



Regulatory and methodological issues in adaptive designs for confirmatory trials

Franz König and Martin Posch

Medical University of Vienna, Austria
Center for Medical Statistics, Informatics and Intelligent Systems
Franz.Koenig@meduniwien.ac.at
www.meduniwien.ac.at/user/franz.koenig

BBS Seminar Adaptive Designs, Basel, 2016-05-05



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-602552 (IDEAL - <http://www.ideal.rwth-aachen.de/>).

- A. Elsässer, J. Regnstrom, T. Vetter, F. Koenig, R. Hemmings, M. Greco, M. Papaluca-Amati, and M. Posch.
Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency.
Trials 15, 383, (2014)
<http://dx.doi.org/10.1186/1745-6215-15-383>
This work has received funding from the Austrian science fund FWF - P23167.
- P. Bauer, F. Bretz, V. Dragalin, F. Koenig, and G. Wassmer.
Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls.
Statistics in Medicine - Early View (2015)
<http://dx.doi.org/10.1002/sim.6472>
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013–602552.

Some History of Adaptive Designs

- 27 years ago Bauer: “Multistage Testing with Adaptive Designs”
- 21 years ago Proschan & Hunsberger: “Designed Extension of Studies Based on Conditional Power”
- 9 years ago EMA Reflection Paper
- 6 years ago FDA Draft Guidance (Drugs and Biologics)
- last year FDA Draft Guidance (Devices, CDRH, CBER)

Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls

P. Bauer, F. Bretz, V. Dragalin, F. Koenig, and G. Wassmer.
Featured Article in *Statistics in Medicine* 35, 325-347, 2016.
<http://dx.doi.org/10.1002/sim.6472> (Open Access)

With invited discussion by HUNG, WANG AND LAWRENCE; MEHTA AND LIU; VOLLMAR; MAURER

What is an Adaptive Design?



A study design is called “adaptive” if statistical methodology allows the **modification of a design element** (e.g. sample-size, randomization ratio, number of treatment arms) at an interim analysis with full **control of the type I error**.

EMA 2007

Guidance for Industry

Adaptive Design Clinical Trials
for Drugs and Biologics

DRAFT GUIDANCE
This guidance document is being distributed for comment purposes only.

A study that includes a **prospectively planned opportunity for modification** of one or more specified aspects of the study design and hypotheses **based on analysis of data** (usually interim data) from subjects in the study.

CBER, CDER FDA 2010

Draft Guidance for Industry and
Food and Drug Administration
Staff

Adaptive Designs for Medical
Device Clinical Studies

A clinical trial design that allows for **prospectively planned modifications based on accumulating study data** without undermining the trial's **integrity and validity**.

CBER, CDRH, FDA, 2015

Minimal Requirements for Confirmatory Adaptive Trials



London, 18 October 2007
Doc. Ref: CHMP/EWP/245002

COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE
(CHMP)

REFLECTION PAPER ON METHODOLOGICAL ISSUES IN CONFIRMATORY
CLINICAL TRIALS PLANNED WITH AN ADAPTIVE DESIGN

Guidance for Industry

Adaptive Design Clinical Trials for Drugs and Biologics

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

*“Using an adaptive design implies that the statistical methods control the pre-specified **type I error**, that **correct estimates and confidence intervals** for the treatment effect are available, and that methods for the assessment of **homogeneity of results from different stages** are pre-planned.”*

EMA REFLECTION PAPER (2007)

*“The chief concerns with these designs are control of the **study-wide Type I error rate**, minimization of the impact of any adaptation-associated **statistical** (see section VII.B) or **operational bias on the estimates** of treatment effects, and the **interpretability of trial results**.”*

FDA DRAFT GUIDANCE (2010)

Where are we now?

- Do sponsors consider adaptive designs in the development plans?
- Which type of adaptive designs are proposed?
- What are frequently identified problems?
- Which issues are still controversial?

European Regulatory Experience with Adaptive Designs

EMA Scientific Advice (SA) and Protocol assistance given by Scientific Advice Working Party (SAWP)

- Multidisciplinary group of 25 experts (complementary scientific competences)
- In the European Union it is not mandatory for sponsors of medical products to request SA
- SA provided is not legally binding with regard to any future MAA of the product concerned, neither for the regulatory agency nor for the company
- Advice to sponsors on all aspects of drug development: quality, non-clinical, clinical
- Advice on non-product related issues (e.g., on new statistical approach or validation of a scale)
- Advice on qualification of procedures, e.g., MCPMod in 2014

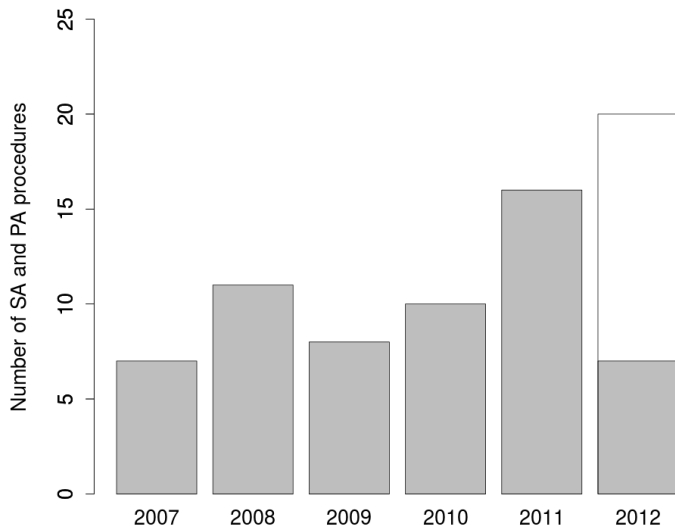
Adaptive Designs in EMA Scientific Advice Procedures

Survey of Scientific Advice/Protocol Assistance Procedures (01/2007- 05/2012)

- Scientific Advice/Protocol Assistance procedures of EMA Scientific Advice Working Party
- Search for Scientific Advice Letters containing terms such as, *adaptive design, flexible design, adaptive interim analysis, ...*
- Exclusion of phase I trials
- 59 procedures identified that contained questions on clinical trials with an adaptive designs
- May not include all procedures addressing adaptive designs (e.g., if sponsors use different terminology).

- A. Elsäßer, J. Regnstrom, T. Vetter, F. Koenig, R. Hemmings, M. Greco, M. Papaluca-Amati, and M. Posch.
Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. Trials 15, 383, (2014) (Open Access)
<http://dx.doi.org/10.1186/1745-6215-15-383>

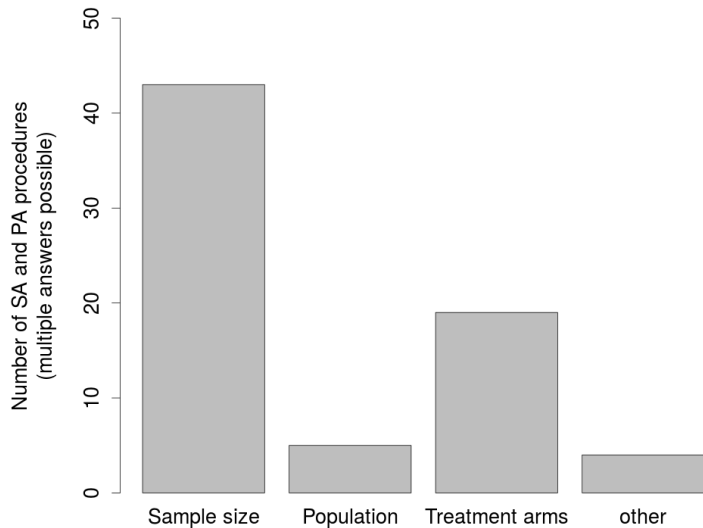
Number of Procedures per Year (n=59)



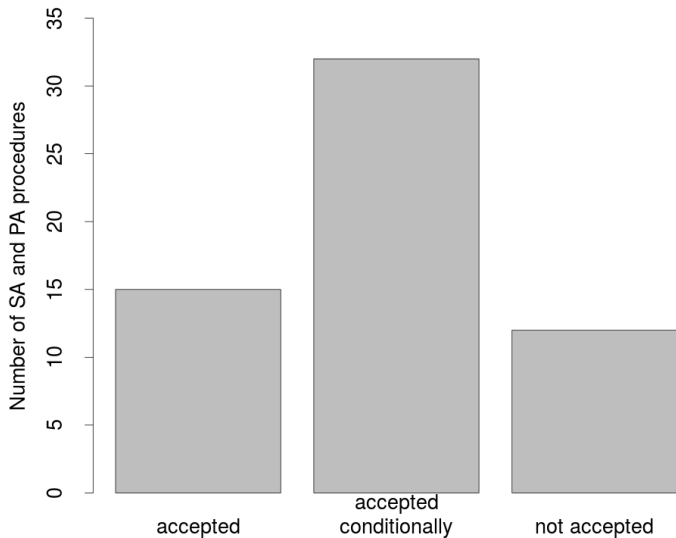
Types of Clinical Trials

- About 60% rare disease (prevalence of $< 5/10,000$),
1/3 applied for orphan designation
- Indications: About 50% oncology
- About 90% phase III or seamless II/III studies. Additionally,
phase II or pediatric studies.
- $\approx 75\%$ proposed as single pivotal trial
- Number of interim analyses:
1 $\approx 70\%$, 2 $\approx 20\%$, $> 2 \approx 5\%$.
- Primary Endpoint: time to event ($\approx 50\%$), binary ($\approx 30\%$),
continuous ($\approx 20\%$).

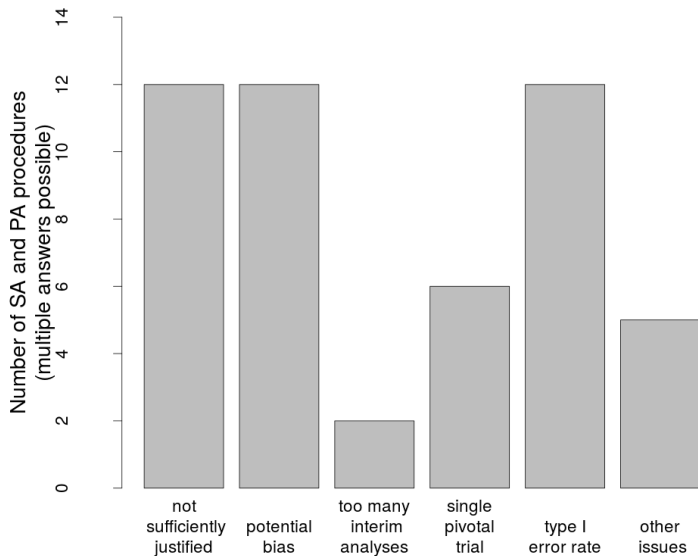
Types of Adaptations (n=59)



Overall Regulatory Response (n=59)



Issues Raised (Years 2009-2012, n=41)



Further Issues Identified in Adaptive Clinical Trial Proposals

- Insufficient sample size for subgroup analyses
- The option for adaptations is not prospectively planned (Post-hoc adaptive trial)
- Issues due to interim analyses (as in group sequential designs)
 - Overrunning
 - Feasibility of interim analyses because of large recruitment rates or delayed endpoints.
 - “Maturity” of survival data in interim analyses
 - Leakage of interim information leading to “silent adaptations”, not captured by the statistical methodology. They may result in issues for the interpretability of results.

Usually, standard estimators, not accounting for the adaptations, are proposed.

- In general, point estimates of adaptive designs will be biased.
- For specific scenarios, the bias can be quantified by simulations.
- The size of the bias will vary, depending on
 - the type of adaptation and specific adaptation rule,
 - the actual treatment effect(s)
 - nuisance parameters
- Adjusted confidence intervals

Several approaches seen:

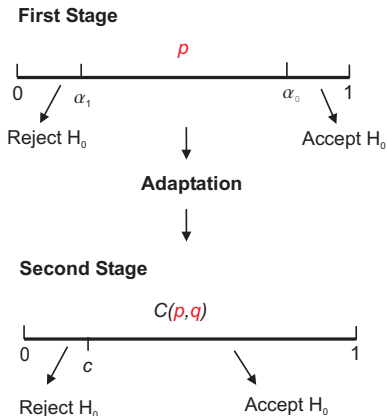
- Adaptive testing procedures (conditional error rate, combination tests)
- “Promising zone” approach.
- Standard analysis not accounting for adaptations.
- Simulation methods to demonstrate type I error control

Several approaches seen:

- Adaptive testing procedures (conditional error rate, combination tests)
- “Promising zone” approach.
- Standard analysis not accounting for adaptations.
- Simulation methods to demonstrate type I error control

Adaptive Two Stage Test based on Combination Tests

(BAUER 1989, BAUER & KÖHNE 1994, ...)



Planning:

- Fix design (only) for Stage 1
- Fix combination function $C(p, q)$ and critical value c
e.g. $C(p, q) = p \cdot q$

Stage 1:

- Compute p-value p from Stage 1 data
- Fix design for Stage 2 based on data from Stage 1

Stage 2:

- Compute p-value q from Stage 2 data.
- Reject H_0 iff $C(p, q) \leq c$.

Type I error control and combination functions

Type I error control

Type I error rate $\leq \alpha$ if we choose critical value c such that

$$P[p \leq \alpha \text{ or } C(p, q) \leq c] = \alpha$$

for independent and uniformly distributed p-values p and q .

- *Fisher product test*: $C(p, q) = p \cdot q$
(BAUER 1989, BAUER & KÖHNE, 1994)
- *Weighted inverse normal method*:
 $C(p, q) = \Phi(w_1 \Phi^{-1}(p) + w_2 \Phi^{-1}(q))$
(LEHMACHER & WASSMER, 1999)

(Remark: Can use critical values of a group sequential trial with interim information fraction w_1).

Clue of the Adaptive Test

- Do not pool the data of the stages, combine the stage-wise p-values.
- Then the distribution of the combination function under the null does not depend on design modifications
- Hence the adaptive test is still a test at the level α for the modified design!
- Applicable also for multiple looks, multiple hypotheses, ...
- Adaptations can depend on all (unblinded) interim data including secondary and safety endpoints.
- For a control of the type I error rate, one need not pre-specify how the Stage 1 data determine the design of Stage 2.

Justification of the Adaptive Design

EMA Reflection paper:

"Adaptive designs would be best utilised as a tool for planning clinical trials in areas where it is necessary to cope with difficult experimental situations."

- Is there a need for an adaptive trial?
- Have less complex design options been considered as well and compared to the adaptive design?
- Is the number of interim analysis justified? More than one interim analysis maybe justified in long term clinical trials.
- Is there a need for unblinding?
- Potential advantages of the adaptive design need to be weighed against potential biases and additional complexities.

Case Study 1: Sample Size Reassessment

- Open-label, two-armed, single pivotal phase III study for an anticancer drug in a rare disease
- Objective: To demonstrate superiority of the drug over a standard treatment for the primary endpoint of overall survival.
- Pre-planned adaptive design with two interim analyses (independent data monitoring committee, IDMC) with Haybittle-Peto stopping boundaries
- Interim analyses at 50% and 80% of events, given a fixed overall sample size
- At the second interim analysis, possibility to increase the number of events by 20% if the interim results show a promising but not overwhelming trend (conditional power arguments).
- No increase of the sample size.
- Proposed analysis: inverse normal method

Case Study 1: SAWP/CHMP Reply

- Design is acceptable from a statistical point of view if the type I error rate is controlled and operational bias is avoided.
- No agreement to the early rejection boundary in the first interim (concerns over the totality of evidence that would be available for a benefit-risk assessment) but agreement to futility stopping.
- Discussion whether primary analysis should be based on the standard fixed sample test statistics. Inverse normal test as sensitivity analysis:
 - If sample size is increased only if a promising interim effect is observed, the fixed sample test controls the type I error rate under certain assumptions ("Promising Zone Approach").
 - The inverse normal method down-weights the second stage treatment effect if the number of events is increased. This is undesirable if the survival curves initially separate but become closer at later time points.
 - A complexity (not explicitly discussed), is the potential inflation of the type I error rate if adaptations are based on information of patients censored at the interim analysis.

Type I Error Control in Adaptive Survival Studies

Patients recruited in the first stage maybe still under risk in the second stage.

- Tests based on the **independent increments property of the log-rank statistics** are in general not valid if adaptations depend on **secondary endpoints**.

POSCH & BAUER, 2004

- Test procedures where the **follow-up time from first stage patients is fixed** control the type I error rate, but **do not include all events** in the test statistics if the trial is extended.

JENKINS ET AL. '11, IRLE & SCHÄFER, '12

- Conservative tests based on all observed data are typically strictly conservative.

MAGIRR ET AL. 2016

Case Study 2: Interim Dose Selection

- Seamless phase II/III designs for two pivotal placebo controlled trials of a new chemical entity for the treatment of diabetic nephropathy
- Objectives:
 - Demonstrate superiority in a surrogate marker of kidney disease progression
 - Select two of three initially tested dose strengths based on an interim analysis of the benefit/risk ratio in both trials.
- Pre-planned interim analyses to be performed by an IDMC after 60% of 420 patients had completed 8 weeks of treatment in the first trial.
- Dose selection based on data from both trials using pre-determined criteria for the primary efficacy and safety parameters.
- Proposed type I error rate control: Bonferroni adjustment to control the familywise error rate adjusting the level for two comparisons only.

Case Study 2: SAWP/CHMP Reply

- The statistical testing procedure was not endorsed, as it was not supposed to control the familywise type I error rate for the three hypotheses initially considered.
- Instead, adaptive combination tests based on the closure principle and adaptive Dunnett test procedures based on the conditional error rate are adequate methods to control the type I error rate.
- The advantage of the proposed design with respect to power should be evaluated as it may be small.
- Safety evaluation may not be possible to support dose selection at the proposed time of interim analysis.

Conclusions

Conclusions (I)

- General inferences about regulatory standards and preferences is difficult
- The assessment depends on the overall quality and the general context:
 - overall drug development program,
 - type of medicinal product
 - indication
 -

Conclusions (II)

Questions that are generally addressed in the assessment

- 1 Is there a good rationale? Have alternative, more standard trial designs been considered?
- 2 Does the proposal fit well in the context of the development program and the data that will be available for the marketing authorization application?
- 3 Can the proposal be implemented without important damage to trial integrity?
- 4 Is the type I error rate controlled?
- 5 Has the potential bias of treatment effect estimates been evaluated?
- 6 Is the proposal practical and feasible?

Conclusions (III)

- Adaptive designs seem well accepted if properly planned and **implemented**
- A range of increasingly complex adaptive designs are proposed, the majority in rare diseases
- Surprisingly, still a lack of methodological knowledge
 - how to achieve type I error control
 - how to assess the efficiency of the design (timing of interim analysis, adaptation rules, power)
- Who should be decide on adaptations at interim, (DMC?, sponsor?, ...)
- Group sequential designs developed in the 70s are now well established - do we still have to wait one decade until the adaptive methodology is common knowledge?

See references of

- Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls
P. Bauer, F. Bretz, V. Dragalin, F. Koenig, and G. Wassmer.
Statistics in Medicine 35, 325-347, 2016
<http://dx.doi.org/10.1002/sim.6472> (Open Access)
(179 references)
- Adaptive designs for confirmatory clinical trials
F. Bretz, F. Koenig, W. Brannath, E. Glimm, and M. Posch
Statistics in Medicine 28, 1181-1217, 2009
<http://dx.doi.org/10.1002/sim.3538>
(77 references)

Backup

Adaptive Tests for Survival Data

- The combination test and the conditional error approach can be extended to survival data and the log-rank test (independent increments property).

WASSMER 2006, SCHAEFER & MUELLER 2001

- Stagewise p-values are calculated from the events occurring in each stage.
- Caveat: This may lead to biased tests if adaptations are based on covariate information or secondary endpoints of first stage patients censored at the time of the interim analysis. E.g., adaptations based on PFS when the primary endpoint is OS.

BAUER & POSCH, 2001

Approaches for Adaptive Tests for Survival Data

To include covariate information/secondary endpoints also from patients censored at the interim analysis

- use test statistics stratified for the covariates.

BRANNATH ET AL. 2009, SCHAEFER AND MUELLER 2011

- calculate stagewise p-values based on all patients recruited in the respective stage (regardless when their events occurred).

JENKINS ET AL. 2011

- modify the conditional error approach and condition on the data of all patients recruited in the first stage.

IRLE & SCHAEFER 2012

For the latter two approaches

- the trial can only be extended but not shortened
- if the trial is extended not all events from first stage patients enter the final test statistics.

Type I Error Control in Adaptive Survival Studies

Patients recruited in the first stage maybe still under risk in the second stage.

- Tests based on the **independent increments property of the log-rank statistics** are in general not valid if adaptations depend on **secondary endpoints**.

POSCH & BAUER, 2004

- Test procedures where the **follow-up time from first stage patients is fixed** control the type I error rate, but **do not include all events** in the test statistics if the trial is extended.

JENKINS ET AL. '11, IRLE & SCHÄFER, '12

- Conservative tests based on all observed data are typically strictly conservative.

MAGIRR ET AL. 2016

Simulation Based Procedures for Type I Error Control

Clinical Trial Simulations

"What, if"-scenarios: How do designs and assumptions affect the performance of trials and the drug development program?

- Controllable: doses, regimes, sampling time, study duration, interim analyses, adaptations, ...
- Uncontrollable: drug characteristics (PK/PD), disease progression, drop-outs, unscheduled adaptations: "dealing with the unexpected" as dropping of an unsafe dose, ...

Simulate operating characteristics for specific trial designs :

- Probabilities of "success" (evaluate different power definitions)
- Probabilities for early trial termination (due to safety, efficacy or futility)
- Probabilities to select "best" dose during clinical development
- Impact on effect estimates (bias?) and MSE
- Expected sample sizes
- Demonstration of Type I error rate control

Type I error estimation by simulation

The adaptive trial is simulated a large number of times under the null hypothesis. The fraction of runs with a rejection of the null hypothesis is calculated.

Straight forward to implement if the trial has

- a single point null hypothesis,
- a fully pre-specified adaptation rule depending on the primary endpoint only,
- no nuisance parameters,
- an adaptation rule that is not too complex such that large number of simulation runs can be performed.

- Precise estimates of the Type I error rate, require **large numbers of simulations**
- How large? For small sample numbers, a selective choice of seed may lead to biased estimates.

Table: Expected number of seeds to obtain one simulated Type I error rate below 0.025 when the actual error rate is 0.026.

# of runs	Expected # of seeds
10^4	4
10^5	43
10^6	8×10^9

Nuisance Parameters and Simulation Studies

It is not sufficient to investigate the global null hypotheses but type I error control has to be shown for

- the global and all intersection null hypotheses
- for all possible (nuisance) parameter values
- all considered adaptation options

For example, one needs to consider

- in multi-armed trials: all combinations of effective and non-effective arms and effect sizes
- in enrichment designs: all combinations of treatment effects in the subgroup and overall population
- with adaptation rules depending on surrogate/safety/secondary endpoints: all effect sizes in these endpoints

Example: Response Adaptive Design

Comparison of rates, $n=30$, comparison of 6 test statistics for comparison of rates

p_1		0.200	0.300	0.500	0.700	0.800
p_2		0.200	0.300	0.500	0.700	0.800
	T_{MW}	0.028	0.045	<u>0.056</u>	0.048	0.035
	T_{Risk}	<u>0.118</u>	0.085	0.058	0.034	0.018
SMLE	T_{MO}	0.004	0.012	<u>0.040</u>	0.034	0.023
	T_{MC}	0.038	0.049	0.056	<u>0.057</u>	<u>0.057</u>
	T_{ML}	<u>0.070</u>	0.065	0.056	0.054	0.062

Simulated Type I error (10.000 runs)

GU & LEE, 2010, TABLE 11

Challenges of Type I Error Control with Simulations

- Can one sufficiently explore the type I error rate in adaptive clinical trials (relying on an abundance of parameters and assumptions) by simulations?
- Has the worst case scenario with respect to the type I error really been identified?
- Have only scenarios with favourable assumptions been investigated and presented by the sponsor?
- How can one convincingly communicate the results of the very extensive simulation work required?

Summary – Simulations

- In principle, clinical trial simulation is a valid tool to study operating characteristics of clinical trials.
- However, often it may not be feasible to cover the whole relevant parameter space to show FWER in the strong sense by simulations.
- Statistical methods for which type I error control can be demonstrated under less restrictive assumptions (e.g., combination tests, conditional error rate based tests) are preferred.
- Still simulations are valuable to assess the power of adaptive tests.
- To investigate bias and MSE of point estimates, simulation studies are proper tools. Additionally, worst case scenarios for the bias are of interest.