

Biomarker prioritization and power analysis

Ivan Krukov

December 15, 2020

Problem statement

- Treatment and placebo groups
- 10 patients each
- Baseline, then 5 days of follow-up
- 4000 biomarkers measured each day

Potential approaches

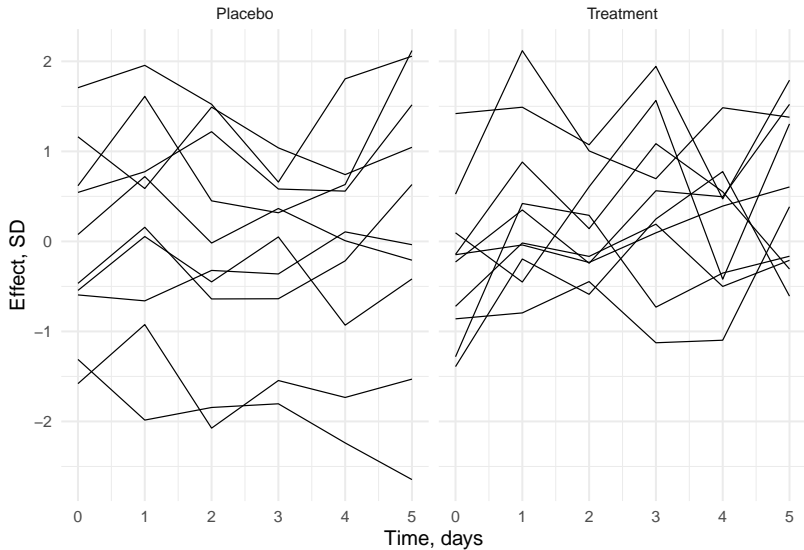
- Repeated measures ANOVA
 - Appropriate, but no effect indication
- Linear mixed model
 - Effect size inference, significance testing
- Bayesian linear model
 - Appropriate with limited data, but priors are hard to choose
- Machine learning
 - Performant, but hard to interpret

Analysis pipeline

Data pre-processing

- Log-transformation
 - Necessary if biomarker varies over degrees of magnitude
- Z-score standardization
 - 0-centered
 - Units of standard deviations

Simulated dataset



Mixed-effects model

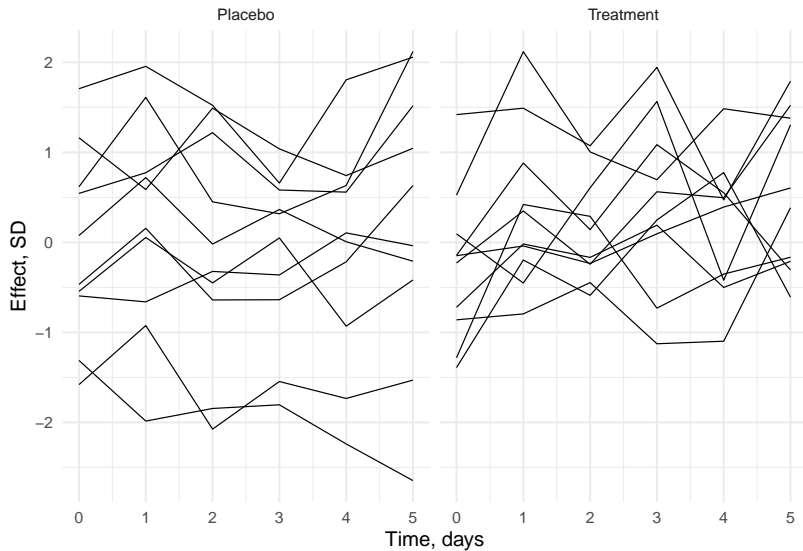
- Fixed effects
 - On average, baseline is zero (standardized)
 - Biomarker activity increases over time (only with treatment)
- Random effects
 - Baseline is different for each patient
 - Individual response to drug is different

Inference

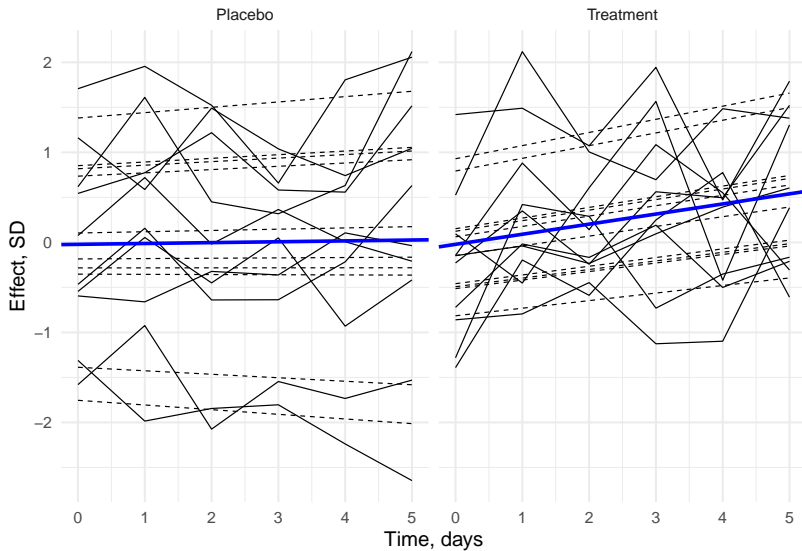
Table 1: Inferred effects

Fixed effect	Estimate	Std. Error	df	t value	p value
(Intercept)	-0.022	0.201	19.284	-0.108	0.915
t:groupPlacebo	0.010	0.040	57.924	0.243	0.808
t:groupTreatment	0.112	0.040	57.924	2.839	0.006

Inference results



Inference results



Multiple testing correction

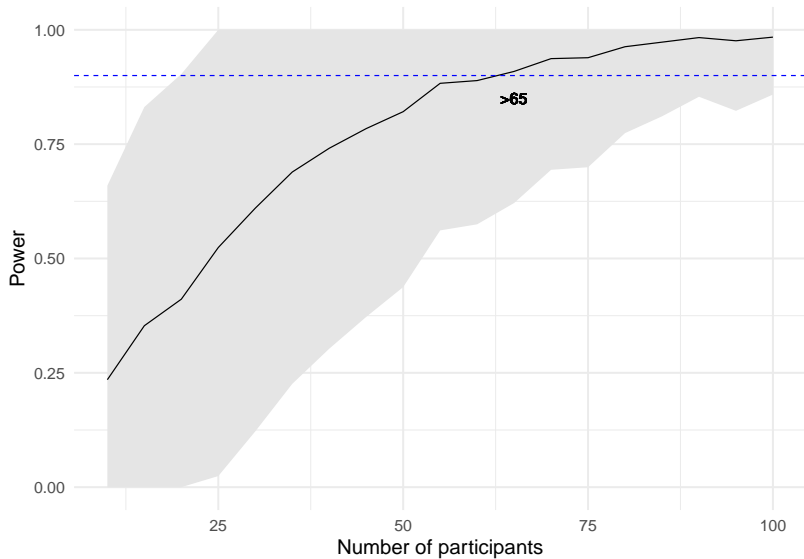
- Benjamini-Hochberg False discovery rate
 - Controls for the false positives in the entire experiment
 - Less stringent than Bonferonni correction

Improving the design

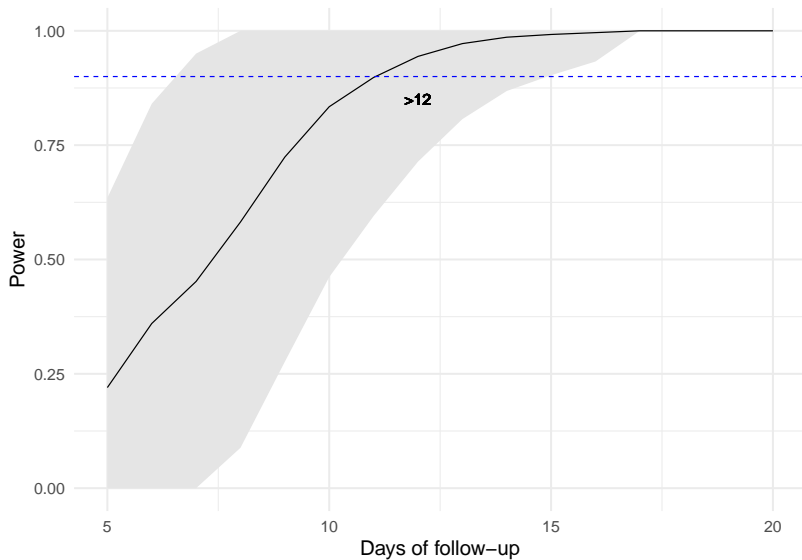
Power of the test

- More data!
- How many patients do we need per group?
- How many days of follow-up do we need?

Participant number



Days of follow-up



Further considerations

- Is the sample truly representative?
- Does the effect increase indefinitely with time?
- Smaller placebo group - 20 participant may be sufficient
- Effect of the drug concentration

Code available at:

github.com/ivan-krukov/biomarker-power

Questions?

Assumptions about effect

- With treatment, biomarker level changes 0.05 SD per day
- 2% improvement per day
- 93rd percentile after 30 days of treatment

```
param <- list(  
  eps_sd = .5, base = 0,  
  trt = list(eff = .05, base_sd = 1, eff_sd = .025),  
  plc = list(eff = 0, base_sd = 1, eff_sd = 0))
```