Current statistical issues in platform trials for the evaluation of multiple treatments

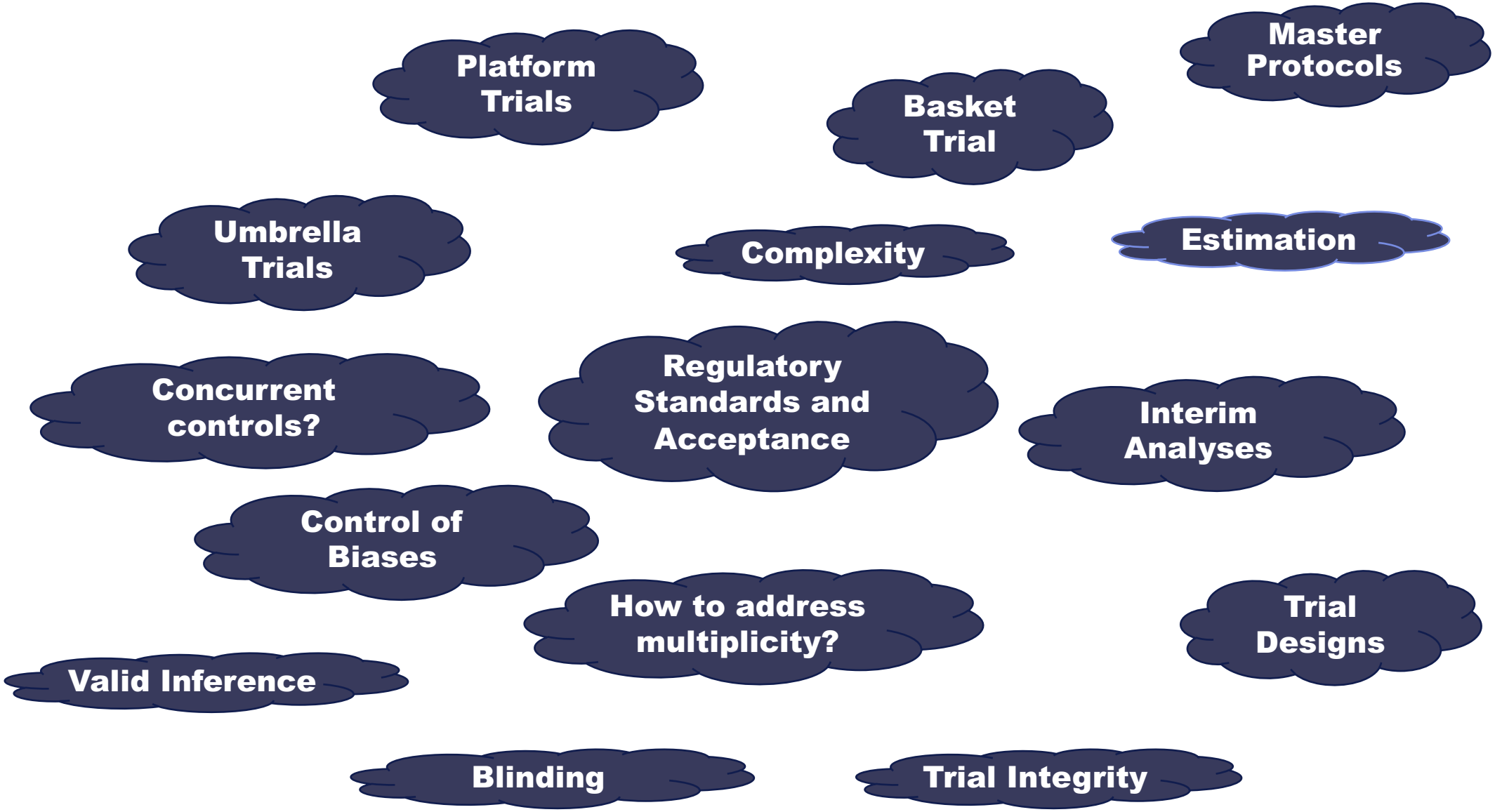
Martin Posch and Franz König
September 2020

MEDICAL UNIVERSITY
OF VIENNA

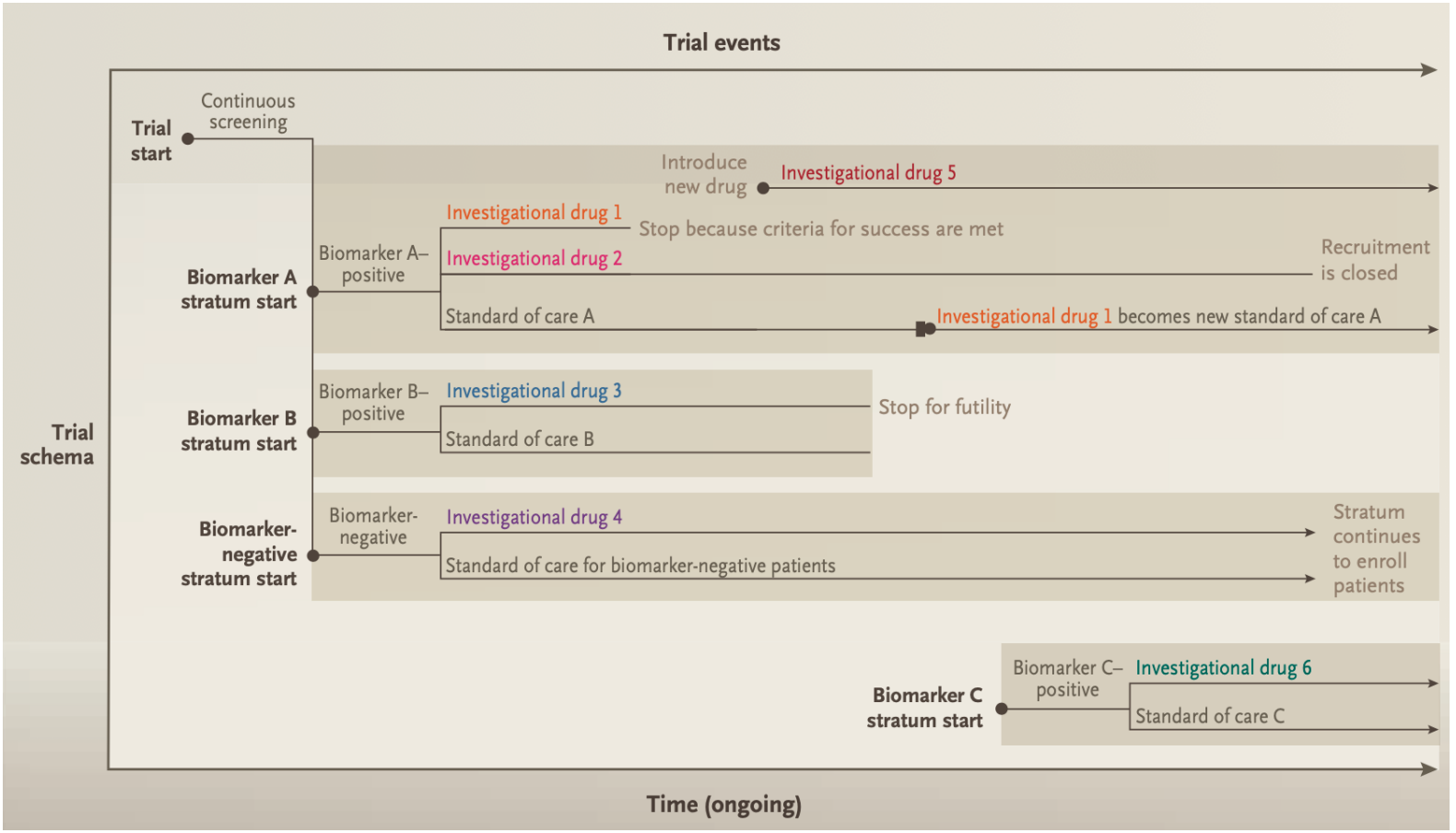
Acknowledgements

Elias Meyer, Olivier Collignon, Carl-Frederik Burman, Benjamin Hofner, Andrew Thomson, Anja Schiel, EU-DIA ADSWG, EMA BSWP Task Force, IMI-EU Pearl

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.


Platform Trials
Master Protocols
Basket Trial
Umbrella Trials
Complexity
Estimation
Concurrent controls?
Regulatory Standards and Acceptance
Interim Analyses
Control of Biases
How to address multiplicity?
Trial Designs
Valid Inference
Blinding
Trial Integrity

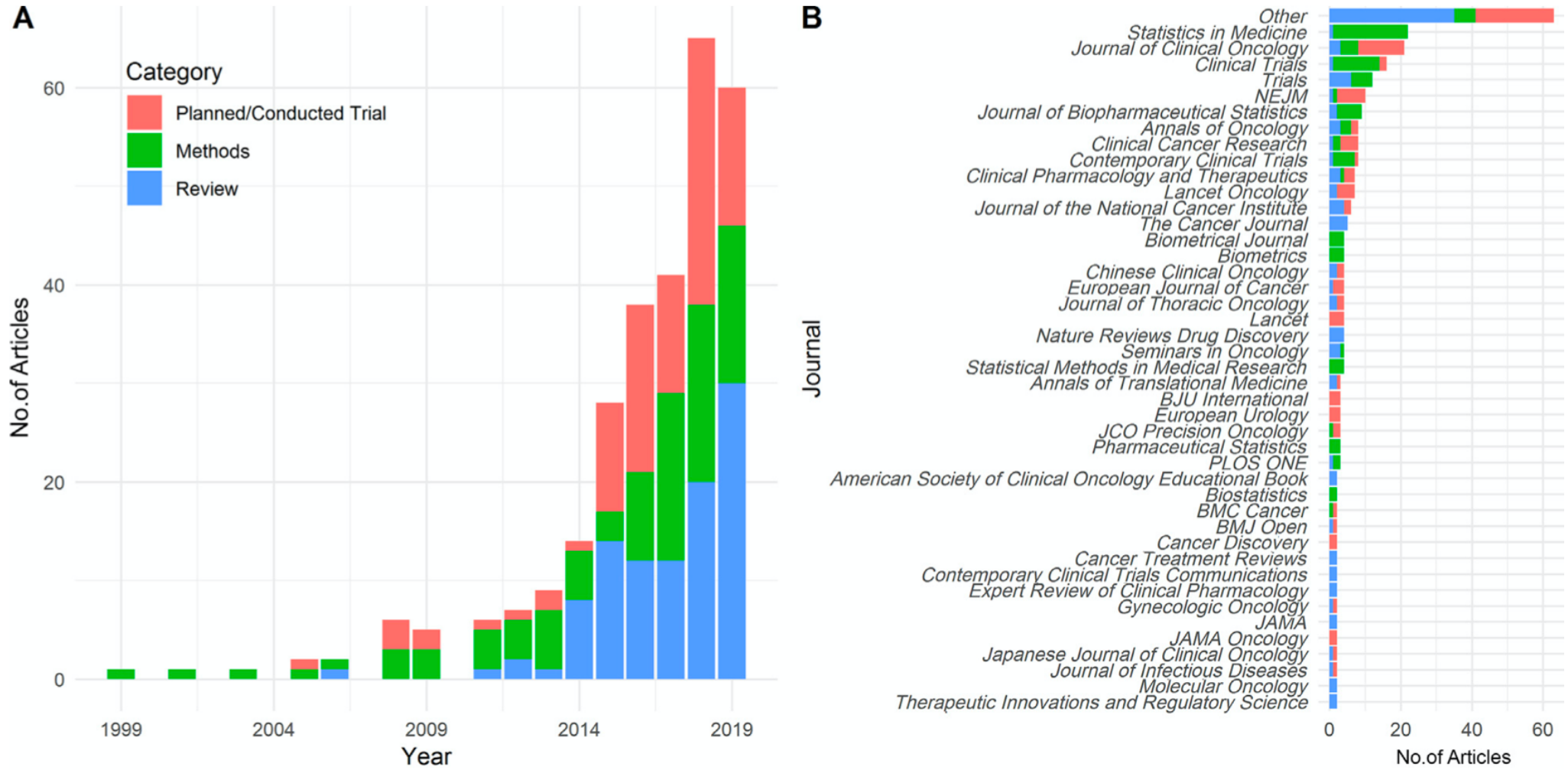
Master Protocols



- **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

Review: Publications on Master Protocols



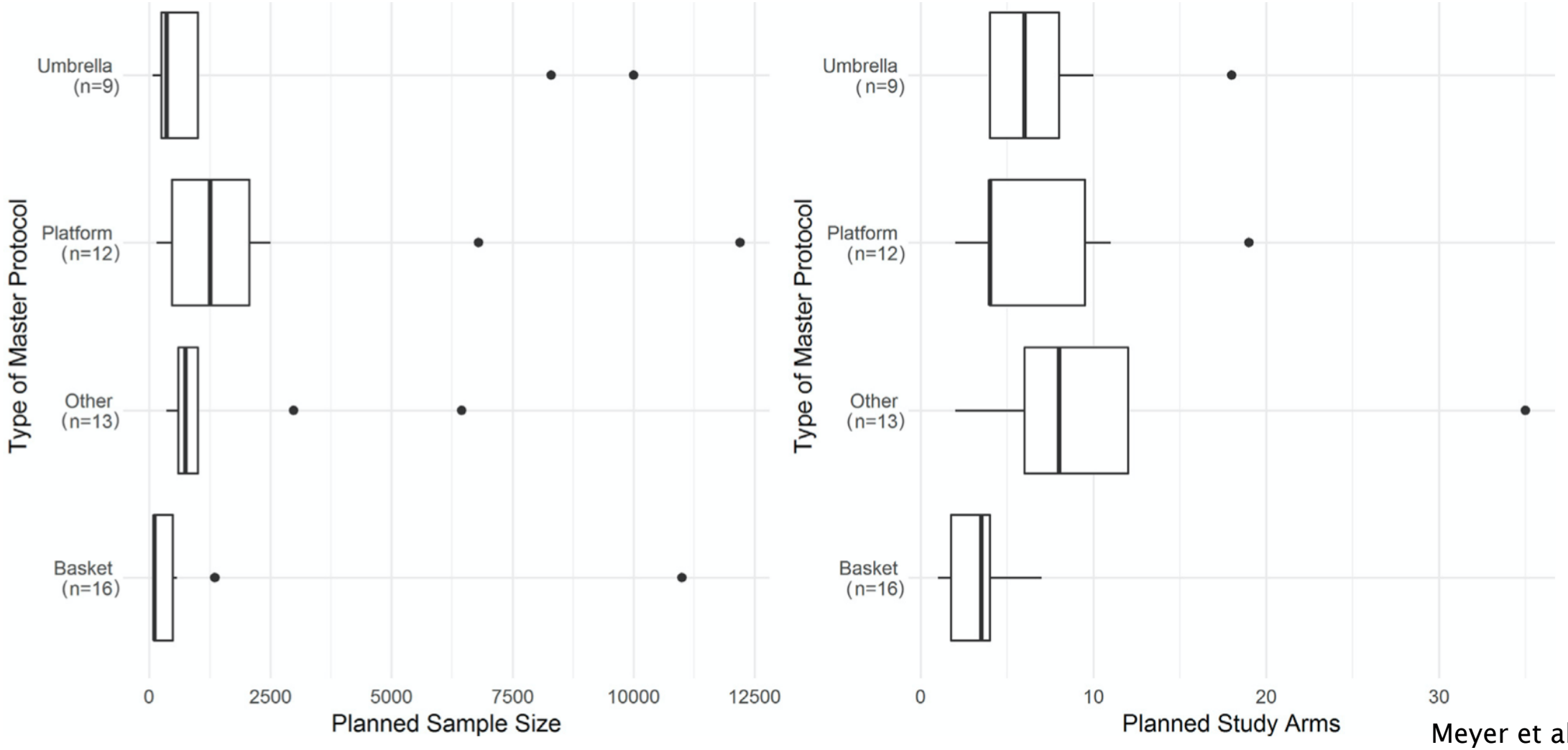
Meyer et al. (2020a)

Literature Review: Published Master Protocols (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**

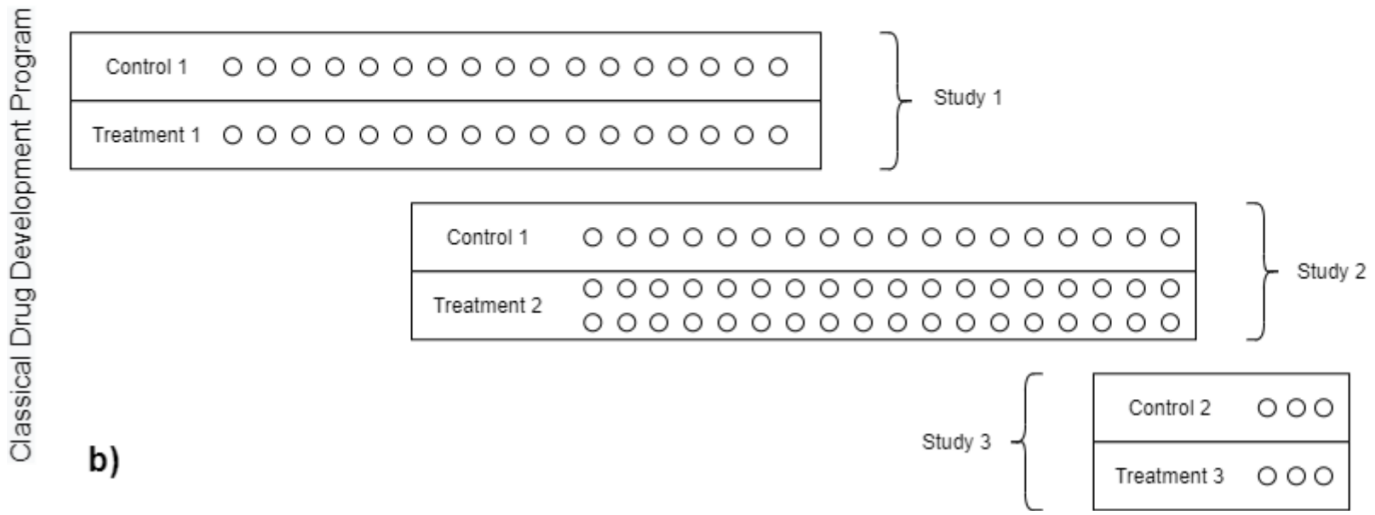
Meyer et al. (2020a)

Sample Size and Number of Study Arms



Meyer et al. (2020a)

Classical Drug Development Programs

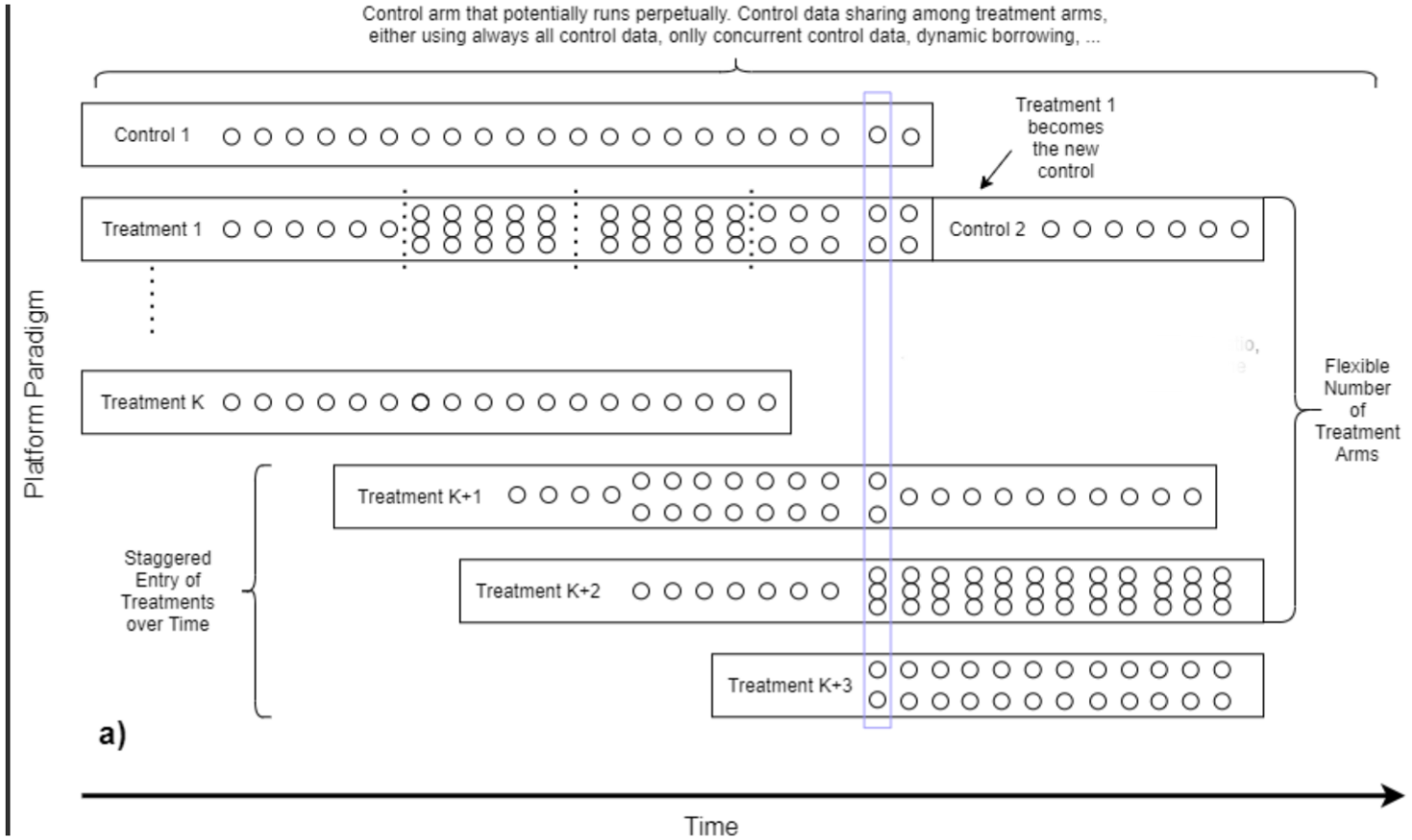


Meyer et al. (2020b)

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



Meyer et al. (2020b)

Potential Benefits of Platform Trials

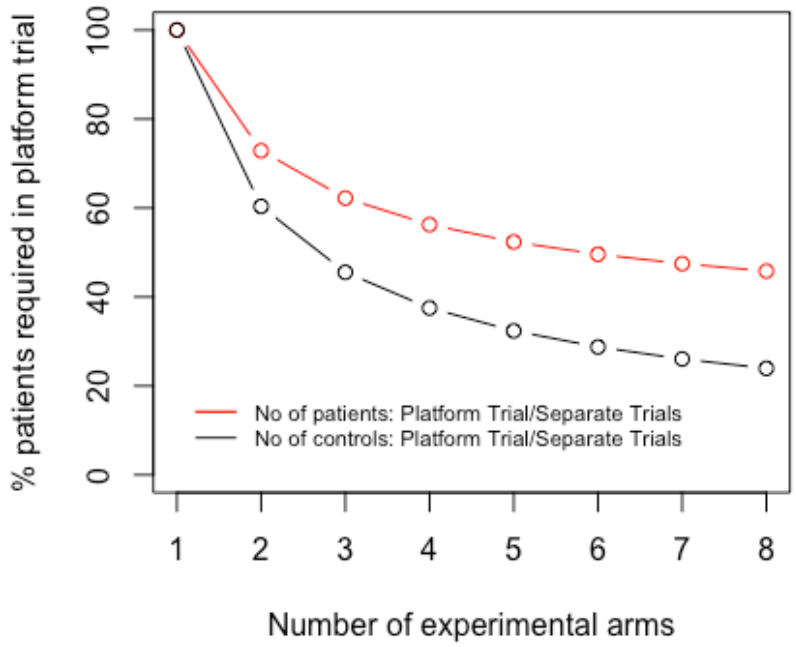
- **Operational advantages**
  - Joint trial infrastructure (savings for the sponsor)
  - Broader inclusion criteria because several treatments are tested (more patients are elgigible for the trial)
  - “Online Learning” ineffective treatments can be stopped early in interim analyses.
- **Better statistical efficiency compared to separate trials**
  - Shared control groups and early stopping potentially reduces the number of patients needed
  - Adaptive allocation of patients to treatments for which more information is required
  - Direct comparisons between treatments.

Fewer Control Patients due to Shared Controls

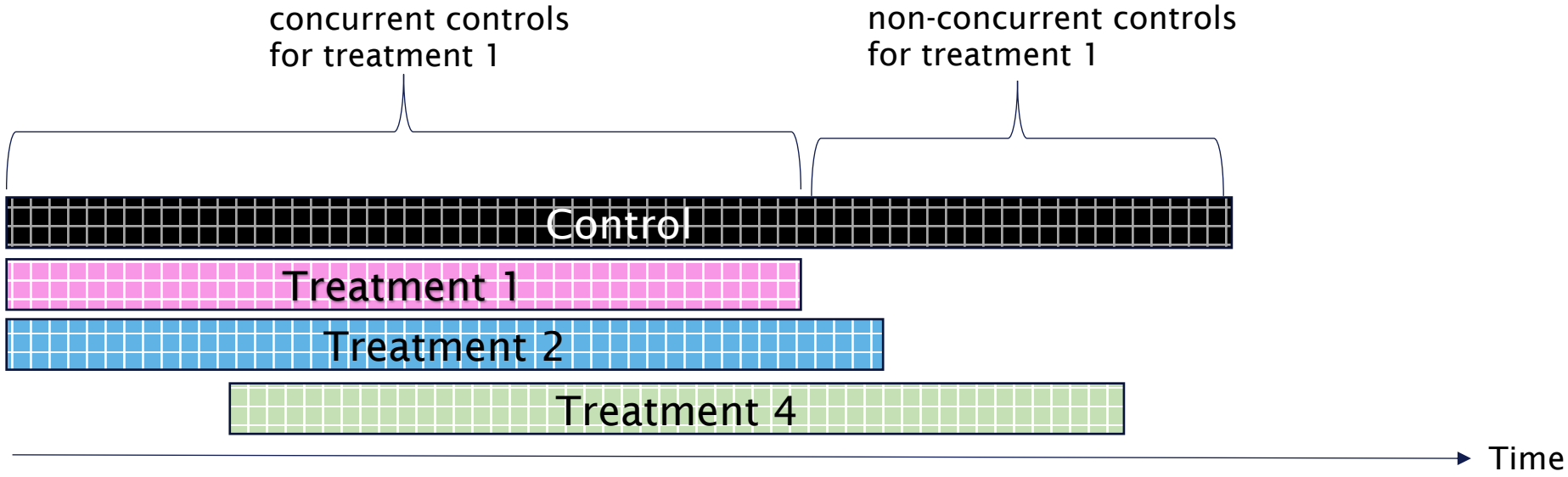
- Classical development for k treatments: k separate trials with 1:1 randomisation and sample size to reach pairwise power 1- β (assume equal treatment effects)

- 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

% of total (control) patients required in a multi-armed trial vs patients required in separate trials



Non-Concurrent Controls.



- If experimental treatments enter or leave the platform over the course of the trial, some of the control patients will be recruited at time periods, where the respective treatment was not part of the platform
- Should these controls be used in the statistical comparisons?

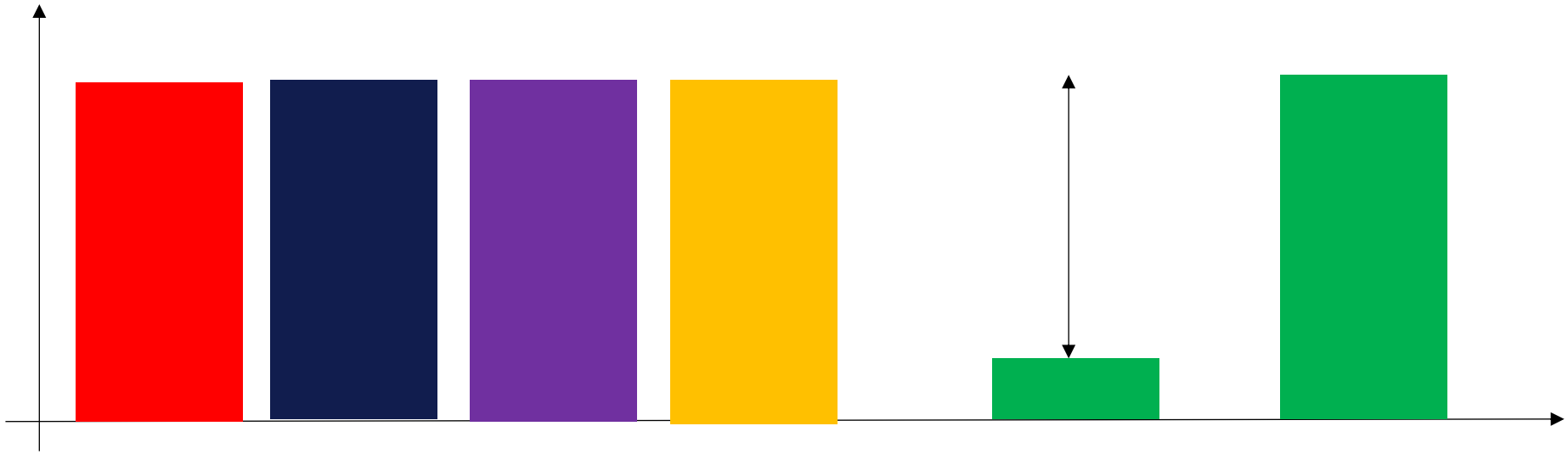
  \+ Increases the statistical power due to larger sample sizes

  - Potential of bias if there are time trends in the population or placebo responses.
    Under suitable assumptions one can adjust for potential time trends. Similar issues as in the use of historical controls apply.
- Tradeoff between power and bias

Saville and Berry, 2016, Eichler et al. 2016, Burger et al. 2020

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

Adjustment for Multiplicity?

- "No adjustment for multiplicity is necessary" if independent hypothesis are tested: 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)

- "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)

Other Sources of Multiplicity

- Multiple Endpoints
- Multiple Interim Analysis
- Adaptations
- Multiple Control Groups
- Subgroups

FWER when Ignoring Multiplicity & Adaptations

- Comparison of *k* treatments with a control
- In a trial with an adaptive interim analysis what **is the most extreme T1E rate** when an unadjusted fixed sample size test is applied at the final analysis **if we allow for sample size reassessment (and selection) at interim**

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

Challenges in Decision Making

- As the number and type of treatments is 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 and may not be appropriate.
- Control of the probability of at least one type I error appears to be too stringent especially in a potentially perpetual trial

Strategies for (Frequentist) Statistical Inference

- Consider **each treatment** in the platform trial as inferentially independent controlling the FWER for the family of hypotheses relating to the treatment (endpoints, doses, treatment regimens, subgroups)
- Consider **each substudy** in the platform trial as inferentially independent study, controlling the study-wise rate of false positive conclusions for each sub-study (e.g., across endpoints, doses, treatment regiments)

Adressing Multiplicity on Platform level

- Controlling for multiplicity on a treatment or sub-study level is a pragmatic solution to integrate platform trials in the current regulatory framework.
- However, from a societal perspective also platform-wise (or indication-wise) error rates are of importance

  Collignon et al. (2020b)
  - What is the probability, that at least one ineffective treatment is declared effective?
  - What is the expected number of ineffective treatments declared effective?
- Bessides FWER control, one approach is to control (or estimate) the False Discover Rate: The expected proportion of false positives among all treatment arms declared effective.

  Wason et al. (2020)
- Other approaches directly account for the *loss* of false positive decisions

An alternative to the FWER – Expected Loss

FWER does not account for the difference in losses induced by different false positive claims or the number of false positive claims

- H_i tested at level $\alpha_i, i = 1, \ldots m$. c_i . . . loss if H_i is incorrectly rejected.
  Assuming additivity, the expected loss is given by

$$L = \sum_{i=1}^{m} c_i \alpha_i$$

- Example Subgroup Testing:

  m disjoint subgroups with prevalences $\lambda_i, \sum \lambda_i = 1$ are tested at level α.
  Assume the losses are proportional to the prevalences: $c_i = c\,\lambda_i$. Then

$$L = \sum_{i=1}^{m} \alpha\, c \lambda_i = c\,\alpha$$

- The concept can be extended to more complex settings, where losses are not additive (e.g., for non-overlapping subgroups)

Xun, Bretz, Glimm, Maurer, 2015
Brannath, 2018
Collignon et al. 2020

A Further Step: Bayesian Decision Theoretic Approaches

- Losses and gains are averaged over the whole parameter space based on a prior distribution.
- For every constellation of effect sizes and trial outcome a utility is defined.
- Compute the expected utility over a prior on the effect sizes.
  - Optimize trial designs to maximize the expected utility, given a conventional multiple testing procedure is used as decision rule.

    Rosenblum ’14, Graf et al. 15, Krisam et al. ’15,’16, Ondra et al.’16,’18, Ballarini et al. ’2020

  - Fully decision theoretic designs: Both, the decision rule and the design are optimized.

    Abrahamyan et al. ’14 Stallard et al. 17, Hee et al. ’17, Miller et al. ’17,Pearce et al.’18

Conclusions

- The increased efficiency of platform trials comes at the cost of additional complexity for statistical inference.
- Several potential sources of bias need to be addressed
- The concept of study-wise T1E rate control is not directly applicable to platform trials, especially if they perpetual in nature
- Risk of false positive decisions should be known for robust inference
- This is important regardless if Bayesian or frequentist inference methods are applied.
- There is no consensus yet on the inferential framework for platform trials

EU-PEARL

EU PATIENT-CENTRIC CLINICAL TRIAL PLATFORMS

SHAPING THE FUTURE OF CLINICAL TRIALS

We are transforming the future of drug development
by creating a sustainable entity
available for industry and academia
to conduct platform trials in any disease area,
co-designed by patients.

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.

EU-PEARL WILL DELIVER

1 A trusted sustainable entity ready to setup and coordinate the operation of Integrated Research Platforms in any disease.

2 A Clinical Trial Platform Framework that can be used for any disease, plus four disease clinical trial platforms ready to operate at the end of the project

3 Four disease trial-ready clinical networks

- Major Depressive Disorder
- Tuberculosis
- Non-Alcoholic Steatohepatitis (NASH)
- Neurofibromatosis

WHO IS INVOLVED?

EUROPEAN UNIVERSITY HOSPITAL ALLIANCE (EUHA) HOSPITALS

Vall d'Hebron Institut de Recerca
Vall d'Hebron
Erasmus MC
Assistance Publique Hôpitaux de Paris
UniSR Università Vita-Salute San Raffaele
Medizinische Universität Wien
King's College London
KU Leuven
Charité Universitätsmedizin Berlin

OTHER HOSPITALS

LMU Klinikum der Universität München

UNIVERSITIES

University of Oxford
Manchester 1824 The University of Manchester
Università degli Studi di Milano
Newcastle University
Tartu Ülikool 1632 Universitas Tartuensis

PATIENT ORGANISATION

EPF European Patients Forum

DATA, STATISTICS

The European Institute For Innovation Through Health Data
Berry Consultants Statistical Innovation
Custodix

REGULATORY

Paul-Ehrlich-Institut
euroscan

PROJECT MANAGEMENT

teamit Research

EUROPEAN RESEARCH INFRASTRUCTURES

eatris
ECRIN European Clinical Research Infrastructure Network

BIOPHARMACEUTICAL COMPANIES/EFPIA/ASSOCIATED PARTNERS

AstraZeneca
Novartis
Janssen Pharmaceutical Companies of Johnson & Johnson
Allergan
novo nordisk
Otsuka
teva
Sanofi
Pfizer
TB Alliance Global Alliance for TB Drug Development
SpringWorks Therapeutics
Children's Tumor Foundation Ending NF Through Research
Servier

DISEASE-AGNOSTIC WORK PACKAGES

WP1 IRP Governance, Quality, Sustainability	WP2 Scientific, Regulatory and Operational Methodology	WP3 Clinical network and patient level data	WP8 Project Oversight, Project Management and Outreach
Information governance and Ethics	Quantitative Methods and Statistical Design	Clinical Network	Project plan
Certification of Implemented ICT platform components	Regulatory Aspects	Patient Data Network	Reporting and timely presentation of deliverables
Legal Framework	Clinical Operational Best Practices (Master Protocol Template)	Deployment and Evaluation	Internal and external communications
Interoperability and Data Quality			Risk Management
Project and Sustainability KPIs			Alliances with other initiatives
Sustainability and Scale Up			
Patient Engagement Platform			

DISEASE-SPECIFIC WORK PACKAGES

WP4 IRP for Major Depressive Disorder (MDD)	WP5 IRP for Tuberculosis (TB)	WP6 IRP for Non-Alcoholic Steatohepatitis (NASH)	WP7 IRP for NeuroFibromatosis (NF)

Define scientific challenges for each disease area.

Design Master Protocol (disease specific).

Establish key operational requirements for the implementation of specific IRPs.

Endorsement of Master Protocols by regulatory and ethics.

Build patients and clinical networks.

Sustainability and dissemination.

Stakeholder Workshop

October 8th, 2020

2-6 PM CET/ 8AM-12 PM EST

EU-PEARL SWS introduction session

This general session will focus on the current and future states of platform trials and the related regulatory guidance. The concept of an IRP will be introduced. The EU-PEARL project along with other relevant case studies will be presented, followed by a panel discussion on current opportunities and challenges

October 22nd, 2020

3-6 PM CET/9 AM-12 PM EST

Breakout session 'Multiplicity Framework for IRPs'

When and how to adjust for multiplicity in platform trials: What are the main sources of multiplicity in complex trial designs? Which error rates and methods are suitable for which study objectives & designs (and which methods are not)?

October 26th, 2020

3-6 PM CET/10 AM-1 PM EST

Breakout session 'Regulatory and Ethic Committee Considerations for IRPs'

Main issues to overcome for alignment on submission principles at various stages of an IRP (Initial submission, during trial execution and end of trial) and other relevant regulatory and ethical aspects of IRPs.

October 29th, 2020

3-6 PM CET/10 AM-1 PM EST

Breakout session 'Patient engagement for IRPs'

Incorporating perspectives of patients and patient advocates on platform trials and IRP design, and understanding potential opportunities and hurdles.

Free registration at http://www.eu-pearl.eu/

Selected References

- Ballarini NM, Burnett T, Jaki T, Jennison C, Kӧnig F, Posch M Optimising subgroup selection in two-stage adaptive enrichment and umbrella designs. Submitted. 2020
- 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, König 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)
- 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.