# Adaptations, Interim Analyses and Multiplicity in Clinical Trials - Ignoring, Bayesian, Frequentist or a little bit of everything?

Franz Koenig
Center for Medical Data Science
franz.koenig@meduniwien.ac.at

# Acknowledgements

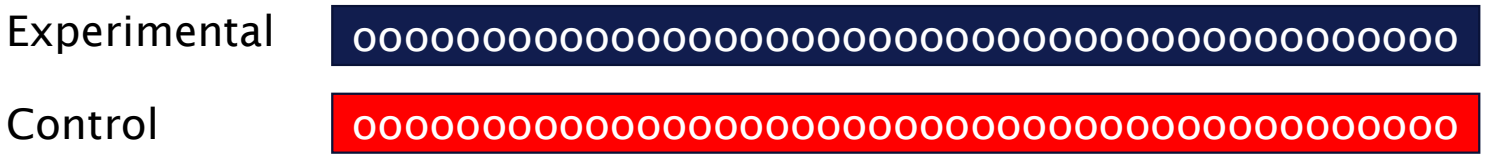

- Marta Bofill-Roig, Pavla Krotka, Elias Meyer, Martin Posch, Sonja Zehetmayer,
- EU-PEARL collaborators: Peter Mesenbrink, Dominic Magirr, Ekkehard Glimm, Tom Parke, Peter Jacko, Katherina Hess, Quynh Lan Nguyen,

EU-PEARL
EU PATIENT-CENTRIC CLINICAL TRIAL PLATFORMS

efpia imi innovative medicines initiative https://eu-pearl.eu/

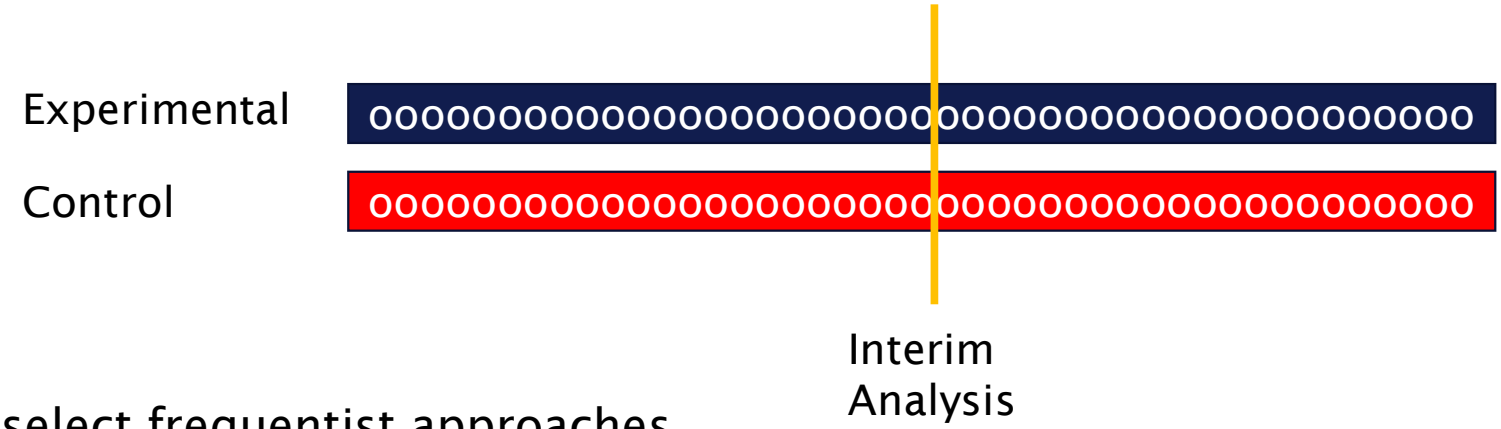
This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853966.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and CHILDREN'S TUMOR FOUNDATION, GLOBAL ALLIANCE FOR TB DRUG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

# 2-arm RCT with a single stage design




- Test of primary endpoint
  - Traditionally frequentist test
  - Is there a point of using Bayesian analysis if one would use non-informative priors?
- Testing of primary and secondary endpoints
  - Multiplicity addressed by closed testing procedures (e.g., hierarchical test)

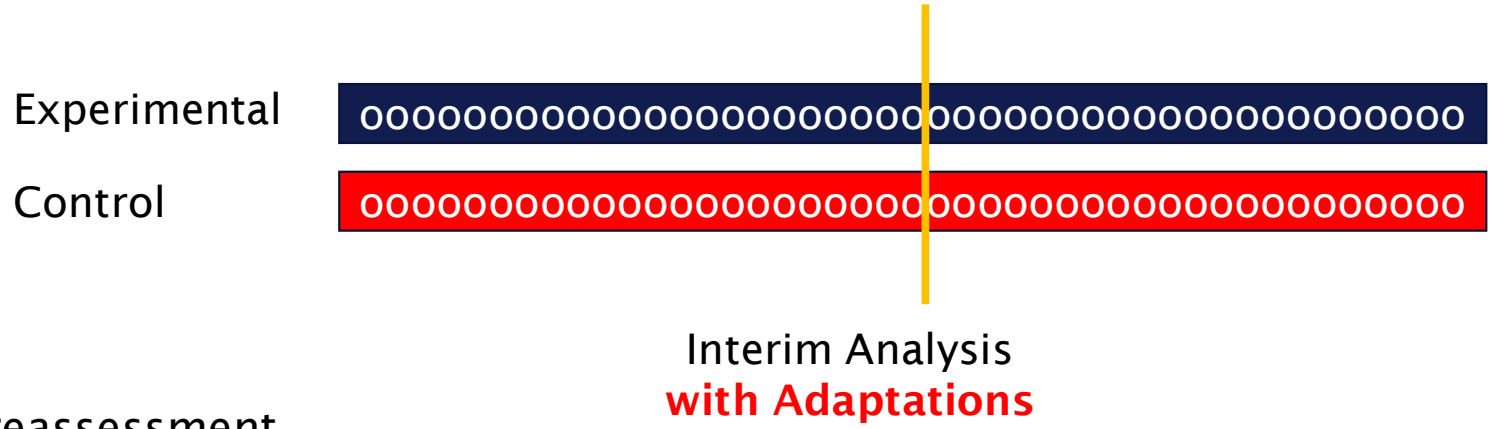
# 2-arm RCT with an interim analysis




Still many will select frequentist approaches

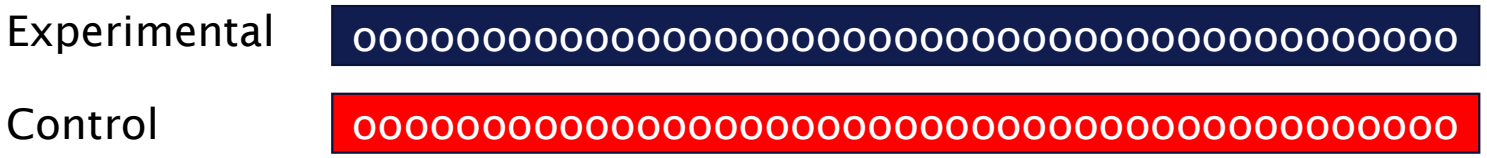
- Early stopping (efficicay / futility): group sequential designs (Pocock 77, OBF 79, Jennison & Turnbull 00, …)
- Testing of primary and secondary endpoints
  - Already more tricky, e.g., cannot simply perform a hierarchical test
  - Inflation of type 1 error if secondary endpoint is tested at full level alpha after primary reached statistical significance (e.g, Hung et al. 2007, Glimm et al. 2010)

# 2-arm RCT with an interim analysis




- Sample size reassessment
  - Blinded (e.g, see papers by Friede et al.)
  - Unblinded: Adaptive frequentist tests (e.g., Bauer et al, 2018)
    - Adaptive combination test (Bauer 89, Bauer & Köhne 99)
    - Conditional Error (Müller & Schäfer 99)
- What about change of allocation ratio?
  - If changed once in a single interim analysis, type I error can be controlled by adaptive tests

# 2-arm RCT with response adaptive randomisation

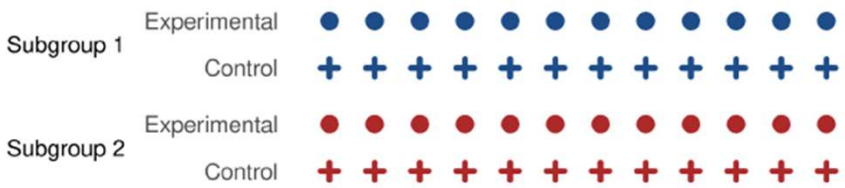



- Response Adaptive Randomisation (RAR)
  - Tricky to control type 1 error with frequentist methods
    - Strict control possible using CE principle conditioning on an artificial design (which actually will never apply)
    - Usually evaluation & calibration of type 1 error rate with simulations
    - Use of Bayesian methods seems more reasonable to comply with very adaptive nature of RAR

# Trial Designs with pre-defined subgroups

- Suppose two biomarker-defined, disjoint subgroups have been identified before starting the trial
- Test of the null hypotheses in each subgroup

$$H_{01} : \theta_1 \leq 0 \text{ and } H_{02} : \theta_2 \leq 0.$$




- Is there a need to adjust for multiplicity?
  - Because two hypotheses are tested, without adjustment the familywise error rate (FWER) will be inflated.
  - Therefore, typically a multiplicity adjustment is required
  - For umbrella (or basket trials), it has been argued that in certain settings no FWER is necessary
    - Collignon, Olivier, et al. (2020). "Current statistical considerations and regulatory perspectives on the planning of conrmatory basket, umbrella, and platform trials." Clinical Pharmacology & Therapeutics 107.5: 1059-1067.
    - Collignon, O., Posch, M., & Schiel, A. (2022). Assessment of tumour-agnostic therapies in basket trials. The Lancet Oncology, 23(1), e8
- What are optimal designs? Single stage or adaptive trials?

# Optimising trial designs under uncertainty

Received: 18 March 2020 | Revised: 11 January 2021 | Accepted: 28 February 2021
DOI: 10.1002/sim.8949

RESEARCH ARTICLE

Statistics in Medicine WILEY

**Optimizing subgroup selection in two-stage adaptive enrichment and umbrella designs**

Nicolás M. Ballarini[1] | Thomas Burnett[2] | Thomas Jaki[2,3] | Christoper Jennison[4] | Franz König[1] | Martin Posch[1]

[1]Section for Medical Statistics, Medical University of Vienna, Vienna, Austria
[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK
[3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[4]Department of Mathematical Sciences, University of Bath, Bath, UK

**Correspondence**
Martin Posch, Section for Medical Statistics, CeMSIIS, Medical University of Vienna, Vienna 1090, Austria.
Email: martin.posch@meduniwien.ac.at

**Funding information**
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 633567; Innovative Medicines Initiative, Grant/Award Number: 853966; Medical Research Council, Grant/Award Number: MR/M005755/1; National Institute for Health Research, Grant/Award Number: NIHR-SRF-2015-08-001

We design two-stage confirmatory clinical trials that use adaptation to find the subgroup of patients who will benefit from a new treatment, testing for a treatment effect in each of two disjoint subgroups. Our proposal allows aspects of the trial, such as recruitment probabilities of each group, to be altered at an interim analysis. We use the conditional error rate approach to implement these adaptations with protection of overall error rates. Applying a Bayesian decision-theoretic framework, we optimize design parameters by maximizing a utility function that takes the population prevalence of the subgroups into account. We show results for traditional trials with familywise error rate control (using a closed testing procedure) as well as for umbrella trials in which only the per-comparison type 1 error rate is controlled. We present numerical examples to illustrate the optimization process and the effectiveness of the proposed designs.

KEYWORDS
Bayesian optimization, conditional error function, subgroup analysis, utility function

- Use Bayesian techniques
- Define a gain function as a measure of overall trial performance, e.g.,

$$\mathcal{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \lambda \mathbb{1}[RejectH_{01}] + (1-\lambda)\mathbb{1}[RejectH_{02}]$$

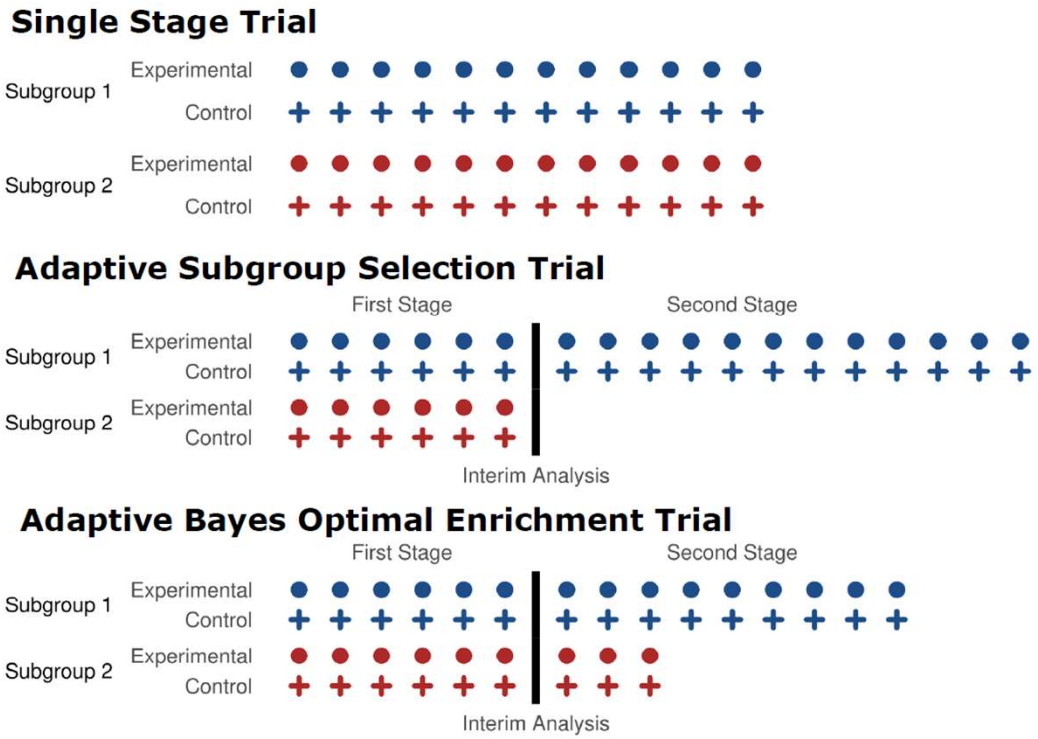
- But one needs to define prior uncertainty on the effect sizes how to design the trial

$$\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\psi_1^2 & \rho\psi_1\psi_2\\ \rho\psi_1\psi_2 & \psi_2^2\end{pmatrix}\right)$$

- Maximise the Bayes expected gain over important design parameters *a*

$$\text{argmax}_a \int_{\boldsymbol{\theta}} \int_{\hat{\boldsymbol{\theta}}} f(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, a)\mathcal{U}(\boldsymbol{\theta}, X)\mathrm{d}\hat{\boldsymbol{\theta}}\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$$

# Use Decision Theoretic Approaches for Design Optimization

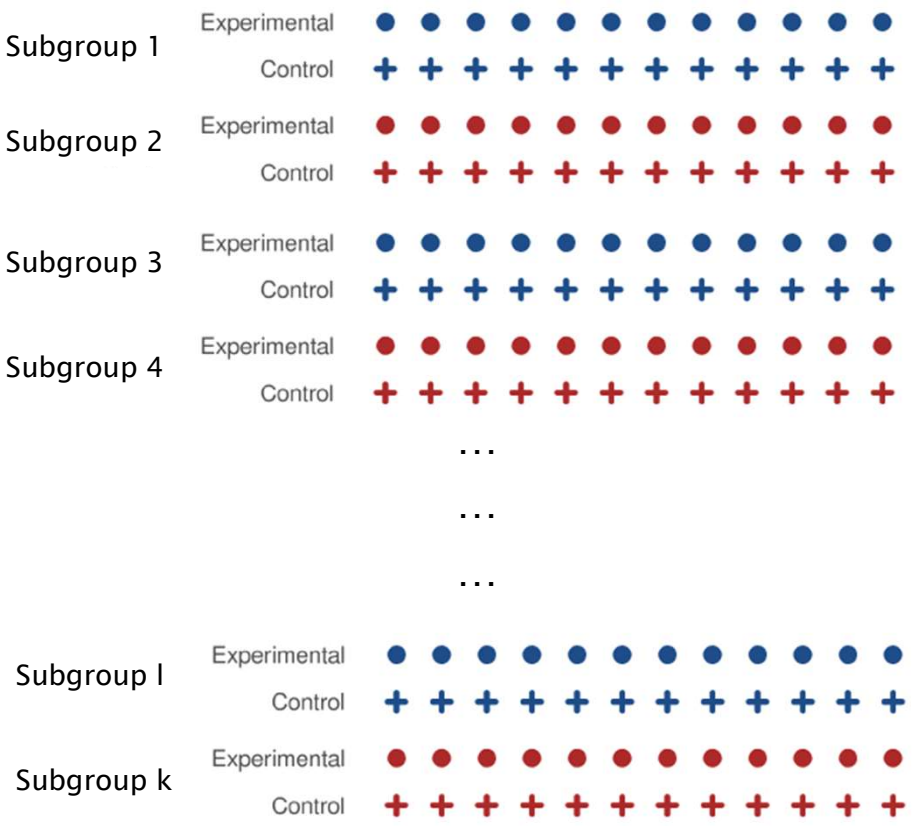



- Trial designs, sample sizes, adaptation rules and multiple testing procedures can be chosen optimal with respect to a utility function.
- Optimal trials depend on prior assumptions on the effect sizes and subgroup prevalence.
- More complex utility functions can be considered, accounting for costs, observed effect sizes, precision of estimates, the true treatment effects etc.

- **To control FWER use adaptive closed tests**
  - Koenig, F., Brannath, W., Bretz, F., & Posch, M. (2008). Adaptive Dunnett tests for treatment selection. Statistics in Medicine, 27(10), 1612-1625.
  - Bauer, Peter, et al. "Twenty-five years of confirrmatory adaptive designs: opportunities and pitfalls." Statistics in Medicine 35.3 (2016): 325-347

# What about having many baskets?

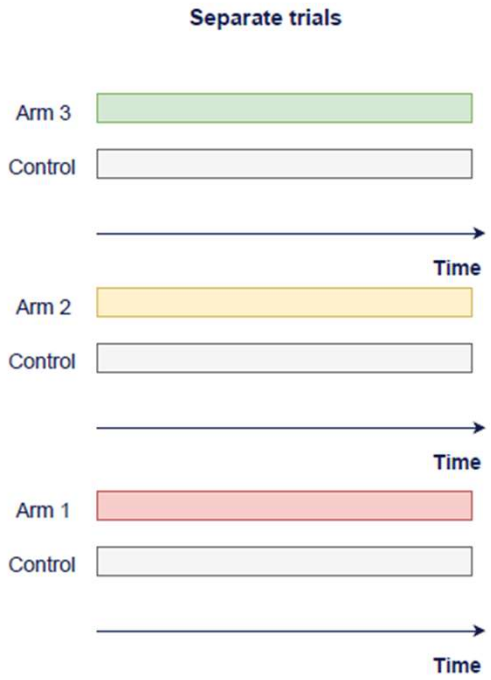
- Natural to borrow strength from „similar“ baskets
- More natural to use Bayesian approaches
- Control of type 1 error less criticial, as basket trials are mainly in exploratory phase II setting. (e.g, see Review by Meyer et al 2020)



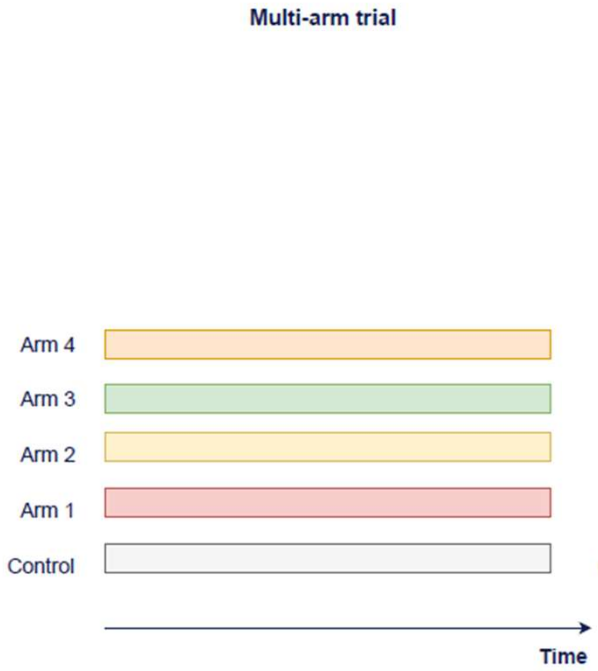
# Trial Design (Adaptive/Basket/Umbrella/Platform/...)

- Bayesian decision theoretic methods to optimize trial designs
  - Sample sizes
  - Allocation ratios
  - Endpoints and testing procedures
  - Number of treatment arms
  - Stopping rules
  - Adaptation rules
- The Bayesian approach is used only to optimize the design of the study.
- The analysis of the trial can still be frequentist (e.g., for pivotal trials) or Bayesian (e.g., Phase IIb in Basket trials).
- Too complex trials may require simulation to evaluate the operating characteristics.

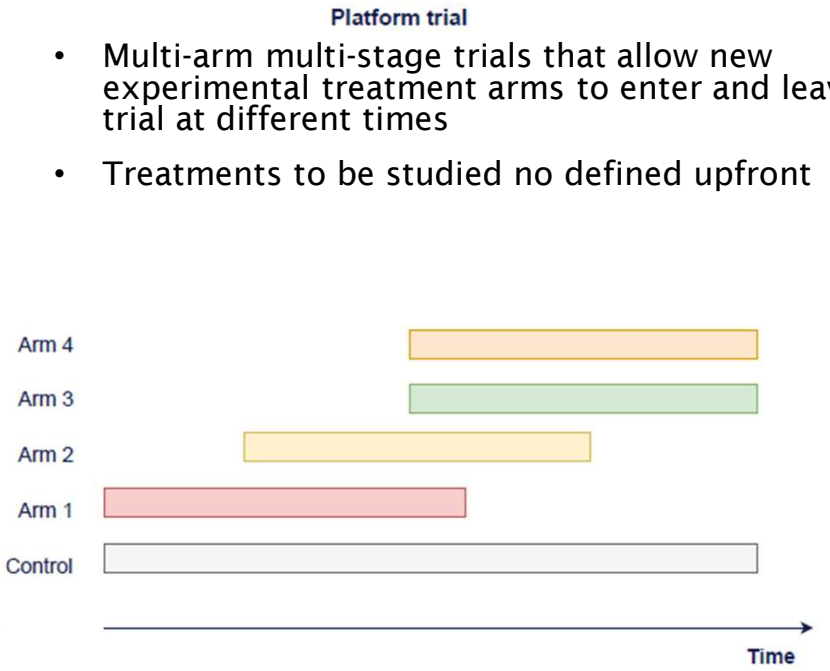
# Separate trials, multi-arm trial and platform trials




No adjustment accross studies




Traditionally adjustment e.g., Dunnett-test




- Multi-arm multi-stage trials that allow new experimental treatment arms to enter and leave the trial at different times
- Treatments to be studied no defined upfront

???

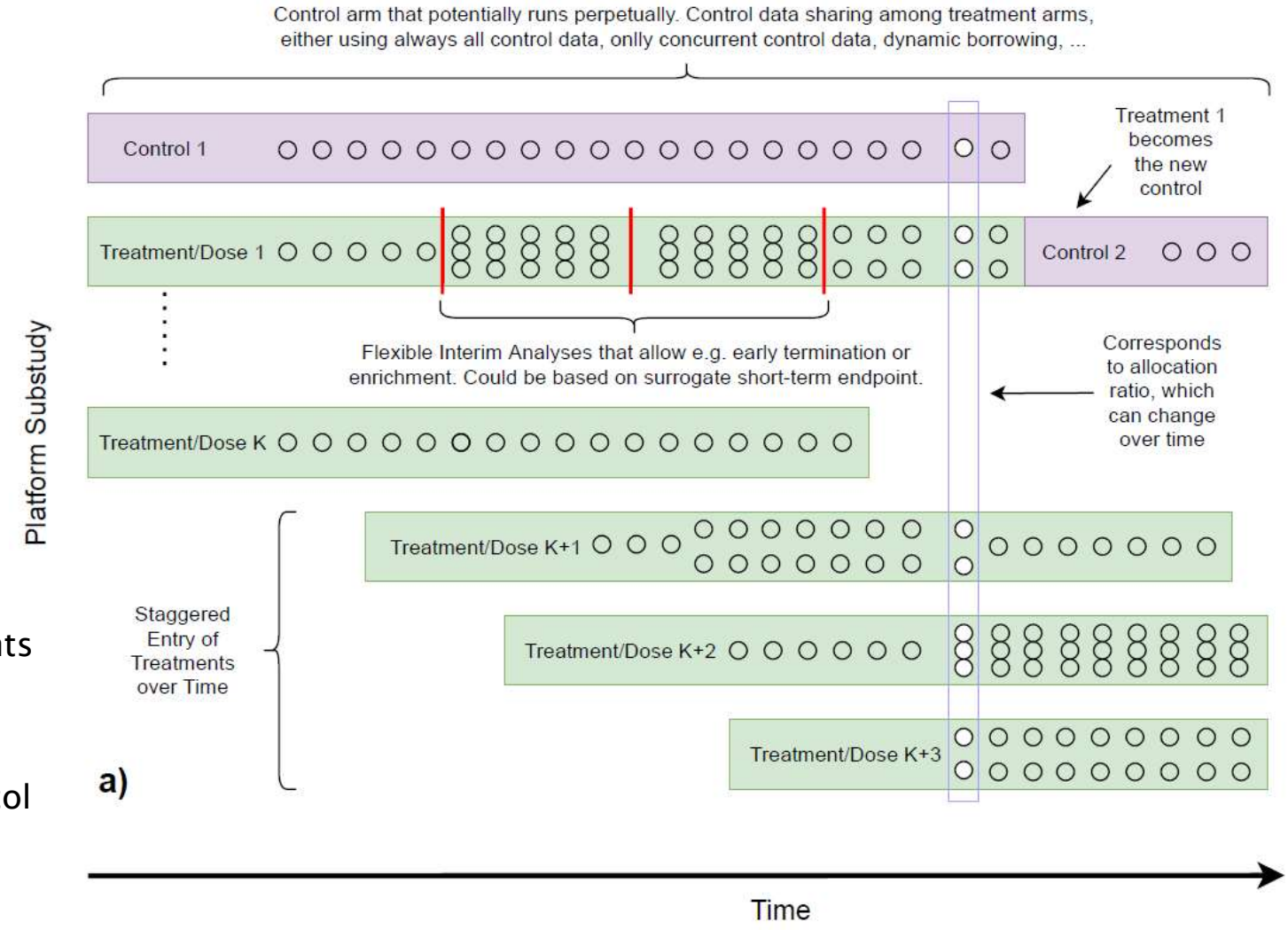
Some debate whether adjustment necessary
[Stallard et al. 2019, Collignon et al. 2020a, 2020b, Park & Weir (2020), Bretz & König (2020), Nguyen et al (2022), Koenig et al (2024) ..]

And if you want to adjust, for how many?
[Online methods, Robertson et, Zehetmayer, …]

# 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



# FWER when ignoring multiplicity & adaptations

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

**What is the most extreme T1E rate:**

- If only a single interim analysis is conducted
- And if SSR* conducted…
- In addition also adapt allocation ratio (unbalanced)
- But analysis not corrected

**Maximum type 1 error inflation:**

| nominal $\alpha$ | $k = 1$ 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*

PLOS ONE

https://doi.org/10.1371/journal.pone.0281674

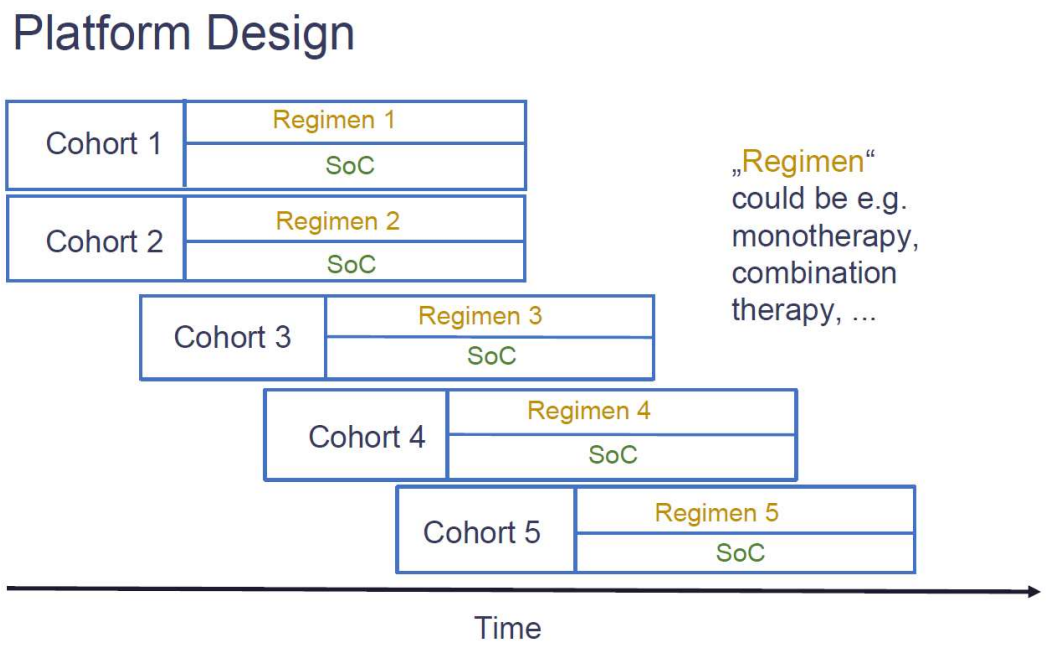
OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## Designing an exploratory phase 2b platform trial in NASH with correlated, co-primary binary endpoints

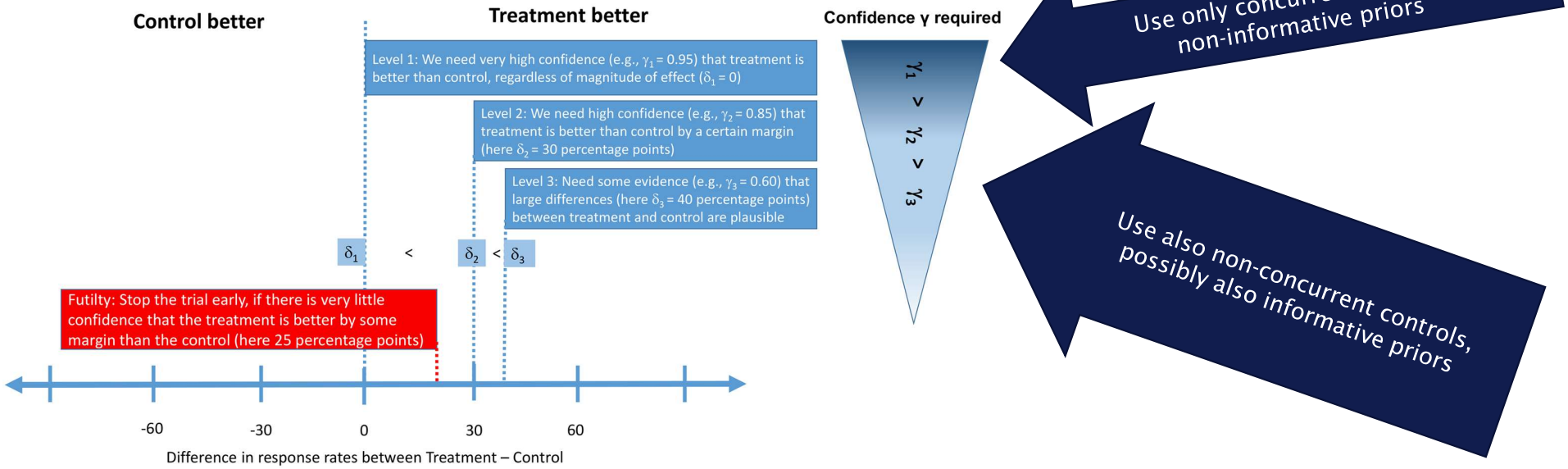
Elias Laurin Meyer, Peter Mesenbrink, Nicholas A. Di Prospero, Juan M. Pericàs, Ekkehard Glimm, Vlad Ratziu, Elena Sena, Franz König, on behalf of the EU-PEARL NASH Investigators

- Platform design with parallel cohorts each consisting of control and active therapy and option to share control data
  - Amount of concurrent control data depends on open cohorts
- Interested in two primary endpoints (NASH resolution and fibrosis improvement)
- Interim analysis (based on surrogate endpoints) for efficacy & futility
- Superiority not sufficient, interested in different level of evidence




For more information see presentation by Elias Meyer on Wednesday

# Different stakeholders interested in different levels of evidence




- Using a closed test and adaptive combination test would become more and more burdensome.
- As exploratory study control of FWER considered not as critical
- Use of Bayesian multi-level decision rules to deal with two endpoints, interim analyses, etc
- Assessment of operating characteristics via simulations
- Early discussion with both EMA and FDA (e.g, see Gidh-Jain et al 2024; Nguyen et al. 2024)

# Enriching the analysis with further data?

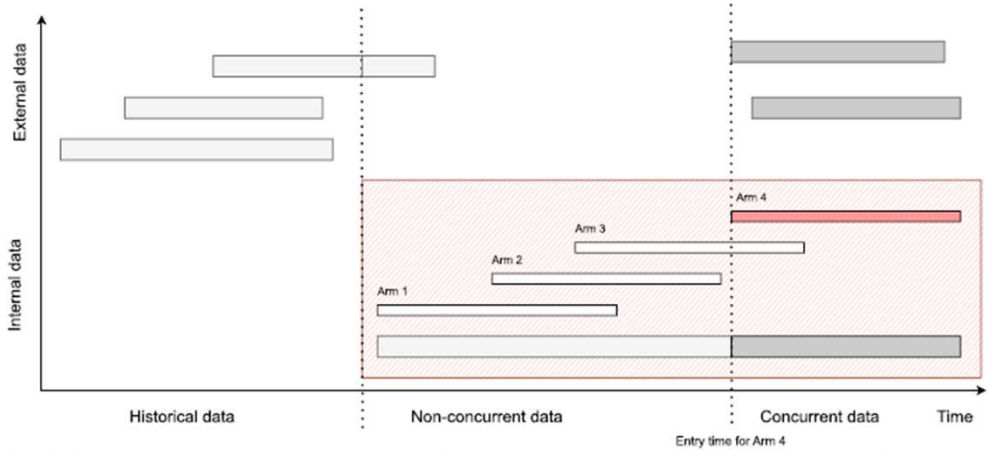



**Fig. 1** Definition of controls depending on the source and time. The data within the red box represents the (internal) data from a platform trial, the data outside the red box represents the external data. Non-concurrent control data for arm 4 is represented in light grey boxes and concurrent controls are represented in dark grey

Bofill Roig et al. Trials (2023) 24:408
https://doi.org/10.1186/s13063-023-07398-7

Trials

REVIEW Open Access

## On the use of non-concurrent controls in platform trials: a scoping review

Marta Bofill Roig[1*], Cora Burgwinkel[1,2], Ursula Garczarek[3], Franz Koenig[1], Martin Posch[1], Quynh Nguyen[2] and Katharina Hees[2*]

**Abstract**

**Background** Platform trials gained popularity during the last few years as they increase flexibility compared to multi-arm trials by allowing new experimental arms entering when the trial already started. Using a shared control group in platform trials increases the trial efficiency compared to separate trials. Because of the later entry of some of the experimental treatment arms, the shared control group includes concurrent and non-concurrent control data. For a given experimental arm, non-concurrent controls refer to patients allocated to the control arm before the arm enters the trial, while concurrent controls refer to control patients that are randomised concurrently to the experimental arm. Using non-concurrent controls can result in bias in the estimate in case of time trends if the appropriate methodology is not used and the assumptions are not met.

**Methods** We conducted two reviews on the use of non-concurrent controls in platform trials: one on statistical methodology and one on regulatory guidance. We broadened our searches to the use of external and historical control data. We conducted our review on the statistical methodology in 43 articles identified through a systematic search in PubMed and performed a review on regulatory guidance on the use of non-concurrent controls in 37 guidelines published on the EMA and FDA websites.

**Results** Only 7/43 of the methodological articles and 4/37 guidelines focused on platform trials. With respect to the statistical methodology, in 28/43 articles, a Bayesian approach was used to incorporate external/non-concurrent controls while 7/43 used a frequentist approach and 8/43 considered both. The majority of the articles considered a method that downweights the non-concurrent control in favour of concument control data (34/43), using for instance meta-analytic or propensity score approaches, and 11/43 considered a modelling-based approach, using regression models to incorporate non-concurrent control data. In regulatory guidelines, the use of non-concurrent control data was considered critical but was deemed acceptable for rare diseases in 12/37 guidelines or was accepted in specific indications (12/37). Non-comparability (30/37) and bias (16/37) were raised most often as the general concerns with non-concurrent controls. Indication specific guidelines were found to be most instructive.

**Conclusions** Statistical methods for incorporating non-concurrent controls are available in the literature, either by means of methods originally proposed for the incorporation of external controls or non-concurrent controls in platform trials. Methods mainly differ with respect to how the concurrent and non-concurrent data are combined

# Enriching the analysis with further data?


**Fig. 1** Definition of controls depending on the source and time. The data within the red box represents the (internal) data from a platform trial, the data outside the red box represents the external data. Non-concurrent control data for arm 4 is represented in light grey boxes and concurrent controls are represented in dark grey

Bofill Roig et al. Trials (2023) 24:408
https://doi.org/10.1186/s13063-023-07398-7

Trials

REVIEW Open Access

## On the use of non-concurrent controls in platform trials: a scoping review

Marta Bofill Roig[1*], Cora Burgwinkel[1,2], Ursula Garczarek[3], Franz Koenig[1], Martin Posch[1], Quynh Nguyen[2] and Katharina Hees[2*]

**Abstract**

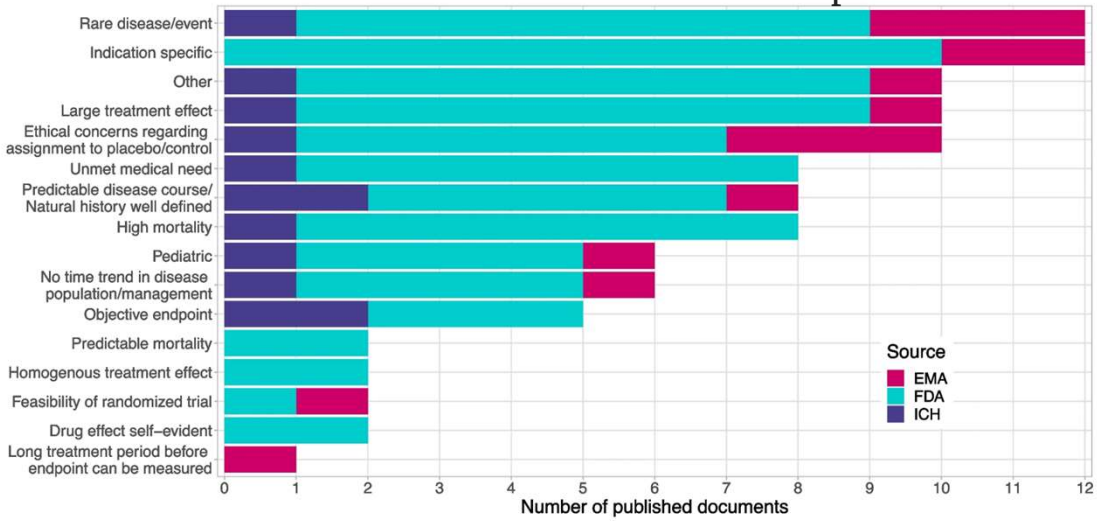
**Background** Platform trials gained popularity during the last few years as they increase flexibility compared to multi-arm trials by allowing new experimental arms entering when the trial already started. Using a shared control group in platform trials increases the trial efficiency compared to separate trials. Because of the later entry of some of the experimental treatment arms, the shared control group includes concurrent and non-concurrent control data. For a given experimental arm, non-concurrent controls refer to patients allocated to the control arm before the arm enters the trial, while concurrent controls refer to control patients that are randomised concurrently to the experimental arm. Using non-concurrent controls can result in bias in the estimate in case of time trends if the appropriate methodology is not used and the assumptions are not met.

**Methods** We conducted two reviews on the use of non-concurrent controls in platform trials: one on statistical methodology and one on regulatory guidance. We broadened our searches to the use of external and historical control data. We conducted our review on the statistical methodology in 43 articles identified through a systematic search in PubMed and performed a review on regulatory guidance on the use of non-concurrent controls in 37 guidelines published on the EMA and FDA websites.

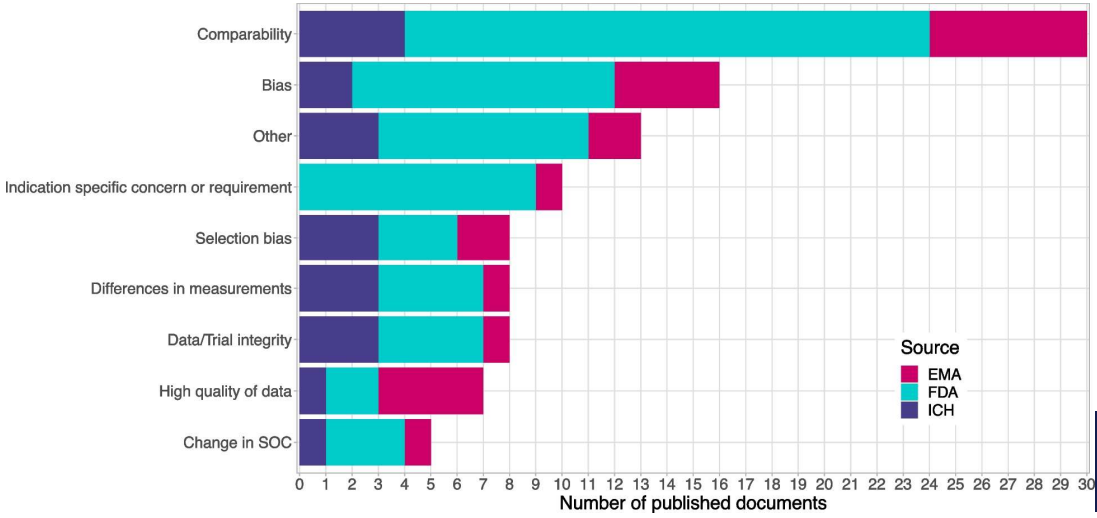
**Results** Only 7/43 of the methodological articles and 4/37 guidelines focused on platform trials. With respect to the statistical methodology, in 28/43 articles, a Bayesian approach was used to incorporate external/non-concurrent controls while 7/43 used a frequentist approach and 8/43 considered both. The majority of the articles considered a method that downweights the non-concurrent control in favour of concurrent control data (34/43), using for instance meta-analytic or propensity score approaches, and 11/43 considered a modelling-based approach, using regression models to incorporate non-concurrent control data. In regulatory guidelines, the use of non-concurrent control data was considered critical but was deemed acceptable for rare diseases in 12/37 guidelines or was accepted in specific indications (12/37). Non-comparability (30/37) and bias (16/37) were raised most often as the general concerns with non-concurrent controls. Indication specific guidelines were found to be most instructive.

**Conclusions** Statistical methods for incorporating non-concurrent controls are available in the literature, either by means of methods originally proposed for the incorporation of external controls or non-concurrent controls in platform trials. Methods mainly differ with respect to how the concurrent and non-concurrent data are combined

## Circumstances in which the use of external/historical/non-concurrent control is recommended or deemed acceptable

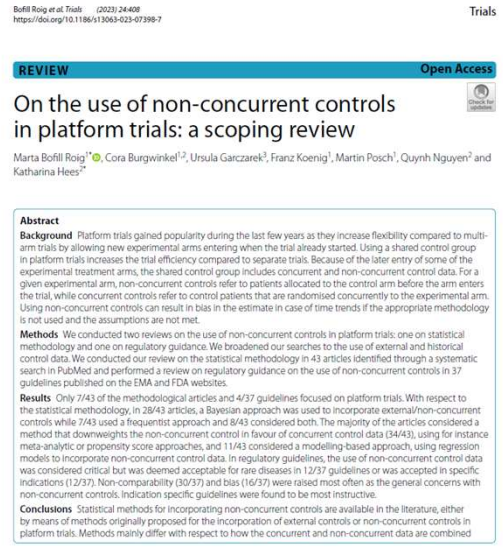



## Concerns raised with the use of historical/external/non-concurrent controls



# Enriching the analysis with further data


**Fig. 1** Definition of controls depending on the source and time. The data within the red box represents the (internal) data from a platform trial, the data outside the red box represents the external data. Non-concurrent control data for arm 4 is represented in light grey boxes and concurrent controls are represented in dark grey


Bofill Roig et al. Trials (2023) 24:408
https://doi.org/10.1186/s13063-023-07398-7

Trials

REVIEW Open Access

On the use of non-concurrent controls in platform trials: a scoping review

Marta Bofill Roig[1*], Cora Burgwinkel[1,2], Ursula Garczarek[3], Franz Koenig[1], Martin Posch[1], Quynh Nguyen[2] and Katharina Hees[2*]

**Abstract**

**Background** Platform trials gained popularity during the last few years as they increase flexibility compared to multi-arm trials by allowing new experimental arms entering when the trial already started. Using a shared control group in platform trials increases the trial efficiency compared to separate trials. Because of the later entry of some of the experimental treatment arms, the shared control group includes concurrent and non-concurrent control data. For a given experimental arm, non-concurrent controls refer to patients allocated to the control arm before the arm enters the trial, while concurrent controls refer to control patients that are randomised concurrently to the experimental arm. Using non-concurrent controls can result in bias in the estimate in case of time trends if the appropriate methodology is not used and the assumptions are not met.

**Methods** We conducted two reviews on the use of non-concurrent controls in platform trials: one on statistical methodology and one on regulatory guidance. We broadened our searches to the use of external and historical control data. We conducted our review on the statistical methodology in 43 articles identified through a systematic search in PubMed and performed a review on regulatory guidance on the use of non-concurrent controls in 37 guidelines published on the EMA and FDA websites.

**Results** Only 7/43 of the methodological articles and 4/37 guidelines focused on platform trials. With respect to the statistical methodology, in 28/43 articles, a Bayesian approach was used to incorporate external/non-concurrent controls while 7/43 used a frequentist approach and 8/43 considered both. The majority of the articles considered a method that downweights the non-concurrent control in favour of concurrent control data (34/43), using for instance meta-analytic or propensity score approaches, and 11/43 considered a modelling-based approach, using regression models to incorporate non-concurrent control data. In regulatory guidelines, the use of non-concurrent control data was considered critical but was deemed acceptable for rare diseases in 12/37 guidelines or was accepted in specific indications (12/37). Non-comparability (30/37) and bias (16/37) were raised most often as the general concerns with non-concurrent controls. Indication specific guidelines were found to be most instructive.

**Conclusions** Statistical methods for incorporating non-concurrent controls are available in the literature, either by means of methods originally proposed for the incorporation of external controls or non-concurrent controls in platform trials. Methods mainly differ with respect to how the concurrent and non-concurrent data are combined

- Methods to incorporate external / non-concurrent controls
  - Test-then-pool approaches; Frequentist and Bayesian regression model approaches; Propensity score approaches and baseline covariates-adjustments; Power prior and commensurate power prior; Hierarchical models; Elastic prior
- Not clear how to incorporate external data with frequentist method if strict control of type 1 error rate required
  - Some will depend on certain assumptions
- Even more lost with interim analyses
- More natural place for using Bayesian methods
  - However, no strict control of type 1 error rate (Kopp-Schneider et al. 2018)

# What about using only data from the platform

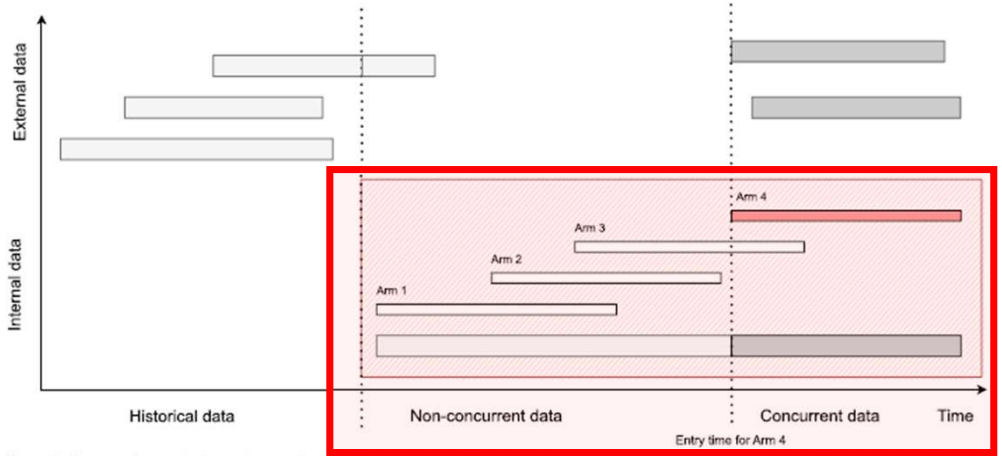



**Fig. 1** Definition of controls depending on the source and time. The data within the red box represents the (internal) data from a platform trial, the data outside the red box represents the external data. Non-concurrent control data for arm 4 is represented in light grey boxes and concurrent controls are represented in dark grey

"Treatment-control comparisons in platform trials including non-concurrent controls". (2024). M. Bofill Roig, P. Krotka, K. Hess, F. Koenig, D. Magirr, P. Jacko, T. Parke, and M. Posch.

"On model-based time trend adjustments in platform trials with non-concurrent controls". (2022). M. Bofill Roig, P. Krotka, CF. Burman, E. Glimm, K. Hess, P. Jacko, F. Koenig, D. Magirr, P. Mesenbrink, K. Viele, and M. Posch. BMC Medical Research Methodology https://doi.org/10.1186/s12874-022-01683-w

# Inference in model based approaches

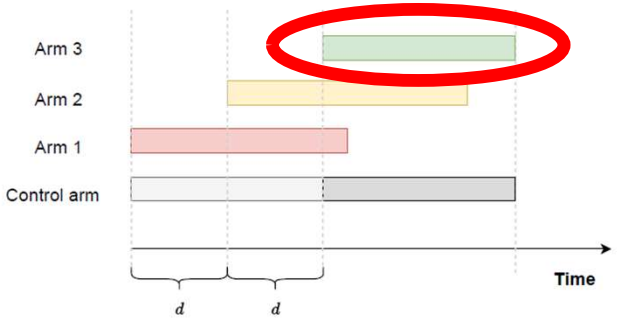
Bofill Roig et al. 2024

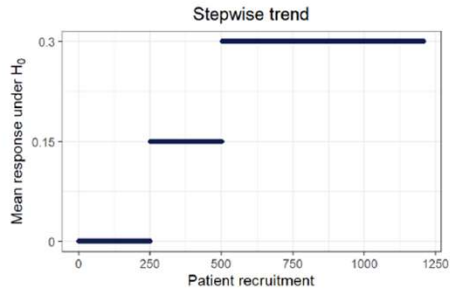
Under which assumptions do the models lead to gain in power and type I error control?

**Simulation settings:**

- 1:1:...:1 allocation ration and block randomization
- Three experimental arms with equal $n = 250$
- New arm enters every $d$ recruited patients
- Time buckets of size 25
- Calibration of $\tau$ assuming small, moderate or large jumps
- Stepwise time trends
- Equal and different time trends

Results for **Arm 3 vs Control**




$\lambda$: strength of the time trend

**Some analysis methods to incorporate NCC**

- **Separate approach:** Analysis using only concurrent controls.
- **Pooled approach:** Analysis pooling concurrent and non-concurrent controls
- **Model-based approaches**[1]
  - Frequentist regression method[2]
  - Bayesian Time Machine[3]



[1] 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

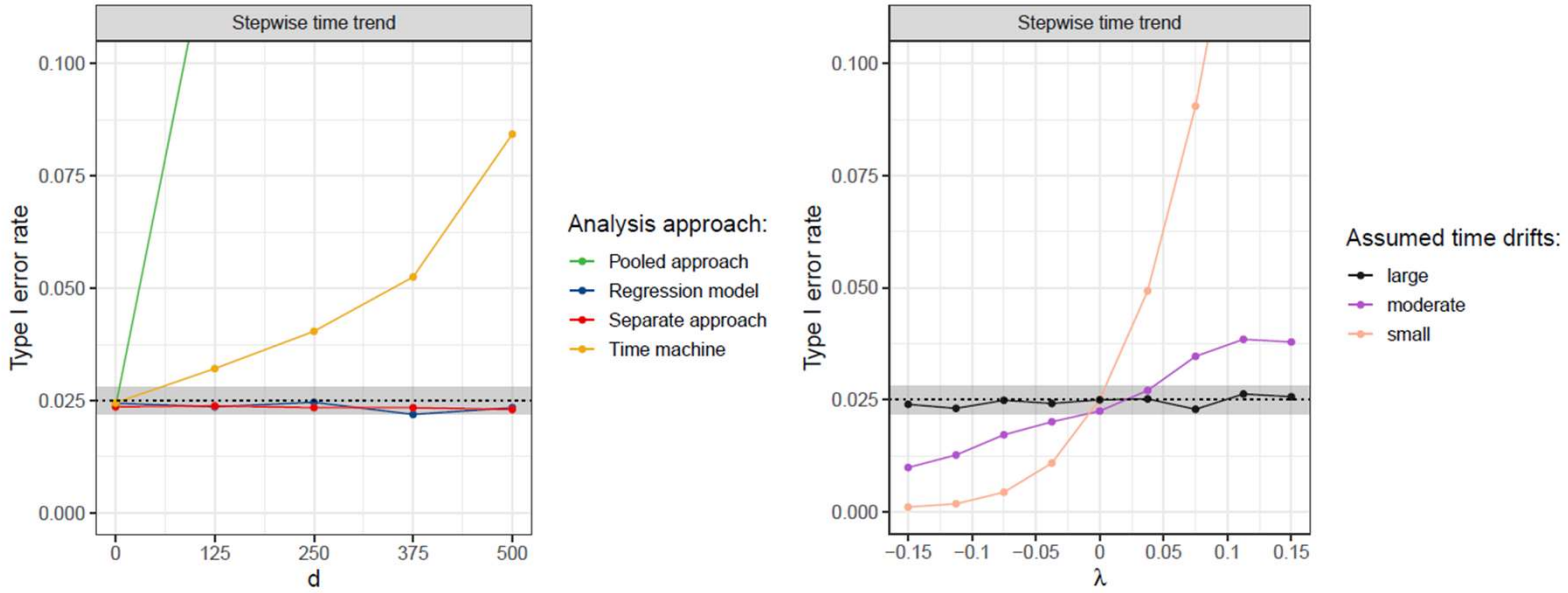
[2] Bofill Roig, M., Krotka, P., et al. (2022). On model-based time trend adjustments in platform trials with non-concurrent controls. BMC Medical Research Methodology

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

# Type 1 error in platform trials with equal time trends across arms

Bofill Roig et al. 2024

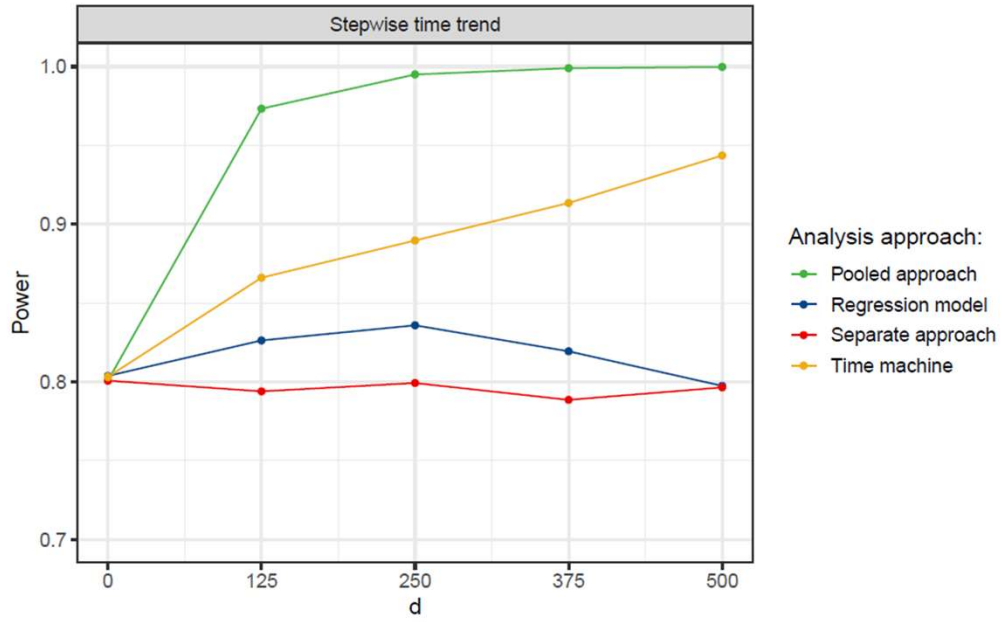
- No treatment effect on arm 3 ($H_0$)
- Same time trend for treatment and control arms ($\lambda$)




Frequentist model controls the type I error (T1E) under the assumption of equal time trends. In addition the T1E control of the Time Machine depends on the assumptions on the prior for the time drift

# Power in platform trials with equal time trends across arms

- Effect on treatment arm 3 ($H_1$)
- Same time trend for treatment and control arms ($\lambda$)




Bofill Roig et al. 2024

Frequentist and Bayesian model-based approaches improve the power as compared to separate analysis using only CC. (The time machine would behave similar to regression model if more conservative prior would have been chosen).

# Conclusions

- The combination of Bayesian techniques and adaptive (frequentist) tests should be way forward to optimize trial designs
- For more complex adaptations (with higher frequency and flexible timing) difficult to implement under frequentist adaptive closed testing framework
- However, insisting on strict control of FWER may prevent the use of alternative methods
- Or require reliance (and acceptance) on simulating important operating characteristics
- It is not Bayesian vs Frequentist, the assumptions an analysis is based matters!
- However, if single arm trials are conducted in certain situations, this should also facilitate the use of Bayesian methods incorporating both adaptations, evidence from 2-arm RCT and non-concurrent control (or external) data.

# Some References


- Ballarini, N. M., T. Burnett, T. Jaki, C. Jennison, F. König, and M. Posch (2021). Optimizing subgroup selection in two-stage adaptive enrichment and umbrella designs. Statistics in Medicine 40(12), 2939-2956.
- Bauer, P., Bretz, F., Dragalin, V., König, F., & Wassmer, G. (2016). Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Statistics in Medicine, 35(3), 325-347.
- Bofill Roig, M., Burgwinkel, C., Garczarek, U., Koenig, F., Posch, M., Nguyen, Q., & Hees, K. (2023). On the use of non-concurrent controls in platform trials: A scoping review. Trials, 24(1), 408.
- Bofill Roig, M., 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), 228.
- Koenig, F., Spiertz, C., Millar, D., Rodríguez-Navarro, S., Machín, N., Van Dessel, A., ... & Hasselbaink, D. (2024). Current state-of-the-art and gaps in platform trials: 10 things you should know, insights from EU-PEARL. Eclinicalmedicine, 67.
- Nguyen, Q. L., Hees, K., Hernandez Penna, S., König, F., Posch, M., Bofill Roig, M., ... & Hofner, B. (2024). Regulatory Issues of Platform Trials: Learnings from EU-PEARL. Clinical Pharmacology & Therapeutics.

# Backup Slide

STATISTICS IN MEDICINE
*Statist. Med.* 2009; **28**:1321–1338
Published online 26 February 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3551

# Optimal choice of the number of treatments to be included in a clinical trial

Nigel Stallard[1,*,†], Martin Posch[2,‡,§], Tim Friede[1], Franz Koenig[2] and Werner Brannath[2]

[1]*Warwick Medical School, University of Warwick, Coventry CV4 7AL, U.K.*
[2]*Section of Medical Statistics, Core Unit for Medical Statistics and Informatics, Medical University of Vienn Spitalgasse 23, A-1090 Wien, Austria*

SUMMARY

It is common for a number of potentially effective treatments to be available for clinical evalua Limitations on resources mean that this inevitably leads to a decision as to how many, and wh treatments should be considered for inclusion in a clinical trial. This paper considers the probler selection of possible treatments for inclusion in a phase III clinical trial. We assume that treatments wi compared using a standard frequentist hypothesis test, and propose a Bayesian decision-theoretic appr that leads to minimization of the total sample size of the trial subject to controlling the familywise ty error rate and the expected probability of rejecting at least one null hypothesis. The method is illustr in the simplest situation, in which two experimental treatments could be included in the clinical exploring the levels of evidence that are required to lead to an optimal trial that includes one or bot these treatments. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: assurance; clinical trial design; decision theory; multiple hypothesis testing

## Assurance

$$\mathbb{E}\{\Pr(\text{reject } H_0 \mid \theta)\} = \int_{-\infty}^{\infty} \Pr(\text{reject } H_0 \mid \theta)\pi(\theta)\,\mathrm{d}\theta$$

## Bayesian Conditional Power

$$\mathbb{E}\{\Pr(\text{reject } H_0 \mid \theta) \mid \theta > 0\} = \frac{\int_0^{\infty} \Pr(\text{reject } H_0 \mid \theta)\pi(\theta)\,\mathrm{d}\theta}{\int_0^{\infty} \pi(\theta)\,\mathrm{d}\theta}$$

# Adaptive designs for subpopulation analysis optimizing utility functions

**Alexandra C. Graf[1,2], Martin Posch*[,1], and Franz Koenig[1]**

[1] Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
[2] Competence Center for Clinical Trials, University of Bremen, Linzer Strasse 4, 28359 Bremen, Germany

Received 1 November 2013; revised 19 August 2014; accepted 24 August 2014

If the response to treatment depends on genetic biomarkers, it is important to identify predictive biomarkers that define (sub-)populations where the treatment has a positive benefit risk balance. One approach to determine relevant subpopulations are subgroup analyses where the treatment effect is estimated in biomarker positive and biomarker negative groups. Subgroup analyses are challenging because several types of risks are associated with inference on subgroups. On the one hand, by disregarding a relevant subpopulation a treatment option may be missed due to a dilution of the treatment effect in the full population. Furthermore, even if the diluted treatment effect can be demonstrated in an overall population, it is not ethical to treat patients that do not benefit from the treatment when they can be identified in advance. On the other hand, selecting a spurious subpopulation increases the risk to restrict an efficacious treatment to a too narrow fraction of a potential benefiting population. We propose to quantify these risks with utility functions and investigate nonadaptive study designs that allow for inference on subgroups using multiple testing procedures as well as adaptive designs, where subgroups may be selected in an interim analysis. The characteristics of such adaptive and nonadaptive designs are compared for a range of scenarios.

*Keywords:* Adaptive design; Enrichment design; Hypothesis selection; Sample size reallocation; Utility function.

Additional supporting information may be found in the online version of this article at the publisher's web-site

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

*Article*

Statistical Methods in Medical Research
2019, Vol. 28(7) 2096–2111

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0962280217747312
journals.sagepub.com/home/smm
SAGE

# Optimized adaptive enrichment designs

**Thomas Ondra,[1] Sebastian Jobjörnsson,[2] Robert A Beckman,[3] Carl-Fredrik Burman,[2,4] Franz König,[1] Nigel Stallard[5] and Martin Posch[1]**

**Abstract**

Based on a Bayesian decision theoretic approach, we optimize frequentist single- and adaptive two-stage trial designs for the development of targeted therapies, where in addition to an overall population, a pre-defined subgroup is investigated. In such settings, the losses and gains of decisions can be quantified by utility functions that account for the preferences of different stakeholders. In particular, we optimize expected utilities from the perspectives both of a commercial sponsor, maximizing the net present value, and also of the society, maximizing cost-adjusted expected health benefits of a new treatment for a specific population. We consider single-stage and adaptive two-stage designs with partial enrichment, where the proportion of patients recruited from the subgroup is a design parameter. For the adaptive designs, we use a dynamic programming approach to derive optimal adaptation rules. The proposed designs are compared to trials which are non-enriched (i.e. the proportion of patients in the subgroup corresponds to the prevalence in the underlying population). We show that partial enrichment designs can substantially improve the expected utilities. Furthermore, adaptive partial enrichment designs are more robust than single-stage designs and retain high expected utilities even if the expected utilities are evaluated under a different prior than the one used in the optimization. In addition, we find that trials optimized for the sponsor utility function have smaller sample sizes compared to trials optimized under the societal view and may include the overall population (with patients from the complement of the subgroup) even if there is substantial evidence that the therapy is only effective in the subgroup.

**Keywords**

Adaptive design, optimal design, enrichment design, precision medicine, subgroup analysis