Statistical and operational challenges with master protocols

Franz König
Section for Medical Statistics
https://cemsiis.meduniwien.ac.at/ms/

franz.koenig@meduniwien.ac.at

EJP RD - European Joint Programme on Rare Diseases - Statistical and operational challenges with master protocols
Rare Disease Training
24 March 2023 Online

Acknowledgements

- Thanks to Martin Posch, Elias Meyer, Marta Bofill Roig, Sonja Zehetmayer, Peter Mesenbrink, Ekkehard Glimm, Tobias Mielke, Tom Parke, Benjamin Hofner, Katharina Hess, Pavla Krotka, Cora Burgwinkel, Cecile Spiertz, Frank Bretz, Carl-Fredrik Burman, Ursula Garczarek, Kert Viele, and all other IMI-EU Pearl Working Group T2.1 members
- Olivier Collignon, David Robertson, James Wason, Nigel Stallard, Thomas Jaki

EU-PEARL
EU PATIENT-CENTRIC CLINICAL TRIAL PLATFORMS

efpia imi innovative medicines initiative

https://eu-pearl.eu/

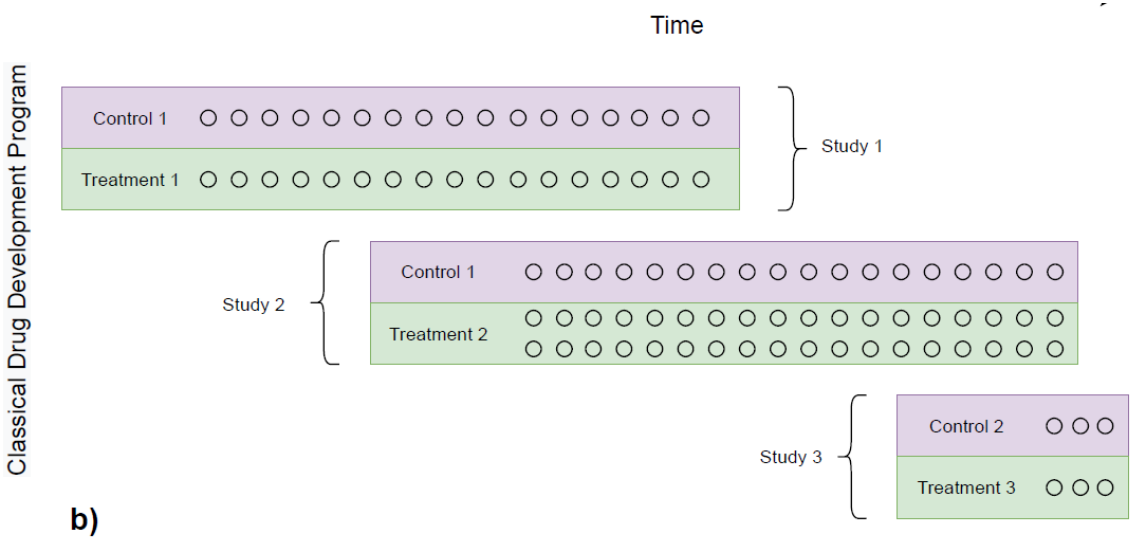
This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853966.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and CHILDREN'S TUMOR FOUNDATION, GLOBAL ALLIANCE FOR TB DRUG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

Overview

1. Introduction
2. Multiplicity Issues in Platform Trials
3. Shared and Non-Concurrent controls
4. Clinical Trial Simulations

Introduction

Classical Drug Development Programs



Traditionally:

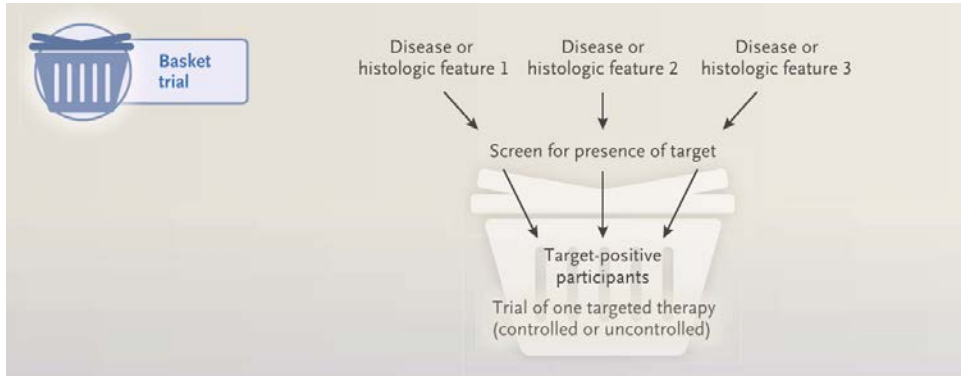
- Type 1 error (T1E) control at study level
- No data-sharing accross studies
- Sample size / power calculations quite simple
- Don´t share information accross studies / indications etc

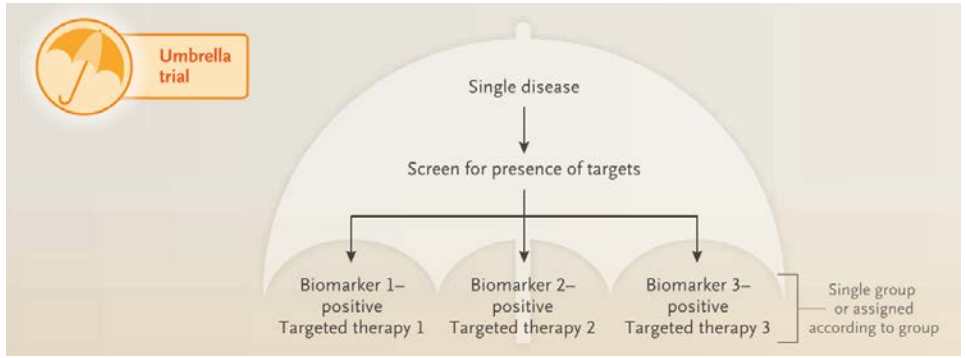
Why is there the wish for something different?

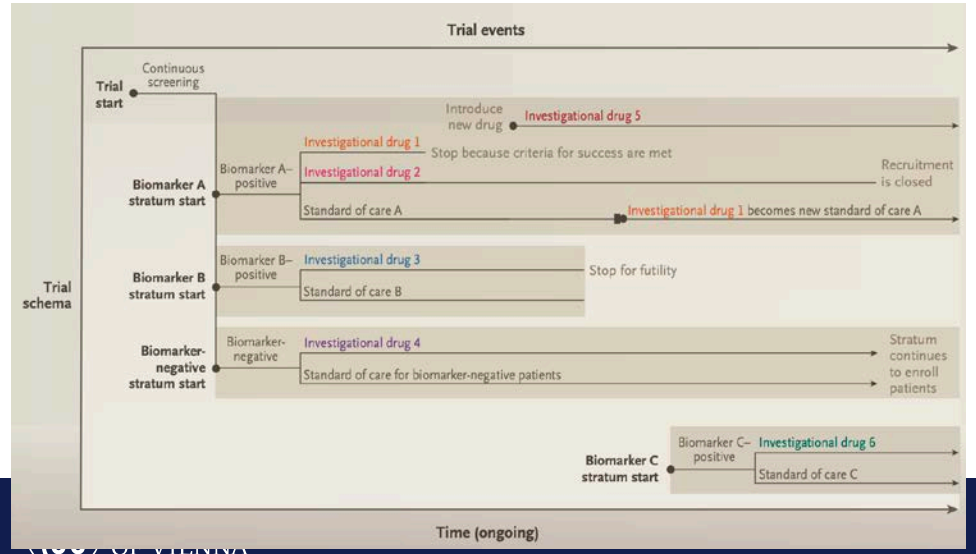
- Inefficient usage of resources
  - Standalone RCTs need their own control group
  - Each time develop new protocol, SAP,
  - Seek ethics & regulatory approval,
  - Look for appropriate trial sites, …
- Advances in personalized medicine lead to massive amount of hypotheses

Meyer et al. (2020b)

Master Protocols







- **Basket trial:** One investigational treatment (combination) is evaluated in the context of multiple diseases or disease subtypes with a common therapeutic target

- **Umbrella Trial**: Multiple investigational treatments (combinations) are evaluated in the context of a single disease, possibly within several substudies for different disease subtypes

- **Platform trial:** Umbrella trial, where drugs (combinations) may enter or leave the trial (e.g., if a new biomarker to identify disease subtypes becomes available)

Woodcock and LaVange ’17

Systematic Literature Review: https://doi.org/10.1016/j.clinthera.2020.05.010

Clinical Therapeutics/Volume 42, Number 7, 2020

The Evolution of Master Protocol Clinical Trial Designs: A Systematic Literature Review

Elias Laurin Meyer, MSc[1]; Peter Mesenbrink, PhD[2]; Cornelia Dunger-Baldauf, PhD[3]; Hans-Jürgen Fülle, MD PhD[3]; Ekkehard Glimm, PhD[3]; Yuhan Li, M.S.[2]; Martin Posch, PhD[1]; and Franz König, PhD[1]

[1]*Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria;* [2]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA; and* [3]*Novartis Pharma AG, Basel, Switzerland*

ABSTRACT

Purpose: Recent years have seen a change in the way that clinical trials are being conducted. There has been a rise of designs more flexible than traditional adaptive and group sequential trials which allow the investigation of multiple substudies with possibly different objectives, interventions, and subgroups conducted within an overall trial structure, summarized by the term *master protocol*. This review aims to identify existing master protocol studies and summarize their characteristics. The review also identifies articles relevant to the design of master protocol trials, such as proposed trial designs and related methods.

Methods: We conducted a comprehensive systematic search to review current literature on master protocol trials from a design and analysis perspective, focusing on platform trials and considering basket and umbrella trials. Articles were included regardless of statistical complexity and classified as reviews related to planned or conducted trials, trial designs, or statistical methods. The results of the literature search are reported, and some features of the identified articles are summarized.

Findings: Most of the trials using master protocols were designed as single-arm (n = 29/50), Phase II trials (n = 32/50) in oncology (n = 42/50) using a binary endpoint (n = 26/50) and frequentist decision rules (n = 37/50). We observed an exponential increase in publications in this domain during the last few years in both planned and conducted trials, as well as relevant methods, which we assume has not yet reached its peak. Although many operational and statistical challenges associated with such trials remain, the general consensus seems to be that master protocols provide potentially enormous advantages in efficiency and flexibility of clinical drug development.

Implications: Master protocol trials and especially platform trials have the potential to revolutionize clinical drug development if the methodologic and operational challenges can be overcome. (*Clin Ther.* 2020;42:1330–1360) © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Key words: adaptive design basket trial, master protocol, multi-arm multi-stage design, platform trial.

INTRODUCTION

Since the late 1940s, randomized controlled trials (RCTs) have served as the gold standard for establishing therapeutic efficacy.[1] However, recent advances in drug discovery and biotechnology have accelerated tremendously the detection of treatment candidates. In addition, diagnostics have become more refined, leading to more precisely defined disease descriptions and hence smaller patient populations for targeted therapies. The classic 2-arm parallel-group RCTs have thus become one of the rate-limiting factors in drug development, and more

Accepted for publication May 11, 2020
https://doi.org/10.1016/j.clinthera.2020.05.010
0149-2918/$ - see front matter

- literature search on PubMed last updated January 01, 2020, search terms such
  - master protocol*[Title/Abstract] OR
  - platform/basket/umbrella trial/stud/design*[Title/Abstract] OR
- Included 164/678 identified papers + 122 manually
- In total 50 planned or conducted trials with master protocol identified

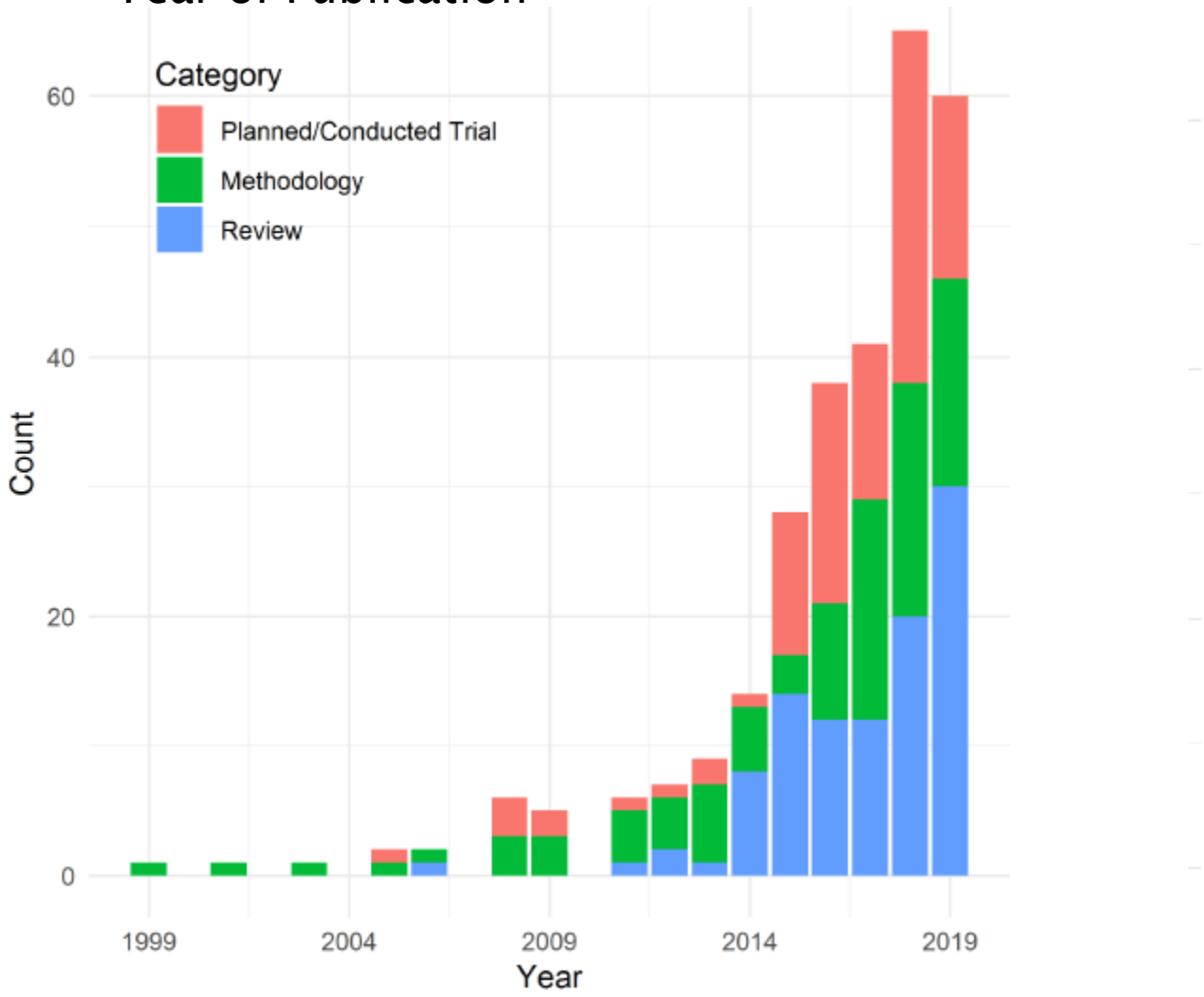
Literature Review Results (1) - Meyer et al. 2020

Feature	Category	Basket	Other	Platform	Umbrella	Sum
Phase	I	1	0	0	0	1
	I/II	1	2	0	2	5
	II	12	11	5	4	32
	II/III	0	0	4	1	5
	III	2	0	1	2	5
	IV	0	0	2	0	2
Indication	Oncology	15	13	6	8	42
	Other	1	0	6	1	8
Endpoint	Binary	10	7	5	5	27
	Binary/TTE	0	5	0	1	6
	Metric	2	0	1	0	3
	Safety/Binary	2	0	0	0	2
	Safety/TTE	0	0	0	1	1
	TTE	2	1	6	2	11
Control	concurrent	3	1	5	4	13
	common	0	1	6	1	8
	no control	13	11	1	4	29
Analysis	Frequentist	16	11	3	7	37
	Bayesian	0	2	9	2	13
Total		16	13	12	9	50

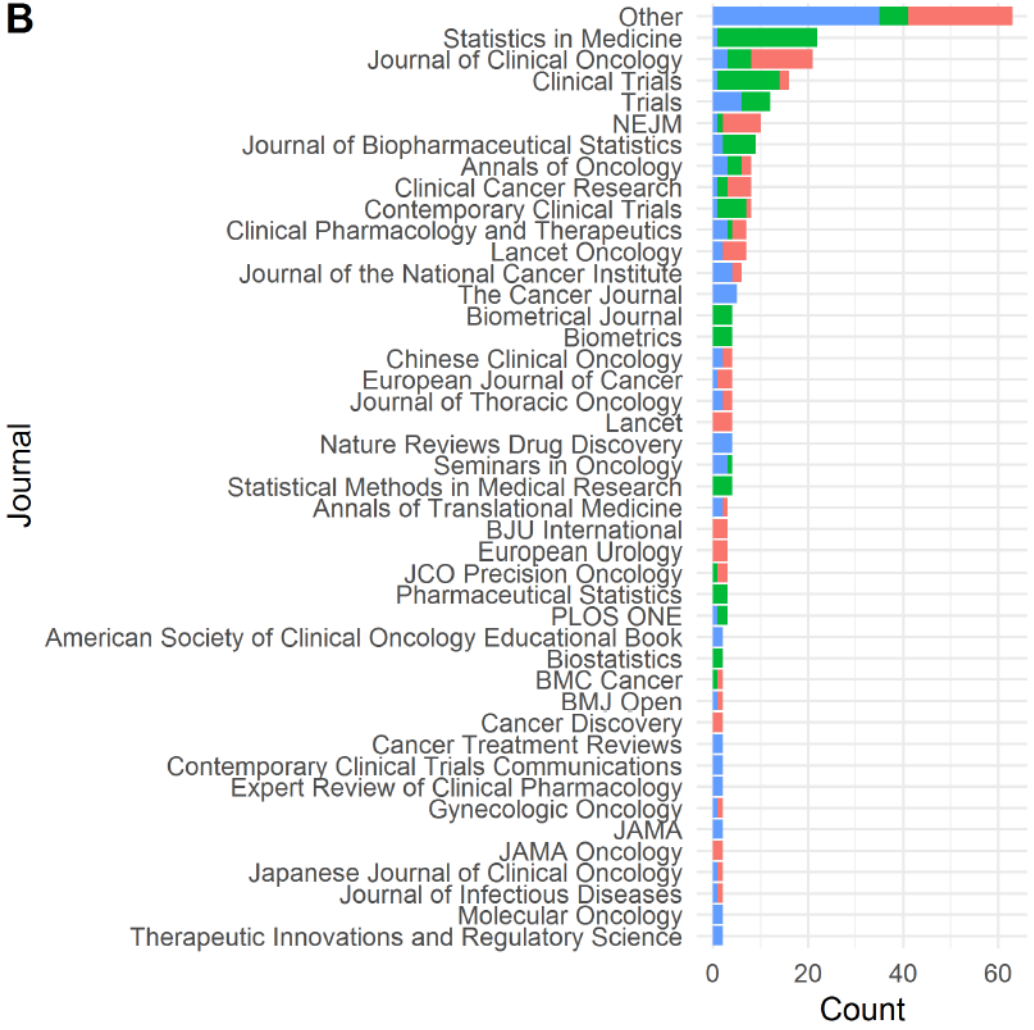
- TTE. . . Time-to-event
- Type of control “common” refers to both concurrent and non-oncurrent controls being used

Literature Review Results (2):

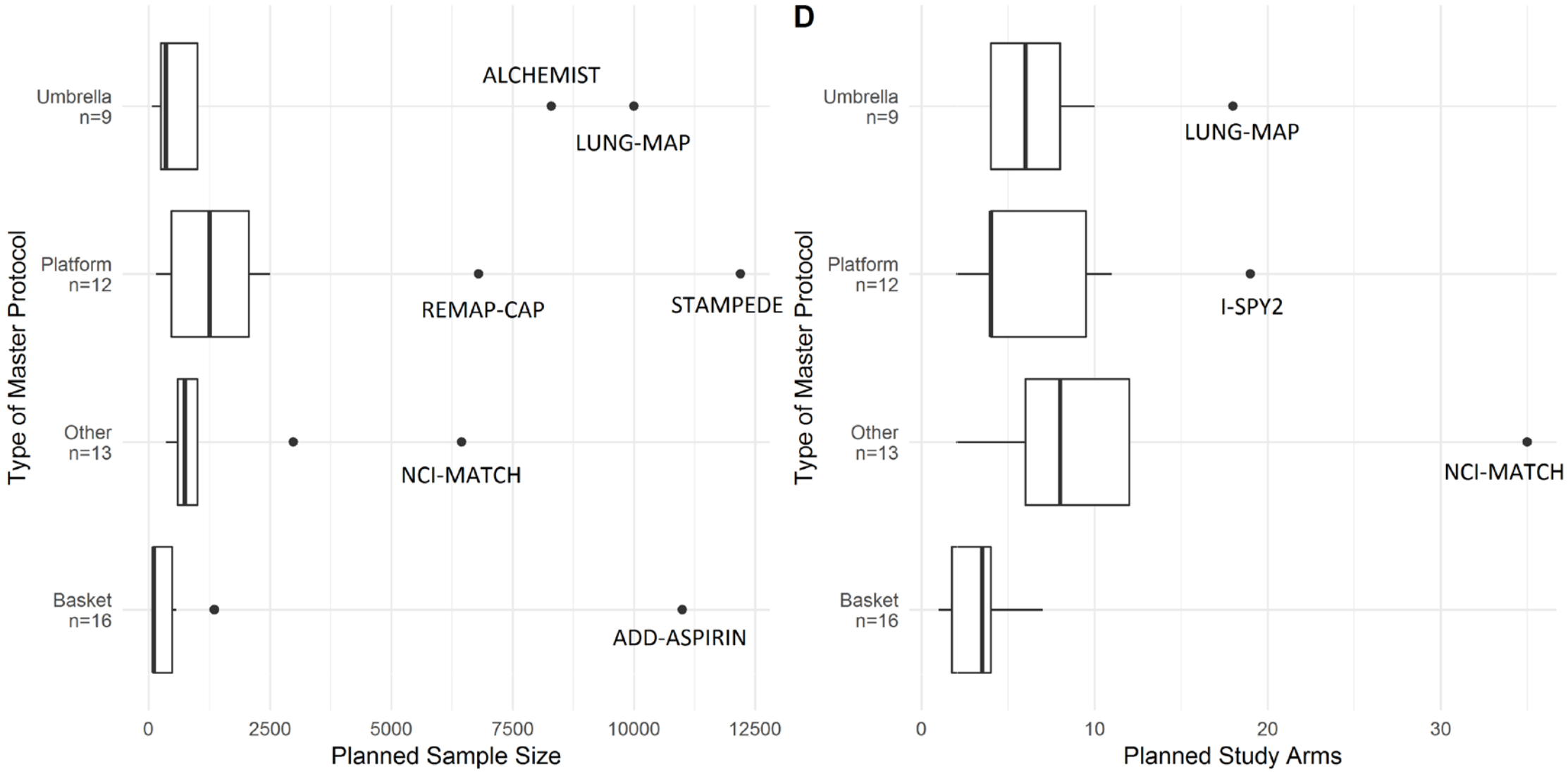
Year of Publication



Journal of publication



Literature Review Results (3): Sample sizes & arms



Some observations of the review

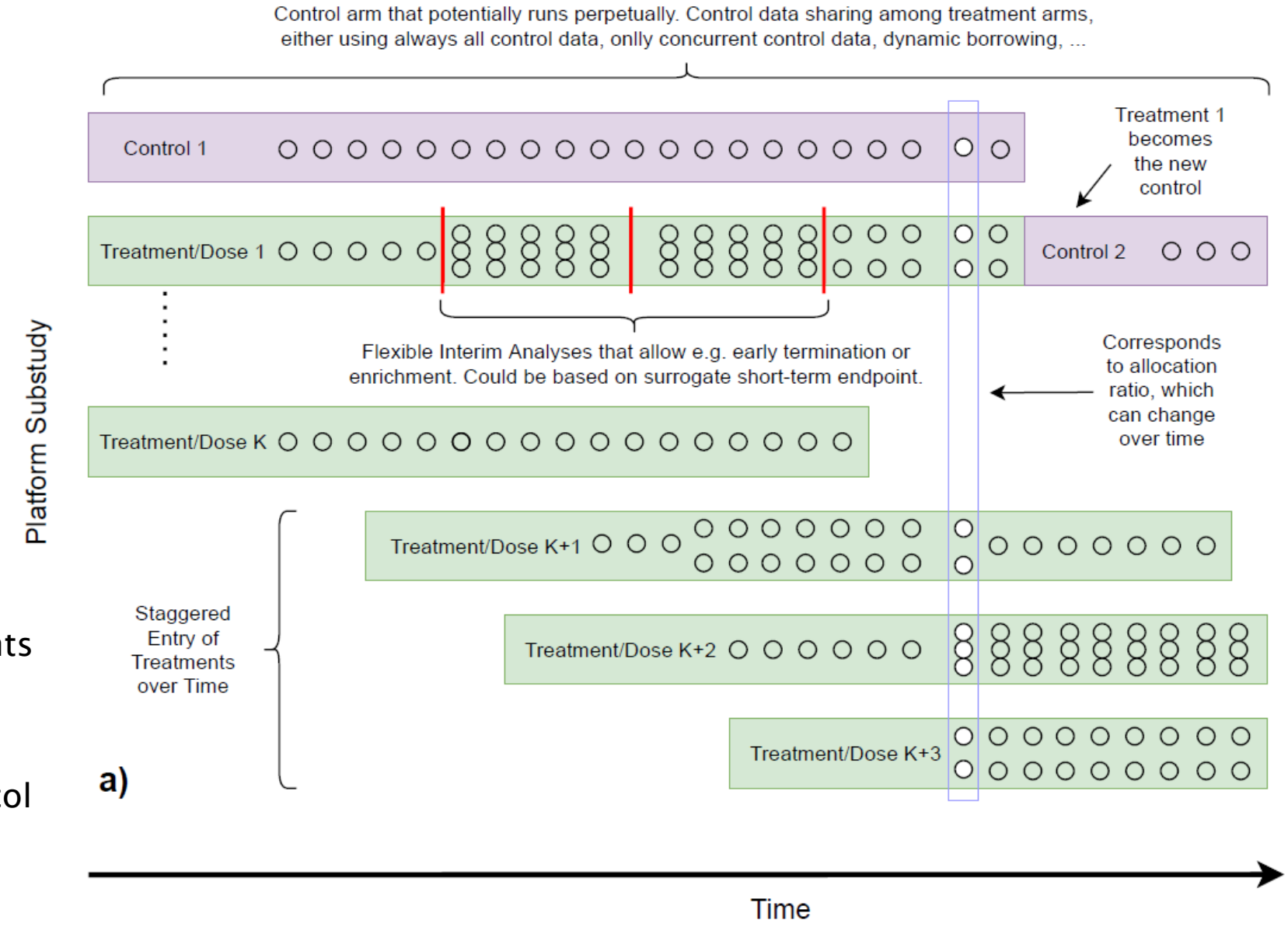
- Exponential increase of publications in this domain in the last years
- Mostly single-arm (n = 29/50), phase II trials (n = 32/50) in oncology (n = 42/50) using a binary endpoint (n = 27/50) and frequentist decision rules (n = 37/50)
- Master protocols provide potentially enormous advantages in efficiency and flexibility of clinical drug development
- Design and associated statistical challenges depend strongly on stage of drug development and require further research

- Now there are many platform trials related to Covid (Recovery, Solidact, Eucovat,…)

Collaborative Platform Trials

Design Characteristics of Platform Trials

- Multi-armed trials
- Interim analyses & adaptations
- Treatments to be studied not defined upfront but may enter during the course of the trial
- Control arm(s) can be shared
- Control arm(s) may change over time
- Populations for the different treatments may not be the same (Umbrella type trials)
- Designed as trial with a Master Protocol with several sub-studies



Potential advantages of platform trial

Operational:

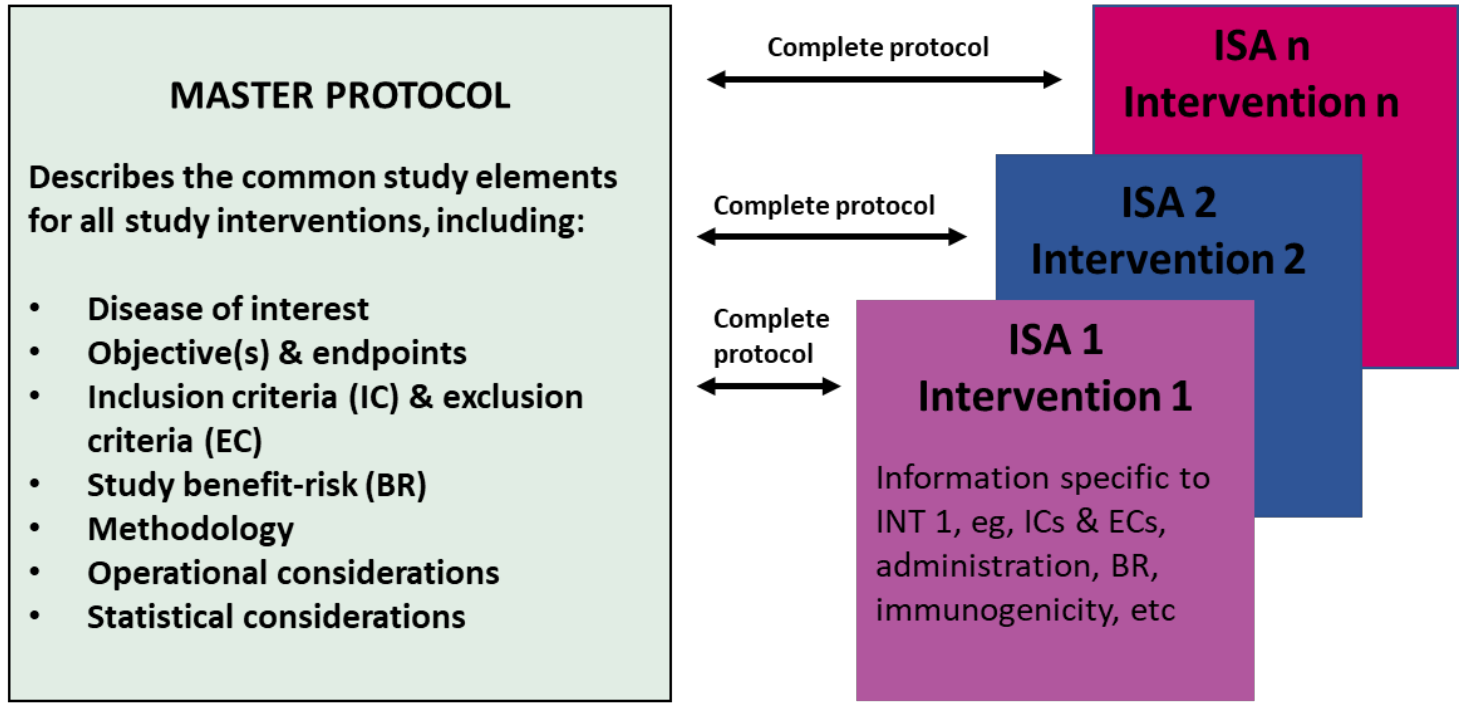
- More patients eligible for trial due to multiple treatments and sub-studies with possibly different inclusion criteria
- Joint trial infrastructure leads to savings in time and money for sponsor(s)

Statistical:

- Multiple hypotheses tested in the same trial (which is also a big challenge)
- Sharing of control data and adaptive decision rules potentially lead to fewer number of patients required
- Direct comparison between treatments allows for adaptive randomization leading to effective treatments “graduating” faster and fewer patients on inefficacious treatments

Master Protocol and Intervention Specific Appendices

- The **Master protocol** governs the entire study and includes the **common key study design elements**
- **Intervention-specific information** is provided in **Intervention Specific Appendices (ISAs),** which are added as interventions become available and are ready to enter the platform study
- Interventions can enter the platform study **simultaneously** or **sequentially** as they become available for study
- **Both protocols** are needed to have all the information needed to conduct the study in an intervention cohort



EU-PEARL will soon release a set of templates for master protocol, ISA, DMC charta, SAP, etc follow https://eu-pearl.eu/

Multiplicity Issues in Platform Trials

Traditionally: Frequentist hypotheses tests for decision making in confirmatory clinical trial

- Traditional decision making in confirmatory clinical trials is based on hypothesis testing
- The null hypothesis **"The experimental treatment is not superior to control"** is tested with a statistical test
- Based on the clinical trial data a **p-value** is calculated
- If **p<0.05** the null hypothesis is rejected and the drug is declared efficacious
- This guarantees that the probability **of a false positive result** (given the treatment does not work) **is lower than 5%**
- However, if **multiple tests** are performed with the same threshold of 0.05, the risk of at **least one false positive conclusion increases**

> "Control of the study-wise rate of false positive conclusions at an acceptable level α is an important principle and is often of great value in the assessment of the results of confirmatory clinical trials."
>
> Points to consider on multiplicity issues in clinical trials, EMA (2002)

Multiplicity & Different Trial Designs
Find the difference

DESIGN 1

One study
1:1:1, e.g. testing:

- Placebo vs. Drug A
- Placebo vs. Drug B

Two chances for success
Typically correct for chance of at least one type-1 error

DESIGN 2

Two separate studies
1:1 and 1:1, e.g. testing

- Placebo vs. Drug A
- Placebo vs. Drug B

Two chances for success
No correction for type-1 error

DESIGN 3

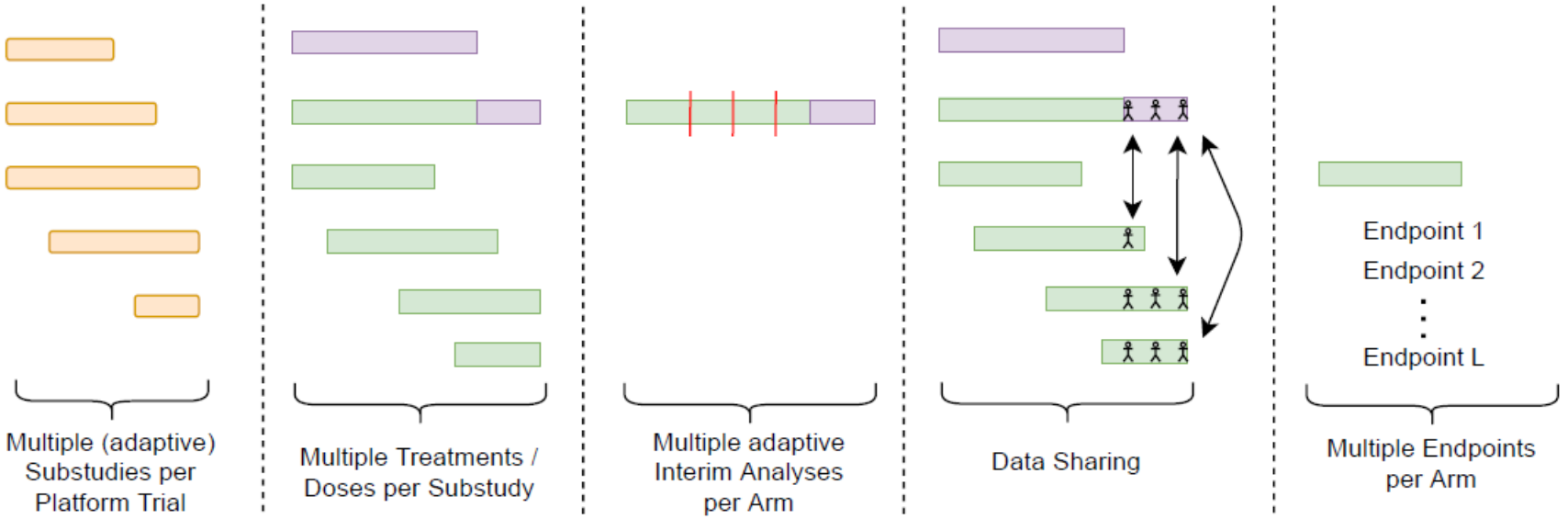
1 Platform with 2 ISAs
1:2 and 1:2, e.g. testing

- Placebo vs. Drug A
- Placebo vs. Drug B

Two chances for success
No correction for type-1 error?

Slide from EU-PEARL -Stakeholder workshop statistics breakout: When and how to correct

Sources of structural multiplicty in Platform trials



Multi-regional different regulatory requirements

Multiple interventions + *Adaptive decision making* + *Multiple Control Groups* + *Multiple endpoints* + *Multiple subgroups* + *Multi-regional*

= Large convoluted multiplicity problem

Error Rates when Testing Multiple Hypotheses

Slides based on Nigel Stallard, EU-Pearl Multiplicity Workshop, 2021

Error rate control in hypothesis testing

Single hypothesis test

Test hypothesis H_0

Control type I error (false positive) rate at level α if

$$P(\text{reject } H_0 \mid H_0) \leq \alpha$$

Error rate control in multiple hypothesis testing (1)

Family of hypotheses

Test hypotheses $H_{01}, \ldots, H_{0m}$

Control familywise error rate (FWER) at level α

in weak sense if

$$P(\text{reject any } H_{0i} \mid H_{01}, \ldots, H_{0m}) \leq \alpha$$

in strong sense if

$$P(\text{reject any true } H_{0i}) \leq \alpha$$

Error rate control in multiple hypothesis testing (2)

	Null Hypotheses		
	True	False	Total
Rejected	V	S	R

- Familywise error rate:
  FWER = $Pr(V \geq 1)$
- False discovery proportion:
  $Q = V/R$ ($Q = 0$ if R=0)
- False discovery rate:
  FDR = $E(Q)$

FDR ≤ FWER (i.e. FWER control is the more stringent requirement!)

For which Family(-ies) of Hypotheses should we Control for Multiplicity?

- **Single Family:** all hypotheses tested in the Platform

  i.e., all treatments, endpoints, subgroups, …

- **Separate Family for each treatment:** all hypotheses tested for a specific treatment

  i.e., for each treatment arm all endpoints, subgroups

- **Separate family for each hypothesis**

  no control for multiplicity!

Is there a need to adjust for multiplicity in Platform trials?

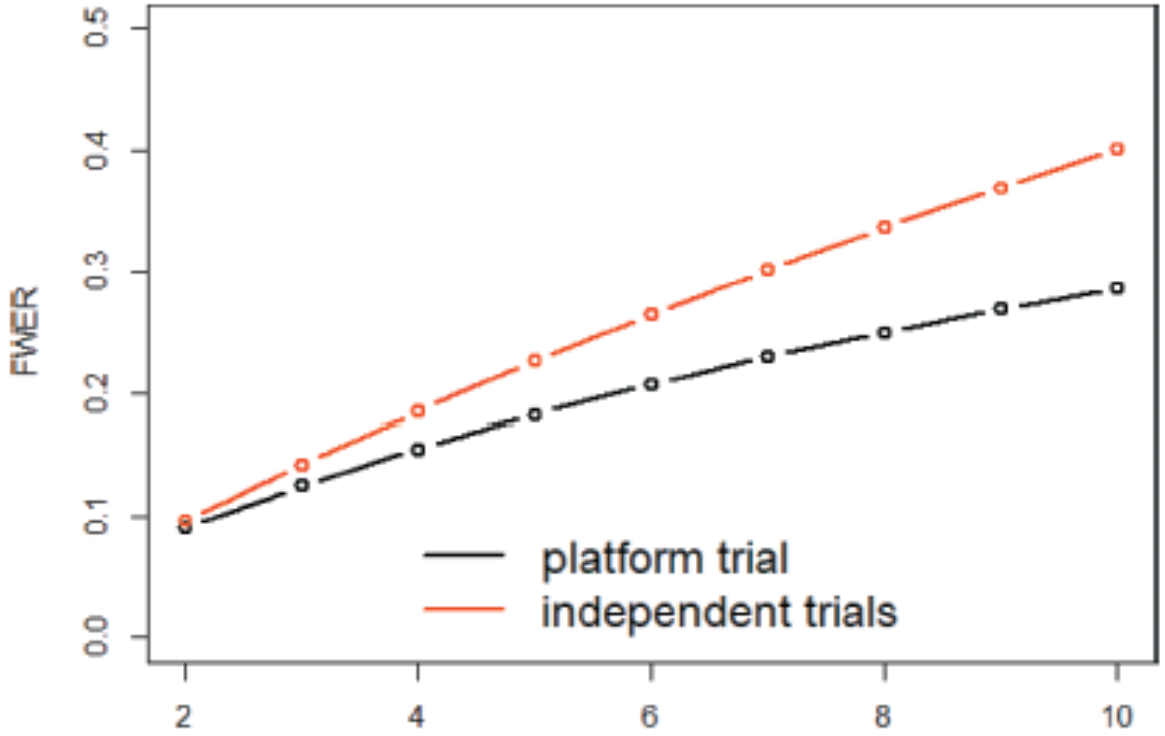
When performing several treatment-control comparisons in a platform trial, is the risk really increased?

- **Given multi-armed 1:1:…:1 allocation:**
- Due to shared controls the test statistics will be positively correlated
- If all treatment control comparisons are tested at nominal level α=0.05 the FWER in single multi-armed study is lower than in a series of independent trials
- … Why to be stricter in platform studies?
  - regulatory risk not increased
  - separate unrelated regulatory claims
- Increasing control allocation:
  - FWER similar to series of independent trials

But : Due to the correlation, the probability to perform several Type 1 Errors simultaneously increases

Stallard et al. 2019
Collignon et al. 2020a, 2020b

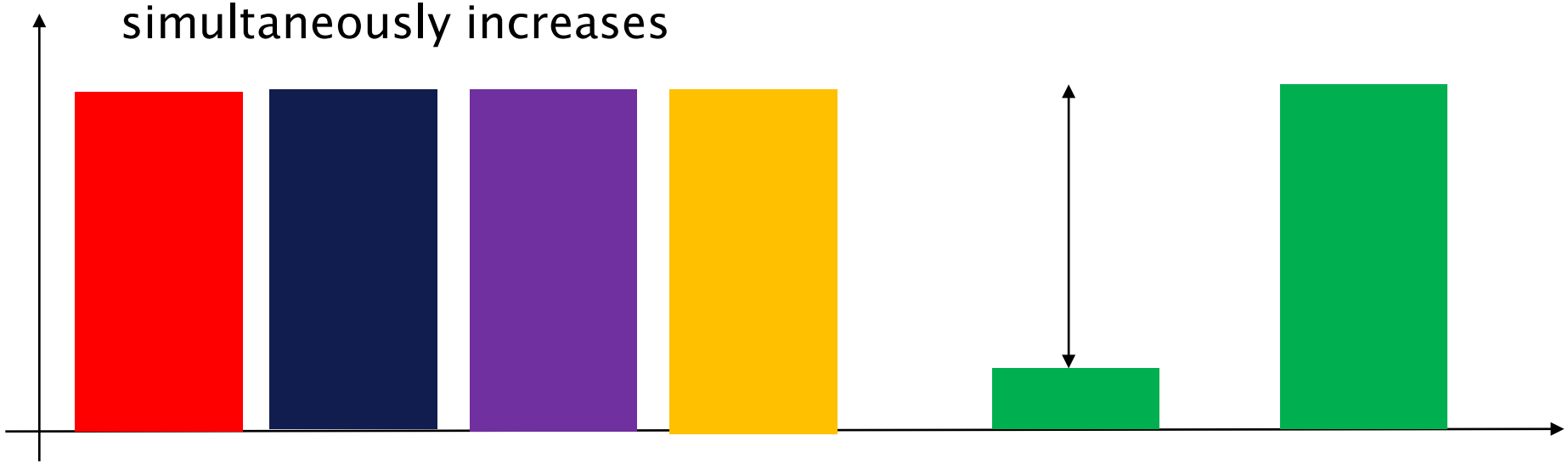
Figure 1 from Bai et al. 2020
Comparison of FWER of platform trial and independent trials.



number of testing arms – all are ineffective

Correlation of Estimates due to Shared Controls

- Due to shared controls the test statistics will be positively correlated
- If all treatment control comparisons are tested at nominal level α=0.05 the familywise error rate (FWER) **is smaller compared to tests in independent trials**.
- Due to the correlation, the probability to perform several Type 1 Errors simultaneously increases



Stallard et al. 2019
Collignon et al. 2020a, 2020b

FWER when ignoring multiplicity & adaptations

What can go wrong: Comparing of k treatments with a control

What is the most extreme T1E rate:

- If SSR* conducted…
- But analysis not corrected

Maximum type 1 error inflation:

nominal α	$k = 1$ balanced[1]	$k = 1$ unbalanced[2]	$k = 2$ unbalanced[3]
0.05	0.115	0.187	0.289
0.025	0.062	0.106	0.170
0.01	0.027	0.049	0.080

1 Proschan and Hunsberger 1995

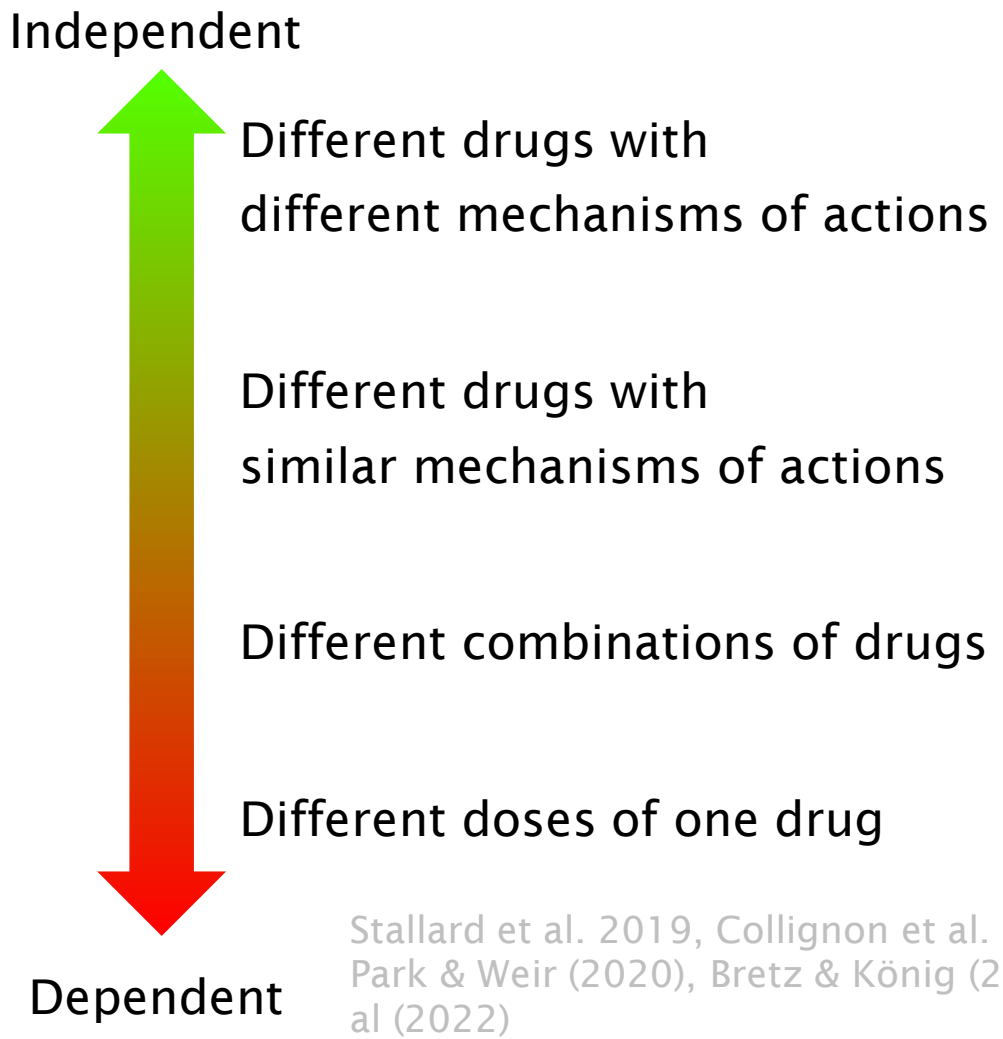
2 Graf and Bauer 2011

3 Graf, Bauer and Koenig 2014

SSR= Adaptive sample size re-estimation on unblinded data*

NO need to adjust WHEN hypotheses are **inferentially** independent

- Hypotheses are inferentially independent, if **the truth or falsehood of one hypothesis is unrelated to the truth and falsehood of the other hypotheses.**
- no extrapolation from one hypotheses to the the other is possible.
- If we did separate trials, we would also not adjust for multiplicity
  (and the shared control group leads to a lower FWER anyway)



Stallard et al. 2019, Collignon et al. 2020a, 2020b, Park & Weir (2020), Bretz & König (2020), Nguyen et al (2022)
EU-PEARL session on multiplicity first stakeholder workshop

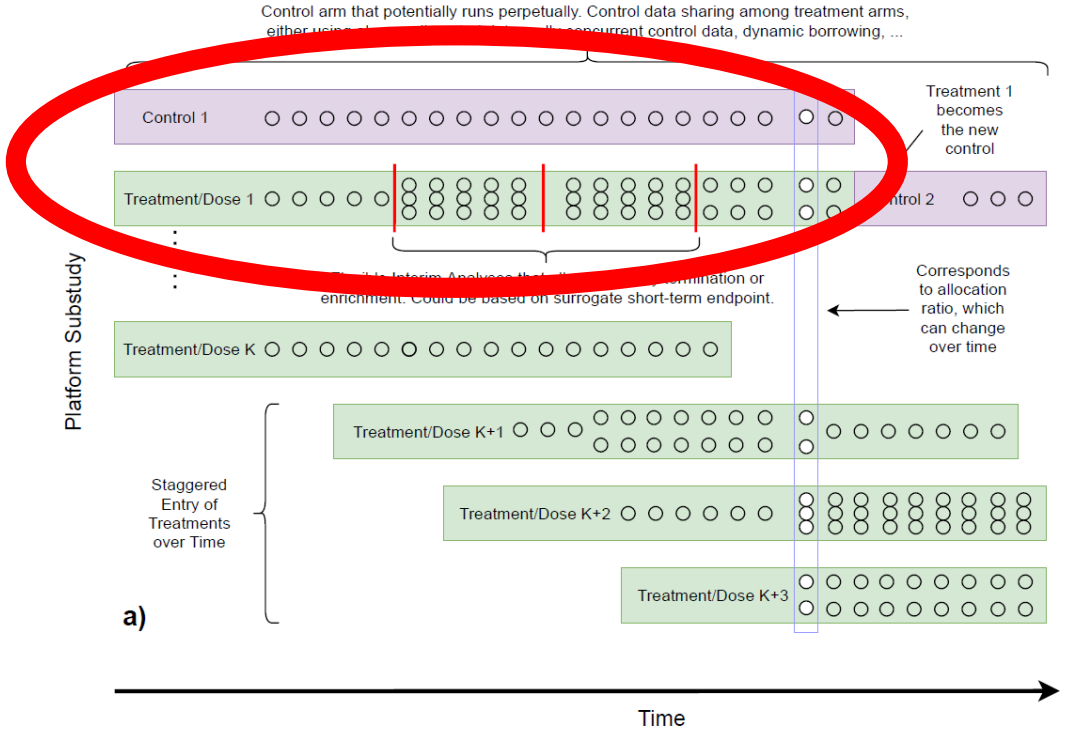
A pragmatic strategy for statistical inference

Each treatment/substudy in the platform trial is considered as an independent separate substudy, each controlling the FWER for the family of hypotheses relating to the treatment/substudy

For each substudy adjust for

- Multiple endpoints
- Multiple doses/treatment regimens
- Multiple subgroups
- Interim Analyses

But no adjustment across substudies



Do the Operating Characteristics (OC) of main interest influence whether we should adjust for multiplicity or not?

OC	Cohort	Platform
Power	Per-Cohort-Power (PCP), probability of a truly effective cohort being declared successful	Disjunctive Power (Disj_Power), probability of at least one truly effective cohort being declared successful
T1E	Per-Cohort-Type-1-Error (PCT1ER), probability of a truly ineffective cohort being declared successful	Family-wise-error-rate (FWER), probability of at least one truly ineffective cohort being declared successful

- Many other OCs possible, e.g. FDR, average power, time until first success, patients allocated to arms superior to SoC, . . .

Meyer et al. 2021

Which OCs (risks, power) are we really interested in?

Makes "independence of claims"—criterion overall error control irrelevant?

- Take COVID-Platform trials with treatments A, B, C, D, … and a shared control
- Are you interested in controlling
  - the risk that
    - Declaring treatment A better than Placebo (if it is not)
    - Any of the ineffective treatments is declared better than control
    - Several inefficient treatments are approved simultanesously (e.g. if the control is on a random low)
  - The proportion of ineffective treatments among the treatments which demonstrated efficacy,
  - …,
  - ?

Do the objectives of the platform trial determine wether to account for multiplicity?

Objective	Adjustment
Find at least one effective treatment.	Control of the probability of at least one false positive decision
For each treatment determine if it is effective.	Unadjusted Analysis
Determine all effective treatments	Unadjusted Analysis/False Discovery Rate

But are there any remaining reasons to adjust for multiple comparisons to a control?

Current regulatory standard

Uphold the principle of study-wise error rate control.

Societal perspective

If many treatment-control comparisons are made either in separate or in a platform trial, **it is relevant to assess if (and how many) treatments are erroneously shown to be effective.** A platform trial can provide a framework to quantify this risk

- by controlling an overall multiple error rate, as the FWER or FDR, at a pre-specified level,
- by estimating, e.g., the FDR of the platform trial to quantify the level of evidence provided.

Best use of ressources

Multi-arm trials have a higher efficiency. Why not **invest some of the efficiency gain for the control of the overall false positive rate**?

Power comparison of separate trials vs multi-arm trials

k experimental treatments, equal effect sizes

- **Separate trials at level $\alpha = 0.025$**

  Per group sample sizes: $n_T = n_C = n$

  Overall sample size: $N_k = 2\,n\,k$

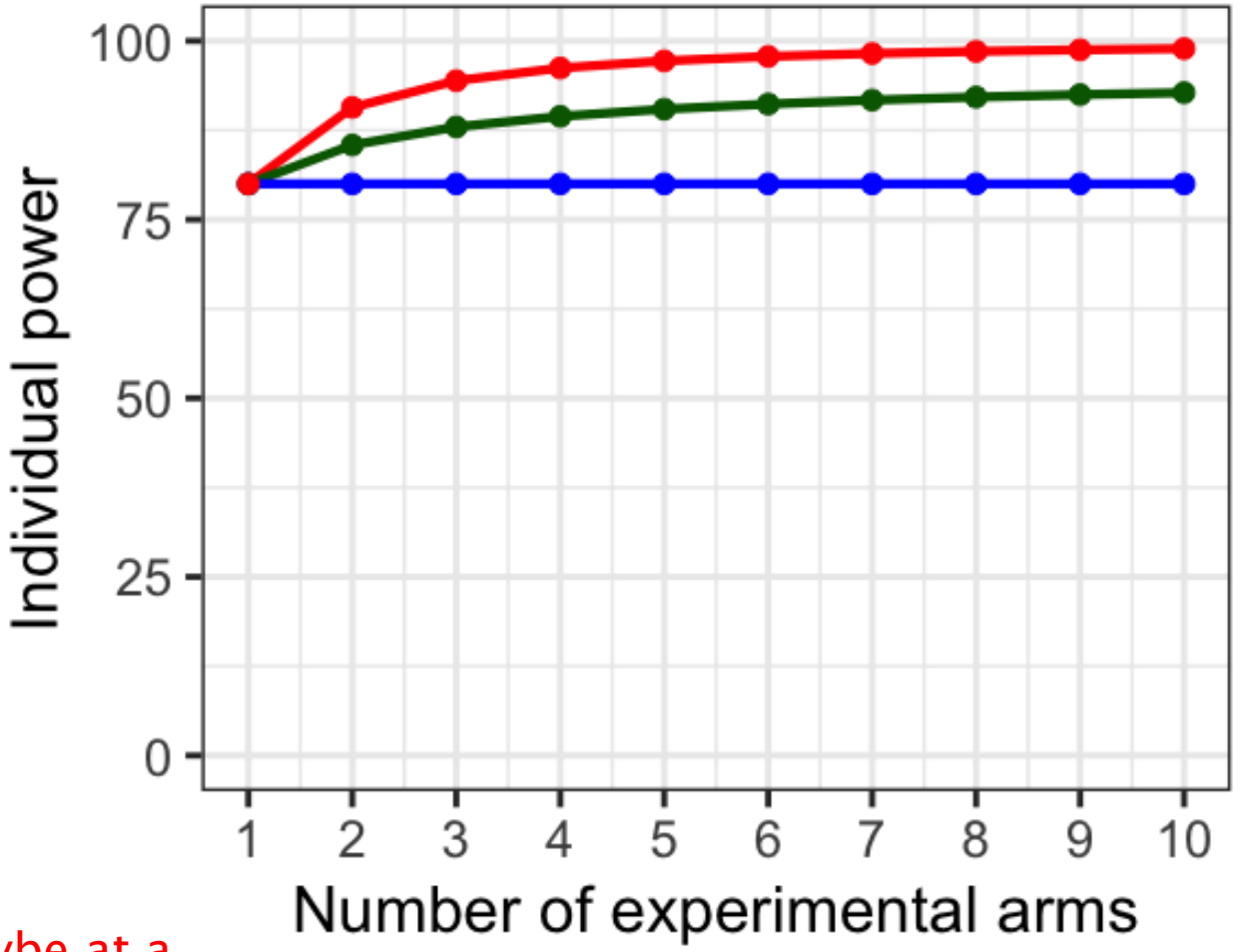
- **Multiarm trials at (unadjusted) level α**

  Overall sample size: N_k

  Allocation ratio: $1:1:1:\ldots 1:\sqrt{k}$

- **Multiarm trials at Dunnett adjusted level**

  (sample sizes as above)



So should we use a multiplicity adjustment, but maybe at a higher level?

Challenges in Decision Making

- As the **number and type of treatments ares not defined upfront**, standard procedures to adjust for multiplicity are not applicable. Methods for "online control" of error rates must be used.

- **Online control of the FWER** can lead to very different significance levels for treatments that enter the platform later

- Control of the probability of at least one type I error appears to be too stringent especially in a potentially perpetual trial.

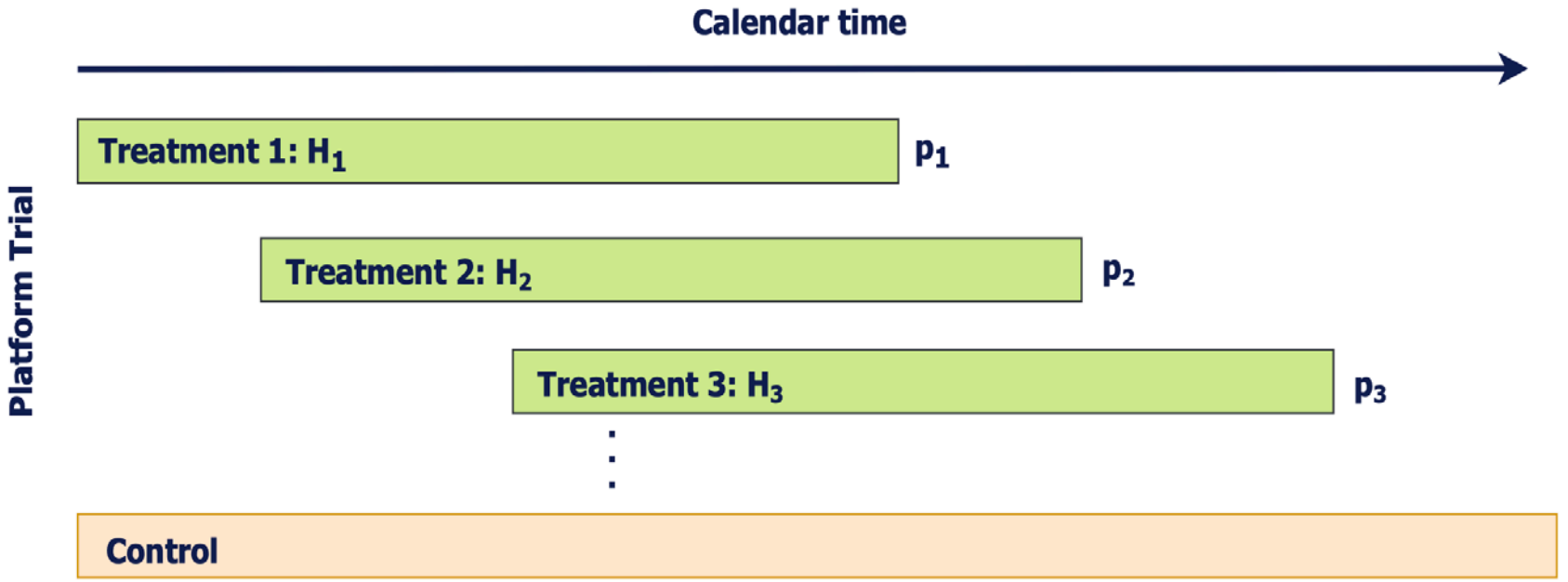
- Besides FWER control, one approach is to control (or estimate) the False Discovery Rate. Other approaches directly account for the *loss* of false positive decisions

Wason et al. (2020)

Online Control of the FDR in Platform Trials

Zehetmayer, Posch and König (2022)
https://journals.sagepub.com/doi/10.1177/0962280221129051

Platform trials



Multiple testing issue

Conventional methods to control error rates assume that

- number of hypothesis tests is fixed
- all p-values are available at time of test decision

Online Error Rate Control of the False Discovery Rate

- Hypothesis tests and test decisions are performed in a pre-defined order. We aim to control the FDR at each step.
- Null hypotheses arrive sequentially $H_1, H_2, H_3, \ldots,$
- Test each null hypothesis when it arrives.
- False discover rate rates after *n* hypotheses have been tested: FDR_n.
- Online control requires that

$$sup_n \, FDR_n \leq \alpha$$

Online control of the False Discovery Rate

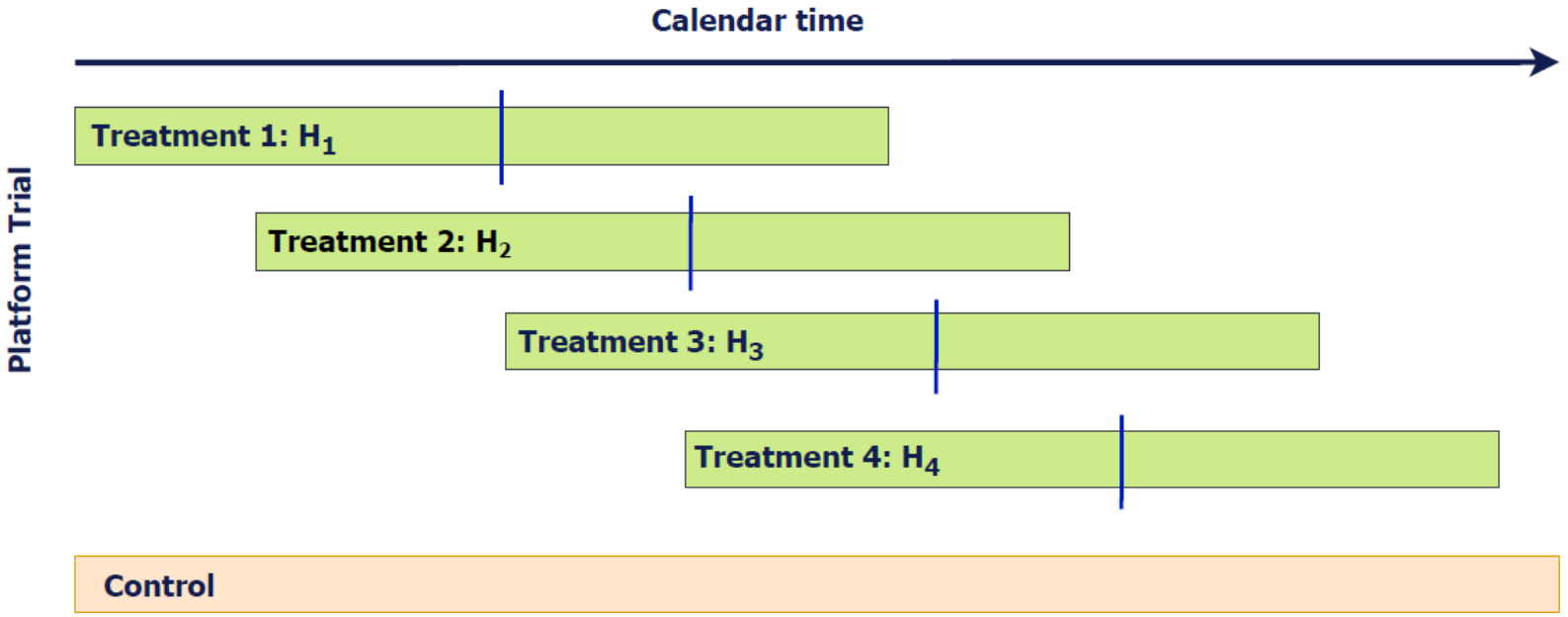
- At each step a decision has to be made if the current null hypothesis should be rejected **based only on previous decisions.**
- For many online FDR methods, control of the online FDR for independent p-values has been proved.
- In platform trials: Due to shared control arm, **positive correlation** of test statistics.

Javanmard and Montanari (2015, 2018), Robertson et al. (2019), Wason and Robertson (2020)

Shoud we use the online FDR in platform trials?

Zehetmayer et al. 2022

False Discovery Rate (FDR) : *The expected proportion of treatments that are falsely declared efficacious, among all treatments that are declared efficacious.*



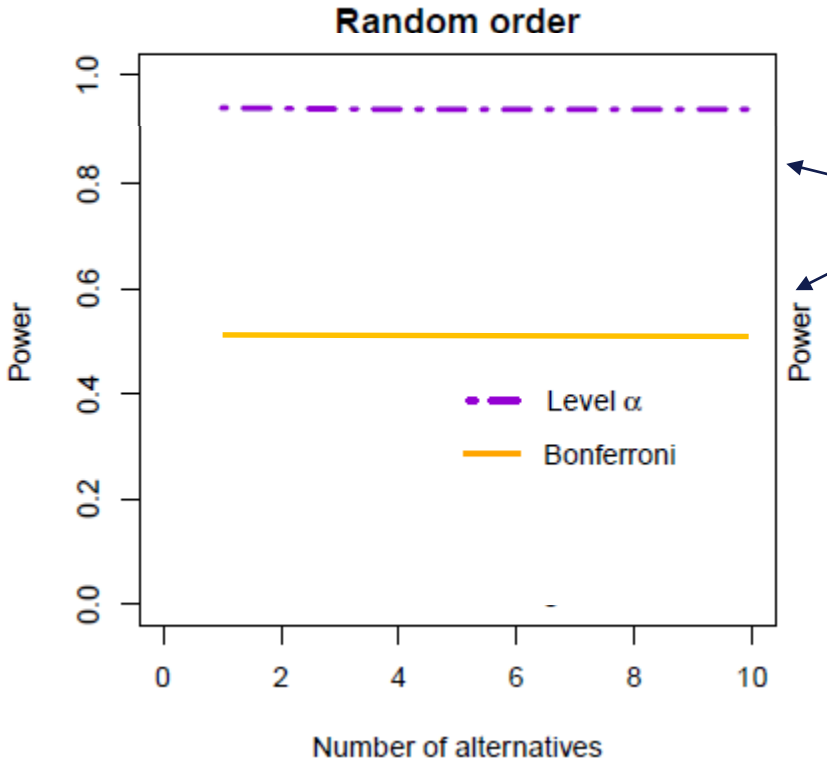
We need to consider two sources of multiplicity:

1. Adjust for the **total number of treatments** with the LOND procedure (Javanmard and Montanari, 2015)
2. Adjust for the **interim analyses** with spending functions ("split" signif. thresholds α_i)

Should we use the online FDR in platform trials?

Zehetmayer et al. 2022

- Average power when 10 arms are compared to common control
- Group sequential at
  - One-sided level alpha 0.025
  - Bonferroni Alpha/100
  - LOND – FDR



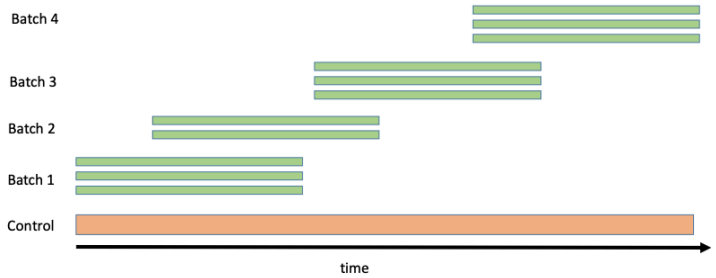
Impact of the adjustment depends on how many further arms we expect in the platform

Or should we just perform unadjusted level alpha tests and report an estimate of the FDR whenever a decision for a treatment is taken?

For online FDR control see also Javanmard and Montanari (2015, 2018), Robertson et al. (2019), Wason and Robertson (2020), Robertson et al. (2023)

Some more remarks

- Other online procedures (LORD, ADDIS, ...) have been proposed for FDR control
- Modifications when some treatments finish at the same time to allow for batch testing

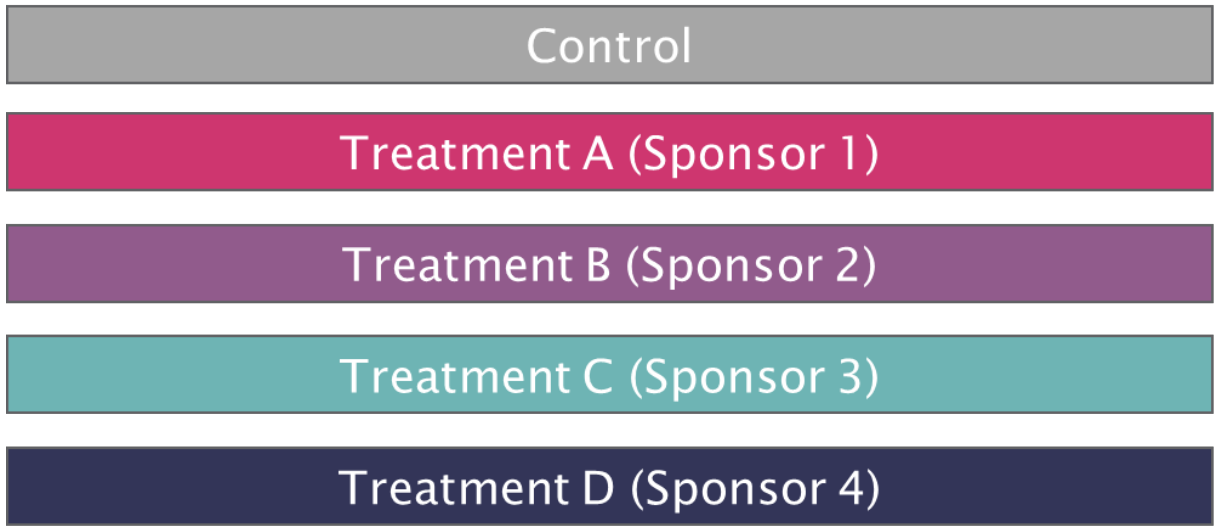


- Due to different allocation and alpha propagation in case of rejections the procedures can quite differ in terms of power in the context of platform trials (see Robertson et al. 22 available at https://arxiv.org/abs/2202.03838)

Summary Online FDR Control

- Online FDR fits exploratory platform trials
- The upper bound for the number of treatments has a strong impact on power.
- FDR of gsLOND was controlled in all considered scenarios.
  - Extensions
  - Optimization of initial allocation of α for online FDR procedure.
  - Use the accumulated data in an on-going platform trial to specify design aspects of new treatment arms
- Zehetmayer, S., Posch, M., & Koenig, F. (2022). Online control of the False Discovery Rate in group-sequential platform trials. *Statistical Methods in Medical Research*, *31*(12), 2470-2485. https://journals.sagepub.com/doi/pdf/10.1177/09622802221129051

Different treatments with different mechanism of action and same control

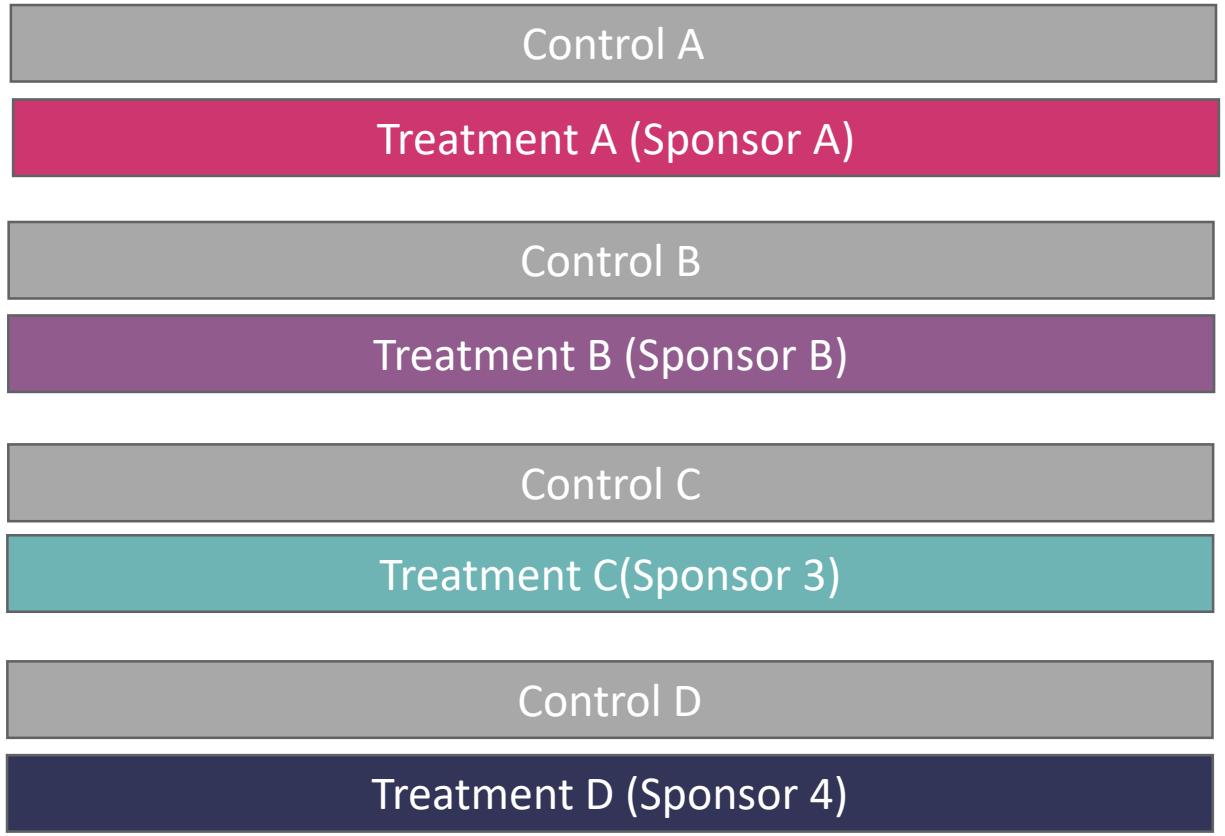


Question for participants:

Should we adjust for several treatment-control comparisons?

YES NO

Different treatments with different mechanism of action and separate controls

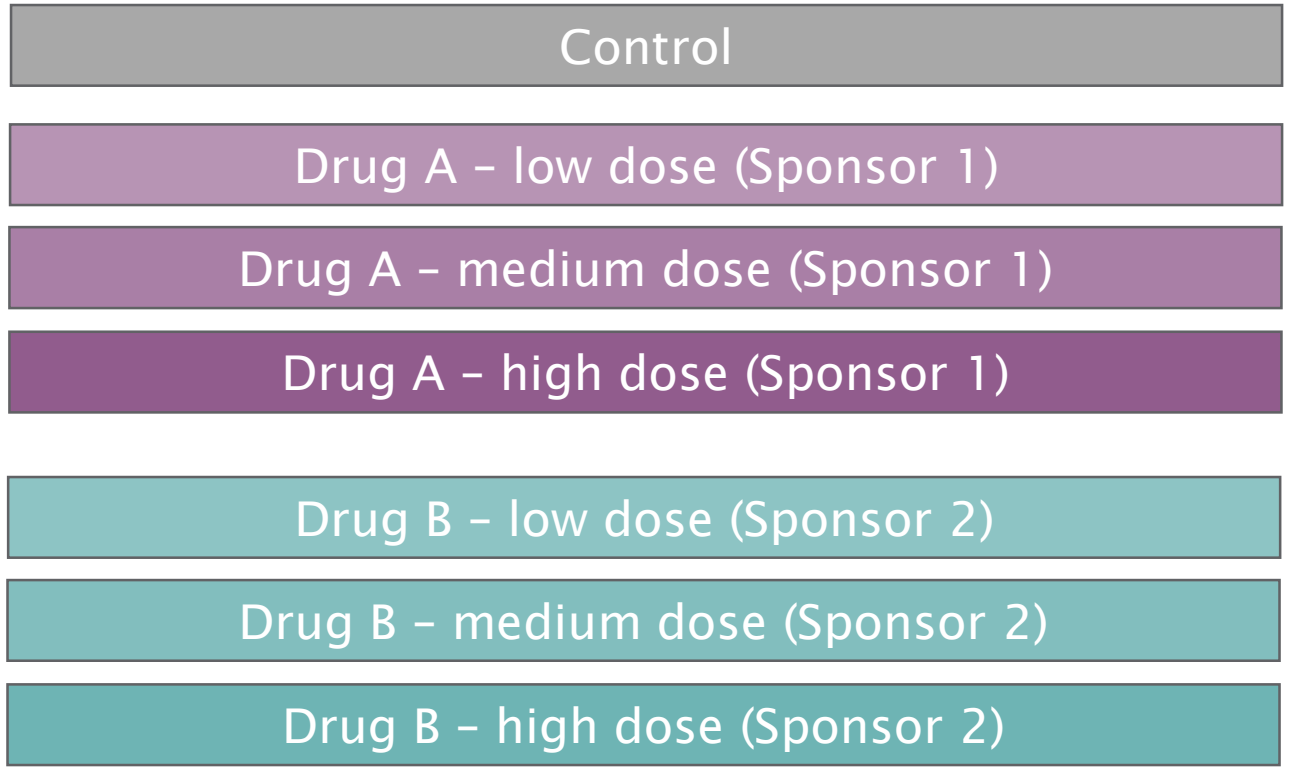


Question for participants:

Should we adjust for several treatment-control comparisons?

YES NO

Platform trial with several treatments and doses



Question:

How would you adjust for multiplicity?

Answer categories:

NO ADJUSTMENT

ADJUST FOR DOSES WITHIN EACH DRUG
(MEANS THAT FOR EACH DOSE YOU CAN SPEND FULL LEVEL ALPHA)

ADJUST FOR THE TOTAL NUMBER OF TREATMENT ARMS

Summary

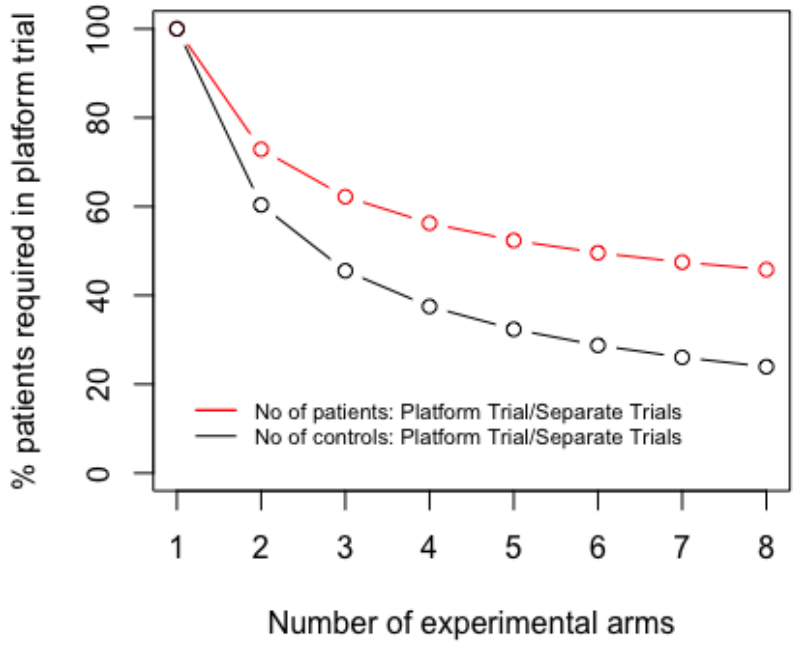
- The concept of study-wise T1E rate control is not directly applicable to platform trials, especially if they are perpetual in nature.
- Control of the FWER rate at treatment or substudy level seems to be a pragmatic approach.
- But is there a consensus on what to consider „independent“?
- Also the overall operating characteristics of the platform trial are of importance. Depending on the trial objective, control of the FDR or FWER (possibly at higher levels) are possible options.
- Other sources of multiplicity (treatments, change of control arms, subgroups, multiple endpoints, interim analysis, adaptations…) and sources of bias (non-concurrent controls, adaptations) need to be taken into account.

Shared and Non-Concurrent Controls

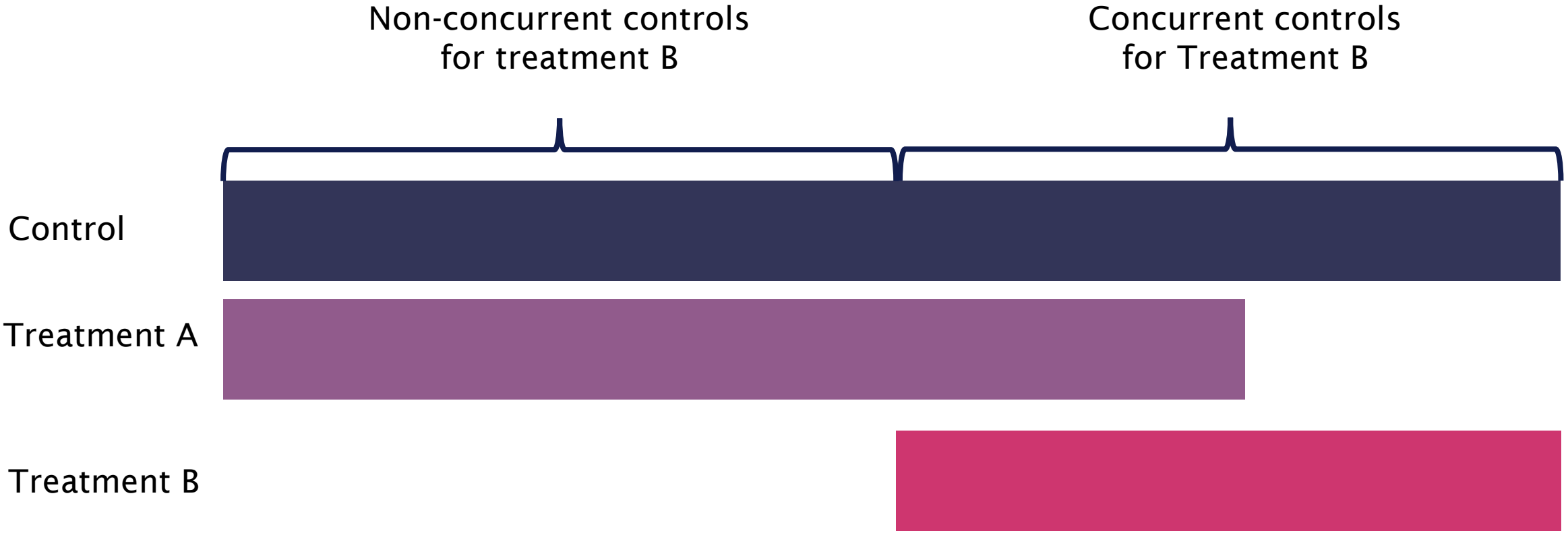
Fewer Control Patients due to Shared Controls

- Classical development for k treatments: k separate trials with 1:1 randomisation and sample size to reach pairwise power 1- β (assume equal treatment effects)

- Multi-armed trial with allocation ratio 1:1:…:1:Sqrt(*k*) (minimizing the overall sample size) and sample size to reach pairwise power 1- β



Can we use ALL control data, which is ALREADY available?



- If platform trials run over a long time period, with multiple treatments entering and leaving the platform over time, incorporating non-concurrent controls can substantially improve the efficiency
- However, non-concurrent controls may introduce bias due to different types of time trends

Non-Concurrent controls = Historical controls in RCT?

Non-concurrent and historical controls share several sources of potential bias
When using historical data for comparisons in clinical trials we accept that strict T1E control is not possible.

Eichler et al. 2016

So in platform trials?

Non-concurrent controls…

- are collected within a framework which has many features standardized (same infrastructure, assessment of endpoints, monitoring, …) and all changes are well documented.
- patients are randomized and blinding is possible

Randomized controlled trials & non-concurrent controls

- Non-concurrent controls can be randomized & blinded but
  - At a **different calendar tim**e such that randomization **does not ensure control on the distribution of prognostic factors** between NCC and experimental arms.
  - patients & investigators **are not blinded with respect to the experimental treatment and the non-concurrent control** it is compared to
- The lack of true randomization can induce time trends

Time Trends due to External and Internal Factors

- **External,** e.g.,
  - Changes in standard of care
  - Patient population
  - Pandemics
- **Internal**
  - Change in **recruiting centers**: an analysis stratified by center is no longer possible if centers enter or leave the platform.
  - Change in **recruitment strategies**, e.g. if promising treatments enter the platform.
  - Change in **inclusion/exclusion criteria** because of other experimental treatments under investigation
  - Change in **assessment of endpoints** (e.g., new diagnostic devices)

Analysis methods for trials with non-concurrent controls

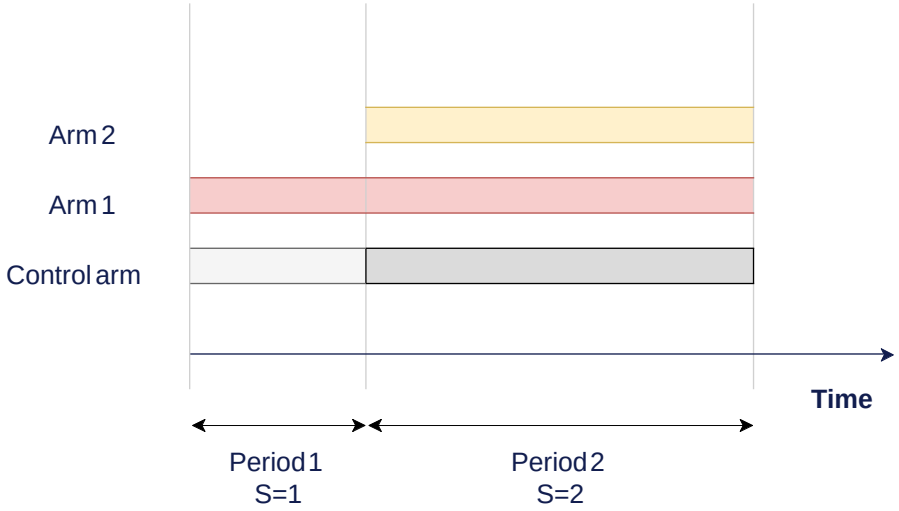
- **Separate approach**: Analysis using only concurrent controls.
- **Pooled approach**: Analysis using concurrent and non-concurrent controls.
- **Model-based approach**: Adjusts for time trends by including time as a covariate in a regression model.

Lee, K. M., & Wason, J. (2020). Including non-concurrent control patients in the analysis of platform trials: is it worth it? *BMC Medical Research Methodology*.

Bofill Roig, M., et al. (2022). On model-based time trend adjustments in platform trials with non-concurrent controls. *BMC Medical Research Methodology*

Saville, B, et al. (2022). The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials. *Clinical Trials*

Frequentist regression methods



Hypothesis testing problem:

$$H_0 : \theta_2 = 0$$
$$H_1 : \theta_2 > 0$$

Model-based approach based on data from all treatment arms and control:

$$E(Y) = \underbrace{\eta_0}_{\text{Control response}} + \underbrace{\sum_{k=1,2} \theta_k \cdot I(T = k)}_{\text{Treatment effects}} + \underbrace{\tau \cdot I(S = 2)}_{\text{Period time effect}}$$

where Y is the outcome, $T = 0, 1, 2$ denotes the treatment and $S = 1, 2$ the period.

Underlying assumptions and properties for the tests

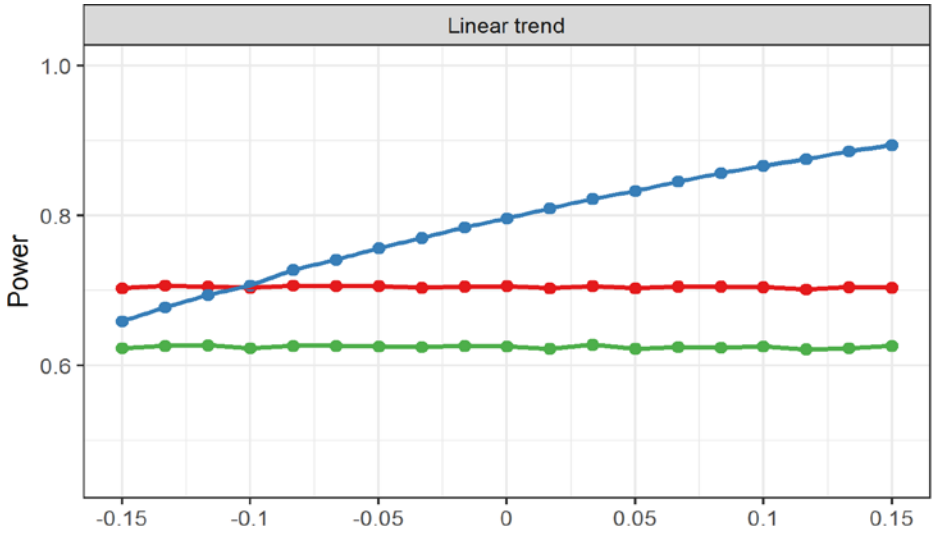
For platform trials without interim analyses or other interactive elements, this model-based approach leads to a valid treatment effect estimator regardless of the functional form of the time trend, if

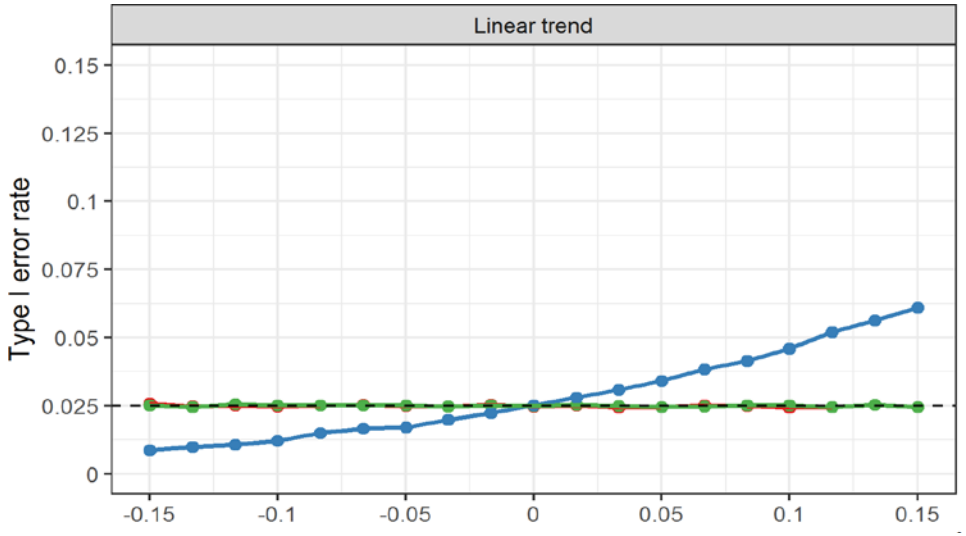
- the time trends in all treatment arms are **equal**
- the time trends are **additive** on the model scale

If block randomization is performed, the corresponding hypothesis test controls under the above assumptions (asymptotically) the type 1 error rate and can substantially improve the power.

Can we use all data?
Problem: Naively pooling control data can lead to error!

Example: 2 experimental arms and a control
Power and type 1 error rate as function of the strength of the linear time trend





- **Separate analysis** using only concurrent controls
- **Pooled analysis** using concurrent and non-concurrent controls
- **Regression model** adjusts for time trends in the model

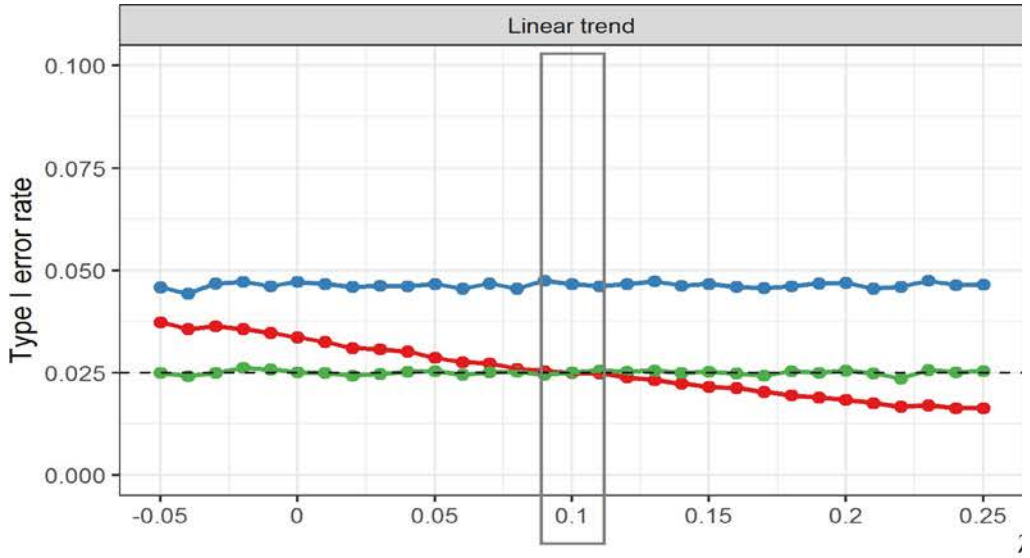
A solution: Bofill et al. (2022):
unbiased treatment effects regardless of the functional form of the time trend if *time trends in all treatment arms are equal and time trends are additive*

EU-PEARL webinar:
https://eu-pearl.eu/workshops/non-concurrent-controls-in-platform-trials/
https://www.youtube.com/watch?v=nYl-lHtVwxA

Analysis approach: Regression model Pooled analysis Separate analysis

T1E for treatment arm 2 (different time trends in groups 1 and 2)

T1E as function of the strength of the time trend λ_1 in arm 1:



However, if time trends differ between treatment arms, estimates may be biased and the type 1 error rate may be inflated.

Bofill Roig, M. B., Krotka, P., Burman, C. F., Glimm, E., Gold, S. M., Hees, 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.

Methods to incorporate non-concurrent controls Tweetorials on by Kert Viele

https://twitter.com/KertViele/status/1562118461157003266

- Frequentist model-based approaches

  https://twitter.com/KertViele/status/1562542200814088192

  Bofill Roig, M. B., Krotka, P., Burman, C. F., Glimm, E., Gold, S. M., Hees, 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.

- Bayesian Time Machine

  https://twitter.com/KertViele/status/1563163753633366016

  Saville, B. R., Berry, D. A., Berry, N. S., Viele, K., & Berry, S. M. (2022). The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials. *Clinical Trials*

- Network meta-analyses

  https://twitter.com/KertViele/status/1563163830225862656

  Marschner, I. C., & Schou, I. M. (2022). Analysis of adaptive platform trials using a network approach. *Clinical Trialsc*

What if previous control data is known when new treatments enter the platform?

- If arms have already left the platform and are published the outcome data from the respective control group is known
- A platform trial with a **control with a random low** in the outcome can be an incentive for sponsors
  - to join the platform
  - to plan an analysis including non-concurrent controls
- Conversely, a platform trial with a **control with a random high** can be
  - a deterrent to join the platform
  - a deterrent to plan for an analysis including non-concurrent controls
- However, making such decisions dependent on the trial data introduces bias!

Summary non-concurrent controls

- Inclusion of non-concurrent controls is a question of variance – bias tradeoff.
- Methods to address potential bias are available, however, they rely on specific assumptions.
- The problem of (the lack of) pre-specification is difficult to address. Keeping control data blinded may not be possible if treatment arms are stopped and results are reported.
- If non-concurrent data are utilized as primary analysis, also the analysis using only concurrent control data should be presented (possibly with a relaxed significance level)

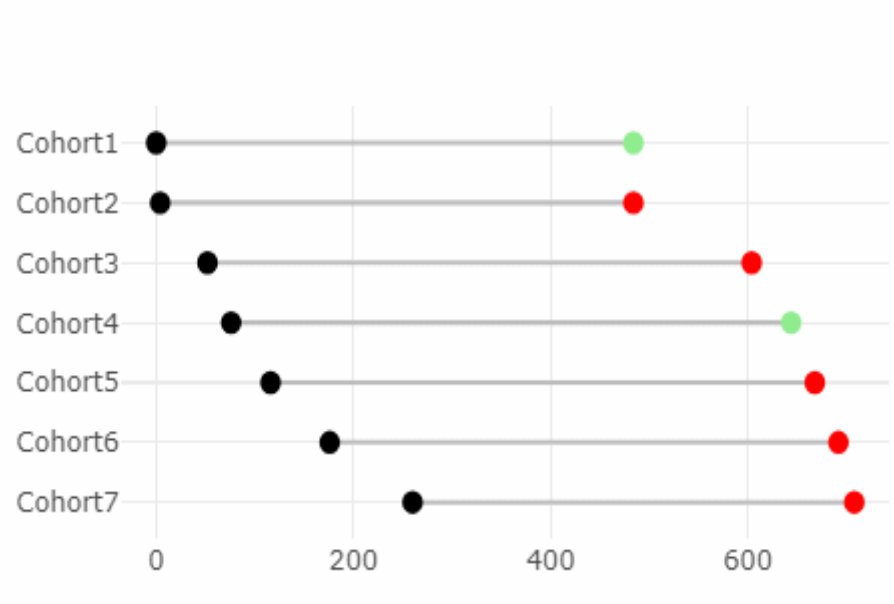
Role of Clinical Trial Simulations

Role of clinical trial simulations

- Platform trials are complex
- Analytic solutions to evaluate OCs (T1E, power) often not available
- Questions in itselfs
  - evaluating of type 1 error via simulations
    - Set of investigated scenarios sufficient? Realistic assumptions and rules?
  - Use of non-concurrent control data
    - Strict type 1 error control (adjusted or not) not possible when using external data (Kopp-Schneider et al., 2020)
- For the acceptance of simulation based methods agreement on „good simulation practices“ needed and validated software

Aims of simulation studies to explore OCs

- **Simulate realistic platform trial trajectories** (a priori timing of analyses, final sample sizes, allocation ratios over time, final number of arms etc. is not known as trial evolves dynamically over time)
- **Compute sensible operating characteristics** that reflect both the interest of sponsors (per-arm operating characteristics) and consortium that runs platform trial (per-platform operating characteristics)
- Be able to **investigate multiple assumptions simultaneously** (e.g. sample sizes, likelihood of new arms entering over time, quality of short-term endpoints at interim, different types of data sharing, treatment effects, etc.)



Developing a master protocol with clincial trial simulations

Hardest is the start => So start with something

The Vanilla Design provides the base – the agreed basic core of the design: the endpoint, phase of design, types of treatment.

To this design we identify a number of design options that could added: interims, adaptive allocation, treatment combinations, early endpoints, etc. etc.

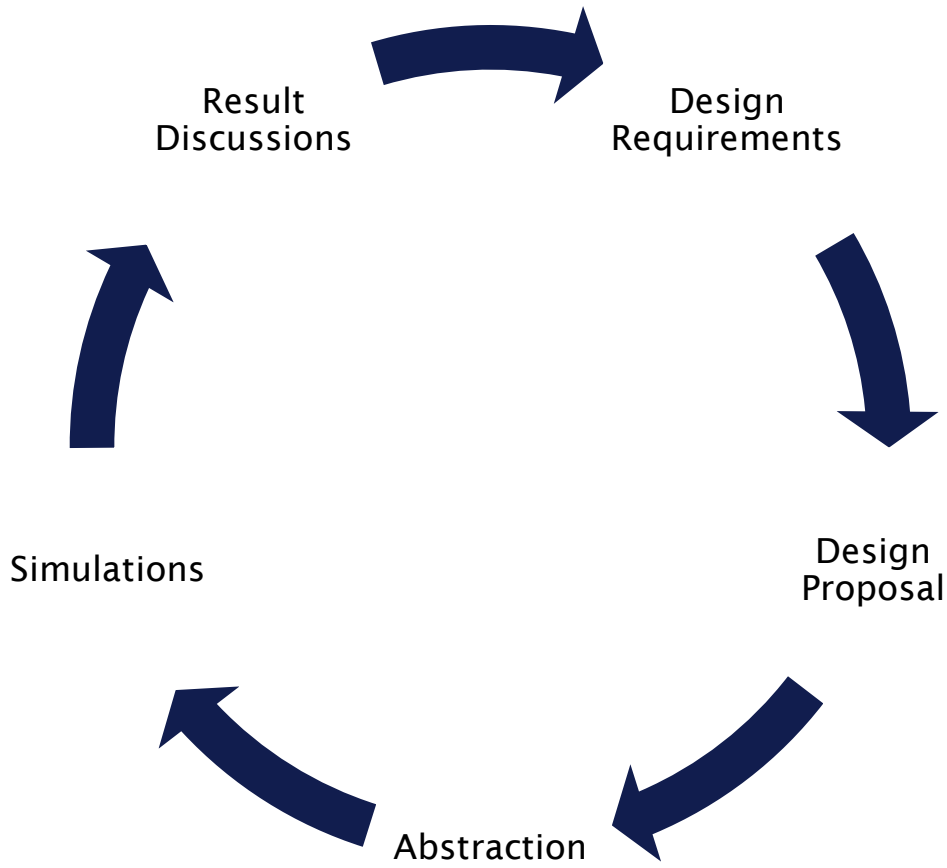
These are easier to assess and prioritise once we have an agreed basic design to add them to.

We can combine them.

But lets not try to go too far....

Iterative Process

Discussions with multiple stakeholders to understand the research problem and design needs



Based on results of simulated trials, stakeholders will identify new design requirements and scenarios

… that's how sprinkles are added

Meyer et al. Trials (2021) 22:183
https://doi.org/10.1186/s13063-021-05130-x

Trials

REVIEW Open Access

Systematic review of available software for multi-arm multi-stage and platform clinical trial design

Elias Laurin Meyer[1], Peter Mesenbrink[2], Tobias Mielke[3], Tom Parke[4], Daniel Evans[5], and Franz König[1*] on behalf of EU-PEARL (EU Patient-cEntric clinicAl tRial pLatforms) Consortium

		Commercial			Open Source Packages				Shiny Apps	
Feature	Description	FACTS	ADDPLAN	EAST	OCTOPUS	nstage	MAMS	asd	HECT	MD Anderson
Staggered Entry	Options regarding the staggered entry of treatments over time, such as pre-planned, randomly, replacing treatments, ...	x	x	x	~	x	x	x	x	(✓)
Surrogate Endpoint	Option to use different endpoint at interim	✓	(✓)	x	~	✓	✓	✓	✓	x
Flexibility Data Sharing	Advanced options for data sharing, such as sharing control data, sharing only concurrent control data, dynamic borrowing, etc.	x	x	x	(✓)	x	x	x	x	x

- Online Shiny Apps:
  - HECT (mtek.shinyapps.io/hect/)
  - MD Anderson Cancer Software Collection (trialdesign.org)
- Most software aimed at MAMS trials, hence lack typical platform trial features

EU-PEARL simulation software developed for

- **Depression**, **NASH**, **TB**, **NF**
- **Methods:** NCC, FDR, Allocation

Freely available

- CRAN, Github and paper supplements
  e.g. https://github.com/MartaBofillRoig/NCC_timetrends
  https://github.com/pavlakrotka/NCC
- https://github.com/el-meyer/simple

Generic simulators

Functional specifications to generic simulator

- SIMPLE TB Simulator
- Rshiny App for visualization

Conclusion

Conclusion

- Better use of resources versus traditional parallel group design
- Operational and statistical advantages, but are also more challenging
- In master protocols it may not be necessary to adjust for all potential sources of multiplicity
  - Control of the T1E rate at treatment or substudy level seems to be a pragmatic approach in platform trial
- Be transparent when using non-concurrent controls
  - may improve the trial's efficiency while decreasing the sample size
  - but can introduce bias due to time trends if not adequately adjusted for
  - Needs early discussions with regulators
- Use tailored methodology to improve efficiency of platform trials
  - Adaptive interim analyses
  - Tailored decision rules
  - Test using clinical trial simulations

Thank you very much for your attention!

References Multiplicity in Platform Trials

- Zehetmayer, Sonja, Martin Posch, and Franz Koenig. "Online control of the False Discovery Rate in group-sequential platform trials." *arXiv preprint arXiv:2112.10619* (2021) (to appear in SMMR).
- Robertson, David S., and James Wason. "Online control of the false discovery rate in biomedical research." *arXiv preprint arXiv:1809.07292* (2018).
- Robertson, D. S., Wason, J., König, F., Posch, M., & Jaki, T. (2022). Online error control for platform trials. *arXiv preprint arXiv:2202.03838*.
- 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.
- Collignon, O., Gartner, C., Haidich, A. B., James Hemmings, R., Hofner, B., Pétavy, F., Posch, M.,... & 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.
- Collignon, O., Burman, C. F., Posch, M., & Schiel, A. (2021). Collaborative platform trials to fight COVID-19: methodological and regulatory considerations for a better societal outcome. *Clinical Pharmacology & Therapeutics*, *110*(2), 311-320.
- Stallard, N., Todd, S., Parashar, D., Kimani, P. K., & Renfro, L. A. (2019). On the need to adjust for multiplicity in confirmatory clinical trials with master protocols. *Annals of Oncology*, *30*(4), 506-509.
- Bretz, F., & Koenig, F. (2020). Commentary on parker and weir. Clinical Trials, 17(5), 567-569.
- Sridhara, R., Marchenko, O., Jiang, Q., Pazdur, R., Posch, M., Redman, M., ... & Binkowitz, B. (2021). Type I error considerations in master protocols with common control in oncology trials: report of an American statistical association biopharmaceutical section open forum discussion. *Statistics in Biopharmaceutical Research*, 1-4.

References Non-Concurrent Controls in Platfrom Trials

- 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), 228. https://doi.org/10.1186/s12874-022-01683-w
- Bofill Roig, M., König, F., Meyer, E., & Posch, M. (2022). Commentary: Two approaches to analyze platform trials incorporating non-concurrent controls with a common assumption. *Clinical Trials*, 174077452211120. https://doi.org/10.1177/17407745221112016
- Meyer, E. L., Mesenbrink, P., Mielke, T., Parke, T., Evans, D., & König, F. (2021). Systematic review of available software for multi-arm multi-stage and platform clinical trial design. *Trials*, *22*(1), 183. https://doi.org/10.1186/s13063-021-05130-x
- Meyer, E. L., Mesenbrink, P., Dunger-Baldauf, C., Fülle, H.-J., Glimm, E., Li, Y., Posch, M., & König, F. (2020). The Evolution of Master Protocol Clinical Trial Designs: A Systematic Literature Review. *Clinical Therapeutics*, *42*(7), 1330–1360. https://doi.org/10.1016/j.clinthera.2020.05.010
- Sridhara, R., Marchenko, O., Jiang, Q., Pazdur, R., Posch, M., Berry, S., Theoret, M., Shen, Y. L., Gwise, T., Hess, L., Raven, A., Rantell, K., Roes, K., Simon, R., Redman, M., Ji, Y., & Lu, C. (2021). Use of Nonconcurrent Common Control in Master Protocols in Oncology Trials: Report of an American Statistical Association Biopharmaceutical Section Open Forum Discussion. *Statistics in Biopharmaceutical Research*, *0*(0), 1–5. https://doi.org/10.1080/19466315.2021.1938204
- Jiao, F., Tu, W., Jimenez, S., Crentsil, V., & Chen, Y. F. (2019). Utilizing shared internal control arms and historical information in small-sized platform clinical trials. *Journal of Biopharmaceutical Statistics*, *29*(5), 845–859. https://doi.org/10.1080/10543406.2019.1657132
- Bai, X., Deng, Q., & Liu, D. (2020). Multiplicity issues for platform trials with a shared control arm. *Journal of Biopharmaceutical Statistics*, *30*(6), 1077–1090. https://doi.org/10.1080/10543406.2020.1821703
- Saville, B. R., Berry, D. A., Berry, N. S., Viele, K., & Berry, S. M. (2022). The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials. *Clinical Trials*, 174077452211120. https://doi.org/10.1177/17407745221112013
- Marschner, I. C., & Schou, I. M. (2022). Analysis of adaptive platform trials using a network approach. *Clinical Trials*, 174077452211120. https://doi.org/10.1177/17407745221112001
- Lee, K. M., Brown, L. C., Jaki, T., Stallard, N., & Wason, J. (2021). Statistical consideration when adding new arms to ongoing clinical trials: the potentials and the caveats. *Trials*, *22*(1), 203.
- Lee, K. M., & Wason, J. (2020). Including non-concurrent control patients in the analysis of platform trials: is it worth it? *BMC Medical Research Methodology*, *20*(1), 165. https://doi.org/10.1186/s12874-020-01043-6

References

- Ballarini NM, Burnett T, Jaki T, Jennison C, Kӧnig F, Posch M Optimising subgroup selection in two-stage adaptive enrichment and umbrella designs. Statistics in Medicine. 2021
- Bai, X., Deng, Q., & Liu, D. (2020). Multiplicity issues for platform trials with a shared control arm. Journal of Biopharmaceutical Statistics, 30(6), 1077-1090.
- Burger, H.U. et al. (2020) The use of external controls: To what extent can it currently be recommended? Submitted (2020).
- Collignon, Olivier, et al. "Current Statistical Considerations and Regulatory Perspectives on the Planning of Confirmatory Basket, Umbrella, and Platform Trials." *Clinical Pharmacology & Therapeutics* 107.5 (2020a): 1059-1067.
- Collignon, Olivier, et al.. (2020b) Collaborative platform trials to fight COVID-19: methodological and regulatory considerations for a better societal outcome. Submitted (2020b)
- Eichler, H. G et al. (2016). “Threshold-crossing”: a useful way to establish the counterfactual in clinical trials?. *Clinical Pharmacology & Therapeutics*, *100*(6), 699-712.
- Meyer EL, Mesenbrink P, Dunger-Baldauf C, Fülle HJ, Glimm E, Li Y, Posch M, Kӧnig F. The evolution of master protocol clinical trial designs: A systematic literature review. Clinical Therapeutics 2020a. https://doi.org/10.1016/j.clinthera.2020.05.010 (open access)
- Meyer EL, Mesenbrink P, Mielke T, Parke T, Evans D, König F. Systematic review of available software for multi-arm multi-stage and platform clinical trial design. Submitted, (2020b)
- Meyer, E. L., Mesenbrink, P., Dunger-Baldauf, C., Glimm, E., Li, Y., & König, F. (2022). Decision rules for identifying combination therapies in open-entry, randomized controlled platform trials. Pharmaceutical Statistics https://onlinelibrary.wiley.com/doi/10.1002/pst.2194
- Stallard N, Todd S, Parashar D, et al. On the need to adjust for multiplicity in confirmatory clinical trials with master protocols. Ann Oncol 2019; 30(4): 506–509.
- Saville, B. R., & Berry, S. M. (2016). Efficiencies of platform clinical trials: a vision of the future. Clinical Trials, 13(3), 358-366.
- Parker, R. A., & Weir, C. J. (2020). Non-adjustment for multiple testing in multi-arm trials of distinct treatments: Rationale and justification. Clinical Trials, 1740774520941419.
  Bretz, F., & Koenig, F. (2020). Commentary on Parker and Weir. Clinical Trials, 1740774520941420.
- Wason, J. M., & Robertson, D. S. (2020). Controlling type I error rates in multi-arm clinical trials: A case for the false discovery rate. Pharmaceutical Statistics.