Adaptive **D**esigns

& **M**ultiple **T**esting **P**rocedures

Workshop June 5-6, 2014

Basel, Switzerland

Organizing Committee:

Ekkehard Glimm (Basel), Dominik Heinzmann (Basel), Eliane Imfeld (Basel), Meinhard Kieser (Heidelberg)

WORKSHOP VENUE

LOCATION:

The workshop "Adaptive Designs and Multiple Testing Procedures 2014" will take place at the Novartis Learning Center Horburg. It can be reached by Tram 8 from Basel SBB train station to the stop "Dreirosenbrücke", and then a short walk along the "Badenweilerstrasse" to the Learning center. Duration: 20 min.

ADDRESS:

Müllheimerstrasse 195
CH-4057 Basel

GET TOGETHER DINNER

The "Get Together Dinner" can be attended on Thursday June 5 at 19:30. Payment is responsibility of the participants. The dinner will take place in the traditional Basel Restaurant "zum Braunen Mutz".

ADDRESS:

Restaurant zum Braunen Mutz
Barfüsserplatz 10
CH-4051 Basel
www.braunermutz.ch

HOW TO GET THERE:

The location of the conference dinner is in Basel downtown.

SCIENTIFIC PROGRAM – OVERVIEW

THURSDAY, JUNE 5

08:30 – 09:00	Registration and Reception
09:00 – 09:10	Welcome Addresses: Ekkehard Glimm (Chair of the working group adaptive designs and multiple testing procedures, Novartis Basel) and Dominik Heinzmann (BBS Council Member, F Hoffmann-La Roche Basel)
09:10 – 11:00	Session 1: Opportunities and challenges for adaptive Designs
11:00 – 11:30	Coffee break
11:30 – 13:00	Session 2: Enrichment designs and subpopulation analysis
13:00 – 14:15	Lunch break
14:15 – 15:30	Session 3: Multiple testing
15:30 – 16:00	Coffee break
16:00 – 17:15	Session 4: Sample size re-estimation
19:30	Get Together Dinner

FRIDAY, JUNE 6

08:30 – 10:30	Session 5: Multiplicity and adaptations in clinical trials
10:30 – 11:00	Coffee break
11:00 – 12:15	Session 6: Multiplicity in regression models
12:15 – 13:30	Lunch break
13:30 – 14:45	Session 7: Group-sequential tests
14:45	End of the workshop

Scientific Program – Detailed Time Schedule

Thursday, June 5

08:30 – 09:00 Registration and Reception

09:00 – 09:10 Welcome Addresses:
Ekkehard Glimm, Chair of the working group adaptive designs and multiple testing procedures, Novartis Basel
Dominik Heinzmann, BBS Council Member, F Hoffmann-La Roche, Basel

09:10 – 11:00 Session 1: Opportunities and Challenges for Adaptive Designs
Chair: Ekkehard Glimm (Basel)

- Peter Bauer: Flexibility in confirmatory clinical trials, what is for free?
- Marc Vandemeulebroecke, Robert L Cuffe, David Lawrence, Andrew Stone: To seamless or not to seamless? Lessons learned from four case studies
- Thomas Jaki and Lisa Hampson: Incorporating feasibility assessment in the design of clinical studies
- Stefan Englert and Meinhard Kieser: Methods for Proper Handling of Over- and Underrunning in Phase II Designs for Oncology Trials

11:00 – 11:30 Coffee break

11:30 – 13:00 Session 2: Enrichment Designs and Subpopulation Analysis
Chair: Meinhard Kieser (Heidelberg)

- Franz König, Alexandra Graf, Martin Posch: Adaptive designs for with subgroup analysis optimizing utility functions
- Johannes Krisam, Meinhard Kieser: Performance characteristics of interim decision rules in adaptive enrichment designs
- Marius Placzek, Simon Schneider, Tim Friede: Comparison of different approaches to enrichment designs with multiple nested subgroups

13:00 – 14:15 Lunch break

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE

14:15 – 15:30 SESSION 3: MULTIPLE TESTING
CHAIR: DOMINIK HEINZMANN (BASEL)

- Robin Ristl, Florian Frommlet, Armin Koch, and Martin Posch: A fallback test for three co-primary endpoints
- Konstantin Schildknecht: A two-stage hierarchical multiple test procedure based on the asymptotically optimal rejection curve
- S. Kropf, S. Weston, D. Adolf: Multiple tests in first-level analyses in functional magnetic resonance imaging

15:30 – 16:00 COFFEE BREAK

16:00 – 17:15 SESSION 4: SAMPLE SIZE RE-ESTIMATION
CHAIR: WILLI MAURER (BASEL)

- Thomas Asendorf, Simon Schneider, Heinz Schmidli, Tim Friede: Sample Size Re-estimation for Repeated Poisson Counts in randomized controlled clinical trials
- Frank Miller and Tim Friede: Sample size re-estimation and continuous monitoring of the variance in longitudinal trials
- Martin Posch, Florian Klinglmüller, Franz König, Frank Miller: Estimation after blinded sample size adjustment

19:30 GET TOGETHER DINNER

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE

FRIDAY, JULY 6

08:30 – 10:30 SESSION 5: MULTIPLICITY AND ADAPTATIONS IN CLINICAL TRIALS
CHAIR: MARTIN POSCH (VIENNA)

- Norbert Benda: The use and discussion of adaptive designs within the European drug approval process
- Geraldine Rauch, Meinhard Kieser: Adaptive designs to improve the interpretation a composite endpoints by addressing the main component or a subcomposite
- Rene Schmidt, Robert Kwiecien, Andreas Faldum, Sandra Ligges: Adaptive Designs for the One-Sample Log-Rank Test
- Georg Ferber, Christine Garnett, Steve Riley, Jim Keirns: Multiplicity in the Model-Based Confirmatory Analysis of the QT-Interval

10:30 – 11:00 COFFEE BREAK

11:00 – 12:15 SESSION 6: MULTIPLICITY IN REGRESSION MODELS
CHAIR: GERNOT WASSMER (COLOGNE)

- Fang Wan, Wei Liu, Frank Bretz and Yang Han : Confidence set for a maximum point of a regression function
- Georg Gutjahr and Björn Bornkamp: Trend Tests Based on Multiple Nonlinear Regression Models
- Giuseppe Palermo and Daniel Sabanés Bové: Dose-escalation using safety and biomarker data: A Bayesian adaptive approach

12:15 – 13:30 LUNCH BREAK

SCIENTIFIC PROGRAM – DETAILED TIME SCHEDULE

13:30 – 14:45 SESSION 7: GROUP-SEQUENTIAL TESTS
CHAIR: FRANZ KÖNIG (VIENNA)

- Svenja Schüler, Meinhard Kieser, Geraldine Rauch: Extended futility boundaries in group sequential designs with two endpoints
- Michael Grayling: Optimally designing group sequential cross-over trials
- Hanna Daniel, Hans-Helge Müller and Nina Timmesfeld: A group sequential version of Fisher's exact test

14:45 END OF THE WORKSHOP

ABSTRACTS

SESSION 1 (1), JUNE 5, 09:10 – 12:30

Flexibility in confirmatory clinical trials – what is for free?

P.Bauer

Medical University of Vienna,

Email:peter.bauer@meduniwien.ac.at

It is discussed which changes of a pre-planned design with a pre-fixed confirmatory statistical analysis can be performed during the running trial so that the original statistical test can still be applied without inflating the type I error rate. Changes exclusively based on information from outside trial, unplanned stopping for futility, blinded sample size reassessment and pre-fixed adaptation rules (generally with adapted critical boundaries) are candidates for such changes where still the original test statistics can be used safely.

The maximum type error rate inflation when applying the original test in case of unconstraint mid-trial design changes modification may become large. However, putting realistic limitations on sample sizes and allocation ratios the option of constraint flexibility may not compromise on the type I error rate at all. If more flexibility is intended the adaptive design methodology may be applied.

SESSION 1 (2), JUNE 5, 09:10 – 12:30

To seamless or not to seamless? Lessons learned from our case studies

Marc Vandemeulebroecke (1), Robert L Cuffe (2), David Lawrence (1), Andrew Stone (3)

1: Novartis, 2: ViiV Healthcare, 3: AstraZeneca

Email: marc.vandemeulebroecke@novartis.com

Background: Inferentially seamless studies are one of the best known adaptive trial designs. Statistical inference for these studies is a well studied problem. Regulatory guidance suggests that statistical issues associated with study conduct are not as well understood. Some of these issues are caused by the need for early pre-specification of the phase III design and the absence of sponsor access to unblinded data. Before statisticians decide to choose a seamless IIb/III design for their programme, they should consider whether these pitfalls will be an issue for their programme.

Methods: We consider four case studies from different pharmaceutical sponsors. Each design met with varying degrees of success. We explore the reasons for this variation to identify characteristics of drug development programmes that lend themselves well to inferentially seamless trials and other characteristics that warn of difficulties.

Results: Seamless studies require increased upfront investment and planning to enable the phase III design to be specified at the outset of phase II. Pivotal, inferentially seamless studies are unlikely to allow meaningful sponsor access to unblinded data before study completion. This limits a sponsor's ability to reflect new information in the phase III portion.

Conclusions: When few clinical data have been gathered about a drug, phase II data will answer many unresolved questions. Committing to phase III plans and study designs before phase II begins introduces extra risk to drug development. However, seamless pivotal studies may be an attractive option when the clinical setting and development programme allow, for example, when revisiting dose selection.

References: Cuffe, Lawrence, Stone, Vandemeulebroecke: "When is a seamless study desirable? Case studies from different pharmaceutical sponsors." Pharmaceutical Statistics, to appear 2014

Incorporating feasibility assessment in the design of clinical studies

Thomas Jaki and Lisa Hampson

Email: jaki.thomas@gmail.com

Many publicly funded clinical trials fail to meet their recruitment timelines, with the consequence that these trials then require an extension of funding in order to complete recruitment. To avoid this scenario, there is a movement by funders towards requiring that larger Phase II and Phase III clinical trials incorporate a feasibility stopping rule, with the aim of establishing early on whether recruitment targets can be met within the planned time frame. The feasibility evaluation is usually based upon factors that are not of primary interest to the trial (i.e., does not concern the endpoint of direct clinical interest) and allows for three different actions: continue as planned; adapt recruitment procedures; or abandon the trial. Efficacy data collected during the feasibility phase of the trial contribute towards the final analysis of efficacy. In this presentation, we will show how ideas from the adaptive designs literature can be used to incorporate feasibility evaluations into the main trial design to ensure that the required type I error rate for testing efficacy is maintained and power is maximised. Simulations are used to illustrate the potential gains in power that follow from using our proposed approach. Optimal boundaries for the feasibility stopping rule are derived which minimise the expected overrun of the trial beyond its planned duration subject to controlling the probabilities of incorrectly allowing a trial to proceed when the recruitment rate is insufficient, and incorrectly abandoning a trial that would have gone on to complete in a timely manner.

Methods for Proper Handling of Over- and Underrunning in Phase II Designs for Oncology Trials

Stefan Englert[1] and Meinhard Kieser[2]

[1] Boehringer Ingelheim Pharma GmbH & Co. KG
[2] University of Heidelberg, Institute of Medical Biometry and Informtics

Email:stefan.englert@boehringer-ingelheim.com

Due to ethical considerations, phase II trials in oncology are typically performed with planned interim analyses. The sample sizes and decision boundaries are determined in the planning stage such that the type I and II error rates are controlled and have to be followed strictly later on. In practice, however, attaining the pre-specified sample size in each stage can be problematic.

The currently available approaches to deal with this problem either do not guarantee that the significance level is kept or are based on assumptions that are rarely met in practice. We propose a general framework for assuring type I error control in phase II oncology studies even when the attained sample sizes in the interim or final analysis deviate from the pre-specified ones.

We will show that the type I error rate must be reduced in case of overrunning to ensure control of the significance level while this does not apply to underrunning. Further, we will discuss both the similarities of our procedure to the conditional rejection principle proposed by Müller and Schäfer and the differences caused by the discrete endpoints used.

Application of the proposed procedure and some of its characteristics are illustrated with a real phase II oncology study.

SESSION 2 (1), JUNE 5, 11:30 – 13:00

Adaptive designs for with subgroup analysis optimizing utility functions

Franz König, Alexandra Graf, Martin Posch

Email:franz.koenig@meduniwien.ac.at

If the response to treatment depends on genetic biomarkers, it is important to identify predictive biomarkers that define (sub-)populations where the treatment has a positive benefit risk balance. One approach to determine relevant subpopulations are subgroup analyses where the treatment effect is estimated in biomarker positive and biomarker negative groups.

Subgroup analyses are challenging because several types of risks are associated with inference on subgroups. On the one hand, by disregarding a relevant subpopulation a treatment option may be missed due to a dilution of the treatment effect in the full population. Furthermore, even if the diluted treatment effect can be demonstrated in an overall population, it is not ethical to treat patients that do not benefit from the treatment when they can be identified in advance. On the other hand, selecting a spurious subpopulation increases the risk to restrict an efficacious treatment to a too narrow fraction of a potential benefiting population. We propose to quantify these risks with utility functions and investigate non-adaptive study designs that allow for inference on subgroups using multiple testing procedures as well as adaptive designs, where subgroups may be selected in an interim analysis. The characteristics of such adaptive and non-adaptive designs are compared for a range of scenarios.

Performance characteristics of interim decision rules in adaptive enrichment designs

Johannes Krisam[1], Meinhard Kieser[1]

[1]*Institute of Medical Biometry and Informatics, University of Heidelberg, Germany*

Email: krisam@imbi.uni-heidelberg.de

During the planning stage of a clinical trial investigating a potentially targeted therapy, there is usually a high degree of uncertainty whether the treatment is more efficient (or efficient only) in a subgroup as compared to the total patient population. Some recently developed adaptive designs allow a mid-course efficacy assessment of both the total population and a pre-defined subgroup thus permitting the selection of the most promising target population based on the interim results (see, e.g., Wang et al., 2007; Brannath et al., 2009; Jenkins et al., 2011; Friede et al., 2012). In order to identify the subset of patients more likely to benefit from a drug, predictive biomarkers are commonly employed. The applied interim selection rule has a crucial impact on the overall characteristics of the design.

The performance of several subgroup selection rules to be applied in adaptive two-stage designs is investigated. We present methods that enable to evaluate the operational characteristics of rules for selecting the target population thus enabling to choose an appropriate strategy. The comparison includes optimal decision rules which take the common situation of uncertain assumptions into account. The uncertainty about parameters such as the treatment effect and the sensitivity and specificity of the bioassay evaluating the biomarker is modeled by prior distributions (Krisam and Kieser, 2014). Additionally, common selection rules proposed in the literature are considered. The performance of these selection rules in adaptive enrichment designs is evaluated by investigating the probability of a correct selection of the target population, Type I error rate, and power.

References:

-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.
-Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine* 2009; **28**:1445–1463.
-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.
-Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine* 2012; **31**:4309–4320.
-Krisam J, Kieser M. Decision rules for subgroup selection based on a predictive biomarker. *Journal of Biopharmaceutical Statistics* 2014; **24**:188-202.

Comparison of different approaches to enrichment designs with multiple nested subgroups

Marius Placzek[1], Simon Schneider[1], Tim Friede[1]

[1] Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

Email:marius.placzek@stud.uni-goettingen.de

Growing interest in personalised medicine and targeted therapies is leading to an increase in the importance of subgroup analyses. Since adaptive designs carry the promise of making drug development more efficient, enrichment designs with adaptive selection of the population (i.e. predefined subpopulation or full population) at interim have gained increased attention. In confirmatory trials under regulatory conditions it is important that the statistical analysis controls the familywise type I error rate. This can be achieved by application of the so-called combination test approach (Brannath et al, 2009; Jenkins et al, 2011). This approach has been extended to the setting of multiple subgroups and implemented in the software package ADDPLAN. More recently, a conditional error function approach to adaptive subgroup selection has been proposed (Friede et al, 2012). To our knowledge comparisons between the combination test approach and the conditional error function approach have focused so far on settings with one subpopulation (Stallard et al, 2014). From similar comparisons in adaptive treatment selection it is known that advantages of the conditional error function approach compared to the combination test approach are more pronounced with larger number of treatments. Therefore we extend the conditional error function methodology to several nested subgroups in this presentation and provide a comprehensive comparison between the combination test approach and the conditional error function approach in simulation studies. The simulations are motivated and illustrated by clinical examples.

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.

-Friede T, Stallard N (2008) A comparison of methods for adaptive treatment selection. Biometrical Journal 50: 767-781.

-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.

-Stallard N, Homburg T, Parsons N, Friede T (2014) Adaptive designs for confirmatory clinical trials with subgroup selection. Journal of Biopharmaceutical Statistics 24: 168-187

SESSION 3 (1), JUNE 5, 14:15 – 15:30

A fallback test for three co-primary endpoints

Robin Ristl[1], Florian Frommlet[1], Armin Koch[2], and Martin Posch[1]

[1]Section for Medical Statistics, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University Vienna, Austria
[2]Centre for Biometry, Medical Informatics and Medical Technology, Hannover Medical School, Germany

Email: robin.ristl@meduniwien.ac.at

When efficacy of a treatment is measured by co-primary endpoints, efficacy is claimed only if for each endpoint an individual statistical test is significant at a local level α. While such a strategy controls the family-wise error rate (FWER) at level α, it may be strictly conservative and have low power. We improve the test of three co-primary endpoints to allow inference also in settings where only two out of the three show a significant result at the local level. While the test does not allow to reject an elementary null hypothesis in this case, it rejects an intersection hypothesis such that an effect in at least one of the endpoints can be inferred and the trial still serves as a proof of principle. We show under the assumption of multivariate normal test statistics with arbitrary correlation matrix that the procedure controls the FWER at level α in the strong sense. Besides the application to tests for co-primary endpoints the result uniformly improves the Rüger test in the setting of tri-variate normal test statistics. The latter rejects if two out of three hypotheses are significant at level 2α/3 but controls the type 1 error rate at level α without the assumption of multivariate normality. We investigate the power of the improved test procedure and compare it to hierarchical and Bonferroni tests for co-primary endpoints. The test procedure is illustrated with a clinical trial for a rare disease. An application of the procedure in the assessment of diagnostic tools is discussed.

SESSION 3 (2), JUNE 5, 14:15 – 15:30

A two-stage hierarchical multiple test procedure based on the asymptotically optimal rejection curve

Konstantin Schildknecht, Weierstrass Institute for Applied Analysis and Stochastics

Email: schildkn@wias-berlin.de

In this work we address the problem of testing a large number of hypotheses in the case where prior knowledge is available to partition the set of all hypotheses into disjoint subsets (families). If the proportion of true null hypotheses differs between families, this structural information can be used to increase statistical power. Such situations are common in modern research, e.g., in studies involving a population (pupils in a school) that can be grouped into families (classes) and the researcher is performing a test on each subject. Hypotheses of interest regard the performance within the whole population and within each subgroup. Additional interest lies on a hypothesis addressing the subgroups, like a partial-conjunction hypotheses.
For a given structure the error measure of interest can be applied separately to each of the families, relaxing the multiplicity problem. In our main result we show that under certain conditions asymptotic control of the false discovery rate (FDR) within each family implies asymptotic control of the FDR with respect to all hypotheses.
We propose a procedure which excludes those families from the analysis without strong evidence for containing true alternatives and show control of the familywise error rate at this step. Then we proceed to test individual hypotheses within each non-excluded family and obtain asymptotic control for the hypotheses within each family and for all hypotheses by applying our result.
In our approach we combine the results regarding an asymptotically optimal rejection curve developed by Finner[1] for asymptotic control of the FDR on the level of the individual hypotheses with the conservativeness of the standard Bonferroni correction on the level of the families. In simulations we demonstrate situations in which we can increase power in comparison with established procedures.

References: Finner, H., T. Dickhaus, and M. Roters (2009). On the false discovery rate and an asymptotically optimal rejection curve. Ann. Stat. 37(2)

SESSION 3 (3), JUNE 5, 14:15 – 15:30

Multiple tests in first-level analyses in functional magnetic resonance imaging

S. Kropf, S. Weston, D. Adolf

Institute for Biometry and Medical Informatics, Otto-von-Guericke University Magdeburg, Germany

Email: Siegfried.Kropf@med.ovgu.de

Typical studies in functional magnetic resonance imaging (fMRI) are typically carried out at two levels. At the first level, a series of brain scans is taken from one proband while he or she has to perform a well-defined schedule of tasks. In second-level analyses the results are summarized over several probands. Here we focus on first-level analyses, where the inference is based on the repeated scans as sample elements, though the physiology of brain processes causes correlations between scans that are taken within short time distances. That requires special adjustments of all statistical procedures.

The activities in the brain are assessed on a three-dimensional grid, so that each scan consists of a large set of about a hundred thousand or more of single measurements at these positions (voxels), which are considered simultaneously. In contrast, a study comprises only a few hundreds of scans. So we have a high-dimensional multiple test problem that is usually analyzed under the claim of familywise error control or control of the false discovery rate.

In a funded research project (DFG grant), we investigated several multiple test procedures with respect to type I error control and power under typical conditions of an fMRI session. These proposals are adaptions of own earlier proposals for ordered or weighted hypotheses and of the Westfall-Young procedure (Westfall and Young, 1993) to the situation of serially correlated sample vectors (Kropf, Hommel, 2004; Westfall et al., 2004). Whereas the own proposals had some local advantages, the modified Westfall-Young procedure generally had a good overall performance. In a re-evaluation of a collection of so-called resting state data, originally analyzed by Eklund et al. (2012), we could show that this procedure modified by blockwise permutation much better controls the familywise error than the conventional analyses assuming autoregressive models for the time course in each voxel.

References

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., Knutsson, H. (2012): Does parametric fMRI analysis with SPM yield valid results? - An empirical study of 1484 rest datasets. *NeuroImage* **61**: 565-578.

Kropf, S., Hommel, G. (2004). New parametric and nonparametric multiple test procedures for high-dimensional data. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **8**, spec. vol., 169-177.

Westfall, P.H., Kropf, S., Finos, L. (2004). Weighted FWE-controlling methods in high-dimensional situations. In *Recent Developments in Multiple Comparison Procedures*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 47, Y. Benjamini, F. Bretz, and S. Sarkar, eds., 143-154.

Westfall P.H, Young S.S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. New York: John Wiley & Sons.

SESSION 4 (1), JUNE 5, 16:00 – 17:15

Sample Size Re-estimation for Repeated Poisson Counts in randomized controlled clinical trials

Thomas Asendorf[1], Simon Schneider[1], Heinz Schmidli[2], Tim Friede[1]

[1] Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

[2] Novartis Pharma AG, Basel, Switzerland

Email: **Thomas**.**Asendorf**@med.uni-**goettingen**.de

In recent years MRI lesion counts have received considerable attention as endpoints in clinical trials in multiple sclerosis (MS) (Nicholas and Friede, 2012). In such trials, MRIs are conducted and lesion counts assessed repeatedly over time. We present a statistical model based on a binomial thinning approach for count data (McKenzie, 1986), which allows for time dependencies. We extended the model to also account for between patient heterogeneity which is particularly prevalent in MRI data. Furthermore, a Wald type test for testing differences in rates between the treatment groups is developed and a sample size formula is derived. An approach to nuisance parameter based sample size re-estimation (Friede and Kieser, 2006) is presented and its applicability to MRI driven trials in MS is discussed.

Keywords: time dependence, count data, sample size estimation, re-estimation, lesion counts

References

Friede T, Kieser M (2006) Sample size recalculation in internal pilot study designs: A review (with discussion). Biometrical Journal 48:537-555.

McKenzie (1986) Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. Advances in Applied Probabiliy 18: 679-705.

Nicholas R, Friede T (2012) Considerations in the design of clinical trials for relapsing multiple sclerosis. Clinical Investigation 2: 1073-1083.

SESSION 4 (2), JUNE 5, 16:00 – 17:15

Sample size re-estimation and continuous monitoring of the variance in longitudinal trials

Frank Miller (Stockholm University) and Tim Friede (University Medical Center Göttingen)

Email: frank.miller@stat.su.se

In many clinical trials, frequent longitudinal data is collected from each patient. For example in chronic pain trials, daily pain measurements of the patients can be collected during several weeks which leads to a large number of highly correlated post-baseline measurements for each patient.

Blinded sample size re-estimation or continuous monitoring of the variance [1] can deal with situations where uncertainty regarding the true variance exists. In trials with longitudinal data, the situation is common that at interim looks a restricted number of patients have completed the study but a large number has started treatment and first post-baseline data is collected but endpoint data is not yet available. Nevertheless, it is reasonable that the partial data available from these patients gives useful information about the variance of the endpoint [2, 3].

In this talk, we first quantify the gain of including partial data from patients when estimating the variance. Variability of sample size is often reduced but the amount of reduction depends on the correlation between measurements. Then, our main interest is to investigate the usefulness of a parametric model assumption for the covariance structure. We quantify the gain from the model assumption when the assumed model is correct and discuss consequences when a wrong model is assumed.

References

[1] Friede T, Miller F (2012). Blinded continuous monitoring of nuisance parameters in clinical trials. Journal of the Royal Statistical Society Series C 61:601–618.
[2] Wachtlin D, Kieser M (2013). Blinded Sample Size Recalculation in Longitudinal Clinical Trials Using Generalized Estimating Equations. Therapeutic Innovation & Regulatory Science 47:460-467.
[3] Wüst K, Kieser M (2003). Blinded sample size recalculation for normally distributed outcomes using long- and short-term data. Biometrical Journal 45:915–930.

SESSION 4 (3), JUNE 5, 16:00 – 17:15

Estimation after blinded sample size adjustment

Martin Posch, Florian Klinglmüller, Franz König, Frank Miller

Email: martin.posch@meduniwien.ac.at

When comparing the means of normally distributed endpoints the sample size to achieve a target power typically depends on nuisance parameters as the variance. It has been shown that superiority trials where the sample size is reassessed based on blinded interim estimates of the nuisance parameter achieve the target power regardless of the true nuisance parameter and the sample size reassessment has no relevant impact on the type I error rate.
While previous work has focused on the control of the type I error rate, we investigate the properties of point estimates and confidence intervals following blinded sample size reassessment.We show that the conventional estimates for the mean and variance may be biased and quantify the bias in simulations. Furthermore, we provide a lower bound for the bias of the variance estimate and show by simulation that the coverage probabilities of confidence intervals may lie below their nominal level, especially when first stage sample sizes are small. Finally, we discuss the impact of the findings for blinded sample size reassessment in clinical trials.

SESSION 5 (1), JUNE 6, 08:30 – 10:30

The use and discussion of adaptive designs within the European drug approval process

Norbert Benda (BfArM)

Email: norbert.benda@bfarm.de

The EMA reflection paper on adaptive designs was adopted 2007. Meanwhile, adaptive designs have become increasingly popular in clinical development programs in preparation of a marketing authorization application (MAA) in the European Union. Within the increasing number of scientific advice procedures at EMA, all kinds of adaptive designs are discussed in detail with respect to their operating characteristics, feasibility and risks. Whereas, classical group sequential designs are still very common, designs with sample size reassessment, Phase II/III combination designs, as well as adaptive dose finding and subgroup selection are more and more part of the discussions on possible study designs. This talk will give an overview on the different proposals, that are discussed with respect to their chances and regulatory challenges.

Adaptive designs to improve the interpretation a composite endpoints by addressing the main component or a subcomposite

Geraldine Rauch[1], Meinhard Kieser[1]
[1]Universität Heidelberg, Deutschland

Email: rauch@imbi.uni-heidelberg.de

Composite endpoints combine several time-to-event variables of interest within a single time-to-first-event analysis. The motivation for the use of a composite endpoint is to increase power by increasing the number of expected events. Sample size calculation for composite endpoints can be particularly difficult, as valid planning assumptions depend on the correct parameter specifications for the individual components. Therefore an adaptive design with sample size recalculation based on the observed effect during the interim analysis is an intuitive approach to solve this problem.
However, even if the sample size was adapted to the observed composite effect, the interpretation of the composite effect in terms of clinical relevance can be difficult if the components effects deviate from the assumptions made in the planning stage.

During interim analysis, we may found out that a particular component that was exclusively added to the composite in order to increase the effect in fact decreased the composite effect. The CAPRICORN Trial [1] is a very illustrative example for this situation.
Another possible scenario would be that the main component which is the most relevant for the patient (e.g. time-to-death) shows a higher effect than originally anticipated. In this situation it might be feasible to base sample size recalculation on the main component in order to improve the interpretation of the trial. In both situations, an adaptive design that allows a change in the primary endpoint *and* a sample size recalculation during the interim analysis would be helpful.

We propose different adaptive design strategies to face the above problems and evaluate and compare them in terms of power and type I error using Monte-Carlo simulations. Applications are illustrated by clinical study examples.

References

[1] The CAPRICORN Investigators. Effect of carvedilol on outcome after myocardial infarction inpatients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001; 357: 1385-1390.

SESSION 5 (3), JUNE 6, 08:30 – 10:30

Adaptive Designs for the One-Sample Log-Rank Test

Rene Schmidt, Robert Kwiecien, Andreas Faldum, Sandra Ligges
Institute for Biostatistics and Clinical Research, University of Münster, Germany

Email: rene.schmidt@ukmuenster.de

Adaptive group sequential designs for the One-Sample Log-Rank test are considered, and an improved method of sample size calculation is provided.
The one-sample log-rank test may be the method of choice if the survival curve of patients under a new treatment is to be compared to that of a historic control. According to the present paradigm proposed by Finkelstein, Muzikansky and Schoenfeld (2003), sample size calculation for the one-sample log-rank test is based on the number of events to be observed in order to obtain a certain power. We propose and study a new stopping criterion to be followed. Both approaches are asymptotically equivalent. Though, a simulation study indicates that the new criterion might be preferred for small sample size.

Reference

Finkelstein DM, Muzikansky A, Schoenfeld DA. Comparing survival of a sample to that of a standard population. Journal of the National Cancer Institute 2003;95: 1434–1439

Multiplicity in the model-based confirmatory analysis of the QT-interval

Georg Ferber[1], Christine Garnett[2], Steve Riley[3], Jim Keirns[4]

Email: Statistik.Georg.Ferber@bluewin.ch

The Thorough QT study proposed in the guideline E14 of the international conference of harmonisation is an example of a confirmatory study to show an aspect of the safety of a drug. The primary analysis is based on the QTc interval, the duration of the QT interval of the ECG, corrected for heart rate, a biomarker for the risk of drug-induced arrhythmia. The analysis proposed in the guideline foresees a test for non-inferiority at each of a number of typically 8 to 15 timepoints To exclude proarrhythmic risk, the null hypothesis of a prolongation of > 10 ms has to be rejected for each timepoint. Therefore reverse multiplicity affects the power, and, as a consequence, the sample size of such studies.

In order to overcome this problem, an analysis based on a mixed effects model relating the plasma concentration of the drug to the change from baseline of QTc has been proposed. In this case, the predicted effect at the geometric mean of the observed C_{max} values across subjects is of primary interest and the null hypothesis stating that it is > 10 ms is the only one to be tested in a confirmatory way. Not surprisingly, such an analysis is much more powerful than the per timepoint analysis.

There is considerable interest in making this type of analysis acceptable as the primary analysis, and to allow for the option to perform it based on ECG data collected during routine Phase I studies. Therefore, the Cardiac Safety Research Consortium (CSRC)and the IQ (International Consortium for Innovation and Quality in Pharmaceutical Development) have launched a study in 6 marketed drugs with known effect on QTc to show the viability of this approach.

However, the gain in power does not come for free. Indeed, the simplest model used is based on a number of assumptions, in particular on linearity of the concentration-effect relationship and on the absence of a delay between plasma concentrations and QTc prolongation. The appropriateness of these assumptions need to be evaluated and, in particular if this analysis is to become the primary one, need to be tested in a prespecified way.

We will present the criteria developed by a group of statisticians and pharmacometricians from industry, CROs and the FDA for the above mentioned study of the CSRC. These criteria are not yet fully formalised, and our presentation will show the state of the discussion on this specific aspect of multiplicity.

[1] Statistik Georg Ferber GmbH Riehen, Switzerland
[2] Certara
[3] Pfizer
[4] Astellas

SESSION 6 (1), JUNE 6, 11:00 – 12:15

Confidence set for a maximum point of a regression function

Fang Wan (1), Wei Liu (1), Frank Bretz (2) and Yang Han (1)

(1) University of Southampton
(2) Novartis Pharma

Email: F.Wan@soton.ac.uk

A confidence set for a maximum point of a regression function provides useful information on where a true maximum point lies, and so has applications in many real problems. In this paper, an exact 1-α confidence set is provided for a maximum point of a linear regression function. It is also shown how the construction method can readily be applied to many other regression models involving a linear function. Examples are given to illustrate this confidence set and to demonstrate that it can be substantially smaller than the only other conservative set that is available in the statistical literature so far.

SESSION 6 (2), JUNE 6, 11:00 – 12:15

Trend Tests Based on Multiple Nonlinear Regression Models

Georg Gutjahr (University of Bremen) and Björn Bornkamp (Novartis Pharma AG, Basel)

Email: georg.gutjahr@gmail.com

Consider a set of nonlinear regression models for the mean vector of normally distributed observations and the hypothesis that at least one of these models fit the data better than a constant model.
For a single sufficiently smooth model, Hotelling showed that the likelihood-ratio test statistic is a monotonous function of the correlation between the observations and the maximum-likelihood prediction; using methods from differential geometry, the exact null distributions of this statistic can be obtained. For multiple models, the best prediction from the multiple models is used in the likelihood-ratio test statistic. The null distribution is determined by volumes of tubular neighborhoods on the unit sphere. We describe how such volumes can be approximated numerically. This approach can also be used to calculate the distribution under alternative hypotheses and it does not require that the models are smooth. To demonstrate the method, we apply it to data from a dose-response clinical trial.

Dose-escalation using safety and biomarker data: A Bayesian adaptive approach

Giuseppe Palermo and Daniel Sabanés Bové
(F. Hoffmann-La Roche, Basel)

Emails: giuseppe.palermo@roche.com; daniel.sabanes_bove@roche.com

In early clinical dose-escalation studies typically the target is to find a dose with a certain toxicity probability, say between 20 and 35%. Therefore, the dose-escalation is only driven by safety data, ignoring potential biomarkers for efficacy. This strategy relies on the assumptions that the efficacy increases monotonically with the dose, and that such levels of toxicity can actually be reached.
However, for targeted monoclonal antibody therapies it is often the case that no dose-limiting toxicity is observed, such that dose selection cannot solely rely on safety, but must take into account pharmacodynamics (PD) data. Therefore we propose a Bayesian adaptive doseescalation framework that also uses a continuous biomarker to find the dose with maximum PD effect within certain safety constraint. Our approach builds on the work by Bekele and Shen (Biometrics, 2005), which uses the probit model to transform the binary safety outcome into a continuous variable, allowing to model safety and biomarker data by a bivariate normal distribution. We compare our approach with alternative dual endpoint designs, and illustrate the performance with simulation results.

Extended futility boundaries in group sequential designs with two endpoints

Svenja Schüler[1], Meinhard Kieser[1], Geraldine Rauch[1]
[1]Universität Heidelberg, Deutschland

Email: schueler@imbi.uni-heidelberg.de

In many clinical trials there are two primary endpoints of interest for both of which the null hypothesis has to be rejected in order to conclude efficacy of the new treatment. A possible application is a study with a composite endpoint including one particularly important component, often given by a harmful event as, e.g., death. As a negative effect in this component would not be acceptable, the efficacy claim could be based on the assessment of both the composite *and* the harmful component to improve interpretation[1,2].

In general, group-sequential designs can be used to stop a trial at an early stage due to proof of efficacy or due to futility. Group sequential designs considering a primary and a secondary endpoint have recently been investigated for hierarchically ordered hypotheses[3,4].

In contrast, we focus on the intersection-union test and develop sequential designs with different futility boundaries for the two endpoints. For the situation of composite endpoints, this takes account of the fact that an early stopping should be possible based on the result for the composite *or* for the component. The proposed group-sequential designs are investigated in terms of power and gain in interpretation. Recommendations for practical application are given and implementation is illustrated by clinical study examples.

References

[1] Rauch G, Beyersmann J. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. Statistics in Medicine 2013; 32:3595-3608.

[2] HUQUE FM, ALOSH M, RAFIA. ADDRESSING MULTIPLICITY ISSUES OF A COMPOSITE ENDPOINT AND ITS COMPONENTS IN CLINICAL TRIALS. JOURNAL OF BIOPHARMACEUTICAL STATISTICS 2011; 21:610-634.

[3] Tamhane AC, Metha CR, Liu L. Testing a primary and a secondary endpoint in a group sequential design. Biometrics 2010; 66:1174-1184.

[4] Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine 2010; 29:219-228.

SESSION 7 (2), JUNE 6, 13:30 – 14:45

Optimally designing group sequential cross-over trials

Michael Grayling

MRC Biostatistics Unit, Cambridge

Email: mjg211@cam.ac.uk

Group sequential designs can significantly reduce the expected sample size of parallel clinical trials, and are thus exploited in many settings to estimate the treatment effect of a new drug. However, for chronic diseases in particular, the cross-over trial remains the design of choice. By administering each patient with multiple treatments, individual effects can be accounted for and the variance of the estimated treatment effects reduced. In theory, a group sequential approach to a cross-over trial promises to bring the same advantages as in a parallel setting. Here, I discuss my work to date on creating a framework for such designs. Utilising the joint distribution of the test statistics, optimal designs in-terms of minimising the expected sample size, subject to required operating characteristics, can be determined via a search over sample size and stopping boundary shape.
Additionally, strongly controlling the Family Wise Error Rate of such trials will be discussed. Using data from the four-treatment four-period TOMADO trial for sleep apnoea devices, the performance of the group sequential cross-over designs will then be detailed.

A group sequential version of Fisher's exact test

Hanna Daniel[1], Hans-Helge Müller[2] and Nina Timmesfeld[1]

1 Institute of Medical Biometry and Epidemiology, Philipps University of Marburg,
2 Institute for Medical Informatics, Biometry and Epidemiology, University of Munich, Germany

Email: nina.timmesfeld@staff.uni-marburg.de

In clinical trials dichotomous endpoints such as occurrence of an event are frequently used. When comparing two groups, data can be summarized in a 2 _ 2 table. If total sample size and event rates are large resulting in large sample size in the four cells, _2 - test can be used to analyse such trials. However, if sample size in the cells is not large enough, exact methods such as Fisher's exact test are more appropriate. In particular, if planned recruitment time is large compared to the time until the endpoint is observed in each patient, group-sequential designs can save sample size and time as well as ful_lling power requirements. For comparing two groups with dichotomous endpoint only group sequential designs based on large sample assumption are available [1, 2].
In this talk we propose a group sequential extension of Fisher's exact test. Similar to Fisher's exact test this group sequential test will be constructed through conditioning on the appropriate su_cient statistics. Hence, this test will control the type I error rate. For di_erent number of stages and sample size we will make a power comparison of our new test and the available tests.

References

1. C. Jennison and B. W. Turnbull. Exact calculations for sequential t, _2 and F tests. Biometrika, 78(1):133{141, 1991. doi: 10.1093/biomet/78.1.133.
2. C. Jennison and B.W. Turnbull. Group sequential methods with applications to clinical trials. CRC Press, 2000.