Looking forward and benefiting from the past: Sample size estimation for new arms in platform trials

Sonja Zehetmayer, Franz König

Medical University of Vienna

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853966.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and CHILDREN'S TUMOR FOUNDATION, GLOBAL ALLIANCE FOR TB DRUG DEVELOPMENT NON PROFIT ORGANISATION, SPRINGWORKS THERAPEUTICS INC.

Benefiting from the past?

New treatment: How can data already collected be utilized to “optimize” the trial?

- **For test decision?**
  - Use of nonconcurrent controls (*Bofill et al., 22*)
  - Modification of testing strategy, e.g. outcome of former treatments influences significance boundaries, e.g., online control of FWER, FDR (*Javanmard et al. 15, Tian et al. 21*)

Benefiting from the past?

New treatment: How can data already collected be utilized to “optimize” the trial?

- **For test decision?**
  - Use of nonconcurrent controls (*Bofill et al., 22*)
  - Modification of testing strategy, e.g. outcome of former treatments influences significance boundaries, e.g., online control of FWER, FDR (*Javanmard et al. 15, Tian et al. 21*)

- **For sample size calculation?**
  - Effect size estimation
  - Variance estimation
  - Significance level

Benefiting from the past?

New treatment: How can data already collected be utilized to “optimize” the trial?

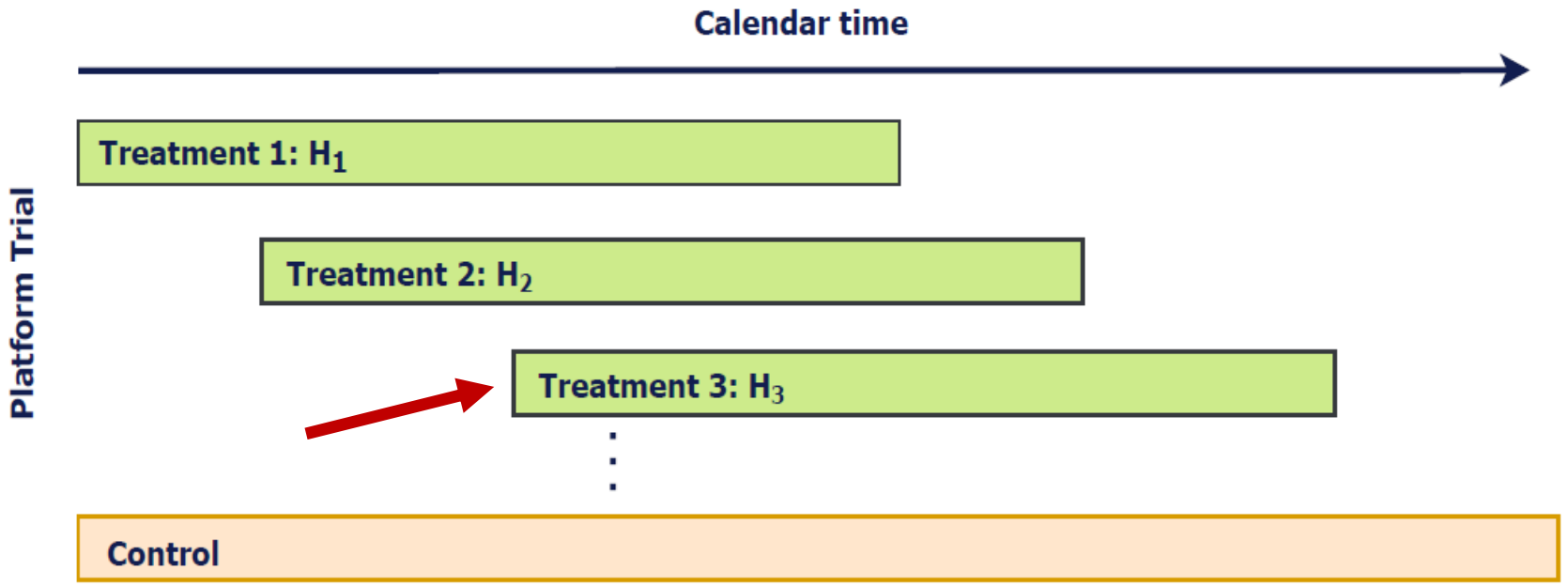
- **For test decision?**
  - Use of nonconcurrent controls (*Bofill et al., 22*)
  - Modification of testing strategy, e.g. outcome of former treatments influences significance boundaries, e.g., online control of FWER, FDR (*Javanmard et al. 15, Tian et al. 21*)

- **For sample size calculation?**
  - Effect size estimation
  - Variance estimation
  - Significance level

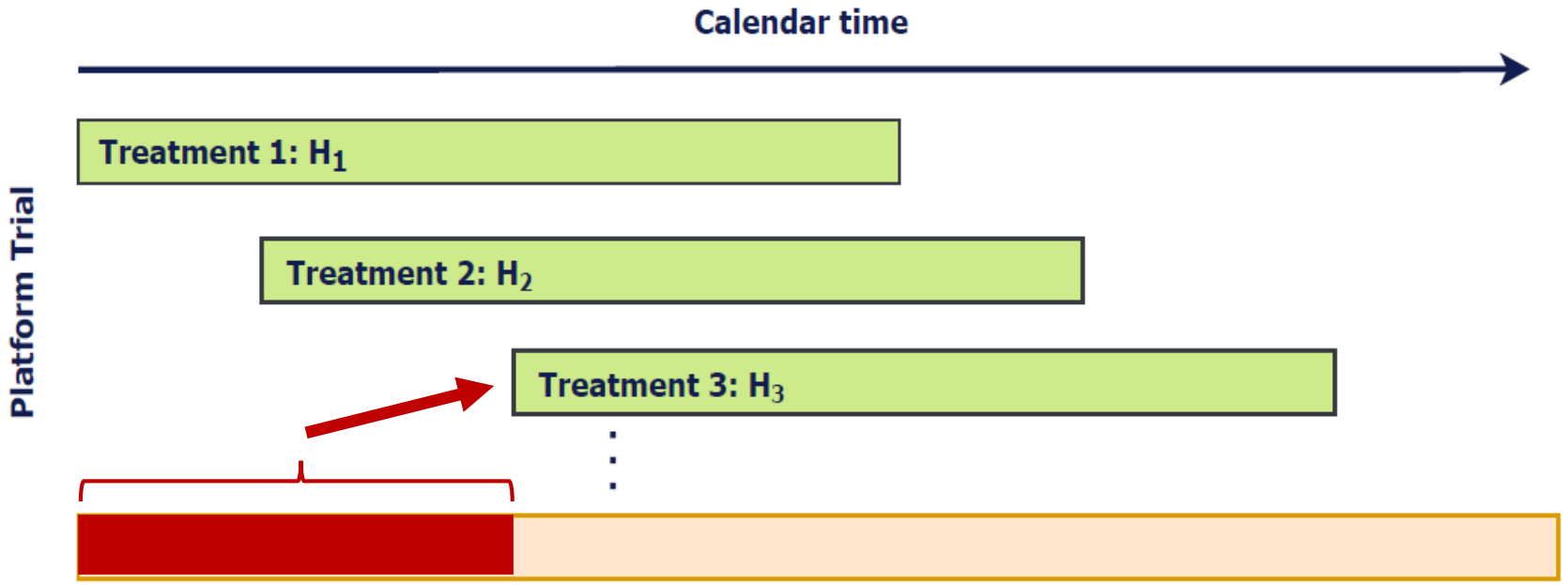
How is this different to historical data?

- Same definition of endpoints
- Same centers and study population
- Same inclusion and exclusion criteria
- ....

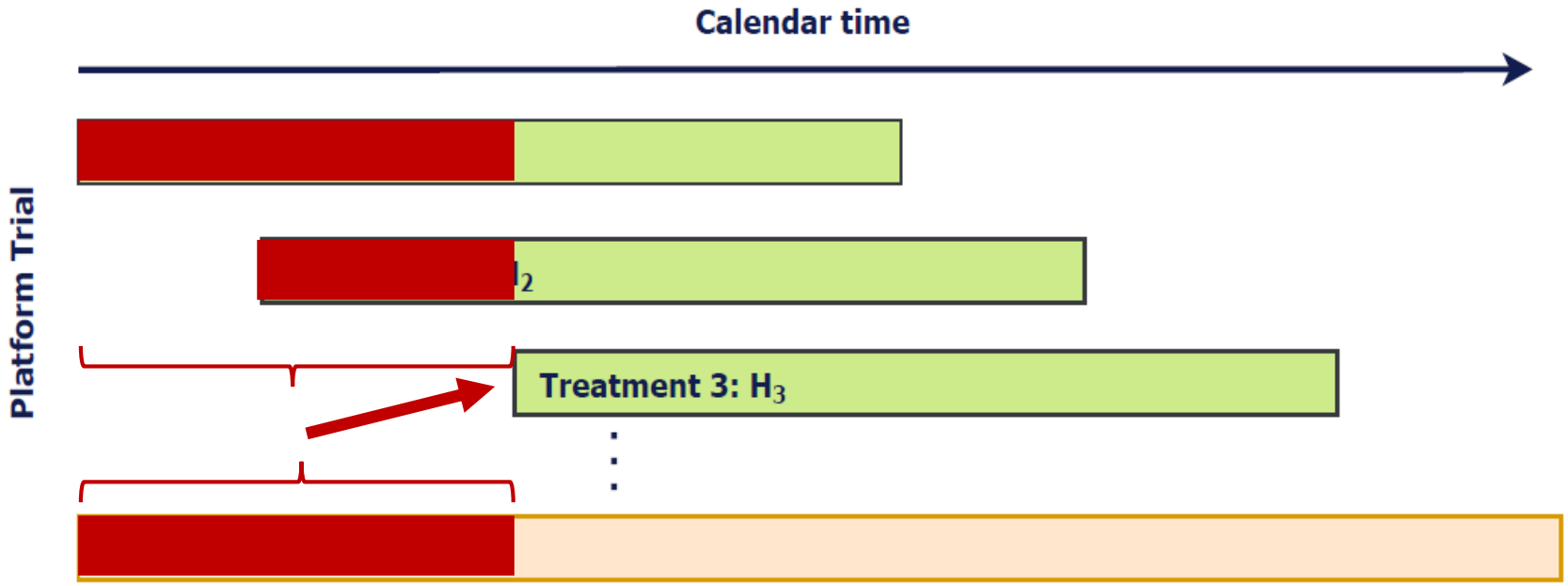
What about sample size calculation for new treatment arms in platform trials?



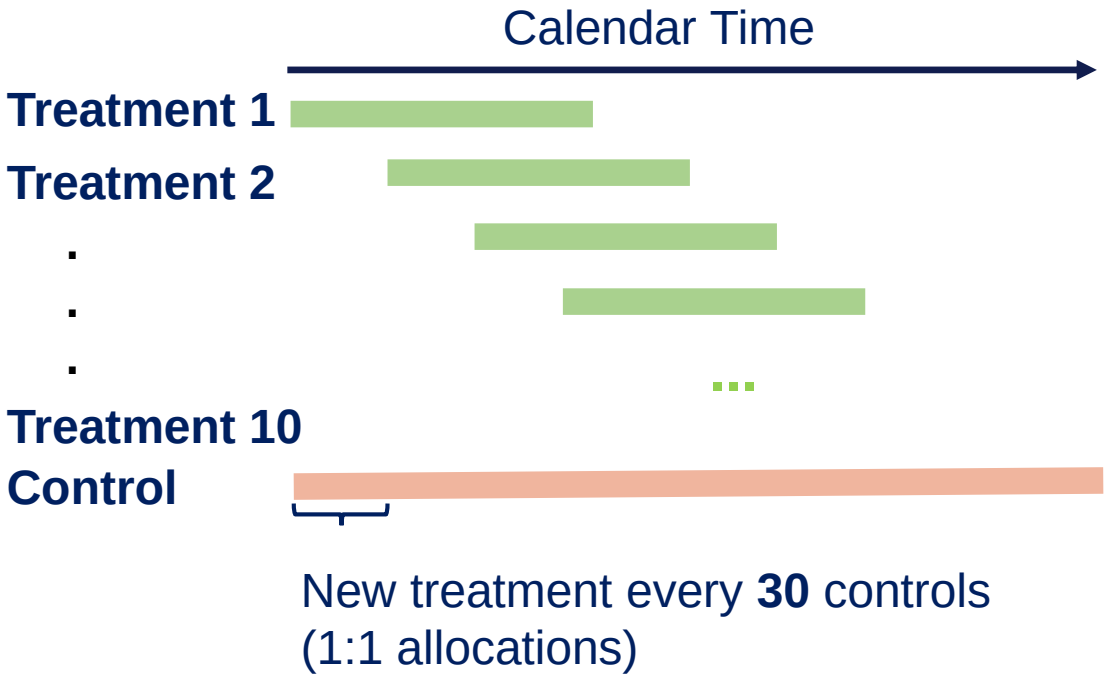
What about sample size calculation for new treatment arms in platform trials?



What about sample size calculation for new treatment arms in platform trials?

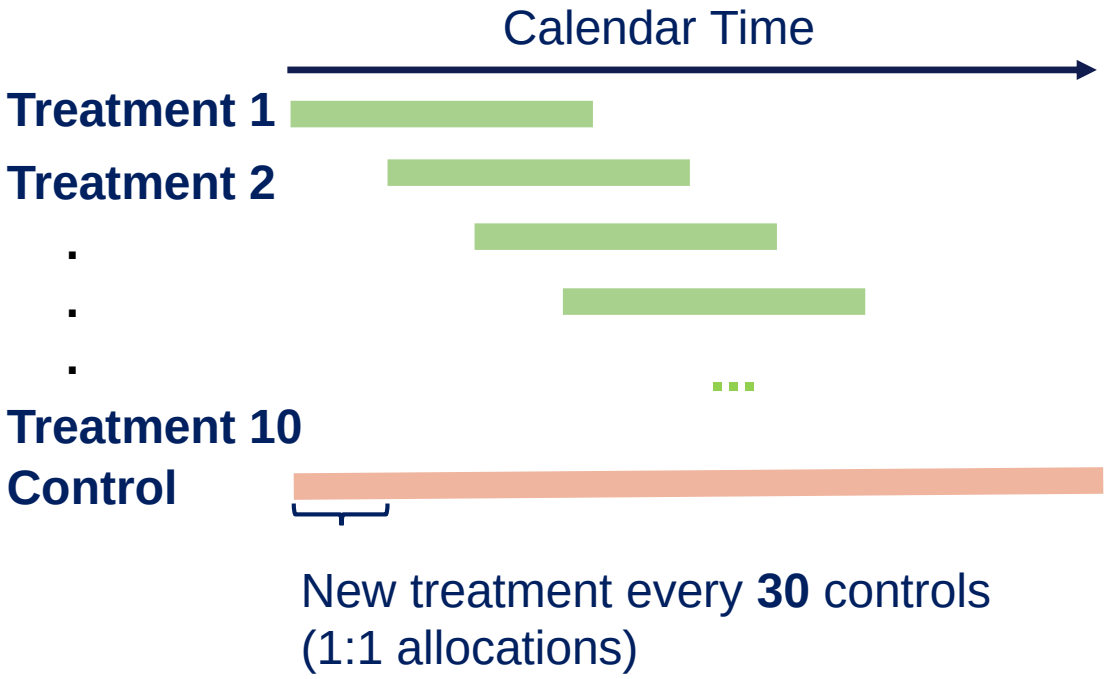


Sample size calculation with an oracle



- N(0,1) data, for alternatives N(Delta=0.5,1)
- T-test
- Concurrent controls for test decision
- α=0.025 (one-sided)
- Level-α tests (e.g., *Collignon et al., 20*)
- Target average power of 0.8

Sample size calculation with an oracle



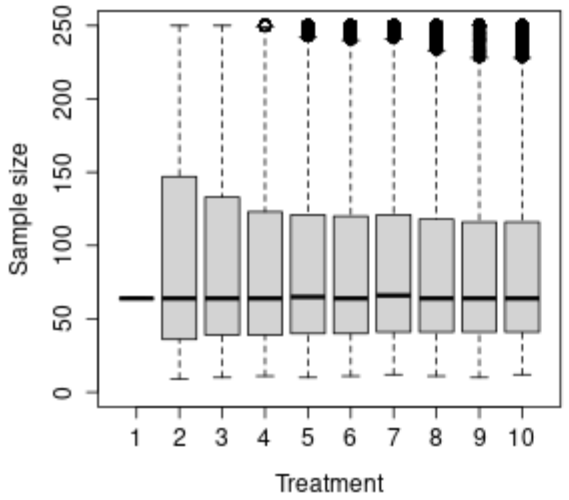
- N(0,1) data, for alternatives N(Delta=0.5,1)
- T-test
- Concurrent controls for test decision
- α=0.025 (one-sided)
- Level-α tests (e.g., *Collignon et al., 20*)
- Target average power of 0.8

If an oracle tells us the true parameters: Sample size of each treatment arm: **n=64**

Naive strategy: Estimate Delta from the observed effect of last treatment and all controls

- Minimum Delta=0.25 (i.e., maximum sample size of 250). Set sd=1
- Simulated sample sizes for treatments 1 to 10:

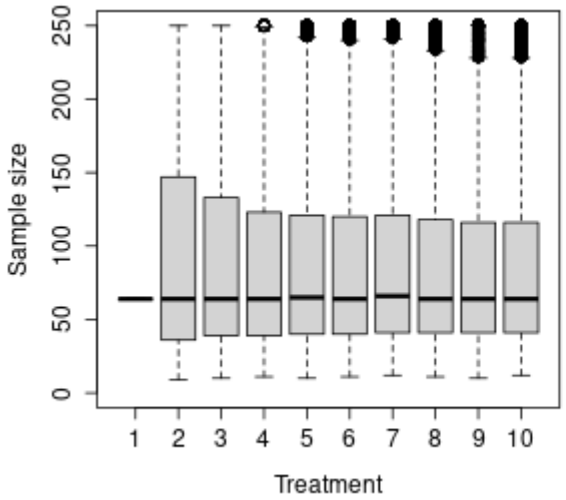
m_1=10 effective treatments



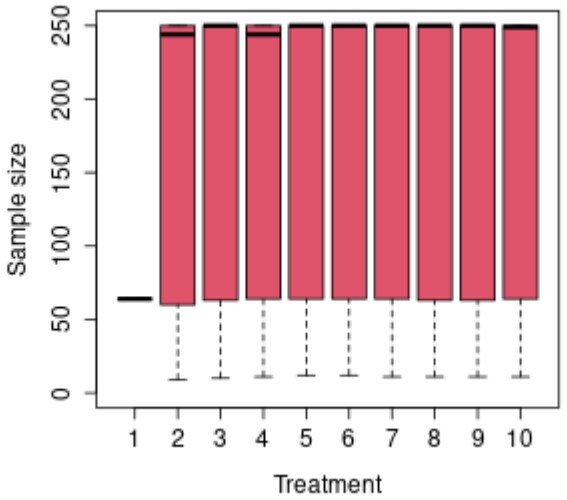
Naive strategy: Estimate Delta from the observed effect of last treatment and all controls

- Minimum Delta=0.25 (i.e., maximum sample size of 250). Set sd=1
- Simulated sample sizes for treatments 1 to 10:

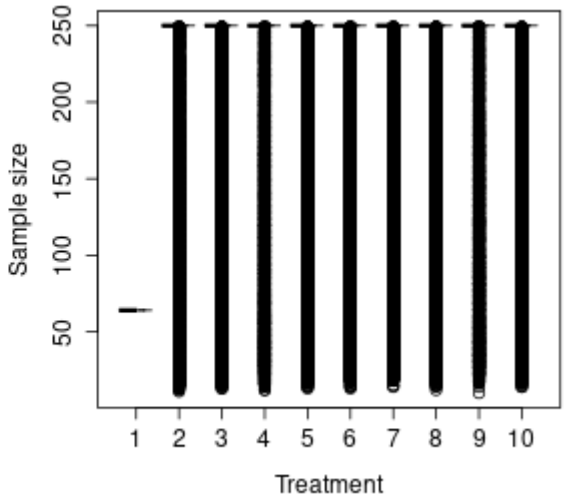
m_1=10



m_1=5



m_1=1



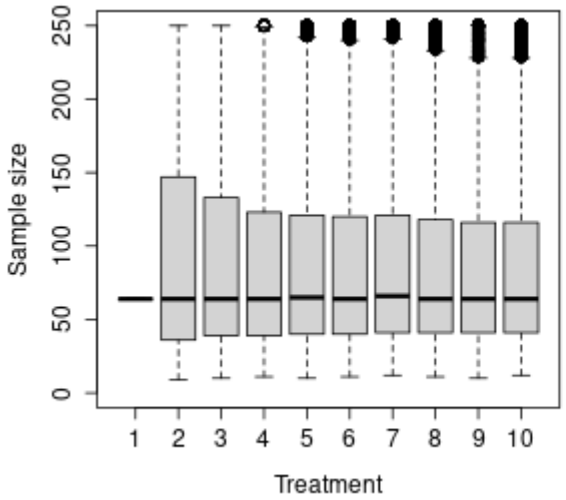
Naive strategy: Estimate Delta from the observed effect of last treatment and all controls

- Minimum Delta=0.25. Set sd=1
- Simulated sample sizes for treatments 1 to 10:

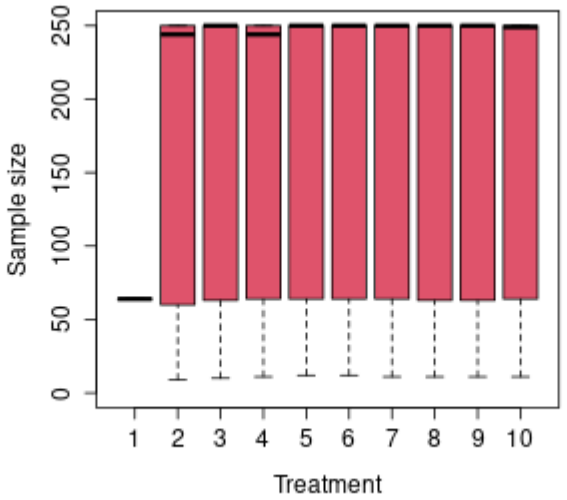
Why should the effect size of interest depend on the (random) outcome of the former treatment?

Set Delta to clinical relevant effect size = 0.5.

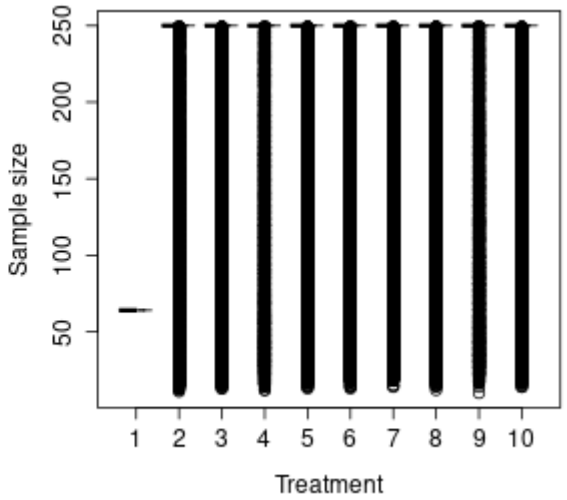
m_1=10



m_1=5

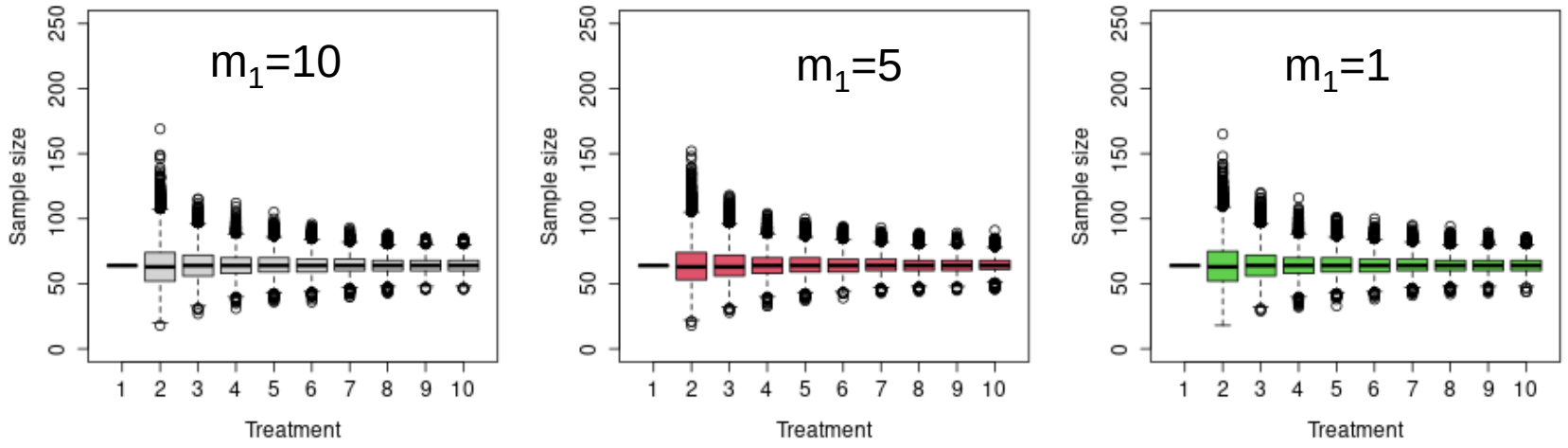


m_1=1



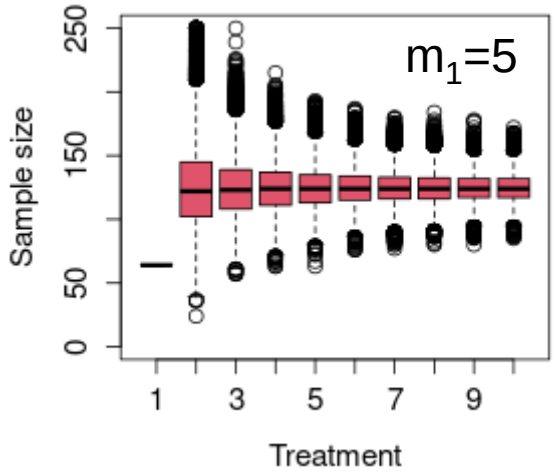
Estimate standard deviation

- Set Delta=0.5 (clinical relevant effect size), true sd=1
- Estimate standard deviation from all so far observed controls
  - Range of sample sizes decreases with time
  - The more data we observe, the more precise is the sample size calculation



False assumption - estimate standard deviation from observed controls

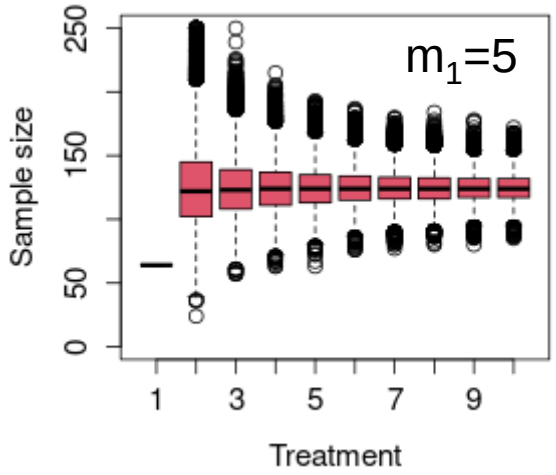
- Assume sd=1 but true sd=1.4



Average power for the scenarios = 0.77

False assumption - estimate standard deviation from observed controls

- Assume sd=1 but true sd=1.4

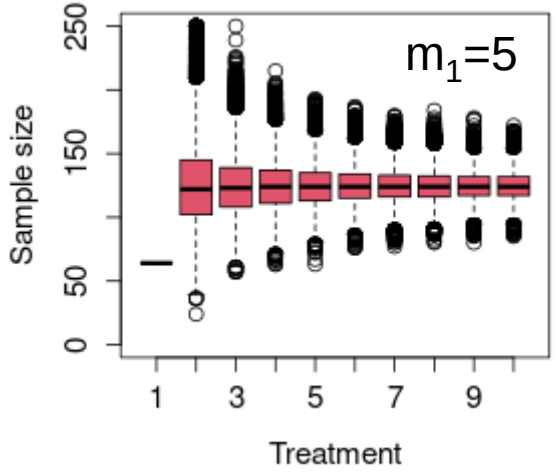


Average power for the scenarios = 0.77

What happens if we start a new treatment after every 5 controls (instead of 30)?

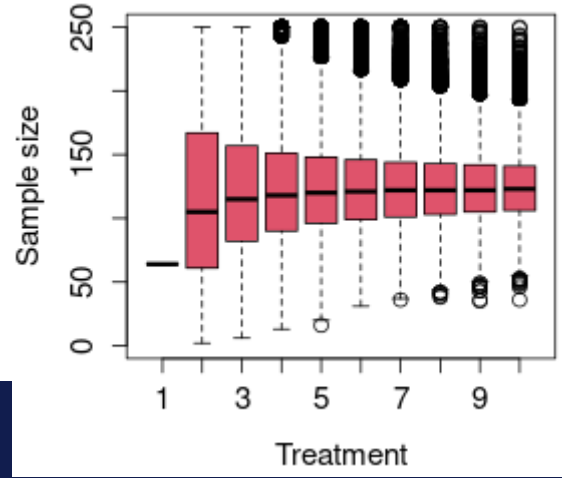
False assumption - estimate standard deviation from observed controls

- Assume sd=1 but true sd=1.4



Average power for the scenarios = 0.77

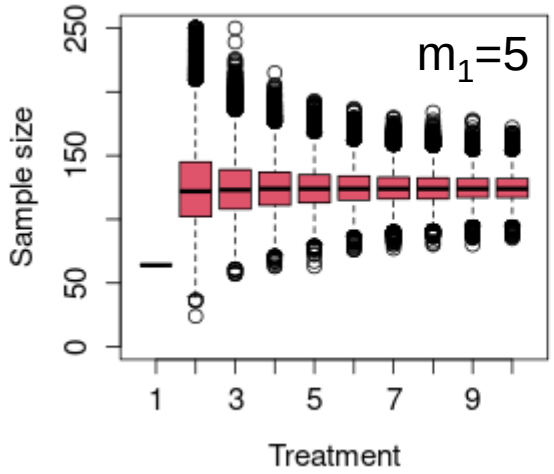
What happens if we start a new treatment after every 5 controls (instead of 30)?

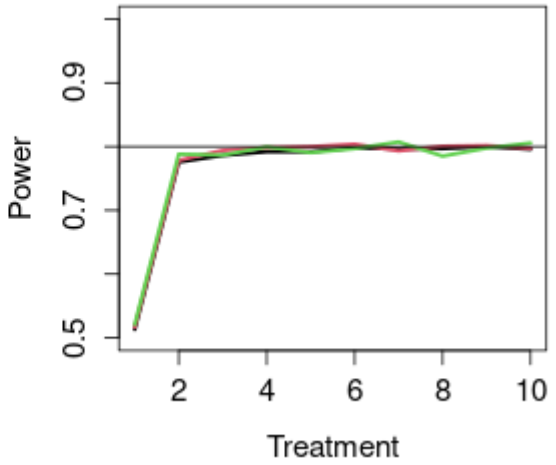


Average power for the scenarios = 0.74

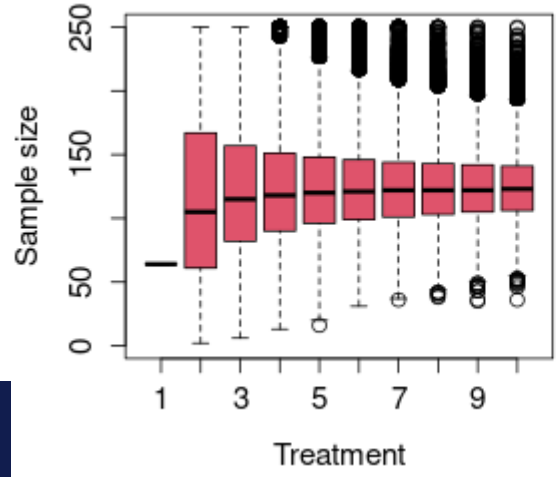
False assumption - estimate standard deviation from observed controls

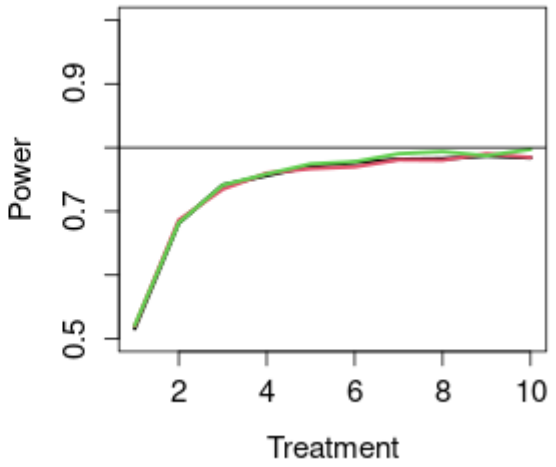
- Assume sd=1 but true sd=1.4





What happens if we start a new treatment after every 5 controls (instead of 30)?



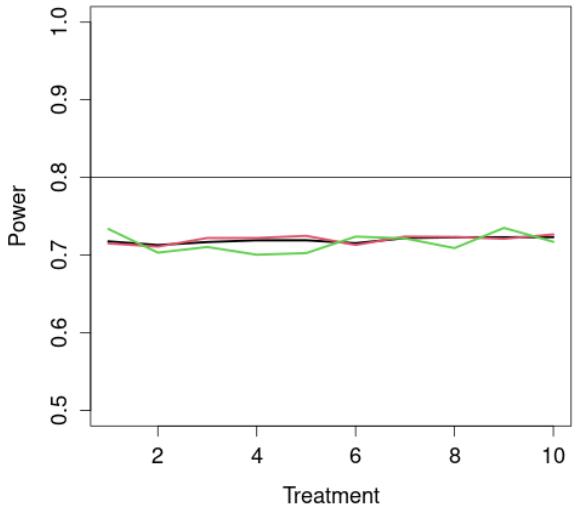


Heteroscedasticity: Treatment arms with higher variation

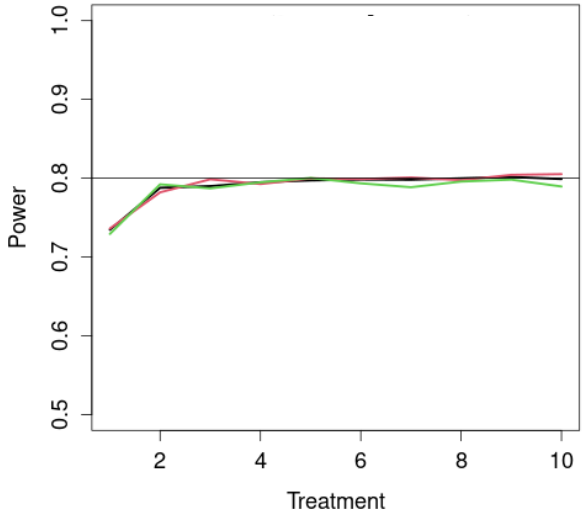
- Controls: sd=1, Treatments: sd=1.2

Estimate sd from data

From controls



From controls and treatments

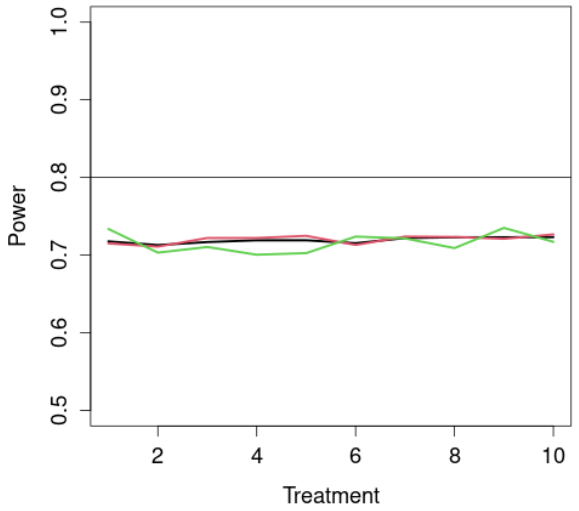


Heteroscedasticity: Treatment arms with higher variation

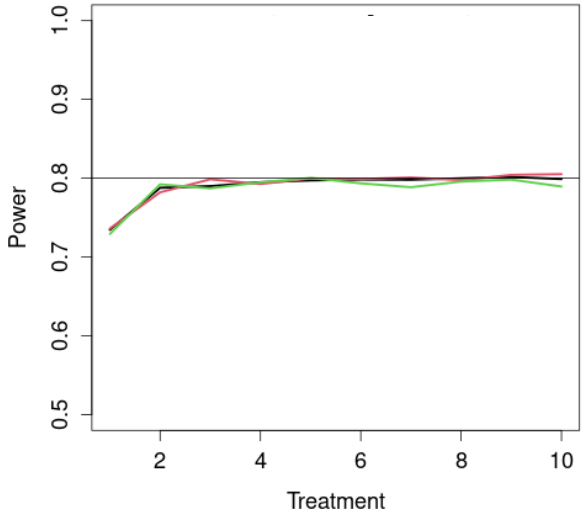
- Controls: sd=1, Treatments: sd=1.2

Estimate sd from data

From controls

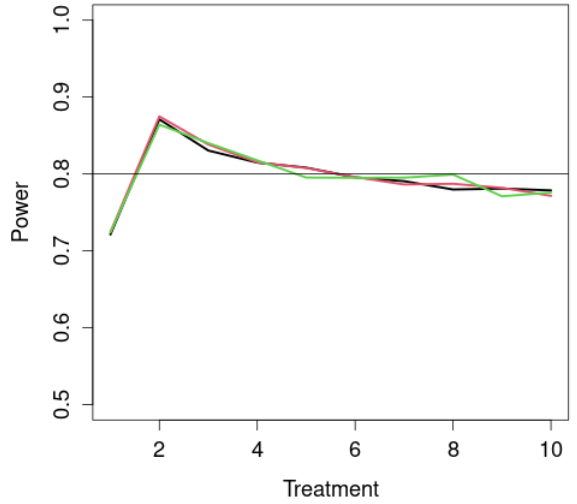


From controls and treatments



Estimate sd with CI

From controls



Wassmer and Kieser, 96: Use upper CI of sd from pilot sample

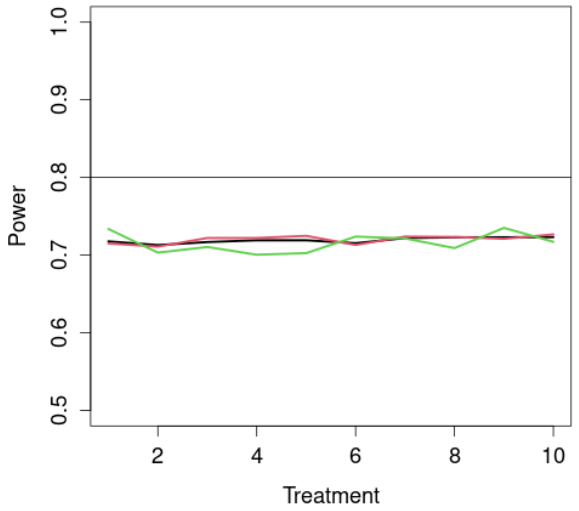
For early treatments, broad CI due to smaller number of previous observations.

Heteroscedasticity: Treatment arms with higher variation

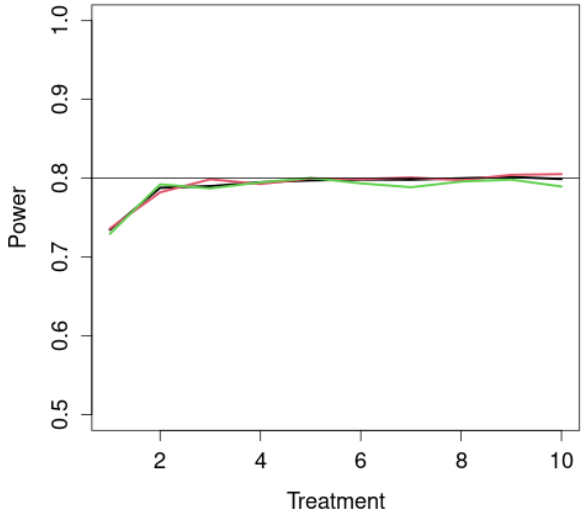
- Controls: sd=1, Treatments: sd=1.2

Estimate sd from data

From controls

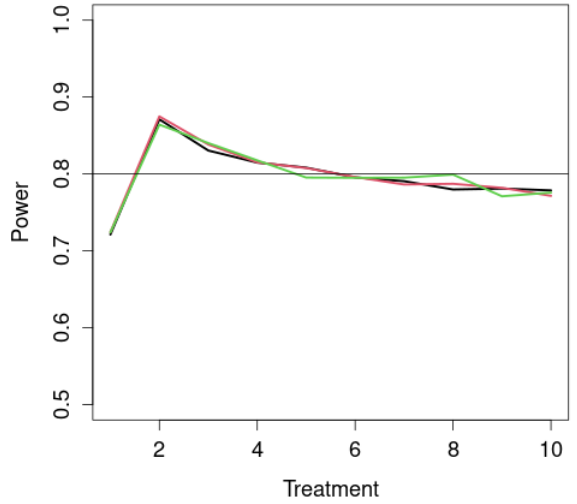


From controls and treatments

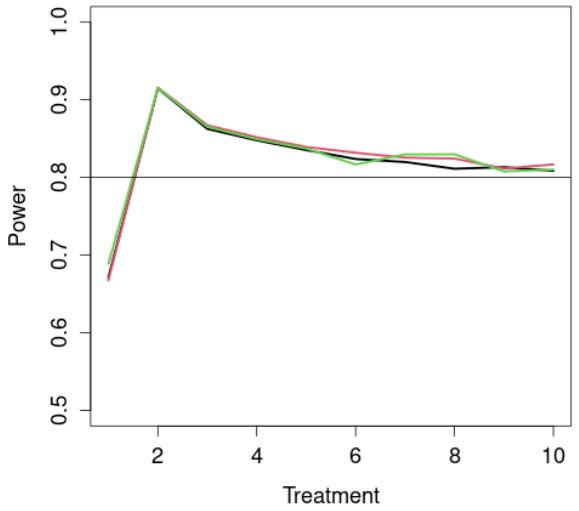


Estimate sd with CI

From controls



From controls and treatments



Wassmer and Kieser, 96: Use upper CI of sd from pilot sample

For early treatments, broad CI due to smaller number of previous observations.

False Discovery Rate (FDR) control

- FDR control for platform trials (e.g, *Robertson et al 2019, 2022, Zehetmayer et al. 2022*)

$$\mathrm{FDR}=\mathrm{E}\left(\frac{V}{\max(R,1)}\right)$$

  - R: number of rejected hypotheses, V: number of wrongly rejected hypotheses

- However, conventional methods for FDR control assume that
  - Number of hypothesis tests is fixed
  - All p-values are available at time of test decison.

- **Online control of the FDR**

Online control of the FDR

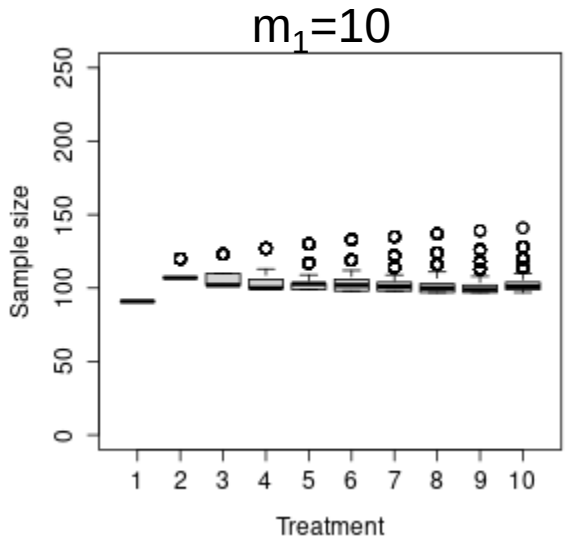
- Predefined order of hypotheses
- At each step a decision is made if the current H_i is rejected based on previous decisions.
- E.g., *Javanmard and Montanari,15, 18*

LOND procedure for online FDR control

- Decreasing significance levels α_1, α_2,... are allocated to sequence of null hypotheses H_1, H_2,...
- In case of previous rejections, the significance levels are increased.

Oracle scenario with online FDR control

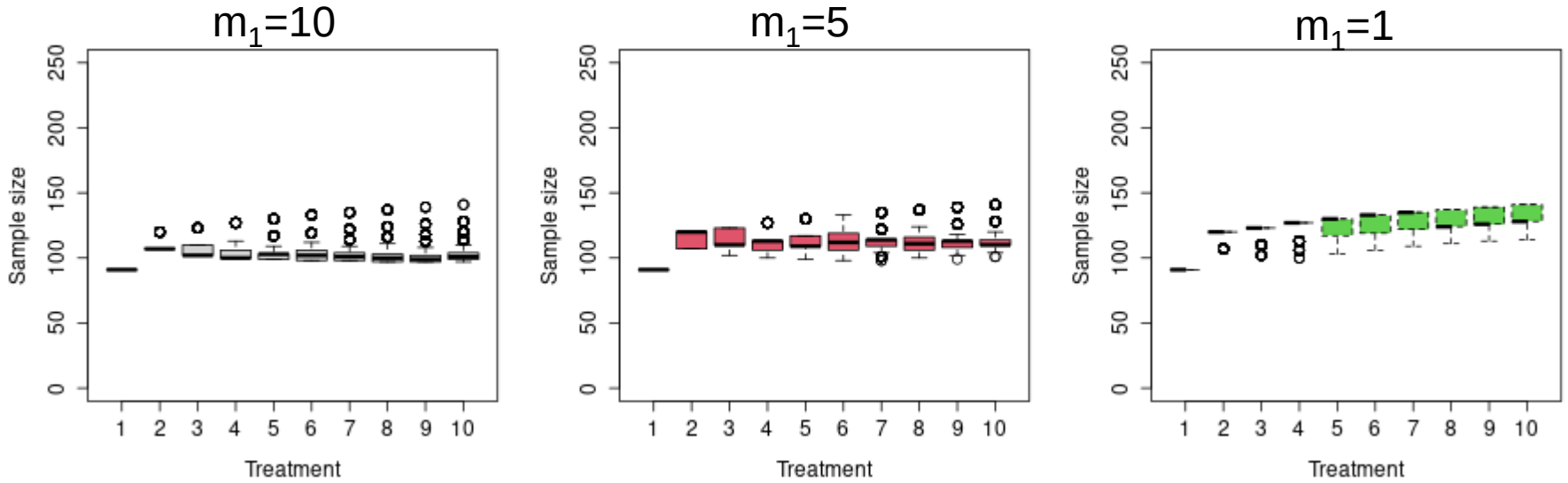
- True sd=1
- Allow for a total of **100** treatment arms



- Remember: For level-α tests, for each treatment **n=64** patients

Oracle scenario with online FDR control

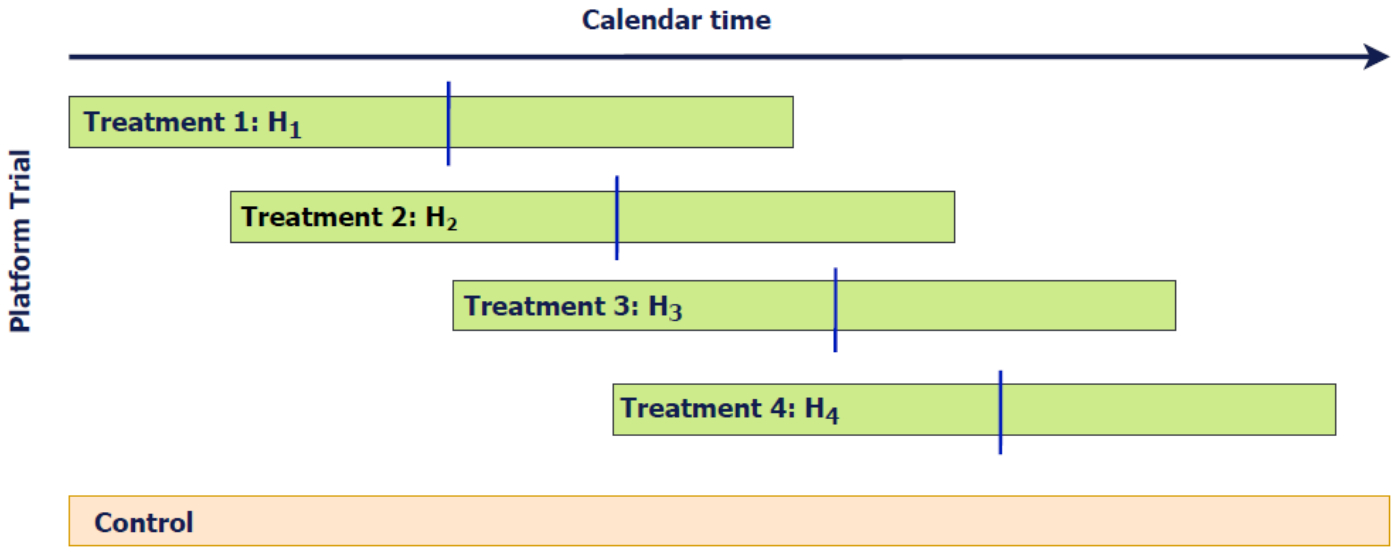
- True sd=1
- Allow for a total of **100** treatment arms



- Remember: For level-α tests, for each treatment n=64 patients
- Sample sizes depend on m_1 and the position of the treatment.

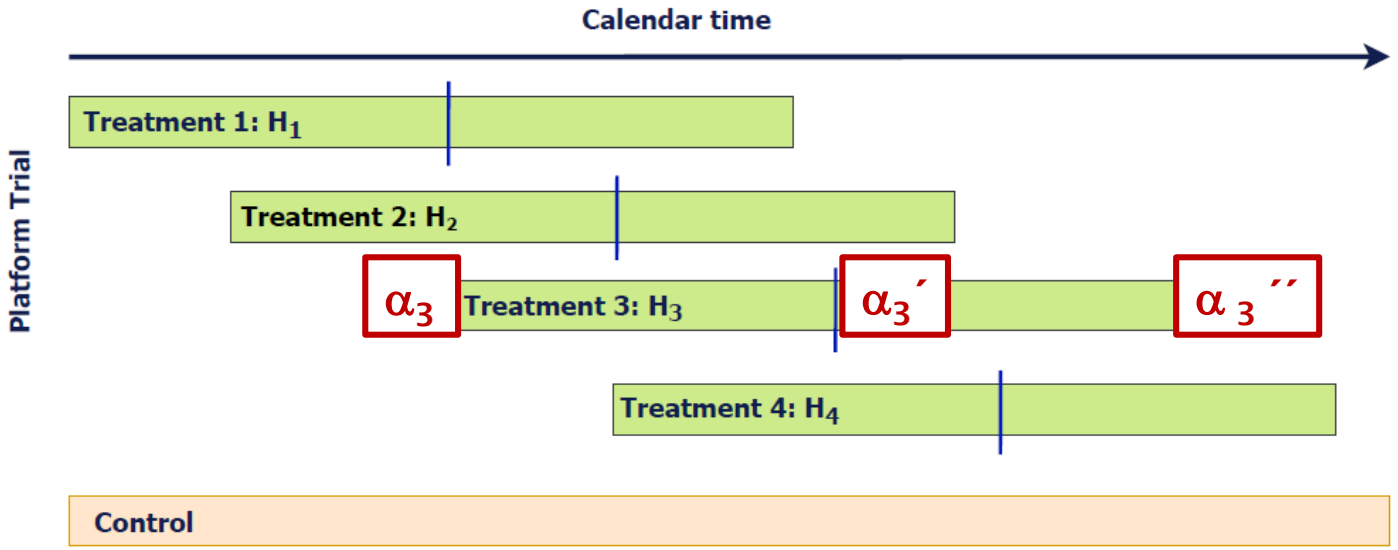
Adaptive design

- In *Zehetmayer et al., 22* we considered a group-sequential LOND procedure (gsLOND) with the option to stop in the interim analysis.
- For gsLOND the significance level α_i might be increased between interim and final analysis.
- Consequently also group-sequential boundaries have to adjusted.



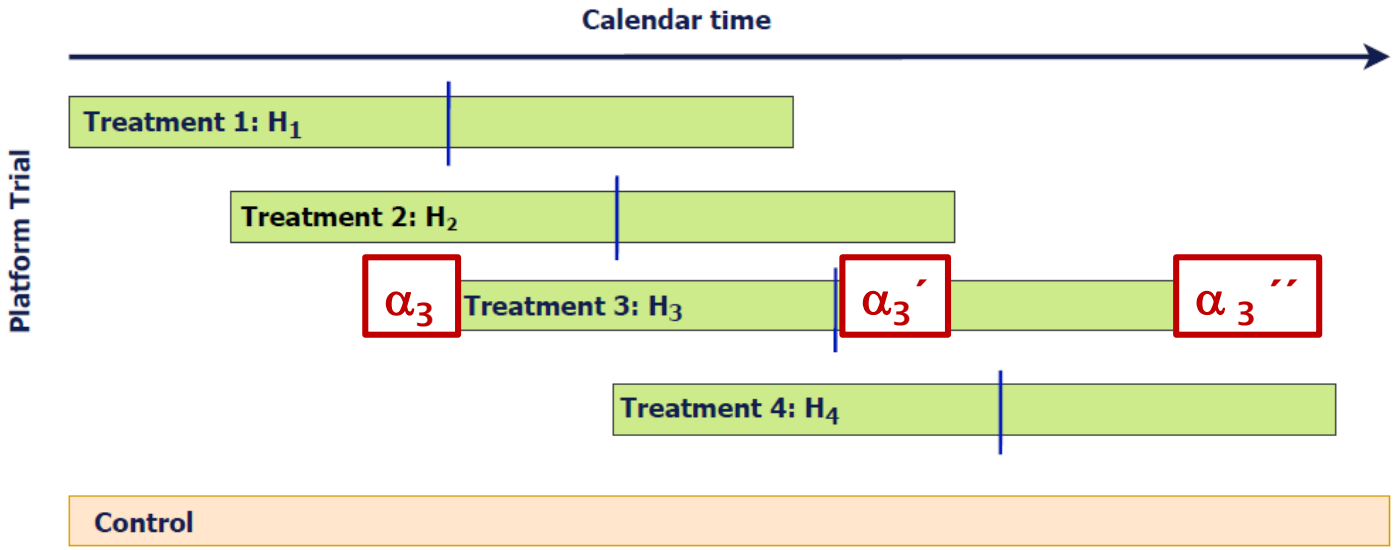
Adaptive design

- In *Zehetmayer et al., 22* we considered a group-sequential LOND procedure (gsLOND) with the option to stop in the interim analysis.
- For gsLOND the significance level α_i might be increased between interim and final analysis.
- Consequently also group-sequential boundaries have to adjusted.



Adaptive design

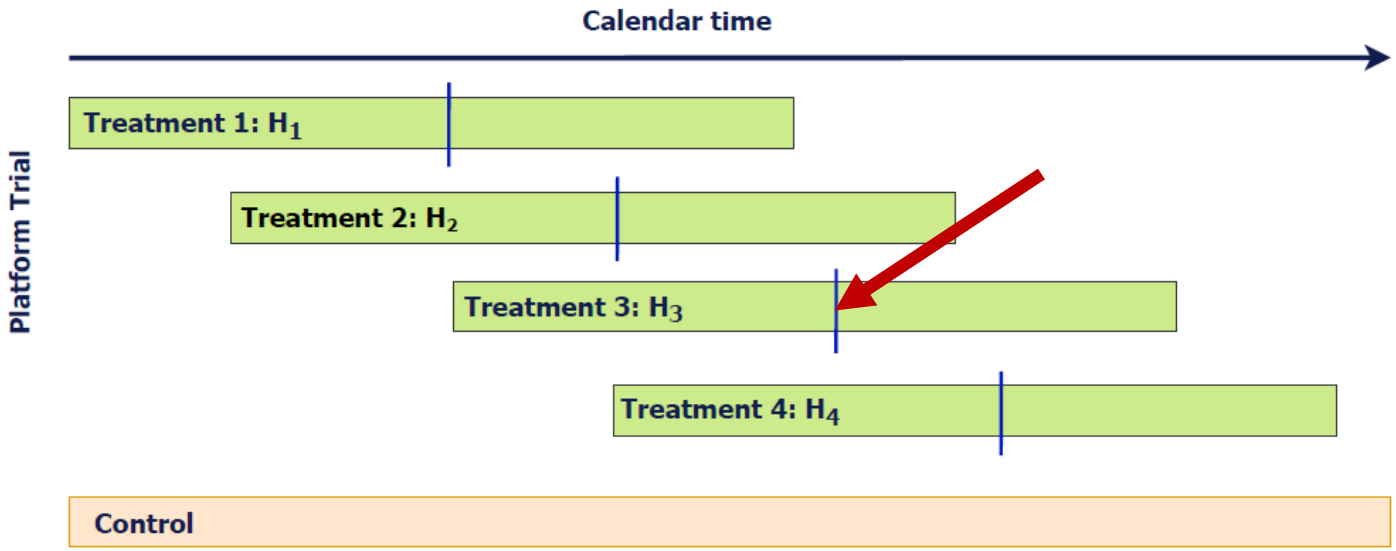
- In *Zehetmayer et al., 22* we considered a group-sequential LOND procedure (gsLOND) with the option to stop in the interim analysis.
- For gsLOND the significance level α_i might be increased between interim and final analysis.
- Consequently also group-sequential boundaries have to adjusted.



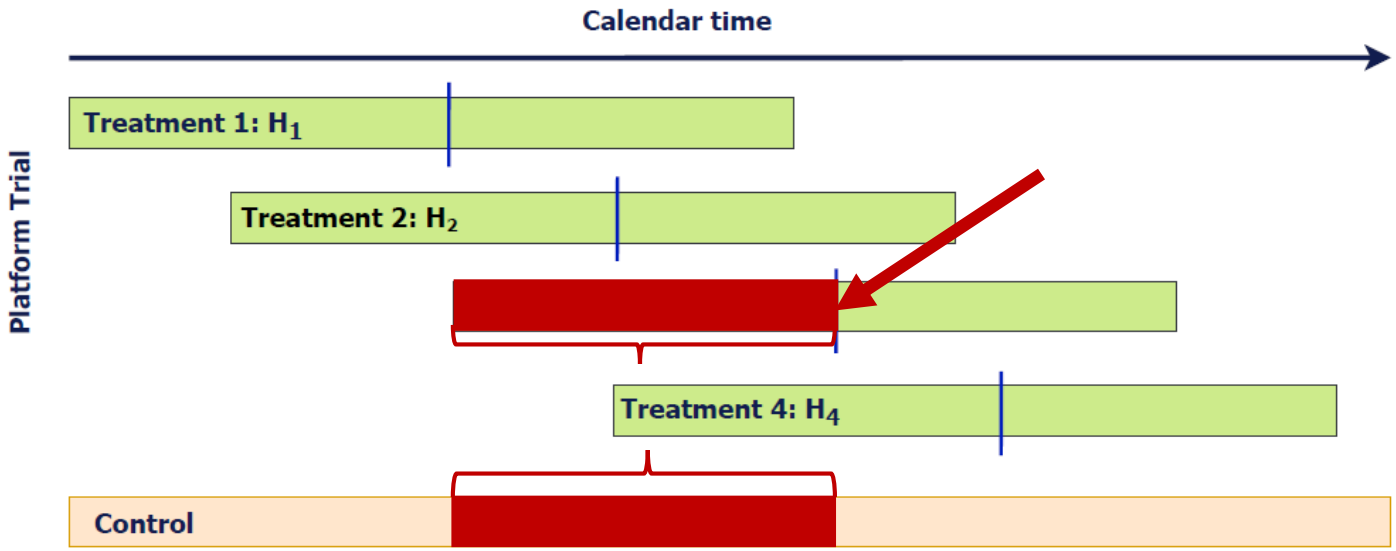
NOW: In interim analysis:

- Stopping for futilty
- Efficacy stopping
- Update α_i
- Sample size reassessment with conditional error function for a conditional power of 0.8

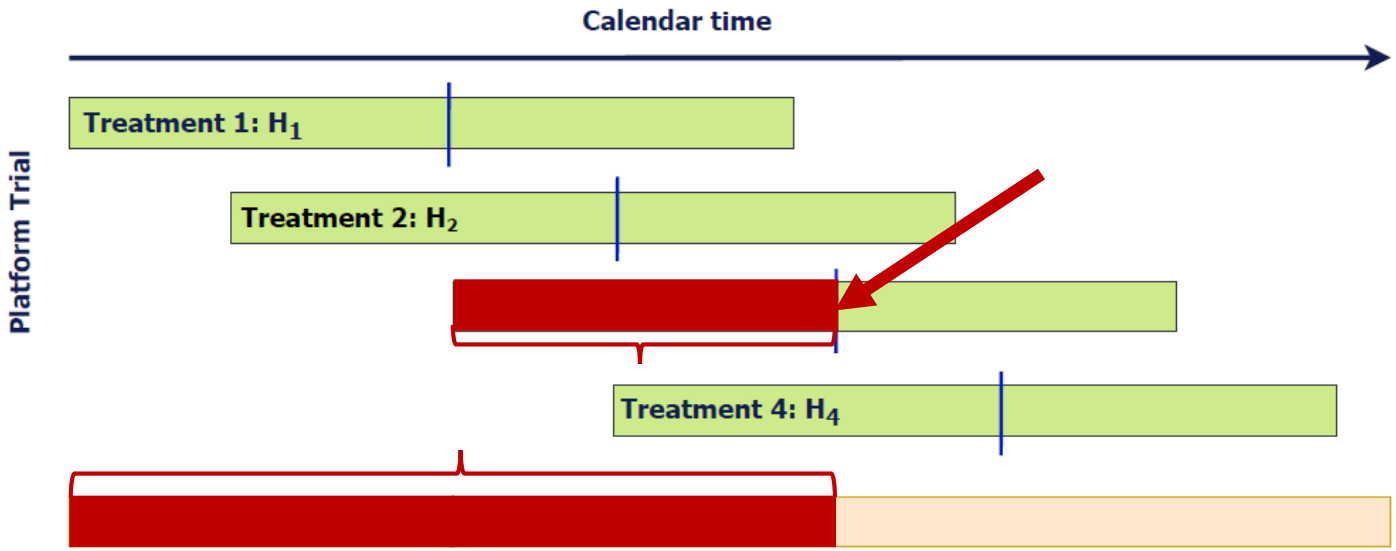
Benefiting from the past – calculation of conditional power



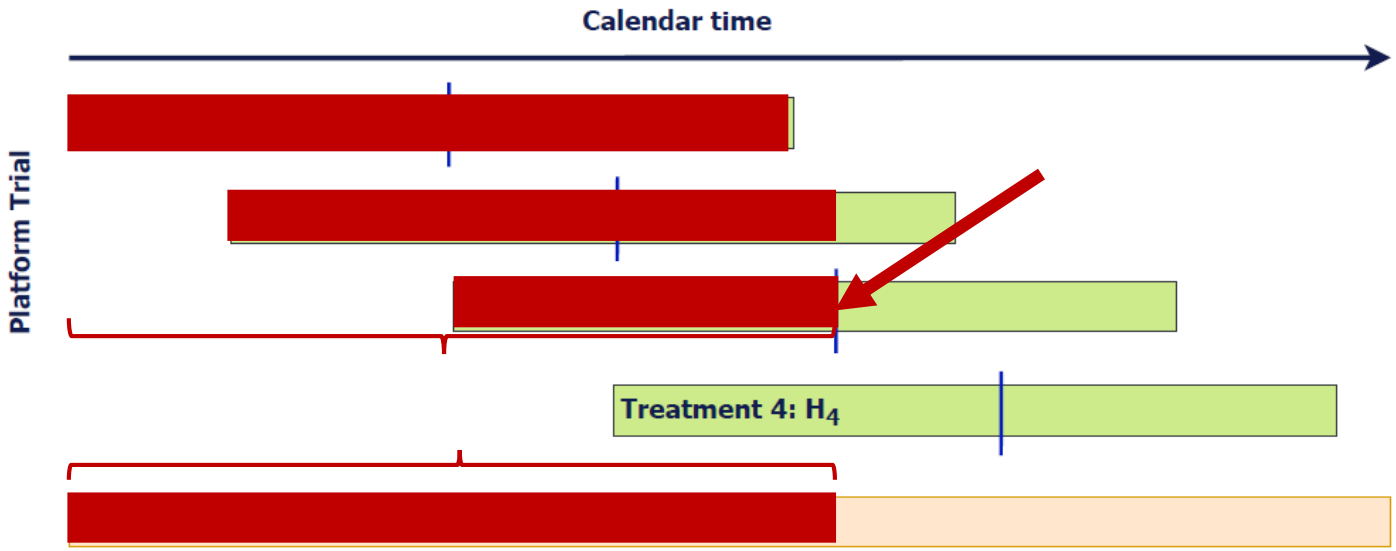
Benefiting from the past – calculation of conditional power



Benefiting from the past – calculation of conditional power



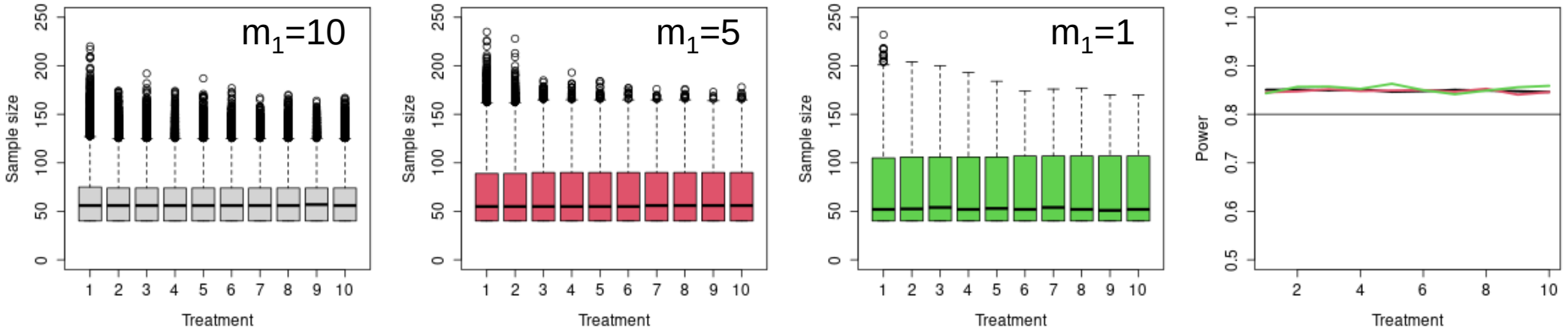
Benefiting from the past – calculation of conditional power



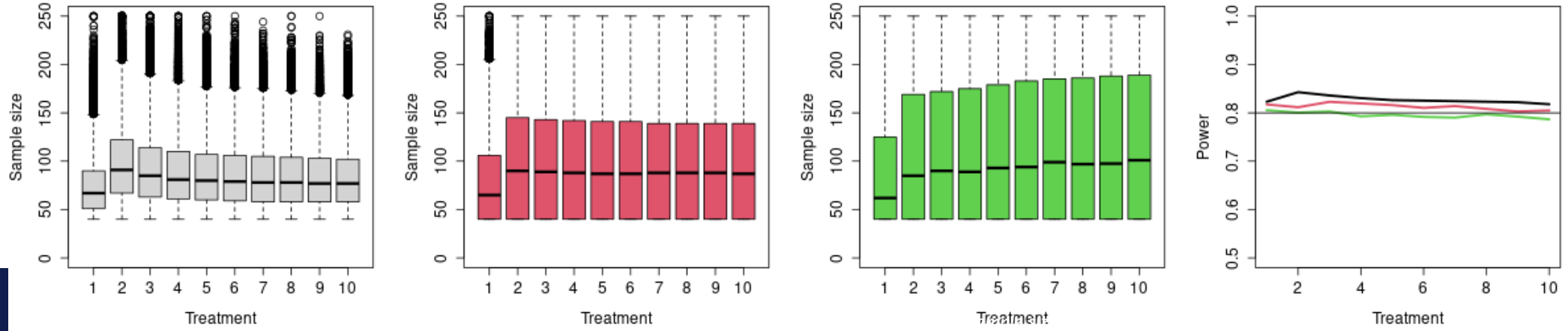
Adaptive design with stopping for futility at α_1=0.5

- O‘Brien-Fleming design for inverse normal method, minimum sample size in stage 2 is 10.
- Estimate sd for sample size reassessment from all controls and interim data

Level-α tests



FDR



Summary

- Benefiting from the past: The structure of platform trials helps to estimate sample sizes as past observations can be used in the planning phase.
  - Overlap of treatments has a high influence.
  - For very early treatments, planning is more difficult.

- **Sample size reassessment** might be a good choice.
  - Extension: reduction of influence of data from earlier treatments

efpia imi innovative medicines initiative

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853966.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and CHILDREN'S TUMOR FOUNDATION, GLOBAL ALLIANCE FOR TB DRUG DEVELOPMENT NON PROFIT ORGANISATION, SPRINGWORKS THERAPEUTICS INC.

References

- *Bofill Roig, M., Krotka, P., Burman, C.F., Glimm, E., Gold, S.M., Hees, K., Jacko, P., Koenig, F., Magirr, D., Mesenbrink, P., Viele, K., & Posch, M. (2022). On model-based time trend adjustments in platform trials with non-concurrent controls. BMC medical research methodology, 22(1), 1-16. https://doi.org/10.1186/s12874-022-01683-w*
- *Collignon, O., Gartner, C., Haidich, A. B., James Hemmings, R., Hofner, B., Pétavy, F., ... & Schiel, A. (2020). Current statistical considerations and regulatory perspectives on the planning of confirmatory basket, umbrella, and platform trials. Clinical Pharmacology & Therapeutics, 107(5), 1059-1067.*
- *Javanmard, A., & Montanari, A. (2015). On online control of false discovery rate. arXiv preprint arXiv:1502.06197.*
- *Javanmard, A., & Montanari, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. The Annals of statistics, 46(2), 526-554.*
- *Kieser, M., & Wassmer, G. (1996). On the use of the upper confidence limit for the variance from a pilot sample for sample size determination. Biometrical journal, 38(8), 941-949.*
- *Robertson, D. S., Wason, J. M., König, F., Posch, M., & Jaki, T. (2023). Online error rate control for platform trials. Statistics in Medicine.*
- *Robertson, D. S., Wildenhain, J., Javanmard, A., & Karp, N. A. (2019). onlineFDR: an R package to control the false discovery rate for growing data repositories. Bioinformatics, 35(20), 4196-4199.*
- *Tian, J., & Ramdas, A. (2021). Online control of the familywise error rate. Statistical Methods in Medical Research, 30(4), 976-993.*
- *Zehetmayer, S., Posch, M., & Koenig, F. (2022). Online control of the False Discovery Rate in group-sequential platform trials. Statistical Methods in Medical Research. 2022;0(0). doi:10.1177/09622802221129051*