Adaptive Designs and Multiple Testing Procedures Workshop - 2016

Abstract Booklet

Department of Statistical Sciences - University of Padua - Room SC140

Thank you

We thank all speakers and chairs for their contributions.

We are grateful to the Department of Statistical Sciences of the University of Padua, for providing us with a fantastic venue as well as logistic and infrastructure support.

We thank the German and Austro-Swiss Regions of the International Biometric Society for generous financial support.

Florian Klinglmmueller, Livio Finos, Monica Chiogna
(Local organizers)

Andreas Faldum
(Working group chair)

General Information

Conference badges & registration materials

Name badges and registration materials must be picked up on-site at the conference registration desk. The registration desk will be open Thursday, April 28, from 12:00. Conference badges must be visibly worn at all times.

WiFi

You can connect to the `adaptive` network. Access does not require any password but not all protocols are supported. Registered guests can log in to the eduroam WiFi network of the Department using the user ID and password of their home institution.

Coffee Breaks & Lunch

Coffee breaks will be served free of charges to fully registered participants. During the lunch break on Friday a light lunch will be served. If you prefer, there are some restaurants nearby which are delicious and affordable (shown by circles in the map below).

- Ai Scarponi (Via Cesare Battisti ,138)
- La Vecchia Padova (via Cesare Battisti, 37)
- Nane della Giulia/Osteria La Sofia (Via Santa Sofia, 1)

Information about the guided tour

We have organized a guided tour of the Pallazzo Bo - the historic part of the University. It will start at 18:20 at the main University building (Bo in the Map below), which is about a 10 minute walk from the conference venue. We will walk there together, leaving from the conference venue directly after the end of the last session. The tour will take about 40 minutes.

Participation is free of cost, due to limited capacities only attendants who have successfully registered for the tour may attend. Please bring the coupon that you receive with your registration materials.

Information about the dinner

The dinner will be hosted at the restaurant 'Isola di Caprera' (Via Marsilio da Padova, 11/15, see map below) and start at 20:00. We will be offered a four course menu of local sea food specialties. The price of the menu is 35 Euro. An accompanying selection of drinks (Wine, Water, Coffee) are included in the price. For convenience we ask you to pay for the dinner in cash at the registration desk.

Other menu options or, if you have any dietary restrictions are available on request. Please let us know in advance, so we can make sure that your needs are accommodated.

Map


Isola Di Caprera
Piazza della Frutta
Bo
Vecchia Padova
Via Altinate
Via Rinaldo Rinaldi
Via Altinate
Via S. Mattia
Via Gabriele Falloppio
Via Santa Sofia
Via Aristide Gabelli
Via Agnusdei
Nane della Giulia
Osteria Ai Scarponi
Department of Statistic...
Via del Santo
Via Roma

Detailed Program

Timetable

Thursday, April 28	Session title	Friday, April 30	Session title
12:50 - 13:00	Welcome	08:50 - 10:30	Adaptive Design 3
13:00 - 14:40	Multiple Testing 1	10:30 - 11:00	Coffee Break
14:40 - 15:10	Coffee Break	11:00 - 12:40	Multiple Testing 2
15:10 - 16:25	Adaptive Design 1	12:40 - 13:50	Lunch Break
16:25 - 16:45	Coffee Break	13:50 - 15:30	Adaptive Design 4
16:45 - 18:00	Adaptive Design 2	15:30 - 16:00	Coffee Break
18:20 - 19:00	Guided Tour	16:00 - 16:20	AG Meeting
20:00 - 23:00	Dinner	16:20 - 17:35	Multiple Testing 3

Thursday, April 28

13:00 - 14:40: Multiple Testing 1

Chair: ***Gerhard Hommel***

1. **Fortunato Pesarin**
   An Analysis of Union-Intersection and Intersection-Union Tests for Equivalence and Non-Inferiority
   University of Padua, Italy

2. **Helmut Finner**, Veronika Gontscharuk
   Two-sample Kolmogorov-Smirnov type tests and local levels
   Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität Düsseldorf, Germany

3. **Veronika Gontscharuk**, Helmut Finner
   Weighted Kolmogorov-Smirnov tests in one- and two-sample problems
   German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich-Heine-University Dusseldorf, Germany

4. **Robin Ristl**, Dong Xi, Ekkehard Glimm, Martin Posch
   Optimal exact tests for multiple binary endpoints
   Medical University of Vienna, Austria

15:10 - 16:25: Adaptive Design 1

Chair: ***Meinhard Kieser***

1. **Silke Jörgens**, Tobias Mielke
   Considerations On Futility Rules For Adaptive Dose-Finding Designs
   ICON Clinical Research, Germany

2. **Deepak Parashar**, Jack Bowden, Colin Starr, Lorenz Wernisch, Adrian Mander
   Optimal designs for group sequential biomarker-enrichment oncology trials.
   University of Warwick, United Kingdom

3. **Federico Andreis**, Marco Bonetti
   An Adaptive Enrollment Strategy for the Identification of Maximum Treatment Effect Regions
   Bocconi University, Milan, Italy

16:45 - 18:00: Adaptive Design 2

Chair: *Gernot Wassmer*

1. **Willi Maurer**, Byron Jones, Ying Chen
   A robust combination test for sample size adaptation in a two-stage cross-over trial for Average Bioequivalence
   Novartis Pharma, Switzerland

2. **Kevin Kunzmann**, Meinhard Kieser
   Optimal Adaptive Two-Stage Designs for Single-Arm Trials with Binary Endpoint: Recent Improvements and Consistent Inference
   Heidelberg University, Germany

3. **Klemens Weigl**, Ivo Ponocny
   ReThink: Adaptive Two-Stage Designs applied with PsyStatAlpha in Psychological Research
   JKU Linz, Austria

Friday, April 29

08:50 - 10:30: Adaptive Design 3

Chair: ***Florian Klinglmueller***

1. **Josephine Khan**, Peter Kimani, Nigel Stallard, Ekkehard Glimm.
   Selection bias and correlation in seamless phase II/III clinical trials with survival data
   University of Warwick, United Kingdom

2. **Matthias Brückner**, Werner Brannath
   Interim Decisions in Adaptive Clinical Trials with Time-to-event Surrogate and Primary Endpoints
   Universität Bremen, Germany

3. **Thomas Asendorf**, Robin Henderson, Heinz Schmidli, Tim Friede
   Blinded Sample Size Reestimation for Time Dependent Negative Binomial Counts with Incomplete Follow-up
   University Medical Center Göttingen, Germany

4. **Roland Gera**, Tim Friede
   Blinded sample size reestimation for Adaptive Enrichment designs with Longitudinal Data
   Universitätsmedizin Göttingen, Germany

11:00 - 12:40: Multiple Testing 2

Chair: ***Jelle Goeman***

1. **Arnold Janssen**
   Martingale approach for multiple testing and FDR control
   Heinrich-Heine University Duesseldorf, Germany

2. **Aldo Solari**, Jelle J. Goeman
   Minimally Adaptive BH: a tiny but uniform improvement of the procedure of Benjamini and Hochberg
   University of Milano-Bicocca, Italy

3. **Jesse Hemerik**
   False discovery proportion estimation by permutations: confidence for SAM
   Leiden University Medical Center, Netherlands

4. **Djalel-Eddine Meskaldji**, Stephan Morgenthaler
   Thresholding of ordered p-values: which error rate is being controlled?
   EPFL, Switzerland

13:50 - 15:30: Adaptive Design 4

Chair: ***Andreas Faldum***

1. **Yida Chiu**
   Designs and Estimation in Clinical trails with Subpopulation Selection
   Lancaster University, United Kingdom

2. **Laura Kohlhas**, Meinhard Kieser
   Timing of subgroup selection in adaptive enrichment designs
   Univeristy of Heidelberg, Germany

3. **Johannes Krisam**, Meinhard Kieser
   Optimal Subgroup Selection Rules in Adaptive Oncology Trials with Time-to-Event Outcome
   University of Heidelberg, Germany

4. **Marius Placzek**, Tim Friede
   Analysis, Sample Size Calculation and Recalculation in Designs with Multiple Nested Subgroups
   University Medical Center Göttingen, Germany

16:20 - 17:35: Multiple Testing 3

Chair: *Livio Finos*

1. **Kornelius Rohmeyer**, Werner Brannath, Sarah Nanninga
   The Populationwise Error Rate - A More Liberal Error Rate for Multiplicity Adjustment in Enrichment Designs
   University of Bremen, Germany

2. **Natalia Sirotko-Sibirskaya**, Prof. Dr. Thorsten Dickhaus, Prof. Dr. Markus Pauly
   Simultaneous Statistical Inference in Dynamic Factor Models (Estimation, Simulation, Application)
   University of Bremen, Germany

3. **Fang Wan**, Wei Liu; Frank Bretz; Yang Han;
   Confidence Sets for Optimal Factor Levels of a Response Surface
   Lancaster University, United Kingdom

List of abstracts

An Analysis of Union-Intersection and Intersection-Union Tests for Equivalence and Non-Inferiority

Fortunato Pesarin
(*University of Padua, Italy*)

Nunnally (1960) wrote: "To minimize type II errors, large samples are recommended. In psychology, practically all sharp or point null hypotheses are claimed to be false for sufficiently large samples so ... it is nonsensical to perform an experiment with the sole aim of rejecting the null hypothesis". Rather than only one point this concept suggests considering the null hypothesis as a closed equivalence interval. To obtain practical solutions, a permutation Union-Intersection (UI) procedure [2] is presented.

The notion of testing for equivalence of two treatments is widely used in clinical trials, pharmaceutical experiments, bioequivalence and quality control. It is traditionally approached by Two One-Sided Tests (TOST) within the Intersection-Union (IU) principle. According to this principle, the null hypothesis is stated as the set of effects lying outside a suitably established interval, and the alternative as the set of effects lying inside the open equivalence interval. The solutions provided in the literature are mostly based on likelihood techniques, which in turn are rather difficult to handle except for cases lying within the regular exponential family and the invariance principle.

The main goals of present communication are: i) to go beyond the limitations of likelihood based methods by working in a nonparametric permutation frame; ii) to provide a parallel analysis of IU and UI solutions; iii) limiting properties, an example and a small simulation study for evaluating their main properties are also presented.

1. Nunnally, J. (1960) EDUC PSYCHOL MEAS
2. Pesarin, F, et al. (2015) STAT COMPUT

Two-sample Kolmogorov-Smirnov type tests and local levels

Helmut Finner, Veronika Gontscharuk
(*Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität Düsseldorf, Germany*)

Local levels can be viewed as an interesting characteristic of union-intersection based overall tests and were recently studied for union-intersection based one-sample goodness-of-fit (GOF) tests. We adopt the concept of local levels to a specific class of union-intersection based two-sample Kolmogorov-Smirnov (KS) type test. Members of this class are e.g. the original Kolmogorov-Smirnov tests, weighted Kolmogorov-Smirnov tests and the supremum version of Anderson-Darling type tests. Such tests are closely linked to 2×2-table tests. For example, an important desirable structural property (Barnard-convexity) of 2×2-table tests turns out to be inherent for two-sample KS type tests. Exact local levels of KS type tests can be computed in terms of the hypergeometric distribution while approximate local levels are typically derived by a normal approximation. We illustrate the behaviour of local levels of various two-sample Kolmogorov-Smirnov (KS) type tests. Furthermore, we propose some new tailored two-sample KS type tests by means of local levels.

Weighted Kolmogorov-Smirnov tests in one- and two-sample problems

Veronika Gontscharuk, Helmut Finner
(*German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich-Heine-University Dusseldorf, Germany*)

We consider a specific class of weighted Kolmogorov-Smirnov (KS) tests. The well-known KS tests and the supremum version of Anderson-Darling type tests represent boundary cases in this class. In this talk we focus on various properties of two-sample weighted KS tests. Thereby, so-called local levels play a crucial role. Local levels are defined as local rejection probabilities and can be viewed as a measure of local sensitivity of a test. We provide the asymptotics of weighted KS statistics and the asymptotics

of the corresponding local levels. It turns out that the one- and two-sample asymptotics may differ. Moreover, we show that the two-sample supremum version of Anderson-Darling type tests asymptotically coincides with a specific minimum p-value test. Finally, we compare finite and asymptotic results in the one- and two-sample cases.

Optimal exact tests for multiple binary endpoints

Robin Ristl, Dong Xi, Ekkehard Glimm, Martin Posch
(*Medical University of Vienna, Austria*)

In confirmatory clinical trials with limited sample size (e.g. in rare diseases), challenges arise because the asymptotic theory may lose accuracy to approximate the distribution of test statistics. Often, non-parametric exact tests are applied instead. However, the distribution of the corresponding test statistics is usually discrete and they may be overly conservative. This phenomenon may become more severe when it is required to control the familywise Type I error rate (FWER) for multiple simultaneous inferences. To overcome this drawback, we propose an optimal multiple testing procedure for binary endpoints to compare a treatment versus a control. The proposed procedure explores the joint distribution of the test statistics for multiple Fisher's exact tests to alleviate the impact of discreteness. The optimal rejection region is then derived under the constrained optimization framework using the linear integer programming technique. The subsequent application of the closed testing principle leads to an optimal multiple testing procedure with strong control of the FWER. Modifications to the procedure provide tests with the properties of consonance or alpha-consistency.

Considerations On Futility Rules For Adaptive Dose-Finding Designs

Silke Jörgens, Tobias Mielke
(*ICON Clinical Research, Germany*)

Phase 2 Dose-Finding studies generally target the description of the dose-response curve in order to allow selection of safe and efficacious doses for following confirmatory Phase 3 studies. For the initiation of costly Phase 3 studies, the efficacy of the compound is of paramount interest and is already examined during Phase 2. Interim adaptations in dose-finding studies frequently target the optimization of patient allocation. Interim analyses allow early development stops, e.g. if further development is considered futile. The probability of false interim decisions depends on the rules considered. Multiple dose groups may increase the probability of false interim decisions, if only pairwise comparisons to control are considered. Dose-response modelling assumptions (e.g. using MCPMod) improve early identification of futile development programs. In the example of seamless PoC & dose-finding designs, this may significantly enhance the study design. A range of different types of futility rules will be examined in this presentation. Model based and model free approaches will be discussed and their implications on the efficiency of adaptive dose-finding designs will be presented.

Optimal designs for group sequential biomarker-enrichment oncology trials.

Deepak Parashar, Jack Bowden, Colin Starr, Lorenz Wernisch, Adrian Mander
(*University of Warwick, United Kingdom*)

Oncology trials based on biomarker-stratified designs are used to establish the effectiveness of a new drug or targeted therapy in specific populations. Targeted or enriched designs are a class of such stratified trial designs that aim to enrich the biomarker-positive sub-population. Jones and Holmgren (JH) [1] have proposed a design to determine whether drug has activity only in target population or the general population in the disease area. Their design is an enrichment adaptation based on two parallel Simon two-stage designs. We study the JH design in detail, establish it's group sequential nature and appropriately control the type I and type II error probabilities that yield novel optimal designs [2]. We also discuss alternative FWER and Individual Hypothesis control in weak as well as strong sense. Our approach provides a robust framework for adaptive enrichment in biomarker-based Phase II trial design. [1] Jones CL,Holmgren E. An adaptive Simon two-stage design for phase 2 studies of targeted therapies. Contemp. Clin. Trials 2007; 28: 654–661. [2] Parashar D,Bowden J,Starr C,Wernisch L,Mander A. An optimal stratified Simon two-stage design. Pharm. Stats, 2016; in press.

An Adaptive Enrollment Strategy for the Identification of Maximum Treatment Effect Regions

Federico Andreis, Marco Bonetti
(*Bocconi University, Milan, Italy*)

Adaptive strategies have received great attention in the literature in recent years as they may improve the current practice in clinical trials [1], [2]. The aim of this work is to contribute to the literature by proposing an enrollment design that incorporates outcome evidence as soon as it becomes available. Consider a two-arm randomized trial to compare two treatments for a disease such that an outcome is available very quickly after randomization. We consider two primary targets: (i) to oversample individuals that are more likely to present a large (small) treatment effect, and (ii) to identify the covariate region where the treatment effect is largest (smallest). A first sample of patients is obtained from the general patient population by means of probability proportional-to-size designs and randomized. After the first sample is drawn, oversampling is sought for by setting inclusion probabilities for the remaining units to be proportional to their expected treatment effect, as estimated from the evidence provided by previously enrolled patients; the new patients are then randomized. The process continues until a desired sample size is obtained or some other stopping rule is satisfied. Identification of the covariate regions of largest (smallest) treatment effect can then be attempted by analyzing the path in the covariate space induced by the adaptive procedure.

[1] Kairalla, Coffey, Thomann and Muller (2012): Adaptive trial designs: a review of barriers and opportunities. Trials, 13:145.
[2] Atkinson and Biswas (2013): Randomised Response-Adaptive Designs in Clinical Trials. CRC press.

A robust combination test for sample size adaptation in a two-stage cross-over trial for Average Bioequivalence

Willi Maurer, Byron Jones, Ying Chen
(*Novartis Pharma, Switzerland*)

Four methods for sample size re-estimation in a two-stage 2x2 cross-over trial for testing for Average Bioequivalence (ABE) were presented in Potvin et al. (2008). However, none of these methods formally controls the Type I error rate of falsely claiming ABE. In fact, the assessment of a possible inflation in the error rate has to be done in an ad hoc way using simulation. We describe an alternative method of sample size re-estimation that is exact and guaranteed to control the Type I error rate. This method uses a new and robust version of the weighted combination of p-values test in conjunction with standard group sequential techniques. The sample size re-estimation step is based on significance levels and power requirements that are conditional on the first-stage results. We compare the operating characteristics of the new method with those of the Potvin et al. (2008). References: Potvin, D., et al. (2008).Sequential design approaches for bioequivalence studies with crossover designs. Pharmaceutical Statistics, 7, 245-262. Kieser, M. and Rauch, G. (2014). Two-stage designs for cross-over bioequivalence trials. Statist. Med., 34, 2403–2416.

Optimal Adaptive Two-Stage Designs for Single-Arm Trials with Binary Endpoint: Recent Improvements and Consistent Inference

Kevin Kunzmann, Meinhard Kieser
(*Heidelberg University, Germany*)

Early oncological trials are often planned with a single arm and a binary endpoint. Adaptive designs account for the uncertainty about the true effect size by determining the final sample size within an ongoing trial after an interim look at the data. Using mixed integer programming, we derive a general approach for finding designs which minimize expected sample size under the null hypothesis for various constraints. The resulting designs improve previous work both in terms of efficiency as well as practical appeal by avoiding pathologies arising from the discreteness of the underlying distribution. We demonstrate how existing approaches to inference fail to fulfill elementary consistency requirements between estimation, p values, and hypothesis test and propose a novel way of deriving point estimates and p values in this situation, which is consistent in the sense that the p value derived from the ordering of the outcome space

induced by the point estimator is smaller than the significance level if and only if the chosen test rejects the null hypothesis. The bias and MSE profiles of the resulting estimators compare favorable to those of alternative estimators.

ReThink: Adaptive Two-Stage Designs applied with PsyStatAlpha in Psychological Research

Klemens Weigl, Ivo Ponocny
(*JKU Linz, Austria*)

The great potential of group sequential and adaptive designs is well documented in medical and pharmaceutical statistics. Forseeably, these highly sophisticated designs may also have a significant impact on future developments of psychological research settings. Psychology is subdivided into several, immensely diverse psychological fields with highly diverging psychological paradigms. For decades, the overwhelming majority of all conducted psychological trials have got "two-sided" statistical testing in common. Hence, group sequential methods and adaptive two-stage designs for "two-sided" testing, respectively, are of great interest for psychological research. Though the number of R users increases worldwide, there exists only a small number of psychological researchers who actually use R and could apply adaptive designs with appropriate R packages. We tackle these challenges and apply adaptive two-stage designs within the context of psychological research settings. For this attempt we apply the especially conceptualized, programmed and user-friendly software "PsyStatAlpha 1.1" on randomly sampled psychological data from a fairly large subjective well-being survey of Austria.

Selection bias and correlation in seamless phase II/III clinical trials with survival data

Josephine Khan, Peter Kimani, Nigel Stallard, Ekkehard Glimm.
(*University of Warwick, United Kingdom*)

In seamless phase II/III trials, at an interim analysis, the 'best' performing treatment is selected for further study in a confirmatory setting. Selection can lead to overly optimistic and thus biased estimates. When the analysis is based on survival data, estimates are correlated due to the common control as well as censoring, as patients who do not experience the event in stage 1 are followed up further in stage 2. Therefore, the common assumption of independence of stage 1 and 2 data in a seamless phase II/III trial is violated. Uniformly minimum variance conditionally unbiased estimators (UMVCUE's) have been developed for phase II/III trials with normally distributed outcomes. They efficiently correct for selection bias, but are based on an assumption of independence between stages. To extend these methods to survival data, we have derived an UMVCUE that corrects for correlation between stage 1 and 2 data. Under different trial settings with a common control, we will present the selection bias and illustrate the UMVCUE's for a range of hazard ratios for which the normality assumption holds. This method will allow robust estimation of survival data in adaptive designed trials.

Interim Decisions in Adaptive Clinical Trials with Time-to-event Surrogate and Primary Endpoints

Matthias Brückner, Werner Brannath
(*Universität Bremen, Germany*)

In phase 3 survival trials the final decision is usually based on overall survival, while in interim analyses a surrogate endpoint might be used for decision making. However, relying on a surrogate endpoint only might be misleading, when the treatment effect in the surrogate endpoint does not closely correspond to the treatment effect in the primary endpoint. Hence the desire to use the combined information from surrogate and primary endpoint in a decision rule, e.g. when selecting subgroups in an adaptive enrichment design or making the phase 3 go/no go decision at the end of a phase 2 trial. We consider joint models for progression-free survival (PFS) and overall survival (OS) in order to derive decision rules based on PFS, OS or on an optimal combination of both. Assessment of these decision rules is based on thesymptotic distribution of the treatment effect estimators from the model and on Monte-Carlo simulations.

Blinded Sample Size Reestimation for Time Dependent Negative Binomial Counts with Incomplete Follow-up

Thomas Asendorf, Robin Henderson, Heinz Schmidli, Tim Friede
(*University Medical Center Göttingen, Germany*)

Sample size determination is a crucial step in planning clinical trials. However, sample size estimates strongly depend on nuisance parameters, which have to be guestimated from previous trials. Blinded sample size reestimation procedures allow for an adjustment of the calculated sample size within an ongoing trial, by accumulating data to estimate relevant nuisance parameters without unblinding the trial (Friede 2010). We consider two models for time dependent discrete observations with marginal negative binomial distribution, data observed e.g. as longitudinally collected MRI lesion counts in RMS trials. First, we consider a binomial thinning model (McKenzie 1986) for statistical inference of time dependent count data and provide sample size estimation and reestimation techniques within this model. Benefits of incorporating incomplete follow up times within the binomial thinning model are illustrated. Further, we consider a Gamma frailty model (Fiocco 2009) which is suited in situations with time point specific event rates. Advantages and disadvantages of both models are discussed. A simulation study is conducted to assess the finite sample properties of the proposed procedures.

Blinded sample size reestimation for Adaptive Enrichment designs with Longitudinal Data

Roland Gera, Tim Friede
(*Universitätsmedizin Göttingen, Germany*)

Adaptive Enrichment Designs (AED) have become an important tool in clinical research, in particular in personalized medicine. The additional features of AED help to investigate populations for potential subgroups in an efficient and time saving manner. In this presentation AED are suggested for clinical trials with longitudinal data. For this setting, the advantages of two-stage AED [1] with blinded sample size reestimation (BSSR) [2] and subgroup selection over common one-stage designs are presented by simulation studies. Operating characteristics are explored through simulation studies with various settings. We find that AED are in many cases superior alternatives to common one-step designs in terms of sample size and power and detection of subgroups. The proposed BSSR procedure makes the AED robust to misspecifications of nuisance parameters in the planning phase. These findings are underlined by our simulations.

[1] Friede T, Parsons N, Stallard N (2012) A conditional error function approach for subgroup selection in adaptive clinical trials. Statistics in Medicine 31:4309-4320
[2] Jung S.-H., Ahn C (2003) Sample size Estimation for GEE method for comparing slopes in repeated measurements data. Statistics in Medicine 22:1305-1315

Martingale approach for multiple testing and FDR control

Arnold Janssen
(*Heinrich-Heine University Duesseldorf, Germany*)

Under martingale dependence the false discovery rate (FDR) of various multiple tests can exactly be calculated. The results are key tools in order to discuss finite sample FDR control of these tests. Some of these results are also new when the p-values are independent. It is shown how the famous Benjamin/Hochberg multiple test can be modified. Martingale models are also helpful for step down multiple tests. The second part of the talk discusses adaptive multiple tests with data dependent critical values. We extend the adaptive multiple test given by Storey. In Heesen and Janssen (2016) it is shown that a large class of adaptive multiple tests allow the finite sample FDR control.

References P. Heesen and A. Janssen (2015) Inequalities for the false discovery rate (FDR) under dependence, Electron. J. Stat. 9, 679-716. P. Heesen and A. Janssen (2016) Dynamic adaptive multiple tests with finite sample FDR control, J. Statist. Plann. Inference 168, 38-51.

Minimally Adaptive BH: a tiny but uniform improvement of the procedure of Benjamini and Hochberg

Aldo Solari, Jelle J. Goeman
(*University of Milano-Bicocca, Italy*)

We define an adaptive procedure for control of the false discovery rate that is uniformly more powerful than the procedure of Benjamini and Hochberg. The power gain is tiny, however, and only appreciable for small numbers of hypotheses. We illustrate the new method with the case of two hypotheses, for which so far no procedure was known that controls false discovery rate but not also familywise error rate under positive dependence.

False discovery proportion estimation by permutations: confidence for SAM

Jesse Hemerik
(*Leiden University Medical Center, Netherlands*)

SAM is a highly popular multiple testing method that estimates the false discovery proportion (FDP), the fraction of false positives among all rejected hypotheses. It does so based on permutations of the data. Perhaps surprisingly, until now this method had no known properties. We extend SAM by providing (1-alpha)100%-confidence upper bounds for the FDP, so that exact confidence statements can be made. As a special case, an estimate of the FDP is obtained that underestimates the FDP with probability at most 0.5. Moreover, using a closed testing procedure, we decrease the upper bounds and estimates in such a way that the confidence level is maintained.

Thresholding of ordered p-values: which error rate is being controlled?

Djalel-Eddine Meskaldji, Stephan Morgenthaler
(*EPFL, Switzerland*)

Many multiple testing procedures are based on cutting ordered p-values. We present new results that indicate, under different assumptions on the p-values, which type I error rate being controlled when an arbitrary non-decreasing threshold sequence is used. We discuss the advantage of some particular threshold sequences under an optimality framework that considers power of detecting true effects, as well as stability and robustness.

Designs and Estimation in Clinical trails with Subpopulation Selection

Yida Chiu
(*Lancaster University, United Kingdom*)

Addressing population heterogeneity necessitates designs and analysis of clinical trials with subpopulation. Several types of designs have been invented for different circumstances and serve as the basis of subgroup analysis. The accuracy and precision of estimation is also crucial to the development of novel treatments and decisions of treatment implementation. In this talk, we first present designs context and outline the associated design procedures by group sequential designs and under the scheme of fixed subgroup sample size according to subgroup prevalence. Then based on simulation studies on single-stage, and two-stage designs, we provide an overview of estimation assessment of the maximum likelihood estimator (MLE) of effect sizes for subgroups under various circumstances on prevalence and underlying effect sizes of subgroups.

Timing of subgroup selection in adaptive enrichment designs

Laura Kohlhas, Meinhard Kieser
(*Univeristy of Heidelberg, Germany*)

With increasing interest in personalized medicine over the last years, the proof of efficacy in specific subgroups of the total patient population becomes more important. For example, in oncology trials, predictive biomarkers are frequently used to identify a subgroup showing potentially a higher benefit from the treatment. Adaptive enrichment designs consider the uncertainty with respect to the treatment effect in the total population and a subgroup by selecting the target population with the most promising benefit based on the results of an interim analysis (see, e.g., [1]).

For the situation of a normally distributed outcome, we investigate the impact of the interim analysis timing on power. We consider interim decisions based on absolute effect estimates [1] and on the estimated effect difference between the two populations. The performance characteristics are investigated for various effect sizes and prevalences of the subgroup.

[1] Jenkins M, Stone A, Jennison C (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical Statistics 10:347-356.

Optimal Subgroup Selection Rules in Adaptive Oncology Trials with Time-to-Event Outcome

Johannes Krisam, Meinhard Kieser
(*University of Heidelberg, Germany*)

When investigating the efficacy of a recently developed therapy, there is often some doubt whether the treatment might be more or even solely beneficial for a subgroup of the target population. Adaptive enrichment designs incorporating a mid-course efficacy assessment have been proposed as a solution (see, e.g., [1]). After an interim analysis, it is decided whether to continue the trial with the total population or only the subgroup. The employed interim decision rule has a crucial impact on the probability of a correct interim decision and the power of the trial [2].

For the situation of a time-to-event variable as primary outcome, exact formulae for optimal decision thresholds are derived which incorporate the uncertainty about treatment effects by modeling knowledge gained from previous trials by a prior distribution. These optimal rules are evaluated regarding their performance characteristics and are compared to ad-hoc rules proposed in the literature. Our methods are illustrated by means of a clinical trial example.

[1] Jenkins M et al (2011) Pharm Stat 10:347–356. [2] Krisam J, Kieser M (2015) IJMS 16:10354–10375.

Analysis, Sample Size Calculation and Recalculation in Designs with Multiple Nested Subgroups

Marius Placzek, Tim Friede
(*University Medical Center Göttingen, Germany*)

Due to the growing interest in personalized medicine and targeted therapies, the importance of subgroup analyses is increasing. Here designs with multiple nested subgroups will be considered. For the analysis we suggest using the joint distribution of standardized test statistics corresponding to each (sub)population. However, this joint distribution varies with the knowledge about the nuisance parameters, i.e. the variances and prevalences in the populations. We will derive multivariate exact distributions (Genz & Bretz 2009), where possible, and provide approximations for various scenarios, e.g. known variance, or same, but unknown variances across subgroups. Additionally we will give a sample size calculation procedure in those cases. Due to uncertainties about the nuisance parameters which are needed for sample size calculations, a sample size review can be performed in order to make the study more robust against misspecifications (Internal Pilot Study Design, Wittes & Brittain, 1990). We will present a method that performs a blinded sample size reestimation and some tricks that can be applied here when dealing with small sample sizes.

The Populationwise Error Rate - A More Liberal Error Rate for Multiplicity Adjustment in Enrichment Designs

Kornelius Rohmeyer, Werner Brannath, Sarah Nanninga
(*University of Bremen, Germany*)

In clinical studies control of the familywise error rate is appropriate when several hypotheses are investigated on the same population. When the population however splits into disjunct subpopulations and each hypothesis only concerns one of these without a claim beyond the subpopulation, the overall study essentially consists of separate trials which share only the same infrastructure. In this case the familywise error rate is unreasonably conservative. In some cases the subpopulations are disjunct by definition (like two groups 'biomarker positive' and 'negative/unknown'), but in many other cases the subpopulations can overlap. For this setting we propose a generalized error rate that takes into account the probability to belong to a certain subpopulation or intersection of subpopulations. This error rate - which we call the populationwise error rate - extends continuously the spectrum from the FWER in the first setting to the unadjusted case for disjunct populations. We start defining simultaneous test procedures with control of the populationwise error rate. We then generalize the closed testing principle and show how to construct more powerful step-down tests. The gain in power and sample size by using the populationwise error instead of the familywise error rate is illustrated by first simple examples.

Simultaneous Statistical Inference in Dynamic Factor Models (Estimation, Simulation, Application)

Natalia Sirotko-Sibirskaya, Prof. Dr. Thorsten Dickhaus, Prof. Dr. Markus Pauly
(*University of Bremen, Germany*)

In their paper 'Simultaneous Statistical Inference in Dynamic Factor Models' Dickhaus and Pauly 2015 introduce a likelihood-based inference technique which allows for simultaneous testing in (exact) dynamic factor models. The newly introduced testing methodology is based on multivariate central limit theorem for empirical Fourier transforms of observable time series which authors prove in their work and which allows for more elaborate testing approach in the context of dynamic factor models. In my work I provide simulations for the methodology introduced in Dickhaus and Pauly 2015 and test it on the empirical data.

Confidence Sets for Optimal Factor Levels of a Response Surface

Fang Wan, Wei Liu; Frank Bretz; Yang Han;
(*Lancaster University, United Kingdom*)

Construction of confidence sets for the optimal factor levels is an important topic in response surfaces methodology. In our earlier work, an exact (1-a) confidence set has been provided for a maximum or minimum point (i.e. an optimal factor level) of a univariate polynomial function in a given interval. In this talk, the method has been extended to construct an exact (1-a) confidence set for the optimal factor levels of response surfaces.

The construction method is readily applied to many parametric and semi-parametric regression models involving a quadratic function. A conservative confidence set has been provided as an intermediate step in the construction of the exact confidence set. Examples are given to illustrate the application of the confidence sets.

CANCELLED: A method for dose-finding based on weighted differential entropy that does not require a monotonicity assumption

Thomas Jaki, Pavel Mozgunov
(*Lancaster University, United Kingdom*)

Combination treatments are common in diseases such as cancer and tuberculosis and adequate identification of the optimal dose and regime is essential. Almost all of the existing methods are constructed for two-agents only and based on the monotonicity assumption of toxicity. Moreover, the parametric setting for the set of toxicities restrict the search of MTD in a particular surface and it is not able to find the dose of interest even for large sample size if model is misspecified.

We propose a dose-escalation method that does not require monotonicity or any pre-specified model dependence between different doses that is suitable for single and multi-agent dose-escalation trial. The method is based on the weighted differential entropy that attracted a considerable attention in the information theory recently. We show that the proposed method is comparable to well-studied methods such as the CRM in the monotonic scenarios and outperforms other methods in non-monotonic ones.

CANCELLED: A comparison of multiple testing procedures (MTPs) for testing both the overall and one subgroup specific effect in confirmatory clinical trials

Hong Sun, Werner Vach, Institute for Medical Biometry and Statistics, Frank Bretz, Novartis AG
(*Institute for Medical Biometry and Statistics, Germany*)

In confirmatory randomized clinical trials, often a single marker is considered relevant for the treatment selection. The treatment effect is tested with two hypotheses, i.e., whether the treatment efficacy exists in the overall population (Ho) and/or in marker-positive subgroup (H+) or not. Many MTPs have been proposed or applied for this purpose. We compare five non-paprametric and parametric procedures with closure principle through simulation studies, i.e., Song-Chi (SC), weighted parametric (WP), weighted-Holm, fallback procedures and weighted Bonferroni test. Three powers – the powers to reject Ho, to reject H+, and to reject Ho or H+ – are considered for all five procedures in different scenarios. From the results of simulation studies, we found that WP obtains highest powers among all procedures under the same setting of weights in most of scenarios. Due to the consistency constraint in SC procedure, SC provides less power than WP, sometimes even lower than non-parametric procedures. It also performs poor when the treatment effect also exists in the complementary subgroup, so we should be more cautious when applying SC procedure.

CANCELLED: A MANOVA test for multivariate lognormal observations with a spike at zero, with application to ecological niches of South Africa

Alessio Farcomeni
(*University of Rome - Sapienza, Italy*)

We develop an asymptotic likelihood ratio test for multivariate log-normal data with a point mass at zero in each dimension. The test generalizes Wilks' Lambda and Hotelling T-test to the case of semi-continuous data. Simulations show that the resulting test statistic attains the nominal Type I error rate and has good power for reasonable alternatives. We conclude with an application to exploration of ecological niches of trees of South Africa.