Adaptive **D**esigns and **M**ultiple **T**esting **P**rocedures Workshop

26th - 27th February 2025

Regensburg, Germany

Scientific and Organising Committee:

David S. Robertson[1], Thomas Jaki[1,2], Cornelia Kunz[3], Marta Bofill Roig[4]

[1]MRC Biostatistics Unit, University of Cambridge, UK, [2]University of Regensburg, Germany,
[3]Boehringer-Ingelheim, Germany, [4]Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Sponsored by:

Instats is a research training platform that freely provides its partners with sophisticated content production and hosting services, allowing them to more efficiently train researchers, generate new revenue streams, and connect with their communities. Learn more about partnering with Instats at instats.org, where you can find leading research training content

Organised by:

Workshop venue

Room H401
Faculty of Informatics and Data Science
Bajuwarenstraße 4
93053 Regensburg
Germany

FIDS
Fakultät für Informatik und Data Science
DIMAS
Department for Interdisciplinary and Multiscalar Area Studies
FORSCHUNGSGRUPPE
Rechtstransfer und Rechtsentwicklung in Zentralasien
Universität

SCIENTIFIC PROGRAM - OVERVIEW

Wednesday 26th February

08:00 - 08:30	Registration and coffee
08:30 – 08:35	Welcome
08:35 – 10:40	Session I: Dose finding and dose selection
10:40 – 11:00	Coffee
11:00 – 12:45	Discussion session on the practical challenge
12:45 – 13:30	Lunch
13:30 – 15:15	Session II: Group sequential designs
15:15 – 15:35	Coffee
15:35 – 17:40	Session III: Platform trials

Thursday 27th February

08:00 – 08:30	Registration and coffee
08:30 – 10:35	Session IV: Response-adaptive randomisation
10:35 – 11:05	Coffee + Working group meeting
11:05 – 12:50	Session V: Estimation
12:50 – 13:35	Lunch
13:35 – 15:20	Session VI: Multiple testing
15:20 – 15:40	Coffee
15:40 – 17:25	Session VII: Complex innovative designs and decision analysis
17:25 – 17:30	Closing

SCIENTIFIC PROGRAM – DETAILED SCHEDULE

Wednesday 26th February

08:00 – 08:30	**Registration and coffee**
08:30 – 08:35	**Welcome**
08:35 – 10:40	**Session I: Dose finding and dose selection** Chair: James Willard 1. Elias Laurin Meyer: *Designing a seamless P1/P2a open enrollment CRM dose escalation study* 2. Weishi Chen: *On the Consistency of Partial Ordering Continual Reassessment Method (POCRM) with Model and Ordering Misspecification* 3. Xijin Chen: *Joint modelling of ctDNA and response in dose-finding designs in oncology* --- Biobreak --- 4. Connor Fitchett: *Applying Bayesian Response Adaptive Randomization to dose-ranging platform trials* 5. Lorenz Uhlmann: *Adding an active dose level to an ongoing randomized controlled phase III study* 6. Cornelia Kunz: *Optimizing Dose Selection Designs for Phase 3 Trials*
10:40 – 11:00	**Coffee**
11:00 – 12:45	**Discussion session on the practical challenge** Chair: Marta Bofill Roig 1. Cornelia Kunz: *Introduction to the practical challenge* 2. Franz König: *Navigating Interim Analyses and Multiple Endpoints: Handling Accelerated Recruitment and Converging Analyses* 3. Fredrik Öhrn: *Adjusting the information fraction by sample-size reassessment* 4. Moritz Fabian Danzer: *Exploitation of multivariate central limit theorems in a closed test procedure for group sequential designs*

	5. Discussant I: Werner Brannath 6. Discussant II: Leonhard Held
12:45 – 13:30	**Lunch**
13:30 – 15:15	**Session II: Group sequential designs** Chair: Rene Schmidt 1. Stephen Schüürhuis: *Group-sequential methods for generalised pairwise comparisons* 2. Fabrice Lotola Mougeni: *Two-Part Models in Group Sequential Designs for Zero-Inflated Data* 3. Tom Parke: *Using Machine Learning to Optimize Trial Design* --- Biobreak --- 4. Carolin Herrmann: *Accounting for delayed responses in group sequential designs – needed or nice to have?* 5. Aritra Mukherjee: *Evaluating the impact of outcome delay on adaptive designs*
15:15 – 15:35	**Coffee**
15:35 – 17:40	**Session III: Platform trials** Chair: Franz König 1. Jiangyue Yao: *Online control of the Family-wise error rate (FWER) for multi-arm multi-stage platform trials* 2. Pavla Krotka: *Adjusted treatment effect estimators for platform trials with interim analyses utilizing non-concurrent controls* 3. Marta Rofill Boig: *Unconditional treatment effect estimates when adjusting for time in platform trials with binary endpoints* --- Biobreak --- 4. Tobias Mielke: *When is a multi-armed trial a platform trial and what are the implications?* 5. Nikita Mozgunov: *Improving implementation of Adaptive Multi-Arm Multi-Stage Trials: The MAMS R Package* 6. Peter Jacko: *A Fair and Efficient Randomization Scheme for Multi-Arm Seamless Two-Phase Clinical Trials*

Thursday 27th February

08:00 – 08:30	**Registration and coffee**
08:30 – 10:35	**Session IV: Response-adaptive randomisation** Chair: David Robertson 1. Lukas Pin: *Revisiting Optimal Proportions for Binary Responses: Insights from Incorporating the Absent Perspective of Type-I Error Rate Control* 2. Tom Parke: *Response Adaptive Randomization – when trying to select the best arm/dose from those being tested* 3. Stef Baas: *Exact statistical analysis for response-adaptive clinical trials: A general and computationally tractable approach* --- Biobreak --- 4. Gianmarco Caruso: *Patient-oriented response-adaptive designs based on a novel information measure in multi-arm trials with quantitative endpoints* 5. Ayon Mukherjee: *Covariate-Adjusted Response Adaptive Designs for Semiparametric Survival Models* 6. Rosamarie Frieri: *Design and inference for multi-arm clinical trials with informational borrowing: the interacting urns design*
10:35 – 11:05	**Coffee + Working group meeting**
11:05 – 12:50	**Session V: Estimation** Chair: Andreas Faldum 1. James Willard: *Covariate Adjustment in Bayesian Adaptive Randomized Controlled Trials* 2. David Robertson: *Confidence intervals for adaptive designs* 3. Enyu Li: *The uniformly most powerful conditional unbiased test and conditional confidence interval in two-stage adaptive enrichment designs* --- Biobreak --- 4. Han Chang Chiam: *Pre-specification and Bias in Hybrid RCTs* 5. Elad Berkman: *Exploratory Adaptive Enrichment Design with Sample Size Re-estimation: A Novel Approach for Clinical Trial Optimization*

12:50 – 13:35	**Lunch**
13:35 – 15:20	**Session VI: Multiple testing** Chair: Fredrik Öhrn 1. Libby Daniells: *Adding baskets to an ongoing basket trial with information borrowing* *2.* Alex Spiers: *Optimising graph-based multiple testing procedures by incorporating clinical considerations into flexible power objectives for FWER control* *3.* Felix Herkner: *Optimizing Endpoint Analysis in the context of Kidney Transplant Studies: Composite vs. Multiplicity-Corrected Approaches* --- Biobreak --- 4. Michael Grayling: *Extensions to a closed testing procedure for assessing efficacy in a prespecified subpopulation* 5. Leonhard Held: *Sequential conduct of clinical trials*
15:20 – 15:40	**Coffee**
15:40 – 17:25	**Session VII: Complex Innovative Designs and Decision Analysis** Chair: TBC 1. Fabian Eibensteiner: *Real-world use of adaptive designs in paediatric clinical trials – A review of the European Medicines Agency's Clinical Trials Information System (CTIS)* 2. Drifa Belhadi: *Bayesian decision analysis for clinical trial design with binary outcome in the context of Ebola Virus Disease outbreak – Simulation study* 3. Boaz Adler: *Communicating Complex Considerations in Dual Endpoint Trial Design – An Oncology Case Study* --- Biobreak --- 4. Valeria Mazzanti: *Assessing the Effects of Additional Investment in Earlier Phase Trials to Enhance Overall Program Probability of Success Through Informed Priors* 5. Raviv Pryluk: *Model-Guided Parameter Optimization for Complex Innovative Trial Designs*
17:25 – 17:30	**Closing**

ABSTRACTS

Session I: Dose finding and dose selection

Wednesday 26th February 08:35 – 10:40

Designing a seamless P1/P2a open enrollment CRM dose escalation study

Elias Laurin Meyer[1], Tom Parke[2]

[1]Berry Consultants, Vienna, Austria; [2]Berry Consultants, Abingdon, UK; elias@berryconsultants.com

Traditionally, phase I dose escalation designs aim to find the maximum tolerated dose (MTD), usually the highest dose whose probability to cause dose-limiting toxicities (DLTs) stays below a certain target toxicity level (TTL), and an adequate dosing scheme. In practice, however, this focus on the MTD when selecting doses to take into registration trials often leads to exposing many trial participants to doses that either produce more toxicity without increased efficacy or severe toxicities that could both limit the options for receiving benefits or lead to premature discontinuation and missed opportunity for continued benefit. Recently, FDA launched Project Optimus with the aim of educating and innovating all stakeholders to, among other goals, move towards designing dose escalation trials that attempt to find “optimal” doses, where optimality comprises safety, tolerability and efficacy.

In this talk, we draw from our in-house experiences of designing first-in-human oncology trials to present a seamless phase 1 / phase 2a dose escalation design using open enrollment guided by the continual reassessment method (CRM) that evaluates both safety and efficacy. We introduce advanced CRM adaptations such as open enrollment, target toxicity intervals and escalation with overdose control (EWOC), early stopping rules, backfilling and frontfilling, ad-hoc rules, etc., that not only vastly improve the performance and customizability of CRM, but also help identify “optimal” doses to take forward into registration trials. Finally, we discuss how to design and simulate such trials with the aid of targeted software and share insights on how to best communicate such designs with a clinical development team.

On the Consistency of Partial Ordering Continual Reassessment Method (POCRM) with Model and Ordering Misspecification

Weishi Chen, Pavel Mozgunov

MRC Biostatistics Unit, University Of Cambridge, UK; weishi.chen@mrc-bsu.cam.ac.uk

One of the aims of Phase I clinical trial designs in oncology is typically to find the maximum tolerated doses. A number of innovative dose-escalation designs were proposed in the literature to achieve this goal efficiently. Although the sample size of Phase I trials is usually small, the asymptotic properties (e.g. consistency) of dose-escalation designs can provide useful guidance on the design parameters and improve fundamental understanding of these designs. For the first proposed model-based monotherapy dose-escalation design, the Continual Reassessment Method (CRM), sufficient consistency conditions have been previously derived and then greatly influenced on how these studies are run in practice. At the same time, there is an increasing interest in Phase I combination-escalation trials in which two or more drugs are combined. The monotherapy dose-escalation design cannot be generally applied in this case due to uncertainty between monotonic ordering between some of the combinations, and, as a result, specialised designs were proposed. However, there were no theoretical or asymptotic properties evaluation of these proposals. In this paper, we derive the consistency conditions of the partial Ordering CRM (POCRM) design when there exists uncertainty in the monotonic ordering with a focus on dual-agent combination-escalation trials. Based on the derived consistency condition, we provide guidance on how the design parameters and ordering of the POCRM should be defined.

Joint modelling of ctDNA and response in dose-finding designs in oncology

Xijin Chen

MRC Biostatistics Unit, United Kingdom; xijin.chen@mrc-bsu.cam.ac.uk

Dose-finding trials aim at coming up with a safe and efficient drug administration in humans. In the last decade, ‘liquid biopsy’ technology has been well developed, which can provide a readout of treatment response much earlier than conventional endpoints. Modern methods for dose-finding in oncology allow the use of novel biomarkers such as circulating tumour DNA (ctDNA) to improve the rapidity and efficiency of dose-finding trials.

There are proposals to incorporate ctDNA, which is considered an early indicator of treatment responses in dose-finding trials, dichotomizes the continuous outcome. In this talk, we will present a design based on one binary efficacy response and one continuous ctDNA measurement per subject. A joint model is factorized into a marginal model for the continuous ctDNA and a conditional probit model for the binary efficacy. The dose with the estimated probability of efficacy closest to the pre-specified target value is administered to the next cohort of subjects. We compare the proposed joint model with the corresponding univariate model taking account of binary outcomes alone under various scenarios of dose-response relationships. We discuss how to extend the proposed joint model to consider multiple ctDNA values via a multivariate probit conditional model.

Simulations show that the proposed approach, with additional and earlier information from ctDNA, can yield a shorter trial duration and an improvement in the identification of the target dose, with the probability of efficacy at a pre-specified target value. Furthermore, we show that this approach often reduces the time individual patients spend on potentially inactive trial therapies.

The early readouts from biomarkers are expected to bring value to different stakeholders. First, current patients can benefit from early termination of treatment in case there is early evidence for a lack of response. Second, investigators can expect a faster trial as we can make an earlier decision about dose determination based on earlier available ctDNA. Finally, the next enrolled patients tend to be administered the target dose based on more available information when making a dose determination.

Applying Bayesian Response Adaptive Randomization to dose-ranging platform trials

Connor Fitchett[1], David Robertson[1], Sofia Villar[1], Ayon Mukhurgee[2]

[1]University of Cambridge, United Kingdom; [2]Merck, Germany; connorfitchett@gmail.com

Platform trials can offer an efficient way to evaluate a wide range of treatments at once, potentially even considering patient subgroups in the allocation and analysis. This is especially applicable to dose-ranging phase II trials, where we have a wide selection of doses and want to find the dose that maximizes efficacy whilst minimizing toxicity: not just simply accepting the maximum tolerated dose. In addition to having a powerful test that controls type one error, we also want our trial to consider patient welfare. Ideally, we would want a trial that gives us statistically valid results whilst allocating more patients to the best treatment.

We consider applying Bayesian Response Adaptive Randomisation (BRAR) to multi-arm platform trials. It prioritizes patient wellbeing by attempting to allocate more patients to the best treatment whilst also allowing us to control for type one error, giving us valid tests. BRAR's ability to handle composite endpoints such as toxicity and efficacy also make it desirable for dose-ranging trials. Ultimately, the flexibility of BRAR and its cohesion with other adaptive designs all whilst maintaining desirable frequentist operating characteristics makes it a powerful technique to consider in platform trial design.

Adding an active dose level to an ongoing randomized controlled phase III study

Lorenz Uhlmann, Johannes Krisam, Daniela Maier, Cornelia Ursula Kunz

Boehringer Ingelheim Pharma GmbH & Co. KG, Germany; lorenz.uhlmann@boehringer-ingelheim.com

Our real-life case study is an open-label, randomized, active-controlled, parallel-group, multi-centre, and multi-regional Phase III trial comparing an active treatment with standard of care (SoC). In the initial design, there was only one dose level planned to be assessed and compared to SoC. The primary endpoint was progression-free survival (PFS) and additional key secondary endpoints comprised overall survival and a patient-related outcome. Furthermore, an interim analysis was planned to assess PFS allowing for an early stop for efficacy or futility. In order to ensure that the type I error is controlled at a one-sided alpha level of 0.025, a testing hierarchy (starting with the primary endpoint followed by the key secondary endpoints) combined with an O'Brien-Fleming type alpha spending function for the interim analysis were implemented.

During the conduct of the study, new information from another internal study investigating the same experimental drug revealed that a higher dose level might lead to an even greater benefit, while the initial dose level was still considered to provide a favorable benefit-risk profile. Thus, the design needed to be updated such that this new dose level could be added without interfering with the running study [1-3]. Different approaches have been discussed based on the following strategies:

1. Add the new dose level as soon as possible, leading to a switch from a two-arm to a three-arm study design.
2. Finalize recruitment as initially planned. Then, add another sub-study under the same protocol with the new dose level, resulting in two separate two-arm sub-studies with two separate SoC arms.

While there are general challenges and complexities that need to be resolved in such an adaptive design, there are advantages and disadvantages to both strategies. In our contribution, we will discuss different scenarios based on the two strategies and their implications. We will also present a flexible Shiny app that was created to facilitate a quick assessment of the power, sample size, and timing of interim/final analyses for a given scenario.

References:

Cohen, D.R., Todd, S., Gregory, W.M. *et al. (2015)* Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. *Trials* **16**, 179 (2015). https://doi.org/10.1186/s13063-015-0697-y

Bennett, M., Mander, A.P (2020). Designs for adding a treatment arm to an ongoing clinical trial. Trials 21, 251 (2020).

Burnett T, König F, Jaki T (2024). Adding experimental treatment arms to multi-arm multi-stage platform trials in progress. Stat Med. 43(18):3447-3462.

Optimizing Dose Selection Designs for Phase 3 Trials

Cornelia Ursula Kunz

Boehringer Ingelheim Pharma GmbH & Co. KG, Germany; cornelia_ursula.kunz@boehringer-ingelheim.com

Phase 2 trials are critical in determining the dose that balances efficacy and tolerability. Ideally, these trials should pinpoint a single candidate dose. However, there are instances where two doses, typically the highest and second highest, emerge as potential options at the conclusion of Phase 2. The highest dose may offer slightly superior efficacy, but it may also be somewhat less tolerated by patients.

In such circumstances, one might contemplate conducting another dose-finding trial. However, given that only two doses remain from the initial pool, an alternative approach could be to advance to a Phase 3 trial with both doses. To conserve development resources and time, various design variants could be explored. Our focus here is on the class of two-stage designs with three arms – a control arm and two treatment arms.

One design strategy could be to implement a two-stage design with unequal randomization as well as the possibility to adapt the randomization at interim. Let the randomization be H:L:C with H, L, and C representing the values for the high dose, the low dose, and the control group respectively. For the first stage, H could be set to a smaller value than L. If at the interim analysis the lower dose does not appear as promising as anticipated, the randomization scheme could be adjusted so that the second stage places more emphasis on the higher dose with the new randomization being H':L':C' with H' larger than L'.

This design presents several extreme cases: for instance, either H' or L' could be set to zero for the second stage of the trial, resulting in a design that permits stopping for futility. Alternatively, one could initiate with a 0:L:C randomization (meaning that no patients are randomized to the higher dose) and allow for H' to be larger than zero (and larger than L') at the second stage. This would mean adding a treatment arm to the ongoing trial. Another variant would be to start with 0:L:C and allow a switch to H':0:C meaning a switch of the treatment arms. The final design variant could also be viewed as two distinct trials with the first trial investigating only the lower dose while the second trial focuses solely on the higher dose.

The various design variants are assessed against each other based on their operating characteristics such as power, expected sample size, and anticipated time to trial completion.

Discussion session on the practical challenge

Wednesday 26th February 11:00 – 12:45

Navigating Interim Analyses and Multiple Endpoints: Handling Accelerated Recruitment and Converging Analyses

Franz Koenig[1], Thomas Jaki[2], Dominic Magirr[3], Martin Posch[1]

[1]Medical University of Vienna, Austria; [2]Universität Regensburg, DE, University of Cambridge, MRC Biostatistics Unit, UK; [3]Novartis, Switzerland; franz.koenig@meduniwien.ac.at

In the presentation we will explore strategic approaches for testing treatment efficacy in adaptive trials with multiple endpoints. First, if EP1 is considered as the single primary endpoint used for confirmatory testing, the full level alpha can be fully allocated to EP1. Furthermore, if the decision at both interim analyses is only to stop for futility based on EP2 and EP3, respectively, then no further adjustment would be warranted for the final analysis of EP1. However, in the design phase an increase of the sample size should be considered to account for potential power loss due to (non-binding) futility stopping depending on the expected effect sizes and correlations between the endpoints. As far as control of the Type 1 Error (T1E) rate is concerned it wouldn´t matter at all, if all three analyses coincide. However, the design becomes trickier if also EP2 and EP3 should be tested in a confirmatory manner. Similarly allowing for early stopping for efficacy would complicate matters. Here we discuss how to use closed testing in combination with group sequential designs. For example, if one of the early endpoints EP2 or EP3 shall be used in the context of drug development for conditional approval and EP1 later on for seeking full marketing authorisation from regulatory agencies such as EMA. We show how to use adaptive graphical procedures to establish tailored testing schemes based on partial conditional error applied together with closed testing (Klinglmüller et al., 2014, Bauer et al., 2016) and discuss tailored error spending strategies for the different endpoints. We will elaborate on how to use adaptive tests relying on the partial conditional error rate principle to address that IA1, IA2 and the final analysis could be separated or (partly) coincide and demonstrate that the procedure controls the Family-Wise Error Rate (FWER) even when the timings of the interim analyses were triggered in a data dependent way. However, depending on the scenario, not all data collected on EP1 can be eventually used for efficacy testing (Magirr et al., 2016).

References:

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.

Klinglmueller, F., Posch, M., & Koenig, F. (2014). Adaptive graph-based multiple testing procedures. *Pharmaceutical Statistics*, *13*(6), 345-356.

Magirr, D., Jaki, T., Koenig, F., & Posch, M. (2016). Sample size reassessment and hypothesis testing in adaptive survival trials. *PloS one*, *11*(2), e0146465.

Adjusting the information fraction by sample-size reassessment

Fredrik Öhrn[1], Tobias Mielke[2]

[1]Johnson and Johnson, Sweden; [2]Johnson and Johnson, Germany; fohrn@its.jnj.com

Group-Sequential designs have been broadly implemented across numerous therapeutic areas over the last decades, with a particularly large uptake in oncology. Those designs allow rejection of the null hypotheses of “no-treatment effect” at any pre-specified interim analysis time, based on prospectively defined alpha-spending rules. In practice concerns have been raised on the choice of the interim analysis time, with regulators pushing back on any interim analyses and in particular on interim analyses conducted early. “Early” is not an absolute statement. Some regulators consider 60% of the targeted primary endpoint information as being too early for rejection of the null hypothesis, while 75% might be considered acceptable. However, it could also be argued that the absolute number of events (or information) is a more relevant metric.

Adaptive event number re-estimation has been considered in the past as a methodology to increase the power in situations where interim results fall short of planning assumptions. We will discuss in this presentation how the adaptive design methodology could be used strategically to start with more aggressive designs, powered for a larger treatment effect and requiring a lower total target number of events. Such an approach would allow for interim analyses to be conducted earlier, while still satisfying the information fraction requirement, and via event re-estimation provide sufficient power in case the aggressive effect assumptions do not hold true. Using this case-study presentation, we plan to trigger a productive discussion on the timing, appropriateness and utility of interim analyses to inform early regulatory decision making, while also emphasizing the importance of clear regulatory design requirements.

Exploitation of multivariate central limit theorems in a closed test procedure for group sequential designs

Moritz Fabian Danzer, Jannik Feld, Andreas Faldum, René Schmidt Rene

University of Münster, Germany; moritzfabian.danzer@ukmuenster.de

Several endpoints may be relevant for the comparison of different treatments. This can lead to a variety of challenges. First and foremost, control of the familywise error rate is required when the rejection of a single null hypothesis alone is sufficient to demonstrate the superiority of a treatment. The endpoints being tested may be on different scales and the speed at which the information arrives is unknown and will possibly differ between endpoints. Different statistical measures may also be of interest for the same endpoint. All of these problems apply in the scenario envisioned. In order to arrive at an efficient, flexible and at the same time easily interpretable solution, a closed test procedure within a group-sequential design is a good option. As was recently shown, the dependency of test statistics across analysis times and different endpoints can also be exploited within such a procedure [1]. In this contribution, we want to show how the aforementioned challenges can be solved within this framework and pay particular attention to how we can exploit the dependency between the test statistics. Within the closed test procedure, we can determine exactly in which aspect the interventions under investigation differ from each other. Endpoint-specific alpha-spending functions can be used to ensure that individual hypotheses are tested exactly when sufficient information has been collected for the respective hypothesis test. For an efficient evaluation, a model-free characterization of the correlation of the test statistics to be used is also required. For this purpose, we can use known results on the joint distribution of rank-based test statistics for different endpoints [2] or the joint distribution of different statistical measures for a survival time endpoint [3].

References:
[1] Anderson KM, Guo Z, Zhao J, Sun LZ. A unified framework for weighted parametric group sequential design. Biometrical Journal 2022; 64(7):1219-1239

[2] Lin DY. Nonparametric sequential testing in clinical trials with incomplete multivariate observations. Biometrika 1991; 78(1):123-131.
[3] Feld J, Faldum A, Schmidt R. Adaptive group sequential survival comparisons based on log-rank and pointwise test statistic. Statistical Methods in Medical Research 2021; 30(12):2562-2581.

Session II: Group sequential designs

Wednesday 26th February 13:30 – 15:15

Group-sequential methods for generalised pairwise comparisons

Stephen Schüürhuis, Frank Konietschke

Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin Berlin, Germany; stephen.schueuerhuis@charite.de

Common nonparametric statistical tests, such as the Brunner-Munzel test, are fundamentally based on pairwise comparisons between observations of a treatment and a control group. Buyse (2010) extended the concept of pairwise comparisons to encompass multiple prioritized outcomes. The resulting methods are essentially based on the effect estimand *Net Treatment Benefit*, which aggregates the contribution of multiple outcomes into a single effect measure and interprets as the net probability of a favorable outcome in the treatment group. By accounting for multiple outcomes, these methods offer a more patient-centered approach for clinical trials.

Careful planning of clinical trials is essential to ensure timely benefits from promising new drugs and the early identification of ineffective treatments. Group-sequential designs are widely recognized for enhancing trial efficiency by incorporating interim analyses, allowing early termination for efficacy or futility while maintaining control over multiplicity. These designs have long been established for continuous, binary, and survival outcomes. More recently, Nowak et al. (2022) extended group-sequential designs to the Brunner-Munzel test, showing that the sequential test statistics asymptotically follow a multivariate normal distribution. However, current group-sequential theory remains focused on effect measures and tests including only a single primary endpoint.

Building on the work by Nowak et al. (2022), we extend their research within the framework of generalized pairwise comparisons introduced by Buyse (2010). This approach enables interim testing in ongoing trials based on multiple prioritized outcomes with varying measurement scales, rather than relying on a single outcome. In this presentation, we demonstrate that test statistics derived from generalized pairwise comparisons asymptotically adhere to the canonical multivariate normal distribution. A simulation study demonstrates that the proposed methods accurately control the type-I error rate, even with rather small sample sizes. By incorporating multiple outcomes, we highlight the potential to increase the power of group-sequential trials. By leveraging data from multiple outcomes, our findings emphasize the utility of these methods in enhancing both statistical power and clinical relevance.

[1] Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Statistics in Medicine, 29(30), 3245-3257.

[2] Nowak, C. P., Mütze, T., & Konietschke, F. (2022). Group sequential methods for the Mann-Whitney parameter. Statistical Methods in Medical Research, 31(10), 2004-2020.

Two-Part Models in Group Sequential Designs for Zero-Inflated Data

Fabrice Lotola Mougeni, Martin Posch, Sonja Zehetmayer

Medical University of Vienna, Austria; fabrice.lotolamougeni@meduniwien.ac.at

In the treatment of worm infections, microfilaria counts are frequently measured as primary outcome. For fixed designs, non-parametric methods like the Mann-Whitney U-test are commonly used to analyze count data. However, a two-part model may be a better alternative when data consists of a mixture of counts and a point mass at zero. This model uses two independent test statistics for the analysis of the count data and the binomial data (zero versus non-zero) and combines them to obtain a p-value. The impact of using test statistics from two-part models on the operating characteristics of group sequential designs with traditional boundaries has received less attention.

This simulation study aims to assess the performance of a two-part model to evaluate operating characteristics for group sequential designs for count data with excess of zeros and variability by applying inverse normal method for the combination of the stage-wise p-values. We compare one treatment to a control arm with an interim analysis after 50% of the preplanned data were observed, using a one-sided test at a significance level of α = 0.025. To adjust for interim looks, we apply O'Brien-Fleming or Pocock boundaries. Zero-inflated data is simulated from a mixture of a Poisson distribution and a mass at the point 0, with varying mixture weights, which are the probabilities that the outcome takes the value 0. Stage-wise p-values are derived using the Mann-Whitney U-test and a two-part model, combining re-sults from the count and binomial parts within a stage via the inverse normal or Fisher's combination method. In the simulation, we explore scenarios where the 2-part model outperforms the Mann-Whitney U-test. It depends on the amount of zero-inflation which procedure has larger power.

Using Machine Learning to Optimize Trial Design

Tom Parke

Berry Consultants LLP, United Kingdom; tom@berryconsultants.com

As trial designs become more complex, they gain parameters and thresholds that need to be set with no simple and obvious method of deriving them. Even for a group sequential trial (one of the oldest and simplest type of adaptive trial), choosing the number and timing of the interim, and the stopping parameter boundaries is largely a matter of custom and practice.

One of the problems of conventional optimization techniques such as simulated annealing and genetic algorithms is that they assume an exact result of evaluating the "function" at any particular set of parameter choices, but trying to optimize a clinical trial design will require simulation and have an approximate solution. Fortunately there are relatively new optimization techniques specifically for optimizing functions with approximate values (such as from simulation) that use Bayesian smoothing, and Python code libraries (in particular "botorch") that implement them.

We will present the results of an experiment in optimizing a group sequential trial design using such an approach.

Accounting for delayed responses in group sequential designs – needed or nice to have?

Stephen Schüürhuis[1], Gernot Wassmer[2], Meinhard Kieser[3], Friedrich Pahlke[2], Cornelia Ursula Kunz[4], **Carolin Herrmann[5]**

[1]Charité University Medicine Berlin, Germany; [2]RPACT GbR, Am Rodenkathen 11, Sereetz, 23611, Germany; [3]University Medical Center Ruprechts-Karls University Heidelberg, Germany; [4]Biostatistics and Data Sciences, Boehringer Ingelheim GmbH & Co. KG, Germany; [5]Heinrich Heine University Düsseldorf, Germany; carolin.herrmann@hhu.de

In some group sequential trials, outcomes are observed immediately. In others, there is a time gap between patient enrollment and the measurement of the outcome. Methods that account for such delayed responses were introduced some time ago (Hampson and Jennison, 2013). Nevertheless, methods for non-delayed responses are primarily applied. In this talk, we present the results of our recently conducted method comparison. In the work, we compared designs that account for delayed responses and those that do not for scenarios where an outcome delay is prevalent (Schüürhuis et al. 2024). We discuss appropriate performance criteria and describe the simulation study, which was based on the R package rpact (Wassmer and Pahlke, 2024, as well as see https://fpahlke.github.io/gsdwdr/ for the code of the simulation study). Specifically, we consider non-binding futility stopping boundaries and two-stage designs. Performance is found to be dependent on the sample size at interim and the amount of data in the pipeline. Overall, our results highlight the importance of practical considerations regarding operational feasibility.

References:

Hampson, L. V., & Jennison, C. (2013). Group sequential tests for delayed responses (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *75*(1), 3-54.

Schüürhuis, S., Wassmer, G., Kieser, M., Pahlke, F., Kunz, C. U., & Herrmann, C. (2024). Two-stage group-sequential designs with delayed responses–what is the point of applying corresponding methods?. *BMC Medical Research Methodology*, *24*(1), 242.

Wassmer G, Pahlke F (2024). *rpact: Confirmatory Adaptive Clinical Trial Design and Analysis*. R package version 4.0.0, https://CRAN.R-project.org/package=rpact.

Evaluating the impact of outcome delay on adaptive designs

Aritra Mukherjee[1], Michael Grayling[2], James Wason[1]

[1]Newcastle University, United Kingdom; [2]Johnson and Johnson; aritra.mukherjee@newcastle.ac.uk

Adaptive designs (AD) are a broad class of trial designs that allow pre-planned modifications to be made to a trial as patient data is accrued, without undermining its validity or integrity. ADs can lead to improved efficiency, patient-benefit, and power of a trial. However, these advantages may be affected adversely by a delay in observing the primary outcome variable. In the presence of such delay, a choice must be made between (a) pausing recruitment until requisite data is accrued for the interim analysis, leading to longer trial completion period; or (b) continuing to recruit patients, which may result in a large number of participants who do not benefit from the interim analysis. In the latter case, little work has investigated the size of outcome delay that results in the realised efficiency gains of ADs being negligible compared to classical fixed-sample alternatives. Our study covers different kinds of ADs and the impact of outcome delay on them.

We assess the impact of delay on the expected efficiency gains of an AD by estimating the number of pipeline patients being recruited in the trial under the assumption that recruitment is not paused while we await treatment outcomes. We assume different recruitment models to suitably adjust for single- or multi-centred trials. We discuss findings for two-arm group-sequential designs as well as multi-arm multi-stage designs. Further, we focus on sample size re-estimation (SSR), a design where the variable typically optimized to characterise trial efficiency is not the expected sample size (ESS).

Our results indicate that if outcome delay is not considered at the planning stage of a trial, this can translate to much of the expected efficiency gains being lost due to delay. The worst affected designs are typically those with early stopping, where the efficiency gains are assessed through a reduced ESS. SSR can also suffer adversely if the initial sample size specification was largely over-estimated.

Finally, in light of these findings, we discuss the implications of using the ratio of the total recruitment length to the outcome delay as a measure of the utility of different ADs.

Session III: Platform trials

Wednesday 26th February 15:35 – 17:40

Online control of the Family-wise error rate (FWER) for multi-arm multi-stage platform trials

Jiangyue Yao[1], David Robertson[2], Thomas Jaki[1,2]

[1]Faculty of Informatics and Data Science, University of Regensburg, Regensburg, Germany; [2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; jiangyue.yao@ur.de

Multiple treatment arms can enter or leave the study anytime in platform trials. Online testing procedures, where multiple hypotheses will be tested over time and the total number of hypotheses can remain unknown, are suitable in this case for error rate control. Interim analyses in a multi-stage design allow us to test the same hypothesis at different time points, giving the opportunity of early stopping for success or futility in clinical trials. Different online testing methods targeting the false-discovery rate (FDR) have been proposed for platform trials [1]. Methods that control for FDR in multi-arm multi-stage platform trials have also been proposed [3]. However, it is not clear how to control the FWER when multiple looks at the data of overlapping treatments are undertaken in platform trials. ADDIS (Adaptive discarding alpha-spending algorithm) is a method that can control the FWER in an online setting [2]. We modify the spending algorithm so that it is possible to add interim analyses to each hypothesis in platform trials. The developed method updates the error level of the hypothesis at all (interim) analyses when any decision on hypotheses has been made. We compare alternative approaches with our method in a simulation study.

References [1] David S Robertson, James MS Wason, Franz Koenig, Martin Posch, and Thomas Jaki. Online error rate control for platform trials. Statistics in medicine, 42(14):2475–2495, 2023. [2] Jinjin Tian and Aaditya Ramdas. Online control of the familywise error rate. Statistical methods in medical research, 30(4):976–993, 2021. [3] Sonja Zehetmayer, Martin Posch, and Franz Koenig. Online control of the false discovery rate in group-sequential platform trials. Statistical Methods in Medical Research, 31(12):2470–2485, 2022.

Adjusted treatment effect estimators for platform trials with interim analyses utilizing non-concurrent controls

Pavla Krotka[1], Martin Posch[2], Marta Bofill Roig[1]

[1]Universitat Politècnica de Catalunya, Barcelona, Spain; [2]Medical University of Vienna, Vienna, Austria; pavla.krotka@upc.edu

Platform trials evaluate the efficacy of multiple treatment arms within a single infrastructure, with treatment arms entering the trial over time as new treatments become available. Interim analyses are often included to further accelerate drug development by allowing for early stopping for futility or efficacy. Treatment arms are usually compared to a shared control arm. For arms entering later, the control data is divided into concurrent and non-concurrent controls (NCC), referring to control patients recruited while the given arm is in the platform, and before it enters, respectively. Analysis using NCC can reduce the required sample size and increase power, but might also lead to bias in the effect estimates and hypotheses tests.

For platform trials with continuous endpoints, a regression model with categorical time adjustment has been proposed to utilize NCC. The time trends are adjusted for by including the factor “period” as a fixed effect, with periods being defined as time intervals bounded by any treatment arm entering or leaving the platform. In trials without interim analyses, this model leads to unbiased effect estimates and asymptotically controls the type I error rate regardless of the time trend pattern, if the time trend affects all arms equally and is additive on the model scale.

In this work, we show that, in group-sequential platform trials using NCC, the currently available regression model leads to a loss of the type I error rate control and bias in the effect estimators. In addition, we describe how the weight of the non-concurrent controls in the treatment effect estimator is stochastically dependent on the outcome in the non-concurrent controls. We will quantify the bias in the point estimation and maximum type I error rate inflation when applying the model with period adjustment in platform trials with interim looks. Moreover, we will investigate adjusted treatment effect estimators that aim to eliminate or reduce the potential bias and resulting type I error rate inflation. Focusing on a simple platform trial with two experimental treatment arms and a continuous outcome, we will present results from a simulation study, where we evaluate the performance of the considered approaches and compare them to current methods.

Unconditional treatment effect estimates when adjusting for time in platform trials with binary endpoints

Marta Bofill Roig[1], Ekkehard Glimm[2], Kelly Van Lancker[3], Martin Posch[4]

[1]Universitat Politècnica de Catalunya, Spain; [2]Novartis Pharma AG; [3]Ghent University; [4]Medical University of Vienna; marta.bofill.roig@upc.edu

Platform trials are multi-arm, multi-stage trials in which treatment arms are allowed to enter or leave at different times during the trial duration. Experimental treatments are usually compared against a common control. Since these trials are often conducted over long periods of time, time trends may occur. For that reason, analyses must be adjusted for time to avoid bias. This adjustment is particularly critical in trial analyses that incorporate non-concurrent controls.

The use of covariate adjustment methods has been recently discussed to adjust for prognostic baseline variables in the analysis of randomized, parallel-group clinical trials. When adjusting for baseline covariates, the estimand should be correctly specified by distinguishing between conditional and marginal treatment effects. Different estimation strategies can be considered beyond unadjusted analysis when considering the unconditional estimand. In the context of randomized clinical trials, covariate-adjusted estimators of unconditional treatment effects have been proposed, such as inverse probability of treatment weighting and model-based standardization (G-computation).

Assuming a platform trial with a primary binary endpoint, we consider logistic regression adjusting for time trends by adding a covariate for period in the model, where periods are time intervals defined whenever an arm enters or leaves the trial. We discuss the appropriate choice of estimands for trials that adjust for time in the analysis and compare these estimands to those used in traditional (fixed) clinical trials, which typically focus on marginal estimands without conditioning on time. Aiming to distinguish estimators that target conditional and unconditional estimands, we approach the estimation using adjusted logistic regression with and without considering G-computation for estimating the average treatment effect.

When is a multi-armed trial a platform trial and what are the implications?

Tobias Mielke[1], Fredrik Oehrn[2]

[1]Johnson & Johnson, Germany; [2]Janssen Cilag, Sweden; tmielke1@its.jnj.com

Various definitions of platform trials are floating around in the scientific community. The EU-PEARL definition of a platform trial is based on Woodcock and Lavange (2017) and the FDA Oncology master protocol guidance (2022) as: “To study multiple targeted therapies in the context of a single disease in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a decision algorithm”. As such, any simple multi-armed trial comparing two interventions A and B vs. a common control C in the context of a single disease, would be regarded as a platform trial, if therapies are allowed to enter or leave the trial. Based on the work by Howard et al. (2018), control for multiplicity has been considered under certain situations as being not required in platform trials by regulatory agencies (FDA draft guidance, 2023) or some staticians at regulatory agencies (EMA paper, 2020). The argument for “no correction” is that the FWER of conducting trials of A vs. C and B vs. C separately would exceed the FWER of a single non-adjusted multi-armed trial. The common argument for correction is that the chance of conducting two errors would be inflated with a single trial, thereby decreasing confidence in regulatory decision making. Given the referenced guidance on master protocols, there seems to be broad agreement that no correction is deemed necessary, if hypotheses are inferentially independent, while the definition of interential independence appears not entirely well-defined. Concerns are being raised in practice when it comes to implementing this guidance for conventional multi-armed trials conducted by a single sponsor. This could be considered as the simplest platform trials possible. Using this presentation, which is motivated by an existing case study, we will discuss inconsistencies in regulatory advice on multiplicity control for multi-armed vs. platform trials. Our aim is to trigger a clarifying discussion on the eventual need and acceptable level of error rate control for multi-armed trials with inferentially independent hypotheses.

Improving implementation of Adaptive Multi-Arm Multi-Stage Trials: The MAMS R Package

Nikita Mozgunov[1], Dominique-Laurent Couturier[1], Michael Grayling[2], Dominic Magirr[3], Philip Pallmann[4], Thomas Jaki[1,5]

[1]University of Cambridge, United Kingdom; [2]The Janssen Pharmaceutical Companies of Johnson & Johnson; [3]Novartis Pharma AG; [4]Cardiff University, United Kingdom; [5]University of Regensburg, Germany; nikita.mozgunov@mrc-bsu.cam.ac.uk

Multi-Arm Multi-Stage (MAMS) trials are adaptive designs that allow the simultaneous evaluation of multiple experimental treatments against a common control over several stages with interim analyses. They offer significant efficiency gains in identifying promising interventions by incorporating adjustments such as early stopping for futility or efficacy. However, the complex statistical methods required - particularly in relation to multiple testing procedures to control for family-wise error rates - present challenges in their design, analysis and implementation.

We present significant advances in statistical methodology and software implementation for MAMS trials through the development of the new version of MAMS R package. It uses a modular approach that allows easy integration of different adaptive methods, including simultaneous stopping, drop-the-losers design, and separate stopping rules for platform trials.

To facilitate practical application, the MAMS R package provides a user-friendly interface accessible to clinicians and trialists with minimal programming experience. It provides functions for calculating key operating characteristics - such as power, expected sample sizes and stopping probabilities. Visualisation are integrated to help users interpret and effectively communicate design characteristics to stakeholders.

Comprehensive documentation, tutorials and online resources are provided to help users understand and apply the advanced methods. By integrating sophisticated statistical methods into accessible software, this work promotes the wider adoption of adaptive MAMS studies in clinical research. MAMS package enables researchers to confidently design and analyse complex adaptive trials, ensuring statistical validity and ethical integrity.

By simplifying the application design and improving the communication, MAMS package addresses key practical challenges in implementing adaptive designs in real-world settings. This will facilitate wider adoption of MAMS trials and ultimately accelerate the identification of effective treatments.

A Fair and Efficient Randomization Scheme for Multi-Arm Seamless Two-Phase Clinical Trials

Peter Jacko[1,2]

[1]Berry Consultants, United Kingdom; [2]Lancaster University, United Kingdom; peter.jacko@gmail.com

Multi-arm and platform clinical trials designed to span Phase II and Phase III have recently gained popularity. Such two-phase designs bring administrative and operational benefits compared to separate trials for each phase, which would normally need to wait for the Phase II to fully complete for all arms (doses, treatments, etc.) before starting to design and set up the phase III trial. They also have potential for a range of inferential benefits, where data from the earlier phase may be used, depending on the particular trial objective and therapeutic area, to, for example: strengthen the control dataset by using non-concurrent controls, speed up the start of the Phase III for a particular arm without the need for the other arms to complete their Phase II, build a longitudinal model for the outcomes allowing for interim decisions, etc. Response-adaptive methods such as permanent arm dropping (using futility thresholds) or temporary arm dropping (using response-adaptive randomization) are often used in such designs in order to more effectively use patient resources, to more effectively treat trial participants and to identify better arms more rapidly. A commonly cited downside of such two-phase trials is the potential estimation bias when pooling data from the two phases when there are different randomization ratios in the two phases. A commonly cited downside of response-adaptive methods in multi-arm trials where the arms have different sponsors is the sponsors' concern that if early data is unfavourable their arm may not get a chance to enrol the desired number of participants overall or enrol at the desired rate. In this talk we present a novel randomization scheme which is not only operationally and inferentially seamless but also allocationally seamless, as both phases of each arm start from the moment the arm joins the trial. The scheme delivers the desired fairness properties between the arms in Phase II as well as mitigates the estimation bias by keeping the randomization ratio constant for the data to be used in the final analysis of Phase III.

Session IV: Platform trials

Thursday 27th February 08:30 – 10:35

Revisiting Optimal Proportions for Binary Responses: Insights from Incorporating the Absent Perspective of Type-I Error Rate Control

Lukas Pin[1], William F Rosenberger[2], Sofía S. Villar[1]

[1]University of Cambridge, UK; [2]George Mason University, USA; lukas.pin@mrc-bsu.cam.ac.uk

This work revisits optimal response-adaptive designs from a type-I error rate perspective, highlighting how and when these allocations exacerbate type-I error rate inflation — an issue previously undocumented. We explore a range of approaches from the literature that can be applied to reduce type-I error rate inflation. However, we found that all of these approaches fail to give a robust solution to the problem. To address this, we derive two optimal proportions, incorporating the more robust score test (instead of the Wald test) with finite sample estimators (instead of the unknown true values) in the formulation of the optimization problem. One proportion optimizes statistical power and the other minimizes the total number failures in a trail while maintaining a predefined power level. Simulations of an early-phase and confirmatory trial provide practical insight into how these new proportions control type-I error rate effectively where other methods fail. While we focused on binary outcomes, the framework offers valuable insights that could be extended to other outcome types, multi-armed trials and alternative measures of interest.

Response Adaptive Randomization – when trying to select the best arm/dose from those being tested

Tom Parke

Berry Consultants LLP, United Kingdom; tom@berryconsultants.com

Using RAR in the multi-arm setting has the more to offer in efficiency gains, but strangely RAR in the two-arm setting seems to have most publications about it. We will present a simple and replicable simulation that shows the degree to which RAR in a multi-arm setting reduces the multiplicity introduced by testing multiple arms. We will also look at how much, in this simple setting, RAR increases bias and risk of wrong arm selection (problems that have been raised in the literature). Lastly we will look at how these results vary depending on the number and timing of the interim looks.

Exact statistical analysis for response-adaptive clinical trials: A general and computationally tractable approach

Stef Baas[1], Peter Jacko[2,3], Sofía S. Villar[4]

[1]University of Twente; [2]Lancaster University; [3]Berry Consultants; [4]University of Cambridge; stef.baas@mrc-bsu.cam.ac.uk

Clinical trials using a response-adaptive (RA) design adjust the allocation of participants to treatments sequentially based on the observed (response) data so far. This is done to reach a certain objective such as increasing expected patient outcomes or statistical power. Proposals for RA designs of clinical trials face greater scrutiny in reviews by regulatory agencies, partly due to concerns surrounding type I error inflation, which can occur for standard tests under currently used RA designs. Furthermore, commonly used test approaches incorporating the RA design have many limitations: they either only work for specific designs and in specific scenarios, have a risk of model misspecification or Monte Carlo error, are conservative, or are computationally intractable.

This paper focuses on type I error control for statistical tests on binary outcomes collected from a trial with a control and treatment group using any RA procedure, either randomized or deterministic. We develop a general approach to construct conditional exact tests for RA designs, extending Fisher's exact test, and an unconditional exact test for RA designs, generalizing Barnard's test. An efficient implementation of forward recursion is used to compute the critical value and operating characteristics for clinical trials with large trial sizes (up to around 1,000 participants on a standard computer).

In the paper, we compare several exact tests on data collected using the randomized dynamic programming RA procedure and consider two real-life applications. The talk will focus on the second application, where we re-analyse a trial that used a blocked Bayesian group-sequential RA design and show that under the assumed optional stopping threshold, there is substantial type I error inflation under misspecification of the success probability, while the proposed tests exactly control type I error.

In conclusion, our method provides a general and computationally tractable method to ensure robust type I error control in complex and real-life RA designs. An important observation is that while the unconditional exact test is often the exact test resulting in the highest power for non-response-adaptive designs with equal allocation, our results show that in non-group-sequential RA designs a conditional exact test often provides a more powerful alternative.

Patient-oriented response-adaptive designs based on a novel information measure in multi-arm trials with quantitative endpoints

Gianmarco Caruso, Pavel Mozgunov

MRC Biostatistics Unit, University of Cambridge, UK; gianmarco.caruso@mrc-bsu.cam.ac.uk

Multi-arm trials are gaining interest in practice given the statistical and logistical advantages that they can offer. The standard approach is to use the fixed (throughout the trial) allocation ratio, but there is a call for making it adaptive and skewing the allocation of patients towards better performing arms. This is motivated by the goal of providing the most benefit to the patients in the trial while maximizing the information collected on the most promising arms. However, among other challenges, it is well-known that these approaches might suffer from low statistical power. We present a response-adaptive design which explicitly allows to control the trade-off between the number of patients allocated to the "optimal" arm and the statistical power. Such a balance is controlled through the calibration of a tuning parameter, and we explore various strategies to effectively perform it. We consider the general setting of a normally distributed endpoint and the design that targets a desirable value of that endpoint. The proposed allocation rule naturally arises from a weighted version of Shannon's differential entropy, a context-dependent information measure which gives a greater weight to those treatment arms which have characteristics close to the pre-specified clinical target. We also introduce a simulation-based hypothesis testing procedure to assess whether the best performing treatment arm is significantly superior to the second best. This emphasizes a primary focus on selecting the optimal arm rather than on a comparison to the control. A simulation study highlights the potential advantage of the proposed class of designs over the considered competitors in an early Phase IIa proof-of-concept oncology clinical trial.

Covariate-Adjusted Response Adaptive Designs for Semiparametric Survival Models

Ayon Mukherjee

Merck KGaA, Germany; ayon.mukherjee@merckgroup.com

Covariate-adjusted response-adaptive (CARA) designs use the available responses to skew the treatment allocation towards the treatment found to be best at an interim stage of a clinical trial, for a given patient's covariate profile. There has recently been extensive research on CARA designs with parametric distributional assumption on the patient responses. However, the range of application for such designs become limited in real clinical trials. Sverdlov,Rosenberger and Ryzenik (2013) has pointed out that irrespective of a specific parametric form of the survival outcomes, their proposed CARA designs based on the exponential model provide valid statistical inference, provided the final analysis is performed using the appropriate accelerated failure time (AFT) model. In real survival trials, however, the planned primary analysis is rarely conducted using an AFT model. The proposed CARA designs are developed obviating any distributional assumptions about the survival responses, relying only on the proportional hazards assumption between the two treatment arms. To meet the multiple experimental objectives of a clinical trial, the proposed designs are developed based on optimal allocation approach. The covariate-adjusted doubly-adaptive biased coin design and the covariate-adjusted efficient randomised adaptive design are used to randomise the patients to achieve the derived targets on expectation. These expected targets are functions of the Cox regression coefficients that are estimated sequentially with the arrival of every new patient into the trial. The merits of the proposed designs are validated using extensive simulation studies assessing their operating characteristics and has also been implemented to re-design a real-life confirmatory clinical trial.

Design and inference for multi-arm clinical trials with informational borrowing: the interacting urns design

Rosamarie Frieri[1], Alessandro Baldi Antognini[1], Giacomo Aletti[3], Andrea Ghiglietti[4], Irene Crimaldi[2]

[1]University of Bologna, Italy; [2]IMT Lucca; [3]University of Milan; [4]Bicocca University; rosamarie.frieri2@unibo.it

We propose a new design methodology for stratified comparative experiments based on interacting reinforced urn systems. The key idea is to model the interaction between urns for borrowing information across strata and to use it in the design phase in order to i) enhance the information exchange at the beginning of the study, when only few subjects have been enrolled and the stratum-specific information on treatments' efficacy could be scarce, ii) let the information sharing adaptively evolves via a reinforcement mechanism based on the observed outcomes, for skewing at each step the allocations towards the stratum-specific most promising treatment and iii) make the contribution of the strata with different treatment efficacy vanishing as the stratum information grows. In particular, we introduce the Interacting Urns Design, namely a new Covariate-Adjusted Response-Adaptive procedure, that randomizes the treatment allocations according to the evolution of the urn system. The theoretical properties of this proposal are described and the corresponding asymptotic inference is provided.

Session V: Estimation

Thursday 27th February 11:05 – 12:50

Covariate Adjustment in Bayesian Adaptive Randomized Controlled Trials

James Willard[1,2], Shirin Golchi[2], Erica Moodie[2]

[1]University of Cambridge, United Kingdom; [2]McGill University, Canada; james.willard@mrc-bsu.cam.ac.uk

In conventional randomized controlled trials, adjustment for baseline values of covariates known to be at least moderately associated with the outcome increases the power of the trial. Recent work has shown particular benefit for more flexible frequentist designs, such as information adaptive and adaptive multi-arm designs. In this work, we investigate the impact of covariate adjustment on flexible Bayesian adaptive designs, focusing on trials which allow for early stopping at an interim analysis given evidence of treatment superiority. We consider both collapsible and non-collapsible estimands, and show how to obtain posterior samples of marginal estimands from adjusted analyses. Through simulation, it is shown that covariate adjustment increases power and the probability of stopping the trials early, and decreases the expected sample sizes as compared to unadjusted analyses.

Confidence intervals for adaptive designs

David Robertson[1], Thomas Burnett[2], Babak Choodari-Oskooei[3], Munya Dimairo[4], Michael Grayling[5], Philip Pallmann[6], Thomas Jaki[1,7]

[1]MRC Biostatistics Unit, University of Cambridge, UK; [2]University of Bath, UK; [3]MRC Clinical Trials Unit at UCL, UK; [4]School of Health and Related Research (ScHARR), University of Sheffield; [5] Johnson & Johnson Innovative Medicine; [6]Centre for Trials Research, Cardiff University; [7]University of Regensburg, Germany; david.robertson@mrc-bsu.cam.ac.uk

Regulatory guidance notes the need for caution in the interpretation of confidence intervals (CIs) constructed after an adaptive clinical trial because such CIs will often have incorrect coverage (among other potential undesirable properties). In this talk, we discuss the potential problems with using standard CIs for adaptive designs and assess the negative impact this can have. We review the currently available methods to construct adjusted CIs, highlighting the advantages and disadvantages that different methods have, and illustrating their computation using a real adaptive trial example. Finally, we offer some proposed guidance around the choice and reporting of CIs following an adaptive design.

The uniformly most powerful conditional unbiased test and conditional confidence interval in two-stage adaptive enrichment designs

Enyu Li[1], Nigel Stallard[1], Ekkehard Glimm[2], Dominic Magirr[2], Peter Kimani[1]

[1]Warwick Medical School, Coventry, United Kingdom; [2]Novartis Pharma AG, Basel, Switzerland; u5500885@live.warwick.ac.uk

Background: With the deepening understanding of underlying biology, it has been recognized that the effect of a treatment can be heterogeneous among patient subpopulations due to the different pathogenic mechanisms. Two-stage adaptive enrichment designs have been proposed as an efficient approach for subgroup analysis which accounts for the treatment heterogeneity. In stage 1, patients are recruited from the full population. Then, the subpopulation that most appears to benefit most from the experimental treatment is selected by an interim analysis based on stage 1 outcomes data. In stage 2, patients are only recruited from the selected population. Data from both stages are used in the final confirmatory analysis. The selection nature of adaptive enrichment designs poses statistical challenges regarding inference for treatment effects. In this work, we develop hypothesis test and confidence intervals that adjust for adaptive enrichment designs.

Method: We consider the adaptive enrichment design proposed by Kimani et al. [1]. Building on the work of Sampson and Sill [2] and statistical theory in [3], we derive the uniformly most powerful conditional unbiased (UMPCU) tests and construct confidence intervals by inverting the UMPCU test for the treatment effect of the selected subpopulation.

Result: Through an intensive simulation study, we verify that, conditional on the selection made, the UMPCU test controls the type I error rate (as is implied by theoretical results). Furthermore, the simulations show that the UMPCU test is more powerful than type-I-error controlling test procedures only using stage 2 data. We also found that methods using the closure principle and p-value combination functions fail to control the type I error in the conditional perspective. Moreover, we demonstrate that UMPCU confidence intervals achieve coverage probability at the nominal level. Using this metric, they outperform existing methods, including double bootstrap confidence intervals and duality confidence intervals. In summary, our proposed inference methods show a superior capability to correct selection bias over existing methods.

Reference:

[1] Kimani, P. K., Todd, S., and Stallard, N. (2015), Estimation after subpopulation selection in adaptive seamless trials. Statist. Med., 34, 2581–2601. https://doi.org/10.1002/sim.6506

[2] Sampson, A.R. and Sill, M.W. (2005), Drop-the-Losers Design: Normal Case. Biom. J., 47: 257-268. https://doi.org/10.1002/bimj.200410119

[3] Lehmann, E.L. and Romano, J.P. (2010). Testing statistical hypotheses. New York: Springer.

Pre-specification and Bias in Hybrid RCTs

Han Chang Chiam, Franz König, Martin Posch

Medical University of Vienna, Austria; han.chiam@meduniwien.ac.at

Hybrid Control RCTs augment data from randomized controlled trials with external controls, including historical or concurrent control data. A wide range of frequentist and Bayesian methods, as the Meta-Analytic-Predictive Prior approach have been proposed to adjust for potential confounding. While these methods cannot guarantee strict type 1 error rate control, they can mitigate biases if the external controls differ systematically from the data in the RCT. A problem that received less attention so far, is the issue of pre-specification of the analysis.

The integrity of statistical conclusions in hybrid RCTs relies heavily on the assumption that the selection of external controls and the design of the prospective RCT is conducted independently of the trial data. However, information on the historical data may be available at the planning stage of a hybrid control trial and therefore may introduce biases, especially when this information influences design decisions, such as the selection of the historical data sources or the sample size of the prospective part of the hybrid trial.

In a simulation study we quantify the bias that may be introduced into Bayesian and Frequentist Hybrid Control trials by an outcome dependent selection of historical controls in a range of scenarios and discuss potential strategies to mitigate this bias. We explore scenarios where the selection of historical trials is influenced by factors such as the variability between studies, the size of the historical study pool, and the number of selected studies. Variability between studies is introduced by differences in sample size and the magnitude of drift across historical controls. We also assess the role of time trends in the outcomes over time (e.g., due to changes in the standard of care). By simulating a known time trend, we illustrate how different approaches handle this time-related bias and provide insights into pre-specification criteria for hybrid control RCTs.

Exploratory Adaptive Enrichment Design with Sample Size Re-estimation: A Novel Approach for Clinical Trial Optimization

Elad Berkman, Raviv Pryluk, Tzviel Frostig, Oshri Machluf, Amitay Kamber, Keren Peri-Hanania, Anna Aizik

PhaseV Trials Inc.; raviv@phasevtrials.com

We present a novel adaptive enrichment design (AED) that integrates sample size re-estimation with data-driven subgroup identification for clinical trials. Our approach distinguishes itself by adjusting for selection bias through a bootstrap-based debiasing scheme. Unlike conventional methods relying on conditional power calculations, our technique uses conditional assurance with non-informative priors to estimate the probability of trial success, thereby accounting for parameter estimation uncertainty. The incorporation of sample size re-estimation adds flexibility while maintaining control over Type I error through simulations of the combination test under the null hypothesis (assuming no signal).

Simulation studies based on a Phase II osteoarthritis trial demonstrate that our approach substantially enhances trial efficiency. With an interim analysis of 80 patients, our design achieves power increases of up to 23% compared to fixed designs, especially in scenarios with high treatment effect heterogeneity. The method effectively identifies responsive subgroups while preserving unbiased effect estimates across various heterogeneity levels. Notably, the design robustly controls Type I error rates across a range of decision thresholds, making it particularly suitable for exploratory Phase II trials where decision-making flexibility is essential.

This method advances adaptive trial design by offering a balanced framework of statistical rigor and practical flexibility, providing trial sponsors with a powerful tool for efficient drug development in heterogeneous populations.

Session VI: Multiple testing

Thursday 27th February 13:35 – 15:20

Adding baskets to an ongoing basket trial with information borrowing

Libby Daniells[1], Pavel Mozgunov[1], Helen Barnett[2], Alun Bedding[3], Thomas Jaki[1,4]

[1]MRC Biostatistics Unit, University of Cambridge, United Kingdom; [2]Department of Mathematics & Statistics, Lancaster University, United Kingdom; [3]Roche Products Ltd, Welwyn Garden City, United Kingdom; [4]Faculty of Informatics and Data Science, University of Regensburg, Germany; libby.daniells@mrc-bsu.cam.ac.uk

Innovation in trial designs has led to the development of basket trials in which a single therapeutic treatment is tested in several patient populations, each of which forms a basket. This trial design allows for the testing of rare diseases or subgroups of patients. However, limited basket sample sizes can cause a lack of statistical power and precision of treatment effect estimates. This is tackled through the use of Bayesian information borrowing.

To provide flexibility to these studies, adaptive features are desirable as they allow for pre-specified modifications to an ongoing trial. In this talk we focus on the incorporation of (a) newly identified basket(s) part-way through a study. We propose and compare several approaches for adding new baskets to an ongoing basket trial under an information borrowing structure and highlight when it is beneficial to add a new basket to an ongoing trial as opposed to running a separate investigation for them. We also propose a novel calibration for the decision criteria in basket trials that is robust with respect to false decision making. Results display a substantial improvement in power for a new basket when information borrowing is utilised, however, this comes with potential inflation of error rates. This inflation is reduced under the novel calibration procedure.

Optimising graph-based multiple testing procedures by incorporating clinical considerations into flexible power objectives for FWER control

Alex Spiers, Adrian Mander

GSK, United Kingdom; alex.d.spiers@gsk.com

Regulators mandate strong familywise error rate (FWER) control for registrational trials, particularly when multiple claims, endpoints, or subgroups are targeted for product labelling. For a new treatment to be commercially viable, multiple hypotheses often need to be rejected in a confirmatory trial to demonstrate clinical benefit over existing therapies. The challenge is to find multiple testing procedures (MTPs) that maximise the probability of trial success while controlling FWER.

Previous literature on optimal MTPs focus on general power metrics like conjunctive power, disjunctive power, and expected number of rejections; but these metrics often fail to capture the complexity of clinical and commercial priorities. Trial objectives may be best represented by composite functions of power, such as meeting secondary hypotheses conditional on the rejection of primary hypotheses.

This talk introduces a method to optimise graph-based MTPs tailored to specific trial needs, flexible enough to optimise complex clinical and commercial composite power objectives. The method accommodates sources of multiplicity from doses, subgroups and/or trial design. Implemented in a novel internal R package, *multigrain*, this method exploits nonlinear optimisation methods to find graph-based MTPs that optimise user-defined power objectives and trial winning conditions while controlling FWER.

We will showcase the flexibility and value of *multigrain* through case studies from pharmaceutical trials, comparing attained power and sample size against traditional methods. By optimizing MTPs based on objective functions aligned with clinical priorities, *multigrain* enables sponsors to use more relevant metrics for assessing the probability of success, leading to increased likelihood of achieving trial objectives.

Optimizing Endpoint Analysis in the context of Kidney Transplant Studies: Composite vs. Multiplicity-Corrected Approaches

Felix Herkner[1,2], Martin Posch[1], Gregor Bond[2], Franz König[1]

[1]Center for Medical Data Science, Medical University of Vienna, Vienna, Austria; [2]Division of Nephrology and Dialysis, Department of Medicine III, Medical University of Vienna, Vienna, Austria; felix.herkner@meduniwien.ac.at

In the context of kidney transplant trials, e.g. the TTV GUIDE IT, a randomized controlled multi-center study investigating a novel biomarker for guiding immunosuppression, several events (e.g., infection, rejection, graft loss and death) reflecting distinct failures of treatment are of importance. Options to handle multiple endpoints, as discussed in the FDA guidance on use of multiple endpoints in clinical trials, include the use of composite endpoints defined as the first of any of the single outcomes to occur. Alternatively, testing of single components as multiple endpoints applying multiplicity correction can be performed, e.g., embedded in a tailored closed testing strategy using weighted tests. We will discuss the pro and cons of different strategies. It will be shown how to modify the testing schemes if certain restrictions are required for the individual endpoints, e.g., no opposing effect in some of them. The operating characteristics will be compared by clinical trial simulations using different assumptions of the (joint) distributions of the single endpoints concerned. The aim of the simulation study is to assess the impact on power of different analysis strategies (e.g., binary or time to event) in various scenarios concerning the distribution of single endpoints and dropout rates.

Extensions to a closed testing procedure for assessing efficacy in a prespecified subpopulation

Michael Grayling, Yevgen Tymofyeyev

Johnson & Johnson; mgraylin@its.jnj.com

In many clinical trials, an expectation of differential treatment effects in specific patient groups leads to a desire to test for efficacy in the whole population as well as one or more predefined subpopulations. If subgroup analyses are to be performed, a method is required in confirmatory settings to ensure control of the family-wise error rate. One such popular method for this is that described by Song and Chi (2007) [DOI:10.1002/sim.2825], which leverages the correlation between specific subpopulation and population level test statistics to create a testing procedure that can have high power.

Song and Chi described implementation to trials with a single endpoint and a single analysis. Here, we describe how their approach can be readily extended to facilitate the inclusion of interim analyses and/or secondary endpoint(s) using established methods for closed testing. Furthermore, as Song and Chi's proposal requires prespecification of a weighting factor, we also assess the impact on power from misspecification of this weight. Observing the potential for substantial power loss, we detail alternative testing options that are more robust, with the cost of at most minimal inflation to the family-wise error rate. We also discuss optimality considerations regarding Song and Chi's method, describing how the power for an associated intersection hypothesis test may be improved.

Finally, we evaluate how particular trial assumptions influence whether Song and Chi's method can increase the chance of achieving success in *either* the sub or the whole population. The value of subgroup analyses is clearly linked to (a) the proportion of the total dataset that comes from the subpopulation of interest and (b) the treatment effect in the subpopulation vs. its complement. Using our work, some general guidance is provided on performing subgroup analyses.

Sequential conduct of clinical trials

Leonhard Held

University of Zurich, Switzerland; leonhard.held@uzh.ch

The two-trials rule for drug approval requires two significant pivotal trials and is the standard regulatory requirement to provide evidence for the efficacy of a new drug. However, there is need to develop suitable alternatives to this rule for a number of reasons, including the possible availability of data from more than two trials. Here I consider a sequential conduct of three trials, which allows not only to stop for success of failure after two trials, but also to continue and conduct a third trial (Held, 2024). I will concentrate on Edgington's p-value combination method (Edgington, 1972) to ensure sufficient partial Type-I error control while maintaining the overall Type-I error rate at the two-trials rule benchmark of alpha^2 (Rosenkranz, 2023). Several stopping rules will be described following the theory of group-sequential methods (Held, 2024, Section 4) and more recent work will be presented to determine optimal characteristics in terms of project power and the expected number of trials. The results also have relevance for the sequential conduct of replication studies outside the area of clinical trials (Held et al, 2024).

Edgington ES. 1972 An additive method for combining probability values from independent experiments. J. Psychol. 80, 351–363. DOI: 10.1080/00223980.1972.9924813

Held L, Pawel S and Micheloud C (2024) The assessment of replicability using the sum of p-values Roy. Soc. Open Sci.11240149 DOI: 10.1098/rsos.240149

Held, L (2024). Beyond the two-trials rule. DOI: 10.1002/sim.10055

Rosenkranz GK (2023). A generalization of the two trials paradigm. Ther Innov Regul Sci. 57:316-320. DOI: 10.1007/s43441-022-00471-4

Session VII: Complex innovative designs and Decision analysis

Thursday 27th February 15:40 – 17:25

Real-world use of adaptive designs in paediatric clinical trials – A review of the European Medicines Agency's Clinical Trials Information System (CTIS)

Fabian Eibensteiner[1,2], Ralf Herold[3], Christoph Aufricht[1], Martin Posch[2], Franz Koenig[2]

[1]Division of Paediatric Nephrology and Gastroenterology, Department of Paediatrics and Adolescent Medicine, Comprehensive Center for Paediatrics, Medical University of Vienna, Vienna, Austria; [2]Institute for Medical Statistics, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria; [3]European Medicines Agency (EMA), Amsterdam, Netherlands; fabian.eibensteiner@meduniwien.ac.at

In paediatric research, adaptive clinical trial designs represent an important but under-explored alternative to traditional trials addressing uncertainties in the planning phase. Contrasting fixed designs, lower expected sample sizes for equal power and potentially identifying ineffective interventions earlier render them promising to overcome barriers in paediatric trials.[1] The European Union (EU) introduced paediatric investigation plans (PIPs) as a regulation to ensure that development of medicinal products potentially used in paediatric patients get integrated in the development for adults. Adaptive PIPs were recently implemented by EMA in a pilot using stepwise PIPs.[2] In 2022, the EU governance for clinical trials (EU 536/2014), came into application prompting EMA to launch Clinical Trials Information System (CTIS). CTIS aims to streamline submission, approval, and transparency of trials across the EU. Despite these regulations, there are concerns that PIPs may delay paediatric clinical trials.

For this review, we acquired and analyzed available information on clinical trials in paediatric age groups (0-17 years) registered in CTIS. For all studies where clinical trial protocols were approved and available through CTIS we conducted an in-depth review focusing on the use of innovative designs and statistical methods, specifically regarding adaptive trial designs.

In total, we were able to retrieve information on 876 unique clinical trials conducted in paediatric age groups. Only 48% of these trials were exclusively with paediatric participants. English language study protocols were available in 16% (144/876) of paediatric clinical trials in CTIS. There is no straightforward way to identify adaptive elements of clinical trials, since such elements are not included in the structure of information provided by the trial registers. By using text mining tools, a preliminary analysis identified that 21/144 (15%) clinical trial study protocols contained the term "adaptive" to describe their methodology. Additionally, we found 6/144 (4%), 31/144 (22%), and 3/144 (2%) protocols referencing advanced design terms such as "master protocols," "platform", and "basket" trials, respectively.

This review provides insights into the current landscape of paediatric trials in the EU, highlighting the real-world use of innovative methodologies, identifying challenges, and discussing emerging standards in trial design and analysis. Based on our findings, interesting applications of adaptive designs will be presented, showcasing recent advances and their impact on paediatric clinical research.

1 Kelly et al. Considerations for adaptive design in paediatric clinical trials: study protocol for a systematic review, mixed-methods study, and integrated knowledge translation plan. Trials. 2018 Oct 19;19(1):572.

2 EMA (Accessed 03.11.2024). https://www.ema.europa.eu/en/human-regulatory-overview/research-development/paediatric-medicines-research-development/paediatric-investigation-plans

Bayesian decision analysis for clinical trial design with binary outcome in the context of Ebola Virus Disease outbreak – Simulation study

Drifa Belhadi[1,2], Joonhyuk Cho[3,4,5], Pauline Manchon[6], Denis Malvy[7,8], France Mentré[1,6], Andrew W. Lo[3,5,9,10], Cédric Laouénan[1,6]

[1]Université Paris Cité, Inserm, IAME, F-75018 Paris, France; [2]Saryga SAS, France; [3]MIT Laboratory for Financial Engineering, Cambridge, MA, USA; [4]MIT Department of Electrical Engineering and Computer Science, Cambridge, MA, USA; [5]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA; [6]AP-HP, Hôpital Bichat, Département d'Epidémiologie Biostatistiques et Recherche Clinique, F-75018 Paris, France; [7]UMR 1219 Inserm/EMR 271 IRD, University of Bordeaux, Bordeaux, France; [8]Department for Infectious and Tropical Diseases, University Hospital Center Pellegrin, Bordeaux, France; [9]MIT Operations Research Center, Cambridge, MA, USA; [10]MIT Sloan School of Management, Cambridge, MA, USA; drifa.belhadi@saryga.com

When designing trials for high-mortality diseases with limited available therapies, the conventional 5% type I error rate used for sample size calculation can be questioned. Bayesian Decision Analysis (BDA) for trial design allows for the integration of multiple health consequences of the disease when designing trials. This study adapts BDA for trials with binary outcomes to calculate optimal sample sizes and type I error rates in the context of an Ebola virus disease outbreak.

We consider a fixed, two-arm randomized trial with a binary outcome and two types of clinical trial loss: post-trial loss, for not approving an effective treatment or approving an ineffective treatment; in-trial loss, for not administrating an effective treatment to patients in the control arm or for administrating an ineffective treatment for patients in the experimental arm. The model accounts for side effects of an ineffective treatment and the burden of Ebola disease. A loss function was defined to summarize the multiple consequences into a single measure, and optimal sample sizes (n) and type I error rates (α) were derived by minimizing this loss function.

Using the mortality rate as the outcome, we varied model parameters to represent different Ebola epidemic scenarios, such as target population size, mortality rate, and treatment efficacy. In most cases, BDA-optimal α values exceeded the conventional one-sided 2.5% rate and BDA-optimal sample sizes were smaller. Additionally, we conducted simulations comparing a BDA-optimized two-arm trial (fixed or sequential) to standard designs (two-arm/single-arm, fixed/sequential) across various outbreak scenarios. Overall, statistical power remained comparable across designs, except when sample size assumptions were incorrect, or when the trial started after the outbreak peak; in these situations, BDA-optimized trials were associated with superior powers.

This BDA adaptation provides a new framework for designing trials with a binary outcome, especially for high-mortality diseases with few treatment options. In an outbreak context, where case numbers decline after the epidemic peak and there is uncertainty around mortality rate and treatment efficacy, BDA-optimized trials offer an interesting approach for evaluating new experimental treatments.

Communicating Complex Considerations in Dual Endpoint Trial Design – An Oncology Case Study

Boaz Adler[1], Valeria Mazzanti[2], Pantelis Vlachos[2]

[1]Cytel Inc., United States of America; [2]Cytel Inc., Geneva, Switzerland; boaz.adler@cytel.com

In this case study, we describe the challenges faced in the design and selection of a dual endpoint clinical trial in an Oncology indication. We highlight benefits and limitations of selecting a dual- versus single-endpoint design, and the discussion of tradeoffs with a cross-functional study team. Furthermore, the case study highlights the value of adding an efficacy stopping boundary at an interim analysis, as well as a second, later interim analysis both leading to benefits such as savings in average sample size and average study duration. Finally, the case study also shows how extensive simulation work using advanced software in study design supports more realistic expectation for study power by incorporating posterior probabilities of various treatment effects.

Assessing the Effects of Additional Investment in Earlier Phase Trials to Enhance Overall Program Probability of Success Through Informed Priors

Valeria Mazzanti[1], Boaz Adler[2], Pantelis Vlachos[1]

[1]Cytel Inc., Geneva Switzerland; [2]Cytel Inc., United States of America; boaz.adler@cytel.com

Phase III trials represent a significant investment for life science organizations. In this proposal, we assess the benefits of running a smaller phase II trial prior to launching a full phase III, accounting for the additional information that this trial can bring, compared to the investment in a less-informed phase III study. In the planning stage, the probability of success computed for this smaller trial can then be used to evaluate the merits of the originally planned phase III trial. As an extension to this, the probability of success for the phase III is computed conditionally on the success of the smaller phase II, which can then be compared to the probability of success for the original phase III trial on its own. When computing the probability of success of the combination of two trials, simulations can also be more realistic as the observed treatment effects from the first trial can be used to inform the treatment effect assumptions for the phase III study. By using simulations to anticipate the probability of success of a phase II trial, combined with the ability to anticipate the probability of success of a program of trials made up of this same phase II trial followed by a phase III trial, sponsors can have much higher confidence in their R&D programs.

Model-Guided Parameter Optimization for Complex Innovative Trial Designs

Raviv Pryluk, Yaron Racach, Amitay Kamber, Roni Cohen, Amitai Levy, Tzviel Frostig, Oshri Machluf, Elad Berkman

PhaseV Trials Inc.; raviv@phasevtrials.com

Optimizing clinical trial design, particularly within adaptive frameworks, presents significant challenges due to the need to balance trial performance, resource utilization, operational considerations, and adherence to statistical error rates. We introduce a model-guided optimization approach tailored for complex adaptive trial designs across many types of designs, including multi-arm seamless phase 2/3 trials. Our methodology utilizes an iterative process involving localized simulations and predictive modeling to efficiently identify near-optimal parameter configurations that satisfy predefined type I and type II error thresholds while minimizing trial costs. This approach is agnostic to the underlying trial structure and can be adapted across various trial designs.

In a case study that will be presented, we will demonstrate the application of our method in a seamless phase 2/3 design with multiple arms, incorporating features such as arm dropping, dose selection, and futility stopping rule. By iterating simulations around a current set of parameters and using local predictive models to estimate error rates and costs, our optimization strategy identifies configurations that significantly reduce sample sizes required and the computational load. Compared to standard non-adaptive and non-stochastic optimization approaches, our algorithm achieves approximately 15% reduction in required patients under alternative hypotheses and a 10-fold decrease in simulation iterations that can save a lot of time in the design process.

Our contributions address key gaps in adaptive trial design by offering a scalable, automated solution that improves both trial efficiency and design quality, with substantial implications for early-phase trials and software development in adaptive methodologies. The solutions can be used both for open-source software, as well as for commercial software.