daptive Designs and Multiple Testing Procedures

Workshop June 29 – July 1, 2022 in Heidelberg

Sponsored by:

Organized by:

Heidelberg Institute of Medical Biometry

HEIDELBERG UNIVERSITY HOSPITAL

Deutsche Region

Conference Website:
www.biometrie.uni-heidelberg.de/workshop

Local Organizing Committee:
Meinhard Kieser, Marietta Kirchner, Johannes Krisam, Jan Meis, Birgit Schleweis

COVID-19 Regulations

Conference venue:
Please note that due to the COVID-19 regulations imposed by the University of Heidelberg, you need to wear a medical protection mask or FFP2 mask at the conference venue in case a minimum distance to another person of 1.5 meters (5 feet) cannot be kept. This will for example be the case during all conference talks.

Public transportation:
Also note that, when using public transportation in Germany (long-distance and short-distance trains, as well as local buses and trams), a medical protection mask or FFP2 mask is currently also required.

Social events:
In case you participate at the guided tour at the German Pharmacy Museum, wearing a medical protection mask or FFP2 mask is recommended, but not required. For the old town city tour and the conference dinner, a medical protection mask is not required.

Information for Speakers

As a speaker, you should either

1. send your presentation to workshop@imbi.uni-heidelberg.de or
2. bring it to the registration desk on a USB-stick

at least **30 minutes prior to their talk**.

The presentation file should either be a PDF or PPTX-File.

CONFERENCE VENUE

LOCATION:

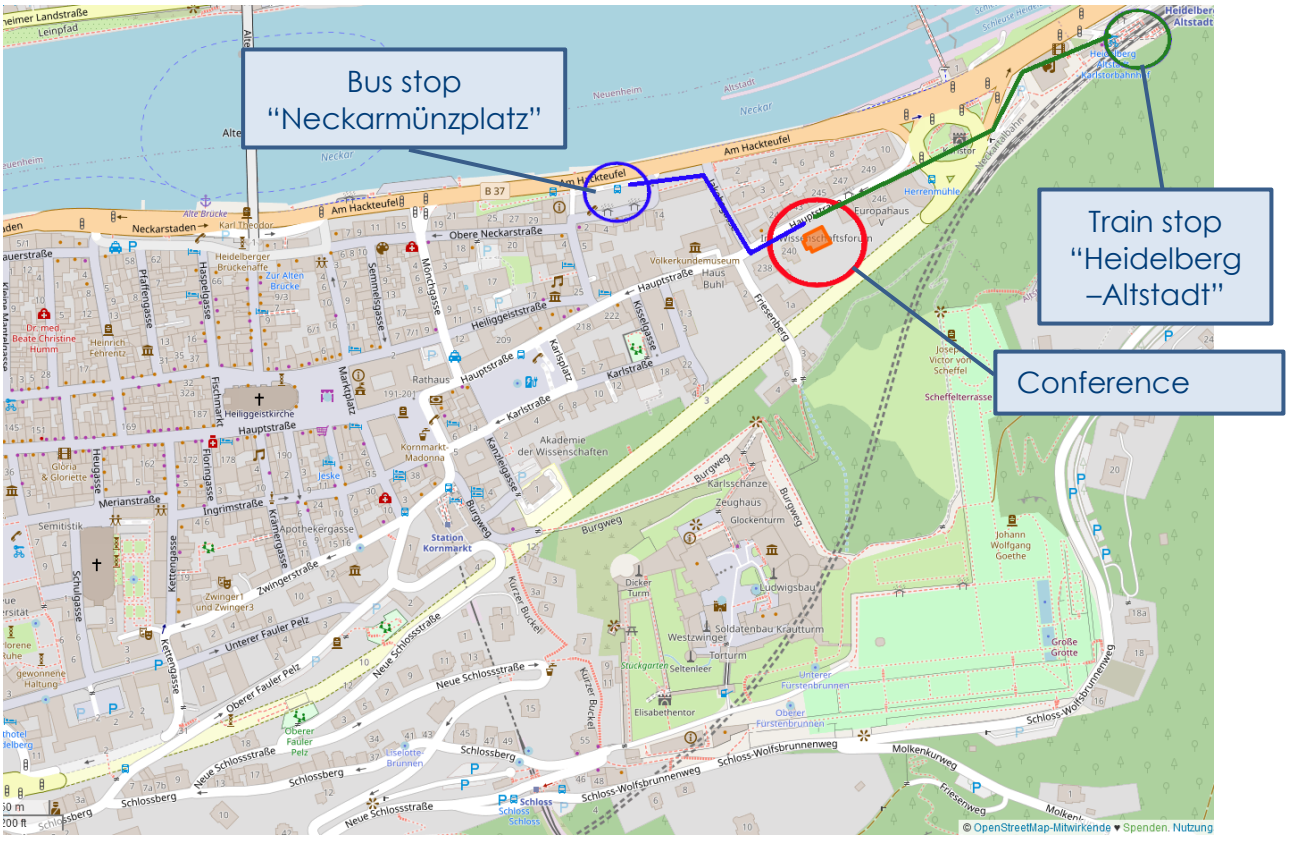
The workshop "Adaptive Designs and Multiple Testing Procedures 2022" will take place at the Internationales Wissenschaftsforum Heidelberg (IWH). It is a centre sponsored by Heidelberg University for scholarly exchange in all areas of science and academic research.

ADDRESS:

Hauptstrasse 242
D-69117 Heidelberg
Tel: +49 (0) 6221 54 36 90
E-mail: iwh@uni-hd.de

How to get there:

- **By regional train ("S-Bahn"):** Exit regional train **S1, S2, or S5** at stop **"Heidelberg-Altstadt"**, then leave the station to the left and walk 500 meters to reach the IWH (green line shown on map).
- **By bus from Heidelberg Central Station:** When arriving at Heidelberg Central Station, use either **bus line 20** (Direction: Heidelberg S-Bahnhof Altstadt) or **bus line 33** (Direction: "Ziegelhausen, Köpfel"). Get off at bus stop **"Neckarmünzplatz"**, to reach the IWH after a 280 meter walk (blue line shown on map)

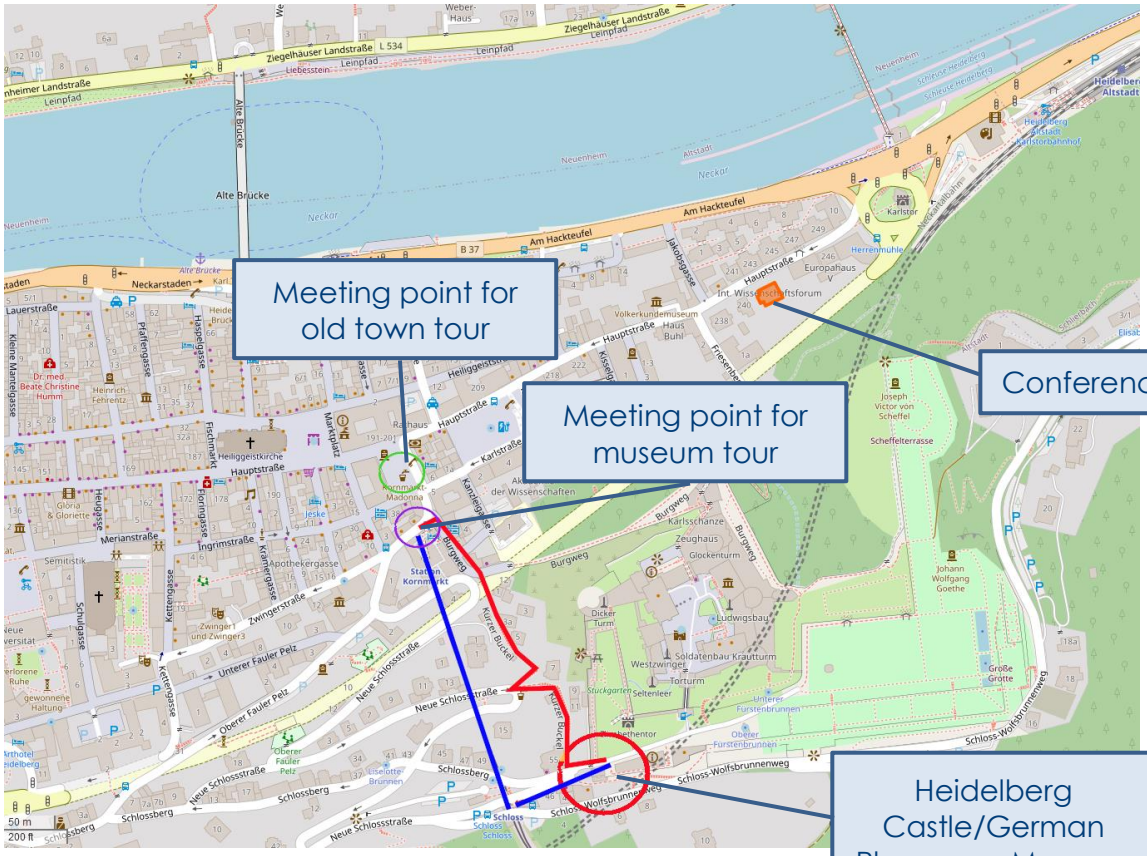


Copyright: openstreetmaps.org

GUIDED TOURS

Old town tour: The old town tour starts at the Saint Mary's Statue on the Kornmarkt (green circle on the map, within walking distance from the conference venue). **Please be there at 17:45 the latest.**

German Pharmacy Museum tour: The meeting point for the tour is at the lower station of the Heidelberg Funicular Railway (Zwingerstraße 20, 69117 Heidelberg, purple circle on the map, within walking distance from the conference venue**). Please be there at 17:30 the latest.** From the station, you can either **climb the 300 steps** to the Heidelberg Castle (red line), or alternatively take the **Funicular Railway** at the costs of 9,-€ (blue line, medical protection mask or FFP2 mask is required).



Copyright: openstreetmaps.org

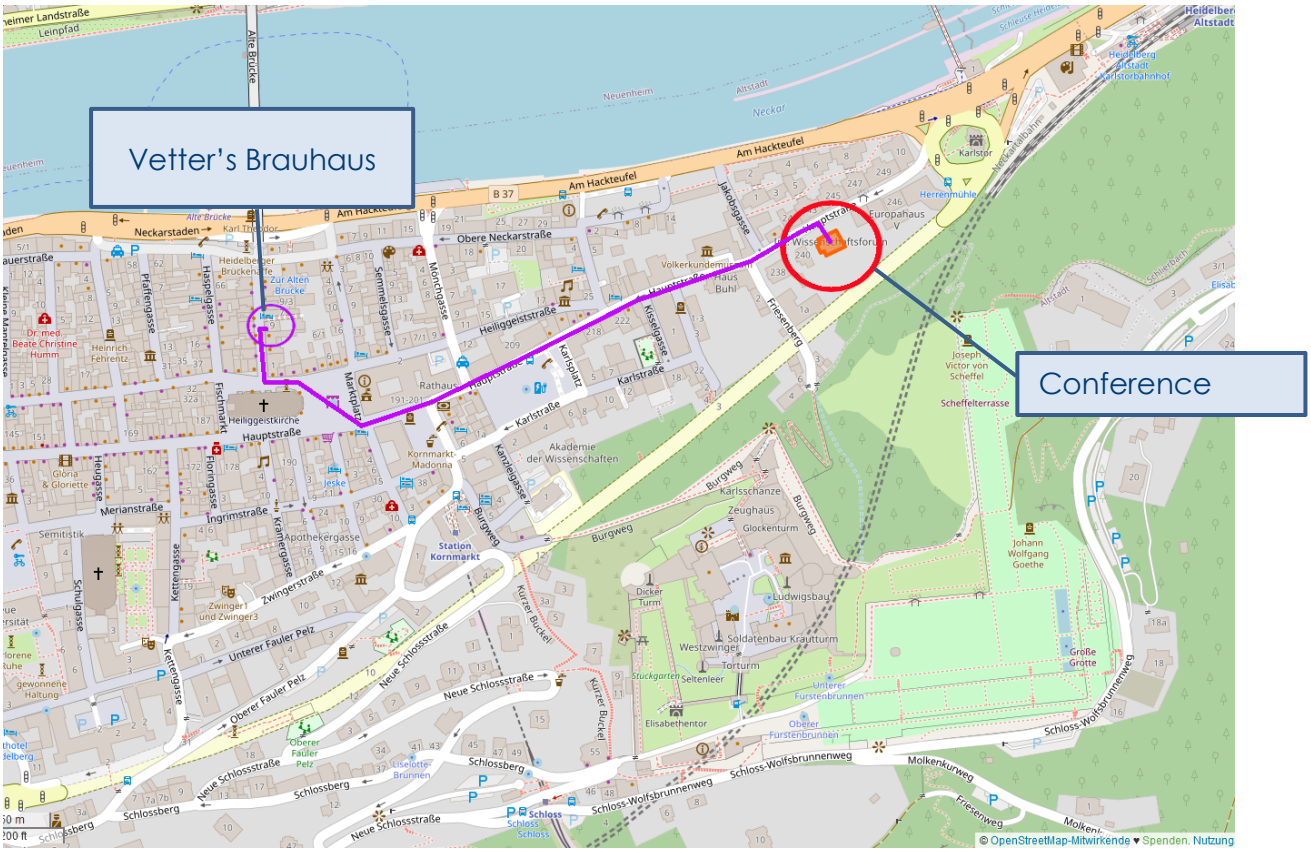
CONFERENCE DINNER

The conference dinner can be attended on June 29 at 19:30. The conference dinner will take place in the restaurant "**Vetter's Brauhaus**".

ADDRESS:

Vetter´s Alt Heidelberger Brauhaus
Steingasse 9
69117 Heidelberg

How to get there:
The location of the conference dinner is in walking distance to the conference venue (550 meters).



Copyright: openstreetmaps.org

SCIENTIFIC PROGRAM – OVERVIEW

WEDNESDAY, JUNE 29

11:30	Registration
12:30 – 13:30	Lunch
13:30 – 13:45	Welcome Address (Johannes Krisam)
13:45 – 15:15	Session I: Optimal designs (Chair: Andreas Faldum)
15:15 – 15:45	Coffee break
15:45 – 17:15	Session II: Multiple Testing (Chair: Maximilian Pilz)
18:00 – 19:30	Social events
19:30	Conference Dinner

THURSDAY, JULY 30

9:00 – 10:00	Invited Session I
10:00 – 10:30	Coffee break
10:30 – 12:30	Session III: Platform trials (Chair: Thomas Asendorf)
12:30 – 13:30	Lunch
13:30 – 14:30	Invited Session II
14:30 – 14:40	Short break
14:40 – 16:10	Session IV: Adaptive and group sequential designs I (Chair: Cornelia Ursula Kunz)
16:10 – 16:40	Coffee Break
16:10 – 16:40	Session V: Adaptive and group sequential designs II (Chair: Lisa Hampson)
17:40	Meeting of the IBS-DR/ROeS Working Group on ADMTP

FRIDAY, JULY 1

9:00 – 10:30	Special Topic Session on Handling overrunning in group sequential clinical trials (Organizers: Lisa Hampson, Cornelia Ursula Kunz, Chair: Gernot Wassmer)
10:30 – 11:00	Coffee break
11:00 – 12:00	Session VI: Bayesian methods (Chair: Tobias Mielke)
12:00 – 12:05	Closing/End of Workshop (Johannes Krisam)

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE (WEDNESDAY)

WEDNESDAY, JUNE 29

11:30 REGISTRATION

12:30 – 13:30 LUNCH

10:30 – 10:45 WELCOME ADDRESS: JOHANNES KRISAM

13:45 – 15:15 SESSION I: OPTIMAL DESIGNS (CHAIR: ANDREAS FALDUM)

Michaela Maria Freitag, Xieran Li, Geraldine Rauch: "Optimal" futility boundaries for binary endpoints

Jan Meis, Maximilian Pilz, Meinhard Kieser: Performance of different estimation methods in optimal adaptive two-stage designs

Maximilian Pilz, Meinhard Kieser: Applying optimal two-stage designs in practice

15:15 – 15:45 COFFEE BREAK

15:45 – 17:15 SESSION II: MULTIPLE TESTING (CHAIR: MAXIMILIAN PILZ)

Helmut Finner, Markus Roters: On positive association of absolute-valued and squared multivariate Gaussians beyond MTP_2

Daniel Fridljand, Nikolaos Ignatiadis, Wolfgang Huber: Better multiple Testing: Using multivariate co-data for hypotheses weighting

Yang Han: Confident and Logical Selection of the Cut-point of a Biomarker for Patient Targeting

18:00 – 19:30 SOCIAL EVENTS

19:30 CONFERENCE DINNER

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE (THURSDAY)

THURSDAY, JUNE 30

9:00 – 10:00 INVITED SESSION I (DISCUSSANT: JOHANNES KRISAM)

Donald Berry: Complex Innovative Designs (CIDs) for Drug Registration including Platform Trials and Basket Trials

10:00 – 10:30 COFFEE BREAK

10:30 – 12:30 SESSION III: PLATFORM TRIALS (CHAIR: THOMAS ASENDORF)

Alexandra Griessbach, Christof Schönenberger, Viktoria Gloy, Ala Taji Heravi, Arnav Agarwal, Stefan Schandelmaier, Perrine Janiaud, Alain Amstutz, Benjamin Speich, Matthias Briel: The planning, development, progression and output of platform trials – a systematic survey

Quynh Nguyen, Katharina Hees, Benjamin Hofner: Platform trials: the impact of shared controls on type 1 error and power

David Robertson, James Wason, Franz König, Martin Posch, Thomas Jaki: Online error rate control for platform trials

Marta Bofill Roig, Martin Posch: Model-based approaches for utilising non-concurrent controls in platform trials

12:30 – 13:30 LUNCH

13:30 – 14:30 INVITED SESSION II (DISCUSSANT: JOHANNES KRISAM)

Donald Berry: Dose-Finding Designs Using Adaptive Randomization and Longitudinal Modeling—Which is More Important?

14:30 – 14:40 SHORT BREAK

14:40 – 16:10 SESSION IV: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS I (CHAIR: CORNELIA URSULA KUNZ)

Leandro Garcia Barrado, Tomasz Burzykowski, Catherine Legrand, Marc Buyse: Using an interim analysis based exclusively on an early outcome in a randomized clinical trial with a long-term clinical endpoint

Svetlana Cherlin, James Wason: Cross-validated risk scores adaptive enrichment design

Moritz Fabian Danzer, Ina Dormuth, Marc Ditzhaus: Adaptive redesigning of combination testing procedures in survival analysis

16:10 – 16:40 COFFEE BREAK

16:40 – 17:40 SESSION V: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS II (CHAIR: LISA HAMPSON)

Björn Bokelmann, Geraldine Rauch, Meinhard Kieser, Carolin Herrmann: Scoring recalculation rules for adaptive designs with binary endpoints

Carolin Hermann, Meinhard Kieser, Geraldine Rauch, Maximilian Pilz: Sample size adaptation – a performance score optimization approach

17:40 MEETING OF THE IBS-DR/ROES WORKING GROUP ON ADMTP

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE (FRIDAY)

FRIDAY, JULY 1

9:00 – 10:30 SPECIAL TOPIC SESSION ON "HANDLING OVERRUNNING IN GROUP SEQUENTIAL CLINICAL TRIALS" (ORGANIZERS: LISA HAMPSON, CORNELIA URSULA KUNZ, CHAIR: GERNOT WASSMER)

Introduction: Motivation for the session (5')

Cornelia Ursula Kunz, Stephen Schüürhuis: Rethinking Delayed Response

Chris Jennison, Lisa Hampson: Analysing over-run data after a group sequential trial

Benjamin Hofner, Elina Asikanius: Updated analyses and overrunning: Lessons learned during the pandemic and a plea to plan ahead

Discussant (Kaspar Rufibach, 10')

Panel discussion (Moderator: Gernot Wassmer, 10')

Audience Q&A (Moderator: Gernot Wassmer, 10')

10:30 – 11:00 COFFEE BREAK

11:00 – 12:00 SESSION VI: BAYESIAN METHODS (CHAIR: TOBIAS MIELKE)

Elja Arjas: Adaptive treatment allocation and selection in multi-arm clinical trials: a Bayesian approach

Miguel Pereira, Giulia Brunelli, Manuel Wiesenfarth, Oliver Shoenborn-Kellenberger: An adaptive Bayesian design for a phase I vaccine clinical trial with adjustment for age group

12:00 – 12:05 CLOSING/END OF WORKSHOP (JOHANNES KRISAM)

ABSTRACTS

SESSION I: OPTIMAL DESIGNS

"OPTIMAL" FUTILITY BOUNDARIES FOR BINARY ENDPOINTS

Michaela Maria Freitag
Charité - Universitätsmedizin Berlin

Xieran Li

Geraldine Rauch

In group sequential designs, the opportunity for an early stop during interim analyses due to efficacy or futility may reduce time and resources. Futility boundaries are routinely integrated but a theoretical justification of the value is often missing, despite the fact that an arbitrary choice of futility boundaries may have a substantial negative impact on performance characteristics.

Schüler et al. (2017) and Li et al. (2020) contributed to address this problem by discussing optimality criteria for non-binding futility boundaries for trials with continuous and time-to-event endpoints. We extend the concept to binary endpoints with one-sided hypotheses. By construction, the provided method of choosing "optimal" futility boundaries maximizes the probability to correctly stop for futility in case of small or opposite effects while timing the interim analysis at an appropriate fraction of the overall sample size and limiting the power loss and the probability of stopping the study wrongly.

The performance of study designs with such an "optimal" futility boundary is then compared to the common Simon's minimax and optimal two-stage designs (Simon, 1989) and modified Simon's two-stage designs, which incorporate control over the probability to wrongly stop for futility and the timing of the interim analysis (Kim et al., 2019).

The results demonstrate the benefit of using our introduced "optimal" futility boundaries, especially compared to the standard Simon's designs by ensuring a high proportion of correct stops for futility and reducing the probability of incorrect interim decisions. Additionally, they provide more flexible non-binding rules compared to Simon's binding rules.

SESSION I: OPTIMAL DESIGNS

PERFORMANCE OF DIFFERENT ESTIMATION METHODS IN OPTIMAL ADAPTIVE TWO-STAGE DESIGNS

Jan Meis
University of Heidelberg

Maximilian Pilz

Meinhard Kieser

The flexibility of adaptive clinical trial designs can offer significant advantages over fixed designs. If design parameters are chosen appropriately, adaptive designs can decrease the expected sample size and save resources in cases where there is early evidence that the continuation of a trial may be futile. A known issue with adaptive designs is that estimators appropriate in a single-stage fixed design setting, such as the maximum likelihood estimator, can be biased due to the dependence structure in the data introduced by the adaptivity. This problem affects point estimators as well as confidence intervals and p-values. Regulatory agencies such as the EMA or FDA recognize this problem and urge researchers that "the extent of bias should be evaluated, and estimates should be presented with appropriate cautions regarding their interpretation" [1].

Over the years, various methods have been proposed to mitigate the bias introduced by adaptive designs. However, estimators often need to fulfill other requirements to be useful in practice, such as having an acceptable variance. In this work, we provide results on the operating characteristics of different estimators in optimal adaptive two-stage designs with normally distributed outcome. In optimal designs, design parameters such as the sample sizes, decision boundaries, and adaptation rules are chosen as the result of an optimization process. The goal of the optimization process is to maximize some metric of design quality, a typical example being the expected sample size required to fulfill certain power requirements under a specified hypothesis. Although optimal adaptive designs have been a topic of recent research, guidance on estimation in this novel kind of designs is still scarce. We compare classical and recently developed point estimators as well as estimation methods for confidence intervals and p-values regarding various performance criteria such as bias, variance, mean squared error, coverage, and consistency with test decisions.

References:
[1] FDA (2019). Adaptive Designs for Clinical Trials of Drugs and Biologics. Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry

SESSION I: OPTIMAL DESIGNS

APPLYING OPTIMAL TWO-STAGE DESIGNS IN PRACTICE

Maximilian Pilz
Fraunhofer Institute for Industrial Mathematics ITWM

Meinhard Kieser

Adaptive two-stage designs are always an option worth to be considered when a clinical trial is planned. This class of study designs can enhance trial characteristics as expected sample size or power due to the increased flexibility induced by an interim analysis. Meanwhile, various work exists on the optimal determination of adaptive two-stage designs. This means that a specified objective criterion, for instance the expected sample size, is optimized while important properties as (conditional) power and the type I error rate are controlled by means of suitable chosen constraints. Although these optimal designs ensure a performance under the chosen objective criterion that cannot be beaten without violating the constraints due to their optimality, optimal adaptive two-stage designs are rarely applied in practice.

In this talk, we give guidance on how to plan an optimal adaptive design that is well-suited for practical application. We show how requirements by trial sponsors can be incorporated into the optimization problem in order to simultaneously satisfy these desires and obtain a high-performing design. In addition, we illustrate cases where frequently desired constraints cannot be fulfilled simultaneously and provide recommendations how to act in those situations.

Furthermore, the tradeoff between complexity and communicability is discussed. We treat this issue by investigating different objective criteria and the behavior of the adaptive designs optimizing them. Finally, we present a class of simplified two-stage designs, namely group-sequential designs with linear critical value function, which may be better communicable to practitioners.

SESSION II: MULTIPLE TESTING

ON POSITIVE ASSOCIATION OF ABSOLUTE-VALUED AND SQUARED MULTIVARIATE GAUSSIANS BEYOND MTP_2

Helmut Finner
German Diabetes Center

Markus Roters

Concepts of positive (negative) dependence associated with probability inequalities are often essential for proving conservativeness of multiple decision procedures. For example, multivariate totally positive of order 2 (MTP_2) as well as the weaker notion of positive association (PA) of random variables yield various probability inequalities useful in multiple testing. In this talk we are concerned with the question which absolute-valued p –dimensional multivariate normally distributed random vectors are positively associated (PA). Around 1980 various authors (cf. e.g. Bolviken (1982), Karlin, Rinott (1983), Rüschendorf (1981)) proved that absolute normals X are MTP_2 if the inverse of the covariance matrix of DX is an M-matrix for some signature matrix D. In this talk we show that this so-called signed MTP_2 condition is not necessary for PA of absolute-valued normals for $p \geq 3$ Hence, there is at least some free space beyond the celebrated but tiny MTP_2 world for absolute-valued normals to be PA. Our main findings are based on the the fact that conditionally increasing in sequence (CIS) implies PA. For $p = 3$ we show that there exist absolute-valued multivariate normals which are CIS (and hence PA) but not MTP_2 iff the underlying covariance matrix satisfies a certain condition. However, for $p \geq 4$, we also show that the existence of a CIS sequence is not necessary for absolute-valued normals to be PA. Finally, our results disprove Theorem 1 in Eisenbaum (2014) and the conjecture that MTP_2, infinite divisibility and PA of squared multivariate normals are equivalent.

References:

Bøviken, E. (1982). Probability inequalities for the multivariate normal with non-negative partial correlations. Scand. J. Statist. 9, 49-58.
Eisenbaum, N. (2014). Characterization of positively correlated squared Gaussian processes. Ann. Probab. 42, 559-575.
Finner, H., Roters, M. (2022). On positive association of absolute-valued and squared multivariate Gaussians beyond MTP_2.Preprint.
Karlin, S., Rinott, Y. (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. Ann. Stat. 9, 1035-1049.
Rüschendorf, L. (1981). Characterization of dependence concepts in normal distributions. Ann. Inst. Stat. Math. 33(3), 347-359.

SESSION II: MULTIPLE TESTING

BETTER MULTIPLE TESTING: USING MULTIVARIATE CO-DATA FOR HYPOTHESES WEIGHTING

Daniel Fridljand
University of Heidelberg

Nikolaos Ignatiadis

Wolfgang Huber

Consider a multiple testing task, where for each test we have access to its p-value and additional information represented by a uni- or multivariate covariate. The covariates may contain information on prior probabilities of null and alternative hypotheses and/or on the test's power. As per several recent proposals, the independent hypothesis weighting (IHW, Ignatiadis and Huber, 2021) framework capitalizes on these covariates for the multiple testing procedure. IHW partitions the covariate space into a finite number of bins and learns weights, used to prioritize each bin a-priori based on the covariate. IHW guarantees false discovery rate control (FDR), while increasing the proportion of correct discoveries (power) compared to unweighted methods such as the Benjamini-Hochberg procedure (BH).

(Ignatiadis and Huber, 2021) use per-covariate quantiles for the partition. Limitations of this are, that the number of quantile combinations explode with multiple covariates and the bins have fixed side length. Here we propose a random forest-based approach (IHW-Forest), where the leaves of the trees form a partition for the covariates. The objective function is chosen such that the splits are sensitive to the prior probability of a hypothesis being true. IHW-Forest scales well to high-dimensional covariates and can detect small regions with signal. IHW-Forest can deal with heterogeneous covariates and ignore uninformative covariates. Latter is useful in practice, when the user does not know which covariates are relevant for the hypotheses under study. This extends the application of IHW by automatic selection of the most relevant covariate. Lastly, IHW-Forest takes advantage of the p-values to construct the partition, yielding homogeneous bins and hence increases power.

We demonstrate IHW-forest's benefits in simulations and in a bioinformatic application. IHWs power vanishes with increasing number of covariate dimensions, while IHW-Forest's power remains stable and well above BH and IHW. With the signal concentrated in a shrinking region, IHW-Forest outperforms BH, IHW and other competing methods in power. We apply IHW-Forest to a hQTL analysis, which looks for associations between genetic variation and histone modifications on the human chromosomes. This resulted in 16 billion tests on the first two chromosomes. We used 16 different covariates, among them the genomic distance and his-tone modifications. Due to an exponential increase of the number of per-covariate quantiles with the number of covariates, IHW is not applicable anymore. The updated package will be available from Bioconductor http://www.bioconductor.org/packages/IHW in release 3.15.

References:
Ignatiadis, N. and Huber, W. (2021) Covariate powered cross-weighted multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol., 83, 720–751. John Wiley and Sons Inc. DOI: 10.1111/rssb.12411.

SESSION II: MULTIPLE TESTING

CONFIDENT AND LOGICAL SELECTION OF THE CUT-POINT OF A BIOMARKER FOR PATIENT TARGETING

Yang Han
University of Manchester

Confidently choosing a cut-point for a continuously valued biomarker to target patients with is challenging because there are two levels of multiplicity: the multiplicity of efficacy in the marker-positive subgroup and in the marker-negative subgroup at each cut-point, and the further multiplicity of searching through infinitely many cut-points. Currently available methods do not strongly control familywise type I error rate (FWER) across both levels of multiplicity. I will present a method that does. Taking a confidence band approach, our method in fact sets forth four principles that we believe every confident biomarker cut-point selection method should strive to adhere to.

For diseases with continuous outcome such as Type II Diabetes and Alzheimer's Disease, our method provides exact simultaneous confidence intervals for efficacy in the marker positive and negative subgroups, simultaneously for all possible cut-point values. I will demonstrate an interactive app for it.

INVITED SESSION I

COMPLEX INNOVATIVE DESIGNS (CIDS) FOR DRUG REGISTRATION INCLUDING PLATFORM TRIALS AND BASKET TRIALS

Donald A. Berry
Berry Consultants

SESSION III: PLATFORM TRIALS

THE PLANNING, DEVELOPMENT, PROGRESSION AND OUTPUT OF PLATFORM TRIALS – A SYSTEMATIC SURVEY

Alexandra Griessbach
University of Basel

Christof Schönenberger
Viktoria Gloy
Ala Taji Heravi
Arnav Agarwal
Stefan Schandelmaier
Perrine Janiaud
Alain Amstutz
Benjamin Speich
Matthias Briel

Background: Randomized controlled trials (RCTs) are considered the "gold standard" of evidence-based clinical research. However, RCTs are often inflexible, inefficient, and costly. Platform trials promise to solve some of the difficulties associated with RCTs.

Aim: In this study, we aim to record all platform trials since their inception and to determine their characteristics, progression, development, and output over time.

Methods: We conducted a systematic search and extracted all randomized platform trial baseline characteristics. We determined the availability of publications and protocols as well as the status of individual interventions/arms. Number of arms which were added or dropped (inclusion criteria), as well as further information on specific platform trial components such as use of a common control arm, non-concurrent control data, additional adaptive designs, recruitment, registration, and statistical framework were also extracted (amongst others).

Results: Our search resulted in 94 randomized platform trials. The first platform trial started in 2005. Preliminary data shows the majority (69.9%) of randomized platform trials were still ongoing, 16.1% were complete, 7.1% were discontinued/terminated/suspended and 3% were in planning. Master protocols were available for 81.7% of all trials. Most commonly, platform trials were conducted in the field of oncology (49.5%), however, a substantial number were dedicated to COVID-19 (34.4%). A common control arm was used in most trials (81.5%). Seamless design (19.4%) followed by sample size readjustment (17.2%), response adaptive randomization (14.0%) and adaptive enrichment (6.4%) were common additional adaptive design elements integrated into platform trials. A Bayesian (36.6%), frequentists (32.3%) or both statistical frameworks (4.3%) were used. However, many (26.9%) did not report this information. The median number of arms at the start of the trials was 3.0 (IQR, [2.0, 5.0]) and the median number of total arms included in trials was 4.0 (IQR, [3.0, 7.0]). One or more arms were added in 50.5% and dropped in 52.7% of all platform trials. Any results were available for 53.2% of all platform trials.

Conclusion: Preliminary results show that platform trials are becoming more common in clinical research, especially during the COVID-19 pandemic. However, it remains to be seen if they can hold their promise to solve known challenges associated with RCTs. Final results and additional outcomes will be available in June 2022.

SESSION III: PLATFORM TRIALS

PLATFORM TRIALS: THE IMPACT OF SHARED CONTROLS ON TYPE 1 ERROR AND POWER

Quynh Nguyen
Paul-Ehrlich-Institut

Katharina Hees

Benjamin Hofner

In platform trials, multiple treatment arms are evaluated with the possibility to add or drop treatments during the ongoing trial at different time points. A common master protocol for the multiple treatment arms can serve as a common logistical framework. These trial designs have gained more attention in recent years, especially in the evaluation of COVID-19 treatments and in oncology. In addition to the logistical advantages, the use of a common control is one of the key features in platform trials. Instead of individual control arms for each treatment, the use of a common control reduces the total sample size if no multiplicity adjustment is performed. However, there is currently no consensus among researchers on the need for adjustment when testing multiple treatment arms with the common control in the platform trial.

Advocates for an adjustment argue that the use of a common control leads to dependencies in the test statistics. A random low or random high control could lead to more false positive or false negative findings. A contrary argument to this is that if we examine multiple drugs in separate trials, we do not adjust over these trials and therefore, should not adjust in platform trials either.

In simulation studies, we investigate the impact of a common control on error rates such as the family-wise error rate (FWER) or the k-family-wise error rate (k-FWER, the probability of at least k false positive findings). We further investigate the impact on power when multiplicity adjustment is performed in platform trials. We consider a fixed platform trial (i.e. all arms enter and exit the platform trial at the same time) and compare these results with those of separate trials per active arm. In both study designs, the FWER is inflated when no multiplicity adjustment is applied. Due to the shared control, the FWER is smaller in platform trials, while the k-FWER is higher in comparison to separate trials ($k \geq 2$). A similar finding is observed for the probability to detect at least one effective treatment (disjunctive power) and the probability to detect all effective treatments (conjunctive power). Finally, we show that in specific circumstances, the platform trial can, even after Bonferroni or Dunnett adjustment, still be beneficial in terms of sample size.

SESSION III: PLATFORM TRIALS

ONLINE ERROR RATE CONTROL FOR PLATFORM TRIALS

David Robertson
University of Cambridge

James Wason

Franz König

Martin Posch

Platform trials evaluate multiple experimental treatments under a single master protocol, where new treatment arms are added to the trial over time. Given the multiple treatment comparisons, there is the potential for inflation of the overall type I error rate, which is complicated by the fact that the hypotheses are tested at different times and are not all necessarily pre-specified. Online error control methodology provides a possible solution to the problem of multiplicity for platform trials where a relatively large number of hypotheses are expected to be tested over time. In the online testing framework, hypotheses are tested in a sequential manner, where at each time-step an analyst decides whether to reject the current null hypothesis without knowledge of future tests but based solely on past decisions. Methodology has recently been developed for online control of the false discovery rate as well as the familywise error rate (FWER). In this talk, we describe how to apply online error control to the platform trial setting, present extensive simulation results, and give some recommendations for the use of this new methodology in practice. We show that the algorithms for online error rate control can have a substantially lower FWER than uncorrected testing, while still achieving noticeable gains in power when compared with the use of a Bonferroni procedure. We also illustrate how online error control would have impacted a currently ongoing platform trial.

SESSION III: PLATFORM TRIALS

MODEL-BASED APPROACHES FOR UTILISING NON-CONCURRENT CONTROLS IN PLATFORM TRIALS

Marta Bofill Roig
Medical University of Vienna

Martin Posch

Platform trials are multi-arm multi-stage trials with the extra feature of allowing new experimental arms to enter and leave the trial at different times. The number of experimental arms is not prefixed, as arms may be added or removed as the trial progresses. Platform trials offer the possibility of comparing the efficacy of experimental arms using a shared control group. Compared to separate trials with their own controls, this increases the statistical power and requires fewer patients. Shared controls in platform trials include concurrent and non-concurrent control data. For a given experimental arm, non-concurrent controls refer to data from patients allocated to the control arm before the arm enters the trial. Using non-concurrent controls is appealing because it may improve the trial's efficiency while decreasing the sample size. However, since arms are added sequentially, randomization occurs at different times. This lack of true randomization over time might introduce bias due to time trends. The challenge is to discern when and how to use non-concurrent controls to increase the trial's efficiency without introducing bias.

In this talk, we review methods to incorporate non-concurrent control data in treatment-control comparisons allowing for time trends. We focus mainly on frequentist approaches that model the time trend. We examine the impact of time trends on the operating characteristics of treatment effect estimators for each method under different patterns for the time trends. We outline under which conditions the methods lead to unbiased estimators and discuss the gain in power compared to trials only using concurrent controls.

INVITED SESSION II

DOSE-FINDING DESIGNS USING ADAPTIVE RANDOMIZATION AND LONGITUDINAL MODELING—WHICH IS MORE IMPORTANT?

Donald A. Berry
Berry Consultants

SESSION IV: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS I

USING AN INTERIM ANALYSIS BASED EXCLUSIVELY ON AN EARLY OUTCOME IN A RANDOMIZED CLINICAL TRIAL WITH A LONG-TERM CLINICAL ENDPOINT

Leandro Garcia Barrado
International Drug Development Institute (IDDI)

Tomasz Burzykowski

Catherine Legrand

Marc Buyse

In randomized clinical trials that use a long term efficacy endpoint, the follow-up time necessary to observe the endpoint may be substantial. To reduce the expected duration and/or sample size of such trials, early-outcome data may be collected to enrich an interim analysis aimed at stopping the trial early for efficacy. We propose a design that includes an interim analysis conducted by using solely early outcome data to expedite the evaluation of treatment's efficacy.

At the interim, we propose to predict the treatment effect on the long-term clinical endpoint based on the estimate of the treatment effect on the early outcome. As no clinical endpoint information is considered at the interim, the prediction depends on the surrogacy characteristics of the early outcome with respect to the clinical endpoint. To allow for any early outcome and long-term endpoint combination, we embed our proposal within the efficient score statistic framework.

We evaluate the potential gain in operating characteristics (power, expected trial duration, and expected sample size) of the proposed design in function of the properties of the early outcome as a surrogate for the long term endpoint.

In the context of an oncology trial considering a time-to-event endpoint and a binary early outcome, results show potentially substantial gains in both the expected trial duration and the expected sample size. A prerequisite, though, is that the treatment effect on the early outcome must be strongly correlated with the treatment effect on the long-term endpoint, i.e., that the early outcome is a valid surrogate for the long term endpoint.

SESSION IV: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS I

CROSS-VALIDATED RISK SCORES ADAPTIVE ENRICHMENT DESIGN

Svetlana Cherlin
Newcastle University

James Wason

Background: Adaptive enrichment clinical trial designs allow the trial to update the inclusion criteria based on the interim analysis. In the second stage, the entry is restricted to the subgroup of patients who are predicted to benefit from the treatment. Current adaptive enrichment methods assume that the subgroup is defined by a known predictive biomarker, which might not be available. With the recent advances in multi-omics technologies, increasingly large numbers of biomarkers are becoming available. Several approaches that utilise high-dimensional data have been proposed, such as the risk scores approach that summarises the high-dimensional information into a single score for each patient. The risk scores are subsequently used for identifying a subgroup of patients who benefit from the treatment.

Method: We propose a design that allows enriching the recruitment with patients who are predicted to benefit from the treatment, based on their high-dimensional baseline covariates. The sensitive group is identified using the risk score approach where each patient is assigned a score constructed from their baseline covariates. The design includes early stopping for futility if no promising treatment effect is identified in the sensitive group and also the difference between the arms in the overall trial population is not significant. We have implemented the new methods in an R package.

Results: We have investigated the performance of the proposed design by applying it to simulated data scenarios with various response rates for the sensitive group and different sample sizes. The design allows to narrow down the eligibility and also achieves this at a smaller expected sample size, in comparison to the non-enrichment alternative. For the null scenario, the design achieves a well-controlled type I error rate with a substantial reduction in the expected sample size (at least 24%). We illustrated our approach on a randomised clinical trial with publicly available high-dimensional gene expression data.

Conclusions: The new method shows a superior performance in terms of the power and the sample size in comparison to the non-enrichment approach. Further work could explore different distributions of outcomes, as well as multiple arms and endpoints.

SESSION IV: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS I

ADAPTIVE REDESIGNING OF COMBINATION TESTING PROCEDURES IN SURVIVAL ANALYSIS

Moritz Fabian Danzer
University of Münster

Ina Dormuth

Marc Ditzhaus

In survival analysis, the assumption of proportional hazards is very common. In many practical scenarios, however, this assumption must be questioned. The possible deviations from this assumption are numerous. In such cases, the standard log-rank test loses its optimality and a differently weighted log-rank type test would be preferable. As the exact shape of the deviation is unknown, the best choice for the weight remains unclear, in particular during the planning stage of a trial.

Recently, testing procedures have been developed that combine the information from differently weighted log-rank type tests. Such procedures have a broader power function than log-rank type tests with a single weight as a large variety of deviations from the null hypothesis can be detected. Of course, this has the disadvantage that the procedure would have less power than the test, whose weight would be optimal in the respective situation.

To narrow this gap between the optimal, but unknown test and the combined approach mentioned here, we propose an adaptive design. A start of the study with a broadly based combination approach can thus be combined with more refined procedures in later stages. For this, we want to use the information collected up to an interim analysis to redefine the testing procedure in terms of the weights of the log-rank tests. At the same time, other commonly used tools from adaptive designs (e.g. sample size recalculations) shall be applicable.

SESSION V: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS II

SCORING RECALCULATION RULES FOR ADAPTIVE DESIGNS WITH BINARY ENDPOINTS

Björn Bokelmann
Charité - Universitätsmedizin Berlin

Geraldine Rauch

Meinhard Kieser

Carolin Hermann

Uncertainty in the parameter values underlying the endpoint distribution is a common problem when planning the sample size for a clinical trial. Sample size recalculation during interim analysis offers a potential remedy to this problem. There is a large number of such recalculation rules to choose from. To guide the choice of a suitable adaptive design with sample size recalculation, previous literature suggests global as well as also a conditional performance score. The global perspective evaluates sample size recalculation approaches overall. The conditional perspective evaluates sample size recalculation approaches when knowing that the interim result actually suggests a recalculation of the sample size. So far, the conditional performance score has only been described for studies with a normally distributed endpoint. However, also binary endpoints are frequently applied in clinical studies. Therefore, we extend the theory of the conditional performance score to binary endpoints. Thereby, we elaborate on the choice of distribution parameters for the recalculation rules as well as on the effects of finite sampling properties on adaptive designs and on our score theory. We illustrate the application by a simulation study.

SESSION V: ADAPTIVE AND GROUP SEQUENTIAL DESIGNS II

SAMPLE SIZE ADAPTATION – A PERFORMANCE SCORE OPTIMIZATION APPROACH

Carolin Hermann
Charité - Universitätsmedizin Berlin

Meinhard Kieser

Geraldine Rauch

Maximilian Pilz

Adaptive designs offer the possibility to adapt design aspects of a clinical trial, such as the sample size. There exist different strategies for adapting the sample size. Many of them are based on attaining a predefined conditional power value [e.g. 1, 2]. Recently, Kunzmann et al. [3] provided an R package for deriving optimal adaptive designs with sample size recalculation in respect of a scoring criterion. The choice of the scoring criterion is a crucial task and different performance measures can be used, e.g. the global power or the average sample size. One possible scoring criterion is the conditional performance score [4] that combines different performances measures, i.e. conditional power and sample size, in terms of location and variation. We use this performance score for deriving optimal adaptive two-stage designs with sample size recalculation [5]. In this talk, we present the resulting adaptive designs and evaluate derivations of the performance score such as different weighting strategies or adding additional optimization constraints. Moreover, we discuss advantages and disadvantages of the approach. The resulting optimal score values can also be used as a benchmark when using the performance score for comparing different adaptive designs in a specific setting.

References:

1. Proschan, M. A., & Hunsberger, S. A. (1995). Designed extension of studies based on conditional power. Biometrics, 1315-1324.
2. Mehta, C. R., & Pocock, S. J. (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statistics in Medicine, 30(28), 3267-3284.
3. Kunzmann, K., Pilz, M., Herrmann, C., Rauch, G., & Kieser, M. (2021). The adoptr package: adaptive optimal designs for clinical trials in R. Journal of Statistical Software, 98(9), 1-21.
4. Herrmann, C., Pilz, M., Kieser, M., & Rauch, G. (2020). A new conditional performance score for the evaluation of adaptive group sequential designs with sample size recalculation. Statistics in Medicine, 39(15), 2067-2100.
5. Herrmann, C., Kieser, M., Rauch, G., & Pilz, M. (2022). Optimization of adaptive designs with respect to a performance score. Biometrical Journal.

SPECIAL TOPIC SESSION ON HANDLING OVERRUNNING IN GROUP SEQUENTIAL CLINICAL TRIALS

RETHINKING DELAYED RESPONSE

Cornelia Ursula Kunz
Boehringer Ingelheim Pharma GmbH & Co. KG

Stephen Schüürhuis

Common statistical theory often assumes that there is not time gap between the enrollment of a patient and the observation of the outcome of interest and that all patients are enrolled at the same time. In practice, however, patients are enrolled successively and there is a latency between the enrollment of a patient and the availability of the outcome measure.

For single-stage designs, the difference between theory and practice only impacts on the trial duration but not on the statistical analysis and the interpretation thereof. For designs with interim analyses, however, the number of patients already enrolled into the trial and the number of patients with outcome measurements available differ which can cause issues regarding the statistical analysis of the data. The two main issues are: (1) at the time of the interim analysis, there are so-called pipeline patients whose data is not used to make a statistical decision (like stopping early for efficacy), (2) the enrollment into the trial needs to be at least paused for every interim analysis to avoid pipeline patients.

Hampson and Jennison (2013) have proposed a group-sequential design for delayed responses that introduced stopping boundaries for the enrollment of patients followed by critical values to reject the null hypothesis in case the stopping boundaries are crossed. Here, we will discuss some other solutions by rethinking delayed responses, including conditional power approaches, sample-size reassessment and group-sequential designs methodology.

Reference:

Hampson LV, and Jennison C (2013): Group sequential tests for delayed responses (with discussion). Journal of the Royal Statistical Society, Series B (75), 3-54. https://doi.org/10.1111/j.1467-9868.2012.01030.x

SPECIAL TOPIC SESSION ON HANDLING OVERRUNNING IN GROUP SEQUENTIAL CLINICAL TRIALS

ANALYSING OVER-RUN DATA AFTER A GROUP SEQUENTIAL TRIAL

Christopher Jennison
University of Bath

Lisa Hampson

When a group sequential stopping rule is applied in a trial with a time to event endpoint and the study is stopped at an interim analysis, it is almost inevitable that additional data will be recorded. Events may occur between the time data are locked for the interim analysis and the time at which the trial is formally concluded. For an event such as disease progression or death from a specific cause, potential events may be awaiting adjudication at the time of the interim analysis. It is then a concern that a trial should be stopped based on evidence for a positive outcome, demonstrating that the new treatment is superior to the control, but data recorded subsequently will lead to a "reversal" of this conclusion.

The group sequential designs for a delayed response proposed by Hampson and Jennison (2013) can be extended to time to event data. We shall describe their use in creating an error spending test for superiority of a new treatment with a non-binding futility boundary. This design is constructed to match the form of the optimal testing procedures derived by Hampson and Jennison. Our design makes efficient use of the over-run data. It allows a reversal of the anticipated outcome in extreme cases when additional data really do show that rejection of the null hypothesis is liable to be a type I error, but the risk of such a reversal is substantially less than that in Whitehead's (1992) deletion method

References:

Hampson LV, and Jennison C (2013) Group sequential tests for delayed responses (with discussion). Journal of the Royal Statistical Society, Series B, **75**, 3–54. https://doi.org/10.1111/j.1467-9868.2012.01030.x

Whitehead, J (1992) Overrunning and underrunning in sequential clinical trials. Controlled Clinical Trials, **13**, 106–121. https://doi.org/10.1016/0197-2456(92)90017-T

SPECIAL TOPIC SESSION ON HANDLING OVERRUNNING IN GROUP SEQUENTIAL CLINICAL TRIALS

UPDATED ANALYSES AND OVERRUNNING: LESSONS LEARNED DURING THE PANDEMIC AND A PLEA TO PLAN AHEAD

Benjamin Hofner
Paul-Ehrlich-Institut

Elina Asikanius

According to the ICH E9 guideline a trial is to be analyzed and the primary hypothesis is to be "tested when the trial is complete" (ICH 1998). However, the primary analysis and the end of trial often do not coincide, e.g. in group sequential designs, which allow confirmatory testing at interim analyses (potentially without stopping the trial) and in time to event trials were the natural end of study is almost never achieved as usually not all participants experience an event. Fixed design trials where updated analyses or a long-term follow-up are foreseen are yet another example. As terminology for designs with multiple analysis time points (both for GSDs and other designs with updated analyses) often differs between studies, we propose a common terminological framework to improve the communication of study designs and results.

During the COVID-19 pandemic updated analyses after the primary confirmatory analysis were presented as part of the primary body of evidence for some of the vaccine trials. The updates occurred within a very short time frame of around two weeks but were based on a substantially increased database with about twice the number of COVID-19 cases.

The added value of updated analyses is not always that straightforward. While the information increases with a more mature data set, the uncertainty due to unplanned data cutoffs and lack of type 1 error control increases as well. Difficulties in adequately defining and aligning the primary hypothesis test and the benefit-risk assessment arise. A slightly different but related issue is overrunning, which occurs e.g. if an endpoint is not immediately observed and hence further events might accrue after the trial was stopped. These data need to be taken into consideration as well at the time of decision making (CHMP 2007).

Both topics, updated analyses and overrunning, raise issues e.g. regarding type 1 error control, appropriate reporting of the uncertainty of the results, and the impact on decision making. We discuss methodological and regulatory considerations regarding the planning, analysis and reporting of updated analyses or overruning especially in the light of the regulatory assessments. The key element is an appropriate pre-specification of analysis time points and methodological considerations in the light of the value of the analyses for the overall procedure.

References:
ICH (1998). ICH E9 – Statistical principles for clinical trials. CPMP/ICH/363/96
CHMP (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. CHMP/EWP/2459/02

SESSION VI: BAYESIAN METHODS

ADAPTIVE TREATMENT ALLOCATION AND SELECTION IN MULTI-ARM CLINICAL TRIALS: A BAYESIAN APPROACH

Elja Arjas
University of Helsinki

Background: Adaptive designs offer added flexibility in the execution of clinical trials, including the possibilities of allocating more patients to the treatments that turned out more successful, and early stopping due to either declared success or futility. Commonly applied adaptive designs, such as group sequential methods, are based on the frequentist paradigm and on ideas from statistical significance testing. Interim checks during the trial will have the effect of inflating the Type 1 error rate, or, if this rate is controlled and kept fixed, lowering the power.

Results: The purpose of this contribution is to demonstrate the usefulness of the Bayesian approach in the design and in the actual running of randomized clinical trials during phase II and III. This approach is based on comparing the performance of the different treatment arms in terms of the respective joint posterior probabilities evaluated sequentially from the accruing outcome data, and then taking a control action if such posterior probabilities fall below a pre-specified critical threshold value. Two types of actions are considered: treatment allocation, putting on hold at least temporarily further accrual of patients to a treatment arm, and treatment selection, removing an arm from the trial permanently. The main development is in terms of binary outcomes, but extensions for handling time-to-event data, including data from vaccine trials, are also possible. The performance of the proposed methodology is illustrated in extensive simulation experiments involving 2-arm and 4-arm trials. An implementation of the methods is provided in the form of a freely available R package 'barts'.

Conclusion: The proposed methods for trial design provide an attractive alternative to their frequentist counterparts.

Reference:
E Arjas, D Gasbarra (2022). Adaptive treatment allocation and selection in multi-arm clinical trials: a Bayesian perspective. BMC Medical Research Methodology 22 (1), 1-18

SESSION VI: BAYESIAN METHODS

AN ADAPTIVE BAYESIAN DESIGN FOR A PHASE I VACCINE CLINICAL TRIAL WITH ADJUSTMENT FOR AGE GROUP

Miguel Pereira
Cogitars GmbH

Giulia Brunelli

Manuel Wiesenfarth

Oliver Shoenborn-Kellenberger

We present an adaptive Bayesian design for dose finding in a phase I vaccine clinical trial. Unlike cancer clinical trials, in vaccine trials the goal is not necessarily to find the maximum tolerated dose. The aim is to find a well tolerated dose that produces a level of immunogenicity that confers protection, and this translates into a different range of accepted vaccine reactogenicity rates. Additionally, vaccine reactogenicity and immunogenicity is expected to change with age and, in this trial, two age groups were defined: younger adults, 18-55 years old, and older adults, ≥65 years old. The adaptive design aimed to perform dose escalation to find the most suitable dose for each age group.

The dose escalation was guided by the Escalation with Overdose Control principle, and we implemented a Bayesian 3-parameter logistic regression model (BLRM) as a modification of the standard 2-parameter model by adding a parameter to account for the age group. This approach allowed the results of one group to inform the decisions in the other group, increasing statistical power while still allowing for separate dose escalation decisions for each age group. We evaluate the model performance in different clinical trial scenarios and its operating characteristics and analyse our experience running the trial with dose decisions informed by the Bayesian model.

ACKNOWLEDGEMENT

We explicitly thank the "Internationales Wissenschaftsforum Heidelberg (IWH)" for hosting this workshop.