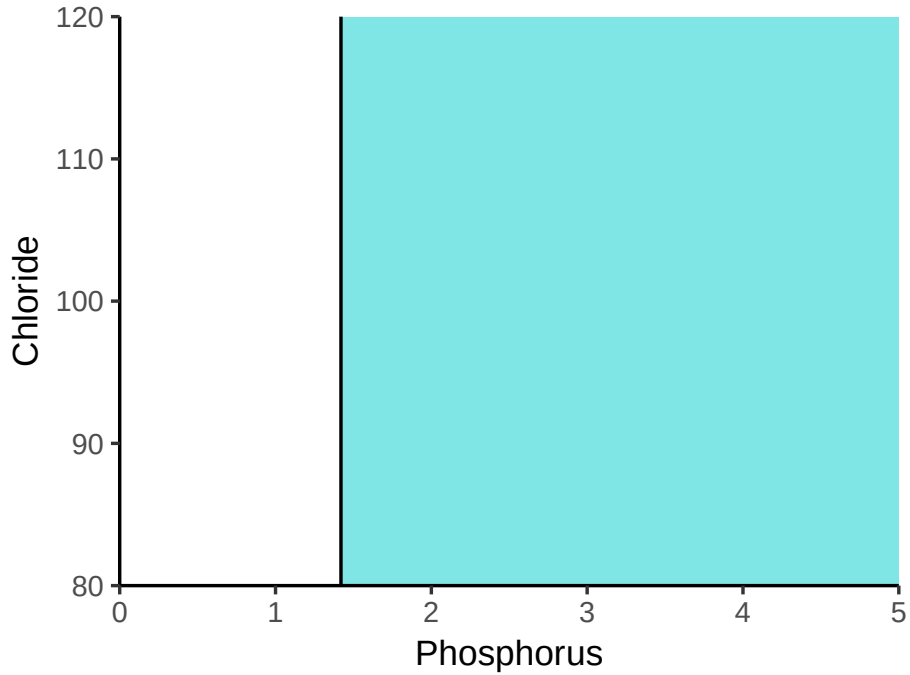
  are selected as target population, referred to as $BM+$
- For ALS the $BM+$ subgroup is selected based on $z(\hat{S}) > 3$

Shape of selected *BM+* subgroup in ALS example

STIMA[1]



IT[1]/MOB /ARDP



All identified *BM+* subgroups include patients with higher phosphorus levels and differ only slightly.

[1]Without originally proposed post-pruning procedure

Simulation - Data generating process

- Data generating model

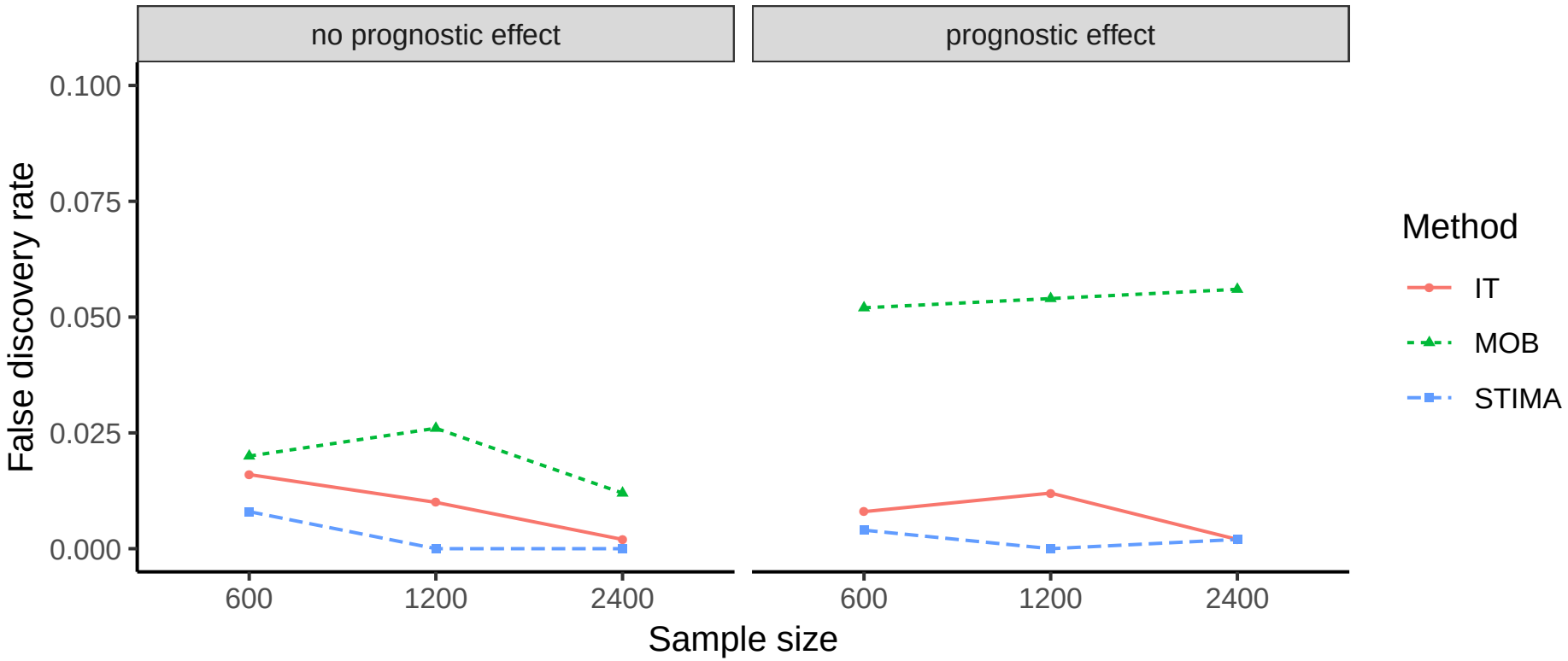
$$Y_i = \mu(T_i, \mathbf{X}_i) + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, 1)$$

- To assess the **rate of falsely identifying a** $BM+$ **subgroup** although the treatment effect is homogeneous across the entire population, the following model is used

$$\mu(T, \mathbf{X}) = 0.2 \cdot T + \gamma \cdot I(X_1 > 0)$$

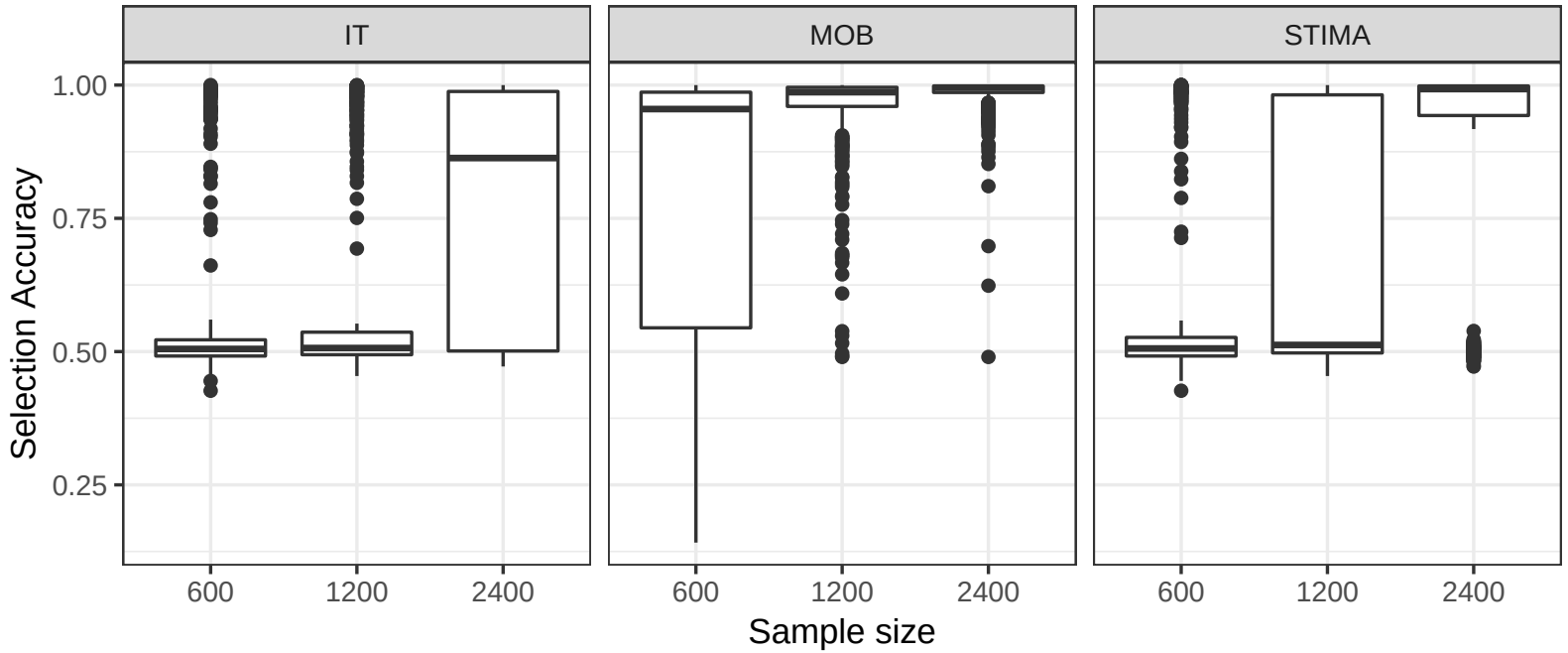
- Distribution of $T \sim \mathcal{B}(1, 0.5)$ and $X_1, X_2, X_3, X_4 \overset{iid}{\sim} \mathcal{N}(0, 1)$
- Number of simulations: 500 per scenario
- Selection of the $BM+$ subgroup is based on the threshold $mintrt = 0.4$
- Sample size n and effect of prognostic variables γ are varied

False discovery rate



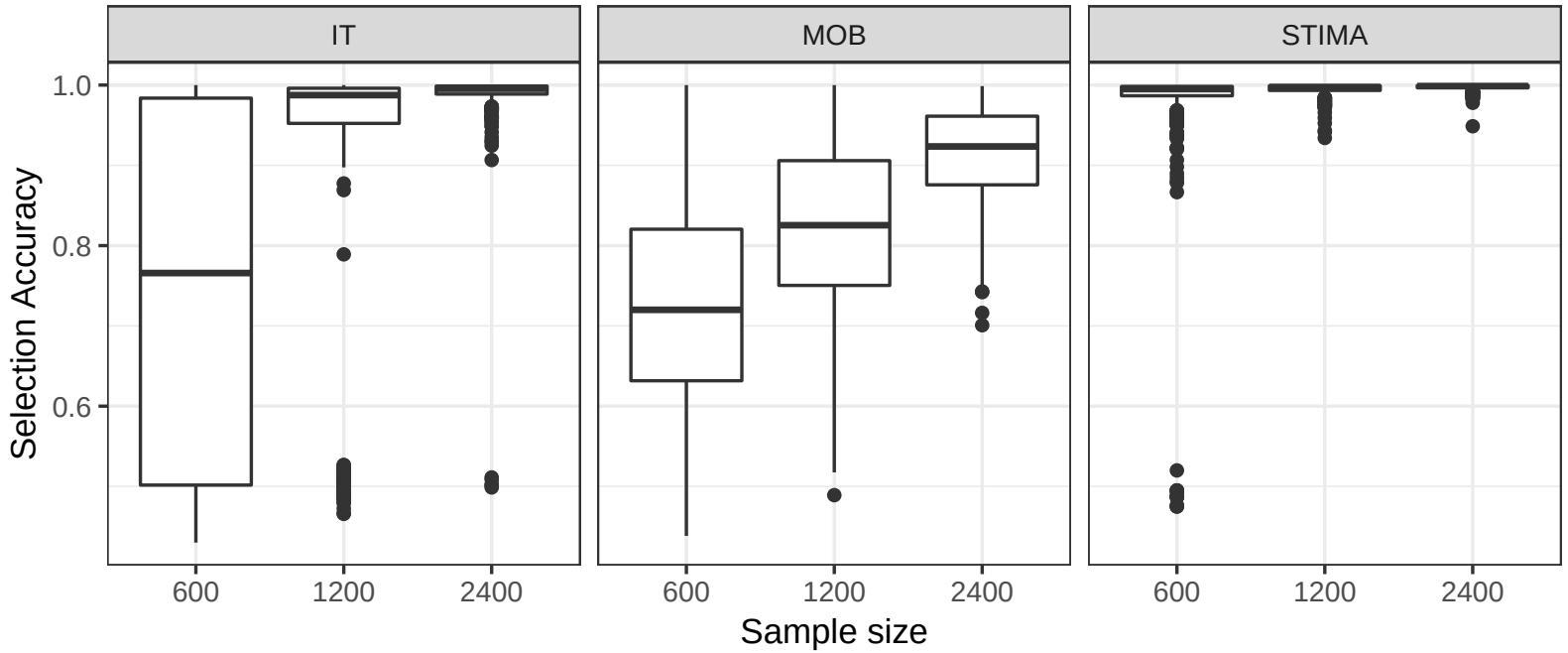
STIMA and IT have false discovery rates below 2% across all settings

Proportion of correctly classified patients
Step function scenario



Data generation based on $\mu(T, \mathbf{X}) = 0.2 \cdot T + 0.5 \cdot T \cdot I(X_1 > 0)$ resulting in treatment effects of 0.2 and 0.7 in subgroups $X_1 \leq 0$ and $X_1 > 0$,respectively

Proportion of correctly classified patients
Prognostic effects scenario



Data generation with a larger interaction based on
$\mu(T, \mathbf{X}) = -0.3 \cdot X_1 + 0.4 \cdot X_2 + 0.3 \cdot X_4 + T \cdot I(X_1 > 0)$

Summary (1/3)

- The erroneous selection of a target population although none exists occurs the least frequently for IT and STIMA
- ARDP is not suitable to select BM+ subgroups when the threshold *mintrt* is chosen to be close to the treatment effect in the overall population (data not shown)
- With SIDES we obtain better BM+ subgroups in the presence of no treatment effect in the overall population compared to settings where an overall treatment effect is present (data not shown)

Summary (2/3)

- MOB, IT and STIMA classify patients well into a target subgroup and its complement when both sample size and treatment-by-subgroup interaction effect size are larger
  Exception: Settings with prognostic markers
- STIMA is the method of choice in scenarios where all considered covariates are prognostic covariates
- MOB is the most promising when we can assume that biomarkers are only predictive or both prognostic and predictive at the same time
  $\rightarrow$ Model parameters in the identified subgroups should be examined in order to decide whether biomarker is only prognostic

Summary (3/3)

- All methods have difficulties with the identification of a target subgroup when data include only 600 (or less) observations and when the treatment effect is smaller.
- Considered sample sizes of ($n \geq 600$) are rarely found in phase II trials with continuous outcome
- Pooling data of multiple trials and accounting for between-trial heterogeneity
  $\Rightarrow$ Some extension for individual patient data meta-analysis available (Mistry et al., 2018; Fokkema et al., 2017; Patel et al., 2016)

Discussion

- Limitations of the simulation study
  - Small number of potential predictive markers
  - Independent covariates
  - Only continuous covariates
- Restriction of the target population might increase probability of success of future trials
- IT, STIMA and MOB helpful for identifying restricted target population
- Adding biological rational based on the drug's mechanism of action adds to credibility of exploratory subgroup findings

Bibliography

Atassi, N., Berry, J., Shui, A., Zach, N., Sherman, A., Sinani, E., et al. (2014). The PRO-ACT database Design, initial analyses, and predictive features. *Neurology*, *83*(19), 1719–1725.

Chio, A., Lagroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., et al. (2009). Prognostic factors in als: A critical review. *Amyotrophic Lateral Sclerosis*, *10*, 310-323.

Dusseldorp, E., Conversano, C., & Van Os, B. J. (2010). Combining an additive and tree-based regression model simultaneously: Stima. *J Comput Graph Stat*, *19*(3), 514-530.

EMA. (2016). Clinical investigation of medicinal products for the treatment of amyotrophic lateral sclerosis.
(Reference number: EMA/531686/2015, Corr.1)

FDA. (2009). Labeling: Rilutek (riluzole) tablets.
(Application Number NDA 20599 S-011 & S-012)

FDA. (2017). Radicava (edaravone injection), for intravenous use.
(Reference ID: 4094543)

Bibliography (continued)

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2017). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behav Res Methods*, 1–19.

LeBlanc, M., Moon, J., & Crowley, J. (2005). Adaptive risk group refinement. *Biometrics*, *61*(2), 370–378.

Lipkovich, I., Dmitrienko, A., & D'Agostino, R. B. (2017). Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med*, *36*(1), 136–196.

Lipkovich, I., Dmitrienko, A., Denne, J., & Enas, G. (2011). Subgroup identification based on differential effect search - a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat Med*, *30*(21), 2601-2621.

Mistry, D., Stallard, N., & Underwood, M. (2018). A recursive partitioning approach for subgroup identification in individual patient data meta analysis. *Stat Med*, *37*(9), 1550-1561.

Ondra, T., Dmitrienko, A., Friede, T., Graf, A., Miller, F., Stallard, N., et al. (2016). Methods for identification and confirmation of targeted subgroups in clinical trials: A systematic review. *J Biopharm Stat*, *26*(1), 99-119.

Bibliography (continued)

Patel, S., Hee, S. W., Mistry, D., Jordan, J., Brown, S., Dritsaki, M., et al. (2016). Identifying back pain subgroups; developing and applying approaches using individual patient data collected within clinical trials. *Programme Grants for Applied Research*, *4*(10), 1-314.

Seibold, H., Zeileis, A., & Hothorn, T. (2016). Model-based recursive partitioning for subgroup analyses. *Int J Biostat*, *12*(1), 45-63.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., & Li, B. (2009). Subgroup analysis via recursive partitioning. *J Mach Learn Res*, *10*, 141–158.

Overview of the methods' properties (1/4)

	IT	MOB	STIMA	SIDES	ARDP
Aim					
Identifying subgroups defined by predictive covariates	yes	yes	yes	yes	yes
Identifying subgroups defined by prognostic covariates	no	yes	no	no	no*

* ARDP initially proposed by (LeBlanc et al., 2005) was developed for identifying prognostic markers only. The here used extension by Patel et al.(Patel et al., 2016) aims at peeling on predictive markers only.

Overview of the methods' properties (2/4)

	IT	MOB	STIMA	SIDES	ARDP
Algorithm					
Recursive partitioning	yes	yes	yes	yes	no
Evaluating splitting criterion at every possible cut-off point for every covariate	yes	no	yes	yes	no
Selection of covariate and cut-off value simultaneously	yes	no	yes	yes	yes
Covariates can be involved in multiple splits	yes	yes	yes	no	yes
Post-pruning procedure	yes	no	yes	N/A	N/A

Statements not reasonably interpretable are marked with N/A.

Overview of the methods' properties (3/4)

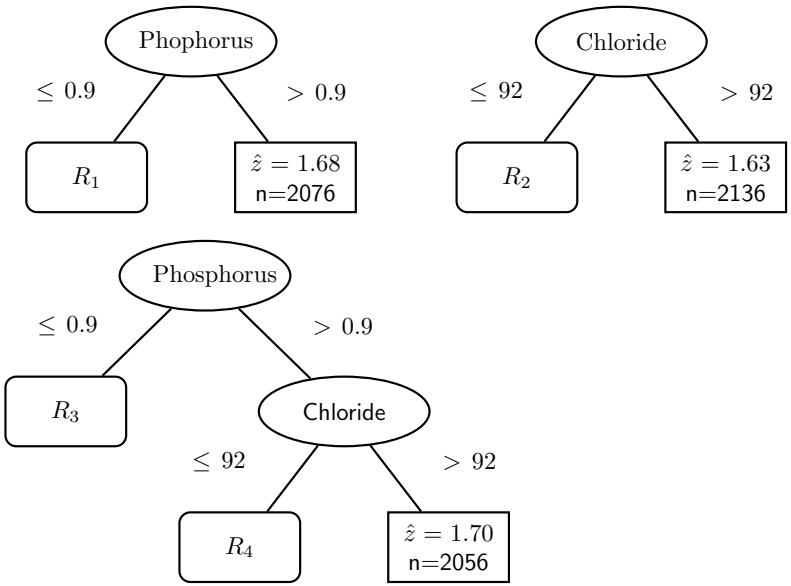
	IT	MOB	STIMA	SIDES	ARDP
Underlying model structure					
Regression model	yes	yes	yes	no	yes
Adjustment for covariate main effects	yes	no	yes	no	yes
- All covariates as main effects	no	no	yes	no	yes
- Dichotomized covariates main effects	yes	no	no	no	no

Overview of the methods' properties (4/4)

	IT	MOB	STIMA	SIDES	ARDP
Results					
Method results in a tree	yes	yes	yes	no	yes
End nodes are the identified subgroups	yes	yes	no	N/A	no
Additional steps needed for obtaining subgroups	no	no	yes	no	yes
Identified subgroups can be overlapping	no	no	N/A	yes	N/A

Statements not reasonably interpretable are marked with N/A.

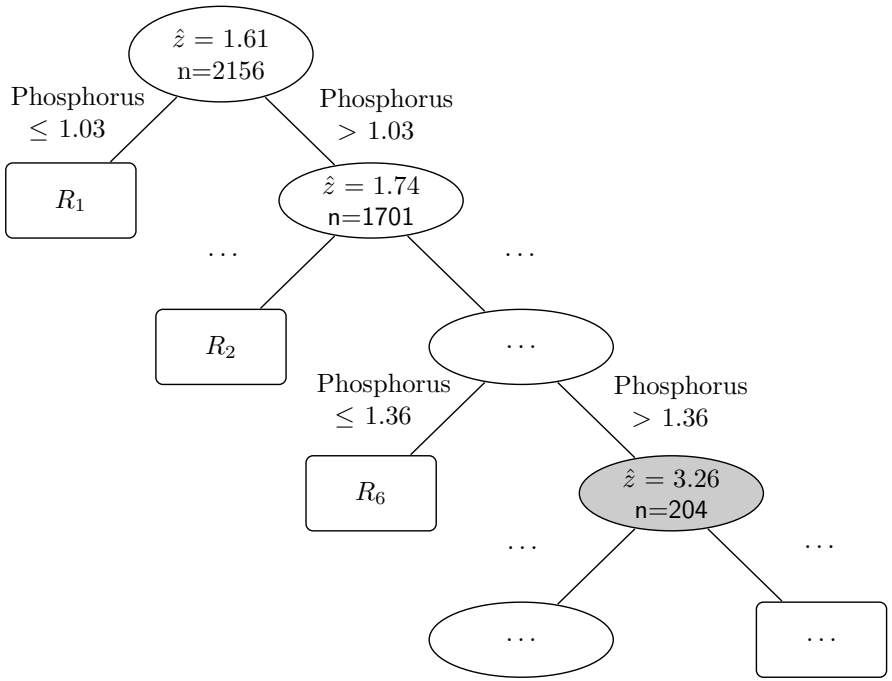
Subgroup Identification based on Differential Effect Search (SIDES)



SIDES applied to ALS data

- Construction of multiple trees
- Each X_j is used at most once within each of the trees
- Nodes with desirable efficacy (based on p-value of treatment effect) are added to a set of candidate subgroups

Adaptive Refinement by Directed Peeling(ARDP)



ARDP applied to ALS data

- **Remove** a prespecified number k of observations of the data in order to improve the treatment effect in the resulting subset
- Removal is based on linear model $E(Y|\mathbf{X}) = \alpha + \beta_0 T + \sum_{j=1}^{p} \beta_j X_j T + \gamma_j X_j$. E.g. For $\beta_j > 0$ remove k subjects with the smallest values of X_j
- ARDP results in sequence of nested subgroups.

Algorithm of SIDES

- **Splitting criterion**: Maximizing the differential effect between the two child groups

$$p_1 = 2\left[1 - \Phi\left(\frac{|Z_{\text{left}} - Z_{\text{right}}|}{\sqrt{2}}\right)\right]$$

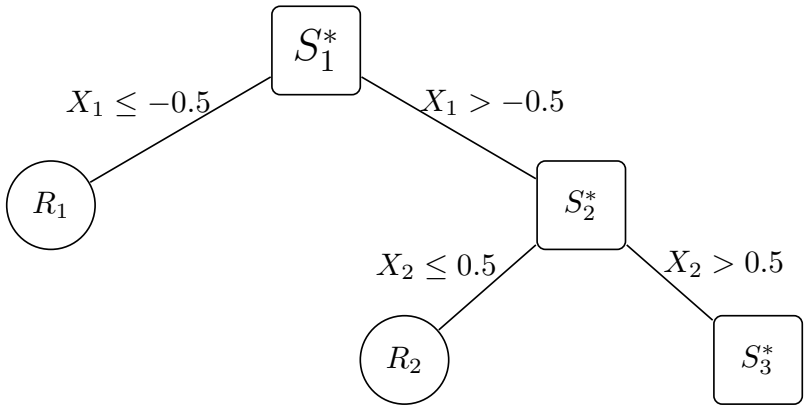
with Z_{left} and Z_{right} denoting the test statistics for a one-sided test of the hypothesis of no differential treatment effect in the left and right child subgroups

- Select the M best pairs of child nodes
- Retain just the child group exhibiting the larger treatment effect from each pair
- If child subgroup meets a continuation criterion further splitting is performed (not involving covariates already used for splitting)
- If selection criterion is met child subgroup is added to a set of candidate subgroups

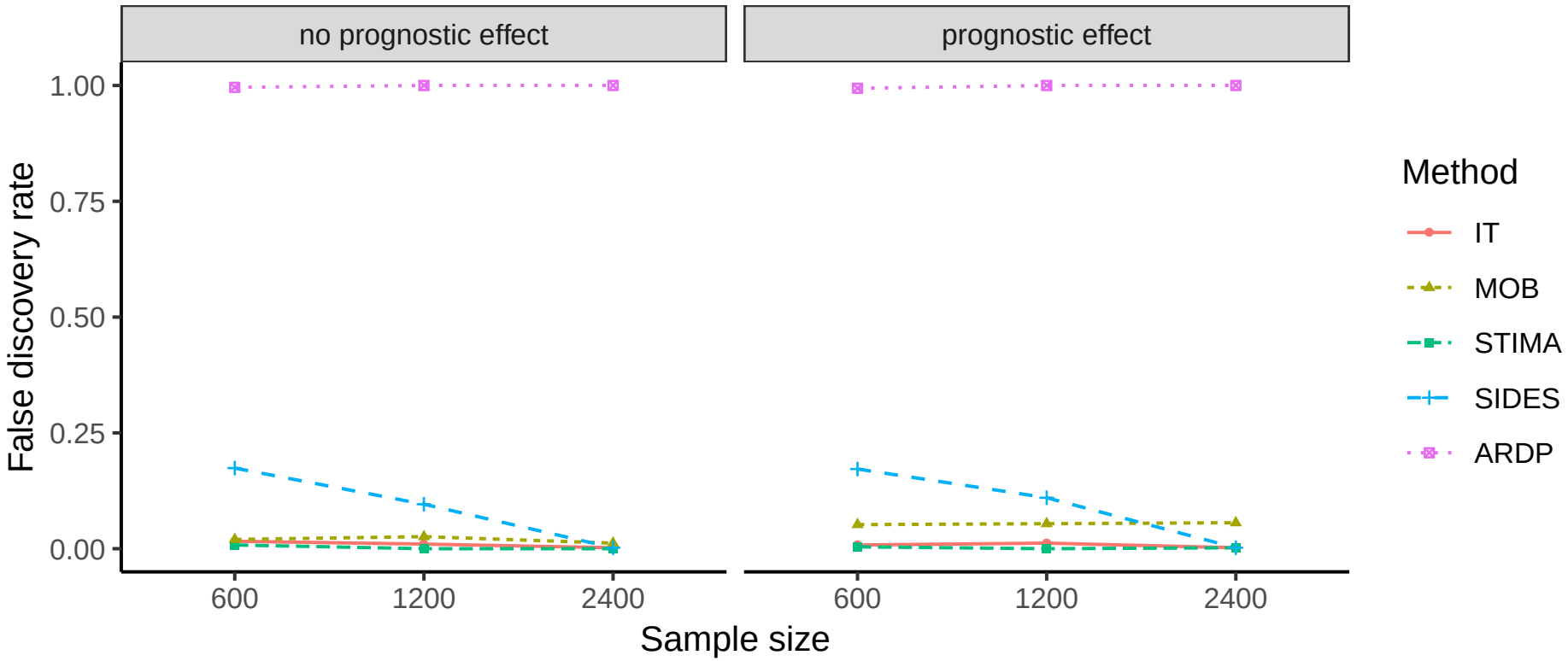
Algorithm of ARDP included in (Patel et al., 2016)

- Illustrated by tree although no recursive partitioning method
- Fit model to each node $E(Y|\mathbf{X}) = \alpha + \beta_0 T + \sum_{j=1}^{p} \beta_j X_i T + \gamma_j X_j$.
- **Remove** a prespecified number of observations for each $X_j, j = 1, \ldots, p$ in the **direction** defined by the β_j (Remove observations with small values of X_j if $\beta_j > 0$)
- Variable achieving the largest improvement of treatment effect compared to the parent node is selected for the split/peeling
- Keep just the child node with the larger treatment effect for further splitting

ARDP results in sequence of nested subgroups.

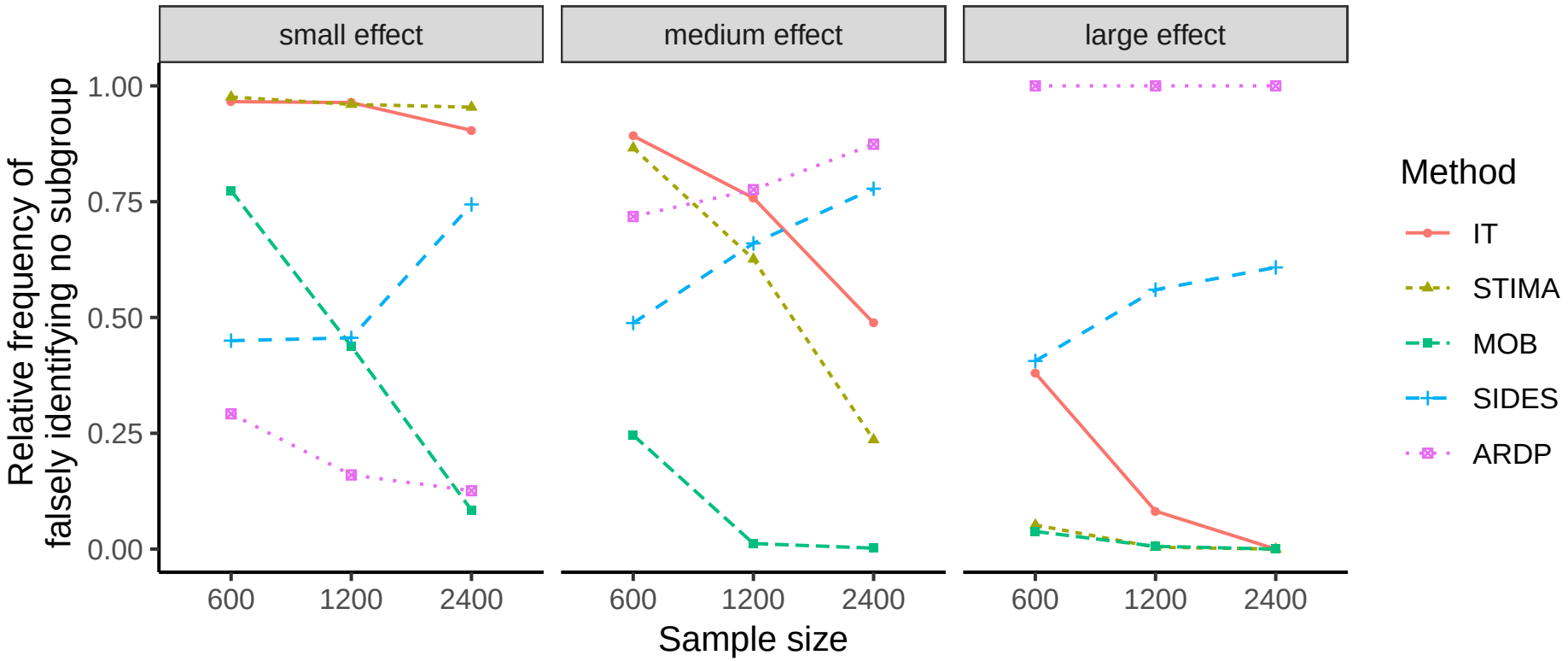


False discovery rate



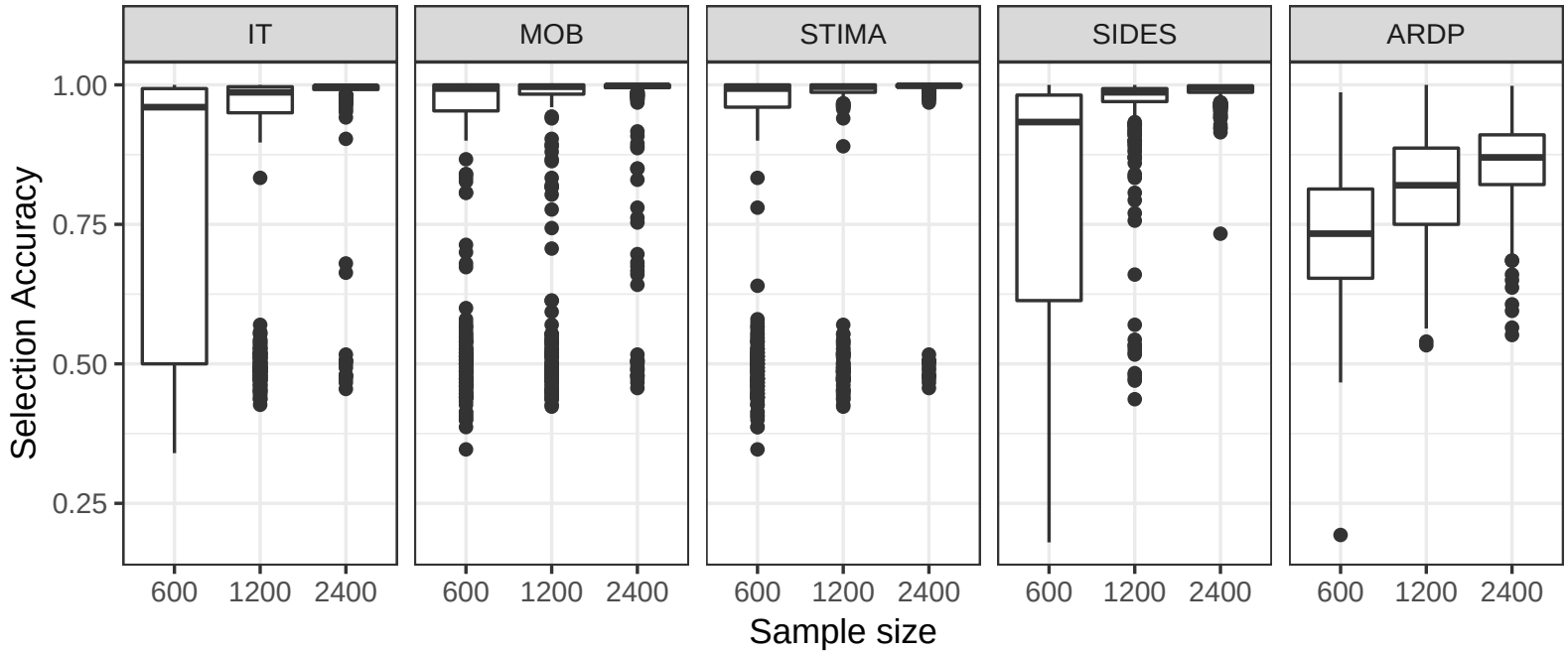
STIMA and IT have false discovery rates below 2% across all settings

Falsely identifying no subgroup although "real" subgroups are present



For data generated with $\mu(T, \mathbf{X}) = 0.2 \cdot T + \beta_1 \cdot T \cdot I(X_1 > 0)$ and $\beta_1 = 0.3$ referring to a small, $\beta_1 = 0.5$ to a medium and $\beta_1 = 1$ to a large effect.

Proportion of correctly classified patients
Qualitative interaction scenario



Data generation based on $\mu(T, \mathbf{X}) = \left\{ I(X_1 > 0) \cdot (1 - \frac{n_{01}}{n}) - I(X_1 \leq 0) \cdot \frac{n_{01}}{n} \right\} T$
with $n_{01} = \sum_{i=1}^{n} I(x_{i1} > 0)$