ADAPTIVE DESIGNS AND MULTIPLE COMPARISON PROCEDURES

Workshop June 24-26, 2015

in Köln

Sponsored by:

International Biometric Society, Deutsche Region

Bayer Pharma AG

Boehringer Ingelheim GmbH

Meda Pharma GmbH & Co. KG

Merck KGaA

Novartis AG

Organizing committee:
Gernot Wassmer (Köln), Martin Hellmich (Köln), Hildegard Christ (Köln), Annemarie Müller (Köln), Werner Brannath (Bremen), Andreas Faldum (Münster)

LOCATION

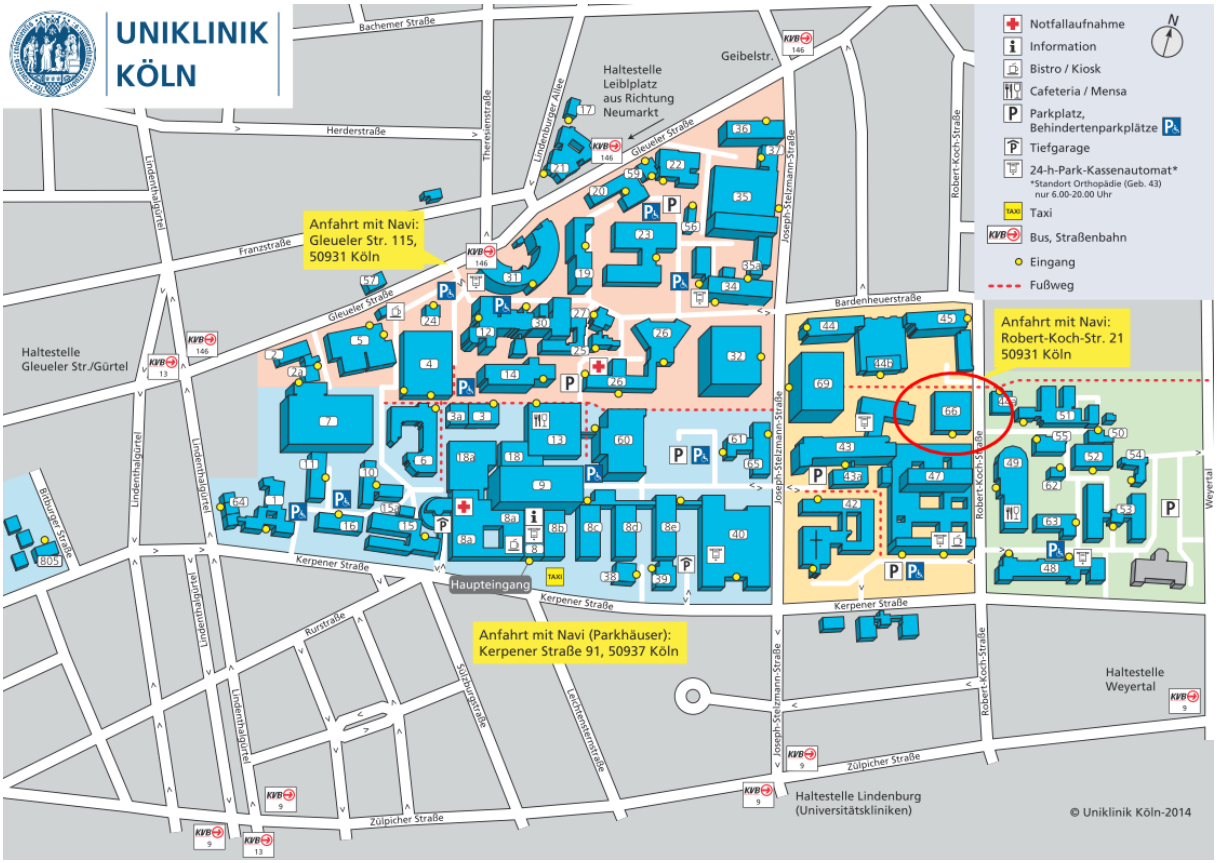
Zentrum für Molekulare Medizin Köln (ZMMK)
ZMMK-Forschungsgebäude (Geb. 66)
Großer Seminarraum
Robert-Koch-Str. 21
50931 Köln - Germany

Public transport:

Tram line 9 (Königsforst - Sülz)
Station: Weyertal or
Station: Lindenburg (Universitätskliniken)

Tram line 13 (Holweide - Sülzgürtel)
Station: Gleuelerstraße/Gürtel

Bus line 146
Bus stop: Geibelstraße



GUIDED TOUR

There will be three guided tours on

Wednesday, June 24, 19:00

It will be on

"Tapas Colonia - Stadtgeschichte Häppchenweise",

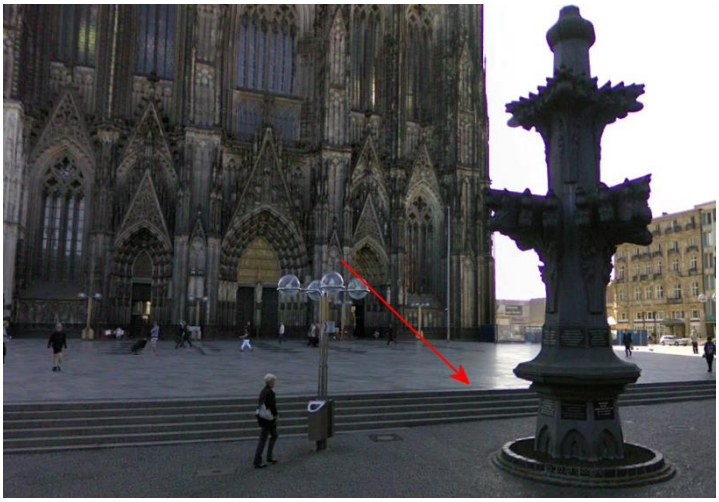
Historisch, Legenden, Milieu, Anekdötchen, Archäologie und eine Absacker-Kölschstation mit Infos zum Kölner Nationalgetränk

also in english, if preferred.

They start at the

Kreuzblume vor den Domtürmen

which is in front of the cathedral.



CONFERENCE DINNER

The conference dinner takes place on

Thursday, June 25, 19:30

in the ”Domstube” (upstairs) at the “Gaffel am Dom” which is opposite to the cathedral or the central train station.

The address is

Gaffel am Dom GmbH
Bahnhofsvorplatz 1
50667 Köln (Altstadt-Dom)
Eingänge über Trankgasse od. Bahnhofsvorplatz

SCIENTIFIC PROGRAM OVERVIEW

Wednesday, June 24

14:00	Registration
14:15 – 14:30	Welcome
14:30 – 16:00	Session 1: Multiple Tests and Confidence Intervals
16:00 – 16:30	Coffee break
16:30 – 18:00	Session 2: Blinded Sample Size Calculation
19:00:	Guided Old City tour

Thursday, June 25

9:10 – 11:00	Session 3: Subgroup Analysis and Enrichment Designs
11:00 – 11:30	Coffee break
11:30 – 13:00	Session 4: Adaptive Designs I
13:00 – 14:00	Lunch break
14:00 – 15:50	Session 5: Multiple Testing Procedures
15:50 – 16:30	Coffee break (16:15 – 16:30: Session of the Working Group)
16:30 – 18:00	Session 6 in honour of Prof. Walter Lehmacher
19:30	Conference dinner

Friday, June 26

9:30 – 11:00	Session 7: Surrogate Variables and Recurrent Event Data
11:00 – 11:30	Coffee break
11:30 – 13:20	Session 8: Adaptive Designs II
13:20	End of workshop

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE

SESSION 1: MULTIPLE TESTS AND CONFIDENCE INTERVALS

Wednesday, 14:30- 16:00

Chair: Helmut Finner

Veronika Gontscharuk and Helmut Finner: Union-intersection based goodness-of-fit tests in terms of local levels

Konstantin Schildknecht: Simultaneous statistical inference for epigenetic methylation data

Thorsten Dickhaus: Uncertainty quantification for the family-wise error rate in multivariate copula models

Ludwig A Hothorn and C. Ritz: Simultaneous confidence intervals by multiple marginal models using R

SESSION 2: BLINDED SAMPLE SIZE RECALCULATION

Wednesday, 16:30- 18:00

Chair: Tim Friede

Marius Placzek and Tim Friede: Nuisance-parameter based sample size reestimation in adaptive enrichment designs with an application in major depression

Cornelia Ursula Kunz, Tim Friede, Nicholas Parsons, Susan Todd, Nigel Stallard: Blinded versus unblinded estimation of the correlation coefficient in adaptive designs with long-term and short-term outcomes

Thomas Asendorf: Blinded sample size reestimation for time dependent negative binomial observations

Anja Sander, Geraldine Rauch, Meinhard Kieser: Blinded sample size re-calculation in clinical trials with binary composite endpoints based on correlation-adjusted local significance levels

SESSION 3: SUBGROUP ANALYSIS AND ENRICHMENT DESIGNS

Thursday 9:10- 11:00

Chair: Werner Brannath

Thomas Ondra, Sebastian Jobjörnsson, Robert Beckman, Carl-Fredrik Burman, Franz König, Nigel Stallard, and Martin Posch: A decision theoretic approach to subgroup selection

Willi Maurer, Frank Bretz, Xiaolei Xun: Decision analytic approach to subgroup analysis

Johannes Krisam, Meinhard Kieser: Performance of subgroup selection rules for a targeted therapy in oncology

Daria Tutschkow, Astrid Dempfle, Nina Timmesfeld: A group-sequential design to test efficacy and inefficency in two subgroups

Heiko Götte, Margarita Donica, Giacomo Mordenti: Improving probabilities of correct interim decision in population enrichment designs

SESSION 4: ADAPTIVE DESIGNS I

Thursday 11:30- 13:00

Chair: Meinhard Kieser

Tobias Mielke: Adaptive treatment arm selection in multivariate bioequivalence trials

Tina van der Horst, Michael Budde, Alexander Strasak, Máximo Carreras: Adaptive seamless phase II/III study in gastric cancer (orphan condition)

Rene Schmidt and Andreas Faldum: On adaptive designs with dependent p-values based on Fisher's combination test

Marc Urban and Andreas Faldum: Adaptive designs and their impact on clinical trials' cost structures – an economic perspective on adaptive designs

SESSION 5: MULTIPLE TESTING PROCEDURES

Thursday 14:00- 15:50

Chair: Gerhard Hommel

Gerald Hlavin, Peter Bauer, Franz König: Adjusting multiplicity using safety data in many-one comparisons

Robin Ristl and Martin Posch: Optimal rejection regions for testing multiple binary endpoints in small samples

Dennis Görlich, Robert Kwiecien, Andreas Faldum: Optimizing graphical multiple testing procedures

Helmut Finner and Klaus Strassburger: On the Simes test under dependence

Arnold Janssen and Julia Benditkis: Martingale methods for the FDR control of multiple tests

SESSION 6 IN HONOR OF WALTER LEHMACHER

Thursday 16:30- 18:00

Chair: Martin Hellmich

Martin Schumacher (Freiburg): Nonparametrics at the University of Dortmund - 40 years ago

Gerhard Hommel (Mainz): Multiple testing - a bridge between Mainz and Neuherberg

Meinhard Kieser (Heidelberg): Multiple endpoints, adaptive designs, and Rhenish mentality

Gernot Wassmer (Köln): 26 years of adaptive designs

Guido Grass (Köln): Statisticians should always be present in ethics committees

SESSION 7: SURROGATE VARIABLES AND RECURRENT EVENT DATA

Friday 9:30- 11:00

Chair: Martin Posch

Silke Jörgens: Adaptive interim analyses with (partial) count data

Matthias Brückner and Werner Brannath: Improving interim decisions in adaptive enrichment designs using short-term or surrogate endpoint data

Antje Jahn, Katharina Ingel, Stella Preussler, Harald Binder: Multiplicity issues in multistate models for recurrent event data subject to a competing terminal event with an application to cardiovascular disease

Katharina Ingel, Antje Jahn-Eimermacher, Harald Binder: On the risk of extrapolating interim results when adjusting the sample size of trials with recurrent event data

SESSION 8: ADAPTIVE DESIGNS II

Friday 11:30- 13:20

Chair: Andreas Faldum

Thomas Jaki, Lisa V Hampson, Roland Fisch, Linh Van: Group sequential designs for verifying whether effective drug concentrations are similar in adults and children

Susanne Urach and Martin Posch: Group sequential designs for clinical trials with multiple treatment arms

Sergey Krasnozhon, D. Schindler, R. Hilgers, W. Rosenberger, F. König: Response-adaptive randomization and adaptive combination test for clinical trials with limited number of patients: practical guide

Georg Gutjahr: Adaptive Dunnett tests based on conditional-rejection-probabilities

Florian Klinglmüller: Visualizing multiple objectives in flexible and group sequential designs

ABSTRACTS

BLINDED SAMPLE SIZE REESTIMATION FOR TIME DEPENDENT NEGATIVE BINOMIAL OBSERVATIONS

THOMAS ASENDORF

UNIVERSITÄT GÖTTINGEN

Sample size estimation procedures are a crucial element in planning clinical trials. However, sample size estimates strongly depend on nuisance parameters, which have to be guesstimated from previous trials, as illustrated in the context of trials in relapsing multiple sclerosis (RMS) in (1). Blinded sample size reestimation procedures allow for an adjustment of the calculated sample size within a running trial, by using gathered data to estimate relevant nuisance parameters without unblinding the trial (2). We consider a model for statistical inference of time dependent count data and provide sample size estimation and reestimation techniques within this model. The model presented allows for time dependent discrete observations with marginal longitudinally collected MRI lesion counts in RMS trials. Procedures will be presented for sample size estimation and blinded sample size reestimation in clinical trials with such data. A simulation study is conducted to assess the properties of the proposed procedures.

References

1. Nicholas et al., 2011. “Trends in annualized relapse rates in relapsing-remitting multiple sclerosis and consequences for clinical trial design”. Multiple Sclerosis Journal, Vol. 17, pp. 1211-1217. SAGE.

2. Friede Tim and Schmidli Heinz, 2010. „Blinded sample size reestimation with count data: Methods and applications in multiple sclerosis“. Statistics in Medicine, Vol. 29, pp. 1145-1156. John Wiley and Sons.

3. McKenzie Ed, 1986. “Autoregressive Moving-Average Processes with Negative Binomial and Geometric Marginal Distributions”. Advances in Applied Probability, Vol. 18, No. 3, pp. 679-705. Applied Probability Trust.

IMPROVING INTERIM DECISIONS IN ADAPTIVE ENRICHMENT DESIGNS USING SHORT-TERM OR SURROGATE ENDPOINT DATA

MATTHIAS BRÜCKNER, WERNER BRANNATH

UNIVERSITÄT BREMEN

In two-stage adaptive enrichment designs with subgroup selection the aim of the interim analysis at the end of the first stage is to identify promising subgroups which have a sufficiently large treatment effect. The decision whether or not to continue with recruitment in a given subgroup is based on the estimated treatment effect.

At the time of the interim analysis long-term endpoints such as overall survival are usually not yet observed for many of the recruited patients resulting in a large percentage of censored observations. Often short-term or surrogate endpoints, such as progression-free survival or tumor response, are available instead. These surrogate endpoint are themselves not observed immediately after randomisation and may be missing for some patients.

We will discuss how estimators of the treatment effect can be constructed using surrogate and primary endpoint information for time-to-event as well as metric endpoints. We investigate how much can be gained from using these estimators in the interim decision. The sensitivity and specificity of the corresponding decision rules is assessed in simulations.

UNCERTAINTY QUANTIFICATION FOR THE FAMILY-WISE ERROR RATE IN MULTIVARIATE COPULA MODELS

THORSTEN DICKHAUS (JOINT WORK WITH JENS STANGE AND TARAS BODNAR)

UNIVERSITY OF BREMEN, INSTITUTE FOR STATISTICS

We construct confidence regions for the realized family-wise error rate (FWER) of certain multiple tests which are empirically calibrated at a given (global) level of significance. To this end, we regard the FWER as a derived parameter of a multivariate parametric copula model. It turns out that the resulting confidence regions are typically very much concentrated around the target FWER level, while generic multiple tests with fixed thresholds are in general not FWER-exhausting. Since FWER level exhaustion and optimization of power are equivalent for the considered classes of multiple test problems, the aforementioned findings militate strongly in favor of estimating the dependency structure (i. e., copula) and incorporating it in a multivariate multiple test procedure. We illustrate our theoretical results with two particular classes of multiple test problems of practical relevance, namely multiple tests for components of a mean vector and multiple support tests. The presentation is based on [1].

Reference

[1] Jens Stange, Taras Bodnar, Thorsten Dickhaus (2014).Uncertainty quantification for the family-wise error rate in multivariate copula models.AStA Advances in Statistical Analysis, online first, http://dx.doi.org/10.1007/s10182-014-0241-5.

ON THE SIMES TEST UNDER DEPENDENCE

HELMUT FINNER, KLAUS STRASSBURGER

DEUTSCHES DIABETES-ZENTRUM AN DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF, LEIBNIZ-ZENTRUM FÜR DIABETES-FORSCHUNG, DEUTSCHLAND

In 1986, R. J. Simes proposed a modified Bonferroni test procedure for testing an overall null hypothesis in multiple testing problems, nowadays referred to as the Simes test. The paper of Simes may be considered as a basic step in the development of many new test procedures and new error rate criteria as for example control of the false discovery rate. A key issue is the validity of the Simes test and the underlying Simes inequality under dependence. Although it has been proved that the Simes inequality is valid under suitable assumptions on dependence structures, important cases are not covered yet. In this talk we investigate p-values based on exchangeable test statistics in order to explore reasons for the validity or failure of the Simes inequality. We provide sufficient conditions for the asymptotic validity of the Simes inequality and its possible strictness. We also show by means of an easy-to-compute counterexample that exchangeability by itself is not sufficient for the validity of the Simes inequality.

References

1. Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289-300.

2. Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. 29, 1165-1188.

3. Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: Asymptotics. Ann. Statist. 35, 1432-1455.

4. Finner, H. and Gontscharuk, V. (2013). Asymptotic FDR control under weak dependence: A counterexample. Statist. Probab. Lett. 83(8), 1888–1893.

5. Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika 73,751-754.

OPTIMIZING GRAPHICAL MULTIPLE TESTING PROCEDURES

DENNIS GÖRLICH, ROBERT KWIECIEN, ANDREAS FALDUM

INSTITUT FÜR BIOMETRIE UND KLINISCHE FORSCHUNG, WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

For clinical trials with multiple endpoints a large variety of multiple testing procedures are available. In certain situations the sequential rejective testing procedures (SRTP) proposed by Bretz et al [1,2] are an appealing tool to plan clinical trials controlling the family-wise-error-rate. In general, the structure of the testing procedure will be developed by the study statistican using information about the importance of hypotheses and other constraints. Nevertheless, often it is not completely clear whether there might exist a better design with the same statistical properties answering the trial hypotheses. We here propose an evolutionary algorithm to optimize graphical statistical trial designs with respect to the procedures power. In particular, we optimize the node and edge weights of the underlying graph. Different measures of power can be used for optimization, i.e. at-least-one rejection, selected hypotheses, and all hypotheses rejected. The evolutionary algorithm is a simple evolutionary (3,1)-strategy without recombination. In each generation the best (= highest power) individual (=SRTP) replaces the worst (=lowest power) and is mutated randomly. Fitness is evaluated by simulations of 10000 clinical trials per generation. To evaluate a SRTP the `gMCP` package [3] in R is used. We can show that, given a planning alternative, SRTPs can be optimized by our algorithm.

References

[1] Bretz F, Maurer W, Brannath W, Posch M (2009), A graphical approach to sequentially rejective multiple test procedures. Stat. Med. 28(4):586-604.
[2] Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K (2011), Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. Biometrical Journal 53 (6):894-913.
[3] Rohmeyer K, Klinglmueller F (2014). gMCP: Graph Based Multiple Test Procedures. R package version 0.8-6. http://CRAN.R-project.org/package=gMCP

IMPROVING PROBABILITIES OF CORRECT INTERIM DECISION IN POPULATION ENRICHMENT DESIGNS

HEIKO GÖTTE[A], MARGARITA DONICA[B]*,GIACOMO MORDENTI[C] *

[A] MERCK KGAA, DARMSTADT, GERMANY
[B] F. HOFFMANN – LA ROCHE LTD (GLOBAL MEDICAL AFFAIRS BIOMETRICS), BASEL, SWITZERLAND
[C] GRÜNENTHAL GMBH, GCD-BM-ST, AACHEN, GERMANY

(*)WAS UNDER EMPLOYMENT OF MERCK SERONO S.A. – GENEVA, SWITZERLAND, WHEN CONTRIBUTED TO THE PUBLICATION WORK.

Choice of target population is essential part at the design stage of clinical trials. Data from earlier clinical development might suggest that the treatment is more effective in a subpopulation but there might not be enough evidence to restrict the target population upfront. Adaptive designs allow modification of the target population based on interim data. Decision for modification should be based on objective decision rules. In a time to event setting, we present decision rules that are based on observed hazard ratios (HR) or conditional power. Potential interim decisions are “go with full population”, “go with subpopulation” or “stop for futility”. In case of small or moderate true effects it might be difficult to make a correct decision with a high probability. Therefore, we focus on true effects with unambiguous interpretation. We show how decision boundaries can be determined that maximize weighted probabilites of correct interim decisions [1]. The weights are determined based on prior assumptions. We compare probabilities of correct interim decision and power from our approach with a decision rule based on the algebraic sign of the observed effect.

Reference

[1] Götte H, Donica M, Mordenti G. Improving probabilities of correct interim decision in population enrichment designs. J Biopharm Stat. 2014 Jun 10. [Epub ahead of print]

UNION-INTERSECTION BASED GOODNESS-OF-fiT TESTS IN TERMS OF LOCAL LEVELS

VERONIKA GONTSCHARUK, HELMUT FINNER

INSTITUTE FOR BIOMETRICS AND EPIDEMIOLOGY, GERMAN DIABETES CENTER, LEIBNIZ

INSTITUTE FOR DIABETES RESEARCH AT HEINRICH-HEINE-UNIVERSITY DÜSSELDORF

Union-intersection based goodness-of-fit (GOF) tests are related to multiple test procedures in the sense the underlying GOF null hypothesis can often be represented as an intersection of suitable elementary hypotheses. An interesting characteristic of such GOF tests are so-called local levels. A local level is defined as the probability to reject a true elementary hypothesis, cf. [1] and [2]. Many GOF tests, e.g., the Kolmogorov-Smirnov test and the supremum version of the Anderson-Darling test as well as GOF tests based on phi-divergences introduced in [3], can be represented in terms of local levels. This allows a simple comparison of tests with respect to regions of sensitivity. Moreover, prior information about possible alternative distributions may be incorporated into the choice of local levels in order to tailor new GOF tests.

In this talk we mainly focus on GOF tests with equal local levels. We provide various representations of these tests and discuss their relationship to several well-known GOF procedures. Thereby, the asymptotics of local levels is of special interest.

References

[1] Gontscharuk, V., Landwehr, S. and Finner, H. [2015] The intermediates take it all: asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. Biom. J., 57, 159-180.

[2] Gontscharuk, V., Landwehr, S. and Finner, H. [2015] Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests. Bernoulli, accepted for publication.

[3] Jager, L. and Wellner, J. A. [2009] Goodness-of-fit tests via phi-divergences. Ann. Stat., 35, 2018–2053.

ADAPTIVE DUNNETT TESTS BASED ON CONDITIONAL-REJECTION-PROBABILITIES

GEORG GUTJAHR

UNIVERSITÄT BREMEN

Timmesfeld discovered an elegant approach to deal with nuisance parameters when applying the Conditional-Rejection-Probability (CRP) principle. In this talk, we use her approach to control the familywise-error-rate after data-dependent treatment-selections or sample-size modifications in Dunnett tests.

ADJUSTING MULTIPLICITY USING SAFETY DATA IN MANY-ONE COMPARISONS

GERALD HLAVIN, PETER BAUER, FRANZ KÖNIG

MEDICAL UNIVERSITY OF VIENNA, WIEN, AUSTRIA

In phase II studies many-one comparisons are performed to determine the most promising treatment(s) to be selected for confirmatory phases. However, when the sample sizes are rather limited such as in rare diseases, only a single trial addressing both treatment selection and confirmatory efficacy testing may be feasible. On the one hand if confirmatory conclusions shall be made, a strict control of the type I error is paramount. On the other hand no alpha level should be wasted for treatment groups no longer of interest, e.g., due to safety issues (Koenig, Bauer, & Brannath, 2006)

We suggest a two-step procedure. First, all treatment groups are screened whether they are safe or not. Only treatments considered to be sufficiently safe are selected for efficacy testing against a common control. We investigate whether it is possible to use multiplicity adjustments adjusting for the number of selected safe treatments only instead of adjusting a-prioiri for all treatments. We show under which association between safety and efficacy strict control of the FWER can be achieved. Furthermore, procedures using the estimated association between safety and efficacy will be developed (Berger & Boos, 1994).

Operating characteristics of the new multiple testing procedures are investigated by simulation studies, e.g., comparing the modified Dunnett and Bonferroni tests to their traditional counterparts.

References

Berger, R. L., & Boos, D. D. (1994). P-values maximized over a confidence set for the nuisance parameter. Journal of the American Statistical Association (89), pp. 1012-1016.

Koenig, F., Bauer, P., & Brannath, W. (2006). An adaptive hierarchical test procedure for selecting safe and efficient treatments. Biometrical Journal(48.(4)), pp. 663-678.

ADAPTIVE SEAMLESS PHASE II/III STUDY IN GASTRIC CANCER (ORPHAN CONDITION)

TINA VAN DER HORST (SUBMITTING AUTHOR), MICHAEL BUDDE, ALEXANDER STRASAK, MÁXIMO CARRERAS

F. HOFFMANN-LA ROCHE LTD. / PHARMA DEVELOPMENT / BIOSTATISTICS (PDBS)

BASEL (SWITZERLAND)

A seamless phase II/III study has been set-up by the company F. Hoffmann-La Roche Ltd. to evaluate the efficacy and safety of trastuzumab emtansine (T-DM1) compared to control in patients with HER2-positive advanced gastric cancer. At the start of the trial, patients were randomized to one of three treatment arms: Dose regimen 1; Dose regimen 2; Control. At the end of the first stage of the study in October 2013, the dose and schedule to be used in the second stage of the study were selected by an independent Data Monitoring Committee (iDMC). Since the study's primary endpoint, overall survival (OS) was expected to be immature at time of regimen selection, surrogate information was provided to the iDMC to be used to improve the regimen-selection process. The statistical inference for the selected group at end of the study will use data from both stages, which is possible due to an innovative design concept and statistical testing procedure as discussed in Jenkins et al (2011). The presentation provides an introduction to the study design and methods used. Pros and cons compared to separate phase II and III studies are discussed.

Reference

Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharm Stat 2011; 10:347−56

SIMULTANEOUS CONFIDENCE INTERVALS BY MULTIPLE MARGINAL MODELS USING R

LUDWIG A HOTHORN (LEIBNIZ UNIVERSITY), C. RITZ, C. (UNIVERSITY OF COPENHAGEN)

Today multiplicity adjustment is used in many low- and high-dimensional contexts.We focus on low-dimensions, such as some dose groups and time points. Very different methods have been proposed (eg. Bonferroni, Holm, claimwise error rate control by user-defined multiple contrasts). We are focusing on the MaxT (min p) test- but not on the common resampling approach acc. to Westfall & Young, 1989. More in detail the special form of an UIT with test statistics following a k-variate t-distribution with a certain correlation matrix, such as for ratio-to-control comparisons (Dilba, 2004) is discussed.Is the explicit formation of the correlation matrix impossible or painful, a nice alternative using multiple marginal models (Pipper et al., 2012) is available- relying on asymptotic results.The great advantage is the availability in the R package (multcomp) (function mmm) and the simplicity in GLMM's.

We demonstrate the usefulness by means of several examples: i)mode-of-inheritance specific genetic association tests, ii) clinical subgroup analysis with demonstration the efficacy of total population and/or targeted and/or complementary subgroups, iii) composite endpoints, iv) multiple endpoints in multi-armed RCT's and v) simultaneous inference on both treatment and time comparisons.

ON THE RISK OF EXTRAPOLATING INTERIM RESULTS WHEN ADJUSTING THE SAMPLE SIZE OF TRIALS WITH RECURRENT EVENT DATA

KATHARINA INGEL, ANTJE JAHN-EIMERMACHER, HARALD BINDER

IMBEI MAINZ

In clinical trials with recurrent event data, as for example relapses in multiple sclerosis or acute otitis media, the required sample size depends on the cumulative hazard at time T, with [0, T] being the patients' follow-up period. As in the planning phase of a trial there might be uncertainty about the recurrent event process over [0, T], internal pilot studies have been proposed to re-estimate the sample size accordingly (Schneider et. al., Statistics in Medicine 2013; Ingel and Jahn-Eimermacher, Biometrical Journal 2014). However, for trials with a long follow-up period relative to a short recruitment period, at the time of the interim analysis patient data might only be available over a shorter period [0, t].

For these situations, we will investigate an extrapolation of the interim estimate for the cumulative hazard to the full follow-up period. In a simulation study different parametric recurrent event models will be applied to blinded interim data and interim estimates will be extrapolated to the end of the planned follow-up T for the purpose of sample size re-estimation. The timing of the interim analysis will be evaluated as one determinant for the accuracy of that extrapolation. Results will be compared with respect to the re-estimated sample size and resulting power, and the validity of the re-estimation design will be investigated.

For simulation, recurrent event data under various hazard functions defined on a calendar time scale are required and we propose a general simulation algorithm for this purpose. Results are used to identify situations, where an extrapolation of interim results can be useful to improve the accuracy of sample size planning in contrast to situations with a potential risk for misleading conclusions.

MULTIPLICITY ISSUES IN MULTISTATE MODELS FOR RECURRENT EVENT DATA SUBJECT TO A COMPETING TERMINAL EVENT WITH AN APPLICATION TO CARDIOVASCULAR DISEASE

ANTJE JAHN, KATHARINA INGEL, STELLA PREUSSLER, HARALD BINDER

INSTITUTE OF MEDICAL BIOSTATISTICS, EPIDEMIOLOGY AND INFORMATICS (IMBEI), UNIVERSITY MEDICAL CENTER OF THE JOHANNES GUTENBERG-UNIVERSITY MAINZ

Systolic heart failure is a disease characterized by recurrent non-fatal events (hospital admissions) and delayed death. There is an ongoing debate to reflect this in the primary endpoint of clinical trials by moving from the commonly applied time-to-first-combined-endpoint model to a multistate recurrent event model [1]. Including recurrent events into the primary analysis allows for a potentially more adequate interpretation of effects and will in general increase the power of a trial, but raises multiplicity issues.

Accordingly, we will propose a multiple testing adjustment for multistate recurrent event models subject to a competing terminal event. Different multistate models are investigated including Markov multistate models and joint frailty models. We use simulations to investigate the effect of the multiplicity adjustment on the power of a trial and to compare this with combined-endpoint-models that do not require multiplicity adjustments. To allow for different simulation models, we propose a recursive algorithm for simulating multistate data following transition hazards, that are defined on the time since study start, e.g. reflecting a continually worsening heart performance [2].

Results confirm, that taking recurrent events into account in general increases the power of a trial. Combined-endpoint models that to not require multiplicity adjustments are shown to be superior with respect to power only if treatment similarly affects all endpoints. For more realistic situations as observed in large clinical trials on systolic heart failure a multistate approach with multiplicity correction increases power and allows a more thorough interpretation of results.

References

[1] Anker S, Murray J: Time to move on from ‘time-to-first’: should all events be included in the analysis of clinical trials? *European Heart Journal* 2012

[2] Jahn-Eimermacher A, Ingel K, Ozga A, Preussler S, Binder H: Simulating recurrent event data with hazard functions defined on a total time scale. *BMC Medical Research Methodology* 2015

MARTINGALE METHODS FOR THE FDR CONTROL OF MULTIPLE TESTS

ARNOLD JANSSEN, JULIA BENDITKIS

HEINRICH-HEINE UNIVERSITY DÜSSELDORF

Under martingale dependence the FDR of various multiple tests can exactly be calculated. The results are key tool in order to discuss finite sample FDR control of these tests. Some of these results are also new when the p-values are independent.

The second part of the talk discusses adaptive multiple tests with data dependent critical values.

GROUP SEQUENTIAL DESIGNS FOR VERIFYING WHETHER EFFECTIVE DRUG CONCENTRATIONS ARE SIMILAR IN ADULTS AND CHILDREN

THOMAS JAKI, LISA V HAMPSON, ROLAND FISCH, LINH VAN

DEPARTMENT OF MATHEMATICS AND STATISTICS, LANCASTER UNIVERSITY

New medicines for children should be subject to rigorous testing while avoiding unnecessary experimentation in children. In particular, paediatric dosing recommendations should be informed by existing relevant data. If the effective concentration of a drug can be assumed to be similar in adults and children, an appropriate paediatric dosing rule may be found by 'bridging', that is, conducting pharmacokinetic studies in children to find doses that produce concentrations therapeutic in adults. However, this strategy may result in children receiving an ineffective or hazardous dose if, in fact, effective concentrations differ between adults and children.

When there is uncertainty about the equality of effective concentrations, some pharmacokinetic-pharmacodynamic (PK-PD) data may be needed in children to verify whether differences between adults and children are small. In this presentation, we develop adaptive procedures that can be used to verify this assumption efficiently. Asymmetric inner wedge group sequential tests are constructed which permit early stopping to accept or reject an assumption of similar effective drug concentrations in adults and children. Asymmetry arises because the consequences of under- and over-dosing may differ. Using exact calculations we compare the efficiency of error spending inner-wedge tests with optimal designs which minimise the expected sample size needed to reach a conclusion. If there is time, we will show how stopping rules can be interpreted in terms of predictive tail area probabilities measuring the consistency of observed paediatric PK-PD data with an assumption of similar effective concentrations in adults and children.

ADAPTIVE INTERIM ANALYSES WITH (PARTIAL) COUNT DATA

SILKE JÖRGENS

ICON PLC, KÖLN

Late phase clinical trials with a limited number of interim analyses and limited projected adaptations have become quite common and have well-explored characteristics. However, in many situations interim analyses do not seem feasible due to a mismatch between individual study participation and projected enrolment: A quick enrolment can lead to (almost) complete randomization at suitable interim analysis time points, thus leaving little room for adaptations. In such cases, basing interim decisions on preliminary data can be an option. Such preliminary data can be, e.g., surrogate data or early observations of a long-term endpoint.

In the case of count data, obtaining preliminary data is especially straightforward: If the endpoint is the number of events of a certain type in a fixed period of time, then partial observations for a fraction of that period of time can be used to predict the total number of events for each subject, and adaptations for subsequent stages of the trial can be performed on this prediction. We present possible approaches for such designs, and give an overview of their limitations and benefits.

VISUALIZING MULTIPLE OBJECTIVES IN FLEXIBLE AND GROUP SEQUENTIAL DESIGNS

FLORIAN KLINGLMÜLLER

SECTION FOR MEDICAL STATISTICS, CENTER FOR MEDICAL STATISTICS, INFORMATICS, AND INTELLIGENT SYSTEMS, MEDICAL UNIVERSITY VIENNA, AUSTRIA

We present adaptive graph-based multiple testing procedures for clinical trials that permit early rejection of elementary hypotheses but also provide the flexibility to perform interim trial modifications based on the unblinded observations from the ongoing trial. The underlying multiple testing procedure can be tailored using an intuitive graphical approach. During the ongoing trial and based on unblinded observations one may reassess and modify the preplanned sample sizes and the multiple testing procedure; introduce additional interim analyses or change the timing of the preplanned ones; add or drop treatments and - or hypotheses. The procedure combines the closed testing principle with the partial conditional error rate approach and therefore controls the family wise error rate in the strong sense. The use of the partial conditional error rate approach does not require the full knowledge of the correlation structure of test statistics involved. However, if present, the procedure will exploit any positive correlation between test statistics.

RESPONSE-ADAPTIVE RANDOMIZATION AND ADAPTIVE COMBINATION TEST FOR CLINICAL TRIALS WITH LIMITED NUMBER OF PATIENTS: PRACTICAL GUIDE

SERGEY KRASNOZHON [1] D. SCHINDLER [2] R. HILGERS [2] W. ROSENBERGER [3], F. KÖNIG [1]

[1] MEDICAL UNIVERSITY OF VIENNA/CEMSIIS, WIEN, AUSTRIA,
[2] UNIVERSITÄTSKLINIKUM AACHEN, AACHEN, GERMANY,
[3] GEORGE MASON UNIVERSITY, FAIRFAX, VIRGINIA, USA

Response-adaptive randomization procedures (e.g., Rosenberger and Lachin [2002]) allow conducting a clinical trial in a more ethical way by updating the probability of assignment of a new patient to one of the treatment groups based on the earlier observed responses and previous allocations. A number of works have explored different randomization procedures (Hu and Rosenberger [2003], Hu and Zhang [2004], Zhang and Rosenberger [2006]) as to their performance, depending on the type of response observed. Among them are urn models, such as the randomized play-the-winner rule and Klein urn design, and sequential estimation designs, such as the efficient randomized-adaptive design and the doubly adaptive biased coin design. But most of the results obtained and conclusions made are based on large-sample size approximations, or based on data obtained from real clinical trials, where the number of patients was also not an issue.

We investigate the four methods above in cases when the number of patients to be recruited is limited, leading to small sample sizes. Continuous and binary responses are considered. The Pearson chi-squared test and the Fisher exact test were used for binary responses and the t-test and the randomization test were used for continuous responses for evaluating performance of the methods as to type I error control as well as for estimating statistical power of procedures. We start with two-arm clinical trials and discuss extensions to multi-arm clinical trials such as the three-arm gold-standard design including an experimental, an active control and placebo.

Instead of performing adaptations after each single observation, we scrutinise adaptive designs using adaptive combination tests (Bauer [1989], Bauer and Köhne [1994]), where design modifications are performed at interim analyses. E.g., as stage-wise p-values have to be combined, there might be a loss of power due to the loss of degrees of freedom when using stage-wise t-tests. We investigate the influence of the number and timing of interim analyses, adaptation of allocation ratios and sample size reassessment on the operating characteristics. As a result, we provide a practical guide for when they are relevant, and if so, how to use such adaptive designs in cases where only a limited number of participants is available.

References

P. Bauer. Multistage testing with adaptive designs. Biometrie und Informatik in Medizin und Biologie, 20(4):130–148, 1989.

P. Bauer and K. Köhne. Evaluation of experiments with adaptive interim analyses. Biometrics, pages 1029–1041, 1994.

F. Hu and W. F. Rosenberger. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. Journal of the American Statistical Association, 98(463):671–678, 2003.

F. Hu and L.-X. Zhang. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. Annals of Statistics, pages 268–301, 2004.

W. F. Rosenberger and J. M. Lachin. Randomization in clinical trials: theory and practice. John Wiley & Sons, 2004.

L. Zhang and W. F. Rosenberger. Response-adaptive randomization for clinical trials with continuous outcomes. Biometrics, 62(2):562–569, 2006.

PERFORMANCE OF SUBGROUP SELECTION RULES FOR A TARGETED THERAPY IN ONCOLOGY

JOHANNES KRISAM, MEINHARD KIESER

INSTITUTE OF MEDICAL BIOMETRY AND INFORMATICS, UNIVERSITY OF HEIDELBERG, GERMANY

The planning stage of a clinical trial investigating a potentially targeted therapy is usually characterized by a high degree of uncertainty whether the investigated treatment is more efficient (or efficient only) in a subgroup as compared to the total population population. Commonly, predictive biomarkers are employed in these trials in order to identify a subset of patients more likely to benefit from a drug. Adaptive designs incorporate a mid-course efficacy assessment of the total population and the subgroup of biomarker-positive patients, and thus allow to select of the most promising target population based on the results of the interim analysis (see, e.g., [1-2]). As has recently been shown, the applied interim decision rule in such a design plays a crucial role in ensuring a successful trial in terms of probability for a correct interim decision [3] and statistical power [4].

We investigate the situation that the primary outcome of the trial is binary or a time-to-event variable. Statistical methods are developed that allow an evaluation of the performance of decision rules in terms of correct selection probability at interim. Additionally, optimal decision rules are derived which incorporate the uncertainty about several design parameters, such as the treatment effects and the sensitivity and specificity of the employed bioassay. These optimal decision rules are evaluated regarding their performance in an adaptive enrichment design in terms of correct selection probability, type I error rate, and power and are compared to ad-hoc decision rules proposed in the literature. Our methods are illustrated by means of a clinical trial example.

References

[1] Wang S J., O'Neill R T, Hung H.M J. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics 2007; 6:227-244.

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

[3] Krisam J, Kieser M. Decision rules for subgroup selection based on a predictive biomarker. Journal of Biopharmaceutical Statistics 2014; 24:188-202.

[4] Krisam J, Kieser M. Performance of biomarker-based subgroup selection rules in adaptive enrichment designs. Statistics in Biosciences 2015, in press, doi: 10.1007/s12561-015-9129-5.

BLINDED VERSUS UNBLINDED ESTIMATION OF THE CORRELATION COEFFICIENT IN ADAPTIVE DESIGNS WITH LONG-TERM AND SHORT-TERM OUTCOMES

CORNELIA URSULA KUNZ[1], TIM FRIEDE[2], NICHOLAS PARSONS[1], SUSAN TODD[3], NIGEL STALLARD[1]

[1] WARWICK MEDICAL SCHOOL, UNIVERSITY OF WARWICK, U.K.

[2] DEPARTMENT OF MEDICAL STATISTICS, UNIVERSITY MEDICAL CENTER GÖTTINGEN, GERMANY

[3] DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF READING, U.K.

Adaptive seamless phase II/III designs with treatment selection at an interim analysis have become increasingly more attractive due to their potential to save development costs and to shorten time-to-market of a new treatment. If the primary endpoint is observed only after long-term follow-up it may be desirable to use short-term endpoint data at the interim analysis to select a treatment. In other cases, some long-term endpoint data might be available at interim analysis which, together with the short-term endpoint data, might be used to estimate the treatment effect and therefore treatment selection decisions. However, the performance of these methods depends in part on the correlation between the short- and long-term endpoint which is usually unknown at the start of the trial. Hence, it might be necessary to estimate the correlation at the time of the interim analysis or at an earlier blinded review.

We compare the properties of several estimators for the covariance and the correlation coefficient based on blinded and unblinded data under various scenarios. The scenarios include simple and block randomization, different block lengths, different sample sizes, and different numbers of groups. While it is possible to derive an unbiased estimator for the covariance based on blinded data, the mean squared error can be substantial depending on the scenario. For the correlation coefficient, estimators are often biased unless specific assumptions regarding the treatment effects hold true.

DECISION ANALYTIC APPROACH TO SUBGROUP ANALYSIS

WILLI MAURER, NOVARTIS PHARMA AG BASEL

FRANK BRETZ, NOVARTIS PHARMA AG, BASEL AND INSTITUTE FOR BIOMETRY, HANNOVER MEDICAL SCHOOL, 30623 HANNOVER

XIAOLEI XUN, NOVARTIS PHARMA, SHANGHAI

We consider the problem of comparing the treatment effect of a new drug against a comparator for two non-overlapping subgroups of patients defined by predictive biomarkers, demographic factors or any other classifier. A decision is to be made if and for which of the two subgroups the respective null hypotheses can be rejected and an advantage of the new drug over the comparator can be claimed. We argue that in this situation traditional methods to control the Type I error rate are too restrictive and that the standard familywise error rate (FWER) does not appropriately reflect losses and gains attributed to possible decisions. Instead, we propose decision procedures that allow us to control the FWER, but for which also upper bounds of expected values for more general loss functions can be derived.

We generalize multiple test procedures that protect the FWER to decision rules for which the expected loss does not exceed an upper bound. For the special case of two non-overlapping subgroups and for selected classes of rules we also determine rules that optimize the expected gain under Bayes or minimax regret criteria while controlling the expected loss.

ADAPTIVE TREATMENT ARM SELECTION IN MULTIVARIATE BIOEQUIVALENCE TRIALS

TOBIAS MIELKE

ICON PLC, KÖLN

Equivalence studies are used in drug development for the development of new formulations for an existing drug and for bringing generic products to the market. Equivalence studies may consist of different test arms, of which at least one should provide equivalence on several endpoints for the success of the trial. Although simultaneous equivalence on multiple endpoints requires no multiplicity correction, the existence of several test arms introduces multiplicity, which needs to be taken into account.

Adaptive treatment selection designs allow to drop treatment arms in an interim analysis. The remainder of the trial will then need only a reduced adjustment for multiplicity, if any. The treatment selection results in an increased power, while the sample size is reduced as compared to designs without treatment selection.

A short introduction on the methodology for adaptive treatment selection in multivariate equivalence trials will be followed by a case study on a cross-over design, which will provide insight into the benefits and limitations of the proposed approach.

A DECISION THEORETIC APPROACH TO SUBGROUP SELECTION

THOMAS ONDRA, SEBASTIAN JOBJÖRNSSON, ROBERT BECKMAN, CARL-FREDRIK BURMAN, FRANZ KÖNIG, NIGEL STALLARD, AND MARTIN POSCH

MEDICAL UNIVERSITY OF VIENNA

An important objective in the development of targeted therapies is to identify the populations where the treatment under consideration has positive benet risk balance. We consider the setting, where the efficacy of a treatment is tested in an overall population as well as a pre-specified subpopulation. Based on a decision theoretic framework we derive optimized trial designs by maximizing utility functions. Features to be optimized include the sample size and the population the trial is performed in (the full population or the targeted subgroup only). The approach accounts for prior knowledge on the efficacy of the drug in the considered populations using a two dimensional prior distribution. The considered utility functions account for the costs of the clinical trial as well as the expected benefit when demonstrating efficacy in the different subpopulations. Examples of optimized trial designs are determined by numerical optimization for several scenarios.

NUISANCE-PARAMETER BASED SAMPLE SIZE REESTIMATION IN ADAPTIVE ENRICHMENT DESIGNS WITH AN APPLICATION IN MAJOR DEPRESSION

MARIUS PLACZEK, TIM FRIEDE

DEPARTMENT OF MEDICAL STATISTICS

UNIVERSITY MEDICAL CENTER GÖTTINGEN

Adaptive enrichment designs offer an efficient way to perform a subgroup analysis while controlling the type-I-error rate. Frequently used testing strategies include, e.g. the combination test approach (Brannath et al, 2009; Jenkins et al, 2011) or the conditional error function approach (Friede et al, 2012). Due to uncertainties about the nuisance parameters which are needed for sample size calculations, a sample size review, which can be carried out with or without breaking the blinding, can be performed in order to make the study more robust against misspecifications of the nuisance parameters in the planning phase (Wittes & Brittain, 1990). Considering normally distributed endpoints we will present methods for blinded and unblinded sample size reestimation in adaptive enrichment designs and will compare them with regard to operation characteristics including type I error rate, power and sample size distribution in simulation studies. The methods and simulations are motivated by a trial in major depression (Frank et. al, 2011) which was recently re-analyzed with the aim of indentifying patient subgroups (Kraemer, 2013).

References

Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. Statistics in Medicine 28:1445–1463.

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.

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.

Wittes J, Brittain E (1990). The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine 9: 65–72.

Frank E, Cassano GB, Rucci P, Thompson WK, Kraemer HC, Fagiolini A, Maggi L, Kupfer DJ, Shear MK, Houk PR, Calugi S, Grochinshi VJ, Scocco P, Buttenfield J, Forgione RN (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. Psychological Medicine 41(1): 151–162.

Kraemer HC (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. Statistics in Medicine 32: 1964-1973

OPTIMAL REJECTION REGIONS FOR TESTING MULTIPLE BINARY ENDPOINTS IN SMALL SAMPLES

ROBIN RISTL, MARTIN POSCH

SECTION FOR MEDICAL STATISTICS, CENTER FOR MEDICAL STATISTICS, INFORMATICS, AND INTELLIGENT SYSTEMS, MEDICAL UNIVERSITY VIENNA, AUSTRIA

The presented work is motivated from the need to make optimal use of data from small clinical trials. In small samples, statistical inference using non-parametric exact tests is often preferred over asymptotic approximations to guarantee strict control of the type I error rate. As a draw-back, exact tests in general do not exploit the nominal significance level due to discreteness. When performing multiple testing adjustments on exact tests, discreteness may be taken into account to avoid overly conservative procedures. A well-known example is Tarone's test, which utilizes the marginal null-distributions from discrete tests to find an adjusted significance level.

However, when testing multiple and possibly correlated endpoints, knowledge on the joint distribution of discrete test statistics can lead to further improvement of rejection regions. For the case of comparing two treatment groups with respect to two binary endpoints, we study the exact joint distribution of the test statistics conditional on the observed data. The distribution is calculated, both, under the global null hypothesis and under parametric alternatives. Optimal rejection regions for one-sided alternatives are found by selecting coherent regions that have maximal probability under the alternative subject to not exceeding the nominal significance level under the null-hypothesis.

Extensions of the proposed approach to include tests for non-inferiority and generalizations to more endpoints are discussed. The unconditional power of the procedure is studied by simulation.

This work has been funded by the FP7-HEALTH-2013-INNOVATION-1 project Advances in Small Trials Design for Regulatory Innovation and Excellence (ASTERIX) Grant Agreement No. 603160.

BLINDED SAMPLE SIZE RE-CALCULATION IN CLINICAL TRIALS WITH BINARY COMPOSITE ENDPOINTS BASED ON CORRELATION-ADJUSTED LOCAL SIGNIFICANCE LEVELS

ANJA SANDER, GERALDINE RAUCH, MEINHARD KIESER

INSTITUTE OF MEDICAL BIOMETRY AND INFORMATICS,

UNIVERSITY OF HEIDELBERG, GERMANY

In clinical trials, composite endpoints are increasingly used as primary outcome. Thereby, the composite endpoint comprises different single outcomes, mostly defined as binary or time-to-event variables. Through combination of different events, the number of expected events is increased with the aim to increase the power of the trial or to reduce the necessary sample size. However, interpretation of the observed effect becomes difficult if the effects in the single components vary in size or even point in opposite directions. Therefore, it is sensible to additionally test at least the (most) important components confirmatorily. This constitutes a multiple testing problem.

Usually, the test statistics of the composite endpoint and respective components are highly correlated. As shown by Rauch and Kieser [1], these correlations (under H0) can be used to define larger local significance levels as compared to a classical Bonferroni correction. To derive the correlations, knowledge about the event rates under H0 is required. In case of uncertainty regarding the event rates, a design with internal pilot study provides the option to estimate overall event rates and correlations mid-course during the trial. These estimates can then be used to specify the multiple testing procedure and to adjust the sample size if necessary [2].

Investigations regarding different designs and analysis strategies have shown that application of an internal pilot study in superiority trials is usually associated only with a slight change in type I error rate, if any. Moreover, it has been shown that the intended power is also achieved in case of misspecifications at the planning stage [3].

We investigate whether this property holds true for the described situation of composite endpoints and whether the estimated correlation together with an adjustment of local significance levels results in a smaller sample size. Type I and II error rates are calculated analytically if computationally feasible. For a range of different scenarios, overall event rates and correlations are estimated within simulation studies, sample size recalculations are performed, and the type I and II error rates are determined. Detailed results are presented and illustrated by clinical trial examples.

References

[1] Rauch G, Kieser M. Multiplicity adjustment for composite binary endpoints. Methods of Information in Medicine 2012; 51:309-317.

[2] Friede T, Kieser M. Sample size recalculation for binary data in internal pilot study designs. Pharmaceutical Statistics 2004; 3:269-79.

[3] Friede T, Kieser M. Sample size recalculation in internal pilot study designs: A review. Biometrical Journal 2006; 48:537-555.

SIMULTANEOUS STATISTICAL INFERENCE FOR EPIGENETIC METHYLATION DATA

KONSTANTIN SCHILDKNECHT

WEIERSTRAß-INSTITUT BERLIN

Epigenetic research leads to complex data structures. Since parametric model assumptions for the distribution of epigenetic data are hard to verify, we introduce in the present work a nonparametric statistical framework for two-group comparisons. Furthermore, epigenetic analyses are often performed at various genetic loci simultaneously. Hence, in order to be able to draw valid conclusions for specific loci, an appropriate multiple testing correction is necessary. Finally, with technologies available for the simultaneous assessment of many interrelated biological parameters (such as gene arrays), statistical approaches also need to deal with a possibly unknown dependency structure in the data. Our statistical approach to the nonparametric comparison of two samples with independent multivariate observables is based on recently developed multivariate multiple permutation tests, see [1]. We adapt their theory in order to cope with families of hypotheses regarding relative effects, in the sense of [2].

Our results indicate that the multivariate multiple permutation test keeps the pre-assigned type I error level for the global null hypothesis. In combination with the closure principle, the family-wise error rate for the simultaneous test of the corresponding locus/parameter-specific null hypotheses can be controlled. In applications we demonstrate that group differences in epigenetic data can be detected reliably with our methodology.

References

[1] Chung EY, Romano JP (2013) Multivariate and multiple permutation tests. Technical report No. 2013-05, Stanford University.

[2] Brunner E, Munzel U (2013) Nichtparametrische Datenanalyse. Unverbundene Stichproben. Heidelberg: Springer Spektrum, 2nd updated and corrected edition.

ON ADAPTIVE DESIGNS WITH DEPENDENT P-VALUES BASED ON FISHER'S COMBINATION TEST

RENE SCHMIDT, ANDREAS FALDUM

INSTITUTE OF BIOSTATISTICS AND CLINICAL RESEARCH,

WWU MUENSTER, SCHMEDDINGSTR. 56, 48149 MUENSTER, GERMANY

Adaptive designs were originally developed for independent and uniformly distributed p-values. However, there are settings where the stage-wise p-values may be dependent with even unknown dependence structure, e.g. in the context of trials with multiple time-to-event endpoints. If the dependence structure is not taken into account adequately, the significance level might in general no longer be preserved. It is thus of interest to consider the most adverse dependence structure maximizing the type I error rate of a given adaptive design (worst case). Using copula techniques, we study the type I error rate in the worst case for adaptive designs without futility stop based on Fisher's combination test [1]. An explicit analytic formula for the type I error rate in the worst case is obtained. We derive a smallest upper bound for the true type I error rate if the design is planned under the assumption of independent and uniformly distributed p-values, but if independence is in truth not fulfilled.

Considerable inflation of the type I error rate can occur, when the dependence structure is not taken into account adequately. This emphasizes that examination of the true dependence structure between the stage-wise p-values and an adequate choice of the conditional error function is crucial when adaptive designs are used.

Reference

[1] Bauer P, Köhne K (1994) Evaluation of experiments with adaptive interim analyses. Biometrics 50: 1029-1041

A GROUP-SEQUENTIAL DESIGN TO TEST EFFICACY AND INEFFICENCY IN TWO SUBGROUPS

DARIA TUTSCHKOW [1], ASTRID DEMPFLE [2], NINA TIMMESFELD [1]

[1] INSTITUTE OF MEDICAL BIOMETRY AND EPIDEMIOLOGY, PHILIPPS-UNIVERSITY MARBURG

[2] INSTITUTE OF MEDICAL INFORMATICS AND STATISTICS, UK S-H, CAMPUS KIEL, CHRISTIAN ALBRECHTS-UNIVERSITY KIEL

Individualized treatments are tailored to subgroups of patients defined on the basis of molecular biomarkers which are predictive of therapeutic response. At present, most proposed designs consider testing efficacy in the biomarker positive subgroup or in the full set. After running such a design and efficacy is shown in the biomarker positive subgroup, it will unclear, if the new treatment is effective in the biomarker negative group. For ethical reason it may be useful to establish the inefficiency of the new treatment in the biomarker negative group.

In this talk, we will proposed a group sequential design, which allows to test hypotheses of efficacy and inefficiency in all subgroups. At each interim analysis efficacy is tested in the biomarker positive subgroup and inefficiency is tested in the biomarker negative. If for example efficacy in the biomarker positive has been proven, efficacy can be tested in the biomarker negative.

GROUP SEQUENTIAL DESIGNS FOR CLINICAL TRIALS WITH MULTIPLE TREATMENT ARMS

SUSANNE URACH, MARTIN POSCH

MEDICAL UNIVERSITY OF VIENNA (AUSTRIA)

For the comparison of several experimental treatments, doses or dose regimens to a control, recently several multi-armed group sequential clinical trial designs have been proposed. They allow one to stop randomization to individual arms early if the corresponding null hypothesis can be rejected in an interim analysis. To minimize the required number of patients, we consider a variant of such designs where the overall trial stops as soon as for any of the arms the null hypothesis of no treatment effect can be rejected. While standard multi-armed group sequential designs control the type I error rate if such a stopping rule is used, they are typically strictly conservative. This can be explained by the fact that treatment arms for which no rejection can be achieved at the interim analysis could be further tested in the final analysis if the trial was continued. For the comparison of two experimental treatments we derive improved stopping boundaries that exhaust the type I error rate for stopping rules where the trial is stopped as soon as a null hypothesis can be rejected. We search for optimized boundaries that minimize the expected sample size while maximizing the probability to show efficacy in both arms. We compare the operating characteristics of these optimized designs to standard multi-armed group sequential designs. Furthermore, several extensions as trials with futility boundaries and group sequential t-tests for small sample sizes are discussed.

ADAPTIVE DESIGNS AND THEIR IMPACT ON CLINICAL TRIALS' COST STRUCTURES – AN ECONOMIC PERSPECTIVE ON ADAPTIVE DESIGNS

MARC URBAN, ANDREAS FALDUM

ZENTRUM FÜR KLINISCHE STUDIEN MÜNSTER, UNIVERSITÄTSKLINIKUM MÜNSTER

INSTITUT FÜR BIOMETRIE UND KLINISCHE FORSCHUNG, WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

Adaptive sequential designs have become very popular in clinical trials over the past few years. Besides methodical challenges and ethical considerations, usually controllability and time saving are mentioned as the main impact and advantage of adaptive trial designs. These characteristics have been intensively investigated during the last decades.

In addition, an intense discussion about costs is taking place and escalating in the public health care sector as well as in the pharmaceutical industry in recent years. In this setting capital is limited and stakeholders are asking for new ideas and methods for the future.

Therefore, a set of three investigator initiated trials have been considered to analyze the impact of interim analyses and adaptive modifications on the total costs of the whole process and trial.

Starting with identifying trial specific work load and the consequences of adaptive modifications on each of them different concepts and examples will be discussed.