

---

RECENT METHODOLOGICAL  
ADVANCES CONTRIBUTING TO  
CLINICAL TRIALS AND REGULATORY  
STATISTICS

---

Joachim Röhmel  
Bremen, Germany

**ICSB 30**  
**Prague, August 25<sup>th</sup> 2009**

---

## Content of my talk

- regulatory medical statistics
  - new methodology is not always appropriate
  - the lag between innovation and practical implementation
  - guidelines/regulatory initiatives
  
- three areas with high activity and innovative ideas
  - adaptive (sequential) clinical trials
    - sample size
    - seamless phase II/III and treatment selection
    - subgroup selection
  - control of the multiple type I error
    - primary and secondary objectives
  - non-inferiority
    - two arm versus three arm trial
    - multiple hypotheses

---

## regulatory medical statistics - statistics used for the evaluation of data from clinical trials

- itt principle and all its variants
  - imputation for missing data
- sample size estimation
  - type II error control
- SAP and pre-definition of hypotheses – primary and secondary
  - multiple objectives
- strict type I error control
  - valid statistical tests and estimation
  - systematic bias, e.g. by using historical references

## new methodology is not always appropriate – an example

- non-parametric evaluation of data from a bioequivalence study
  - median unbiased estimation (Hodges-Lehmann)
  - exact confidence intervals
  - invariant against (monotone) data transformation
  - robust
  - insensitive against outlier
  - **!!!! insensitive against outlier !!!!**
    - the bioequivalence study is in many situations the only study with the new (generic) drug treatment.
    - Outlier may indicate that the study or the generic treatment is not without problems.
    - Outlier necessitate a careful investigation for the biological reasons.

## the lag between innovation and practical implementation

- after a new method has been published and is deemed attractive for application in clinical trials, statistical assessors
  - must become aware of it
  - need time to understand
  - need to discuss (within agency and in public) properties in applications different from those for whom the method is primarily developed
  - need time to check correctness (often enough minor or major errors are detected, not exclusively by statistical assessors, but by the community of potential applicants)

## guidelines/regulatory initiatives

- Format and Content of Clinical and Statistical Sections of Application (1988)
- ICH E9 (1998)
- Notes for Guidance (CPMP/CHMP)
  - non-inferiority
  - missing data
  - multiplicity
  - ...
- Guidance to Industry (FDA)
  - adaptive design
  - non-inferiority
  - missing data
  - ...

## FDA's Critical Path Initiative (CPI)

- 5 areas\*
  - active control non-inferiority clinical trials
  - multiple endpoints, primary and secondary
  - missing data due to patient drop out and withdrawal
  - adaptive study design
  - simulation of clinical trial

\* R.T. O'Neill, Biom J 48, 2006, 559-564

## adaptive (sequential) clinical trials

- Peter Bauer (1989, 1994) and colleagues from Vienna, Marburg, Cologne
  - combination test
- Michael Proschan and Sally Hunsberger (1995)
  - conditional power
- Great enthusiasm (~2000-2006, in particular after CPI has been issued):
  - We can change everything at any time
- EMEA (2007)
  - adaptive designs should not be seen as a means to alleviate the burden of rigorous planning of clinical trials

---

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox  
sample size re-estimation

- ❑ rescue of underpowered trials
- ❑ some problems with suboptimal weighing
- ❑ blinded sample size re-estimation possible

---

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox  
seamless phase II/III - treatment selection

- ❑ main issue in Phase II is finding the therapeutic dose range
- ❑ main feature in phase III is confirmation of efficacy (and safety)
- ❑ interweaving these phases by combining them into one single study conducted in two stages
- ❑ treatment selection at interim
- ❑ type I error control
- ❑ multiple testing in adaptive designs
- ❑ rigorous application of the closure principle

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox  
seamless phase II/III - treatment selection

- the regulatory component.
  - Quotations from an internet site
    - “At the end of phase II, the manufacturer meets with FDA officials to discuss the development process, continued human testing, any concerns the FDA may have, and the protocols for phase III, “
- How does this fit with the meaning of “seamless”
- Oxford Biomedica\*
  - “The FDA invited submissions of adaptive Phase II/III trial designs in metastatic colorectal cancer and will work with Oxford BioMedica and the Company's collaborators to prepare suitable protocols for submission and review. “

\* Posted July 6, 2009

## adaptive (sequential) clinical trials (cont.) – more serious promises from the adaptive toolbox: subgroup selection –enrichment designs

### □ references

- Song Y, Chi GYH (Statist.Med. 2007: 3535-3549) A method for testing a prespecified subgroup in clinical trials
- Simon R, Maitournam A (3rd FDA/DIA Statistical Forum, 2009): planned subset analysis in the primary analysis plan for pivotal clinical trials.
- Dmitrienko A, Zhao Y, Tamura R (3rd FDA/DIA Statistical Forum, 2009): Design of Clinical Trials with a sensitive subgroup
- Alosch M, Huque MF (Statist Med 2009:3-23): A flexible strategy for testing subgroups and overall population
- Wang SJ, Hung HM, O`Neill RT: Adaptive patient enrichment designs in therapeutic trials. Biom J 51, 2009, 358-374
- Brannath W et al (Statist. Med. 2009; **28**:1445–1463): Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology.

---

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox:  
subgroup selection –enrichment designs

- Wang SJ, Hung HM, O'Neill RT:
  - “The adaptive enrichment approaches permit assessment of treatment effect that may be applicable to specific nested patient (sub)sets due to heterogeneous patient characteristics and/or differential response to treatment, *e.g.* a responsive patient subset versus a lack of beneficial patient subset, in all patient (sub)sets studied.”
  - “We emphasize the need of additional studies to replicate the finding of a treatment effect in an enriched patient subset. “

---

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox  
subgroup selection –enrichment designs

- Brannath et al. describe the traditional approach as follows
  - (i) a hypothesis generating (exploratory) study to identify a subpopulation,
  - (ii) the confirmation of the sensitivity of this sub-population in an independent second (e.g. phase II) study, before
  - (iii) running a phase III study in the selected target population
- and propose to combine (ii) and (iii) in an adaptive trial, while the information needed in (i) is coming from a parallel exploratory (biomarker) trial.

---

adaptive (sequential) clinical trials (cont.) – more serious  
promises from the adaptive toolbox  
subgroup selection –enrichment designs

- in the setting of a survival trial, Brannath et al. propose
  - to use Bayesian decision tools to select the population of interest at an interim analysis
    - predictive power probabilities for study success
    - posterior probabilities for treatment effects
  - to apply appropriately stratified logrank tests in the adaptive test procedure
- this guaranties
  - overall type I error control
  - unbiased estimation.
- comprehensive simulations necessary to illustrate the operating characteristics of the interim decision rules

Question mark: randomisation may not be stratified in stage I

---

control of the multiple type I error

primary and secondary objectives

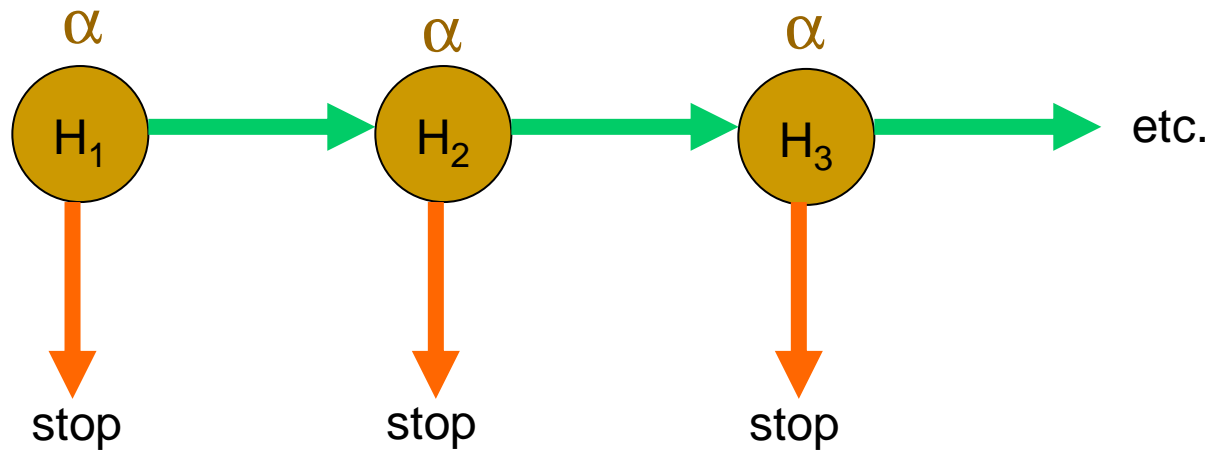
p-value based weighted Bonferroni/procedures

- serial gatekeeping (Maurer, Hothorn, Lehmacher (1995))
- PAAS (Le Moye, 2003)
- fall back (Wiens, Dmitrienko (2003, 2005))
- coherence principle (Hommel, Bretz (2006))
- parallel gatekeeping (Dmitrienko, Tamhane(2007))
- graphical procedures (Bretz et al, Burmann, Sonesson, and Guilbaud (2009))
- confidence sets (Bretz and Strassburger (2007), Guilbaud (2008, 2009))

control of the multiple type I error

primary and secondary objectives

serial gatekeeping

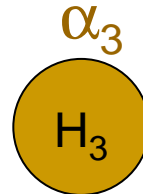
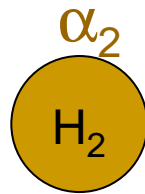
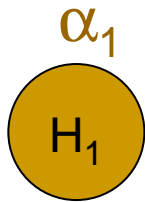


---

control of the multiple type I error

primary and secondary objectives

fall back



etc.

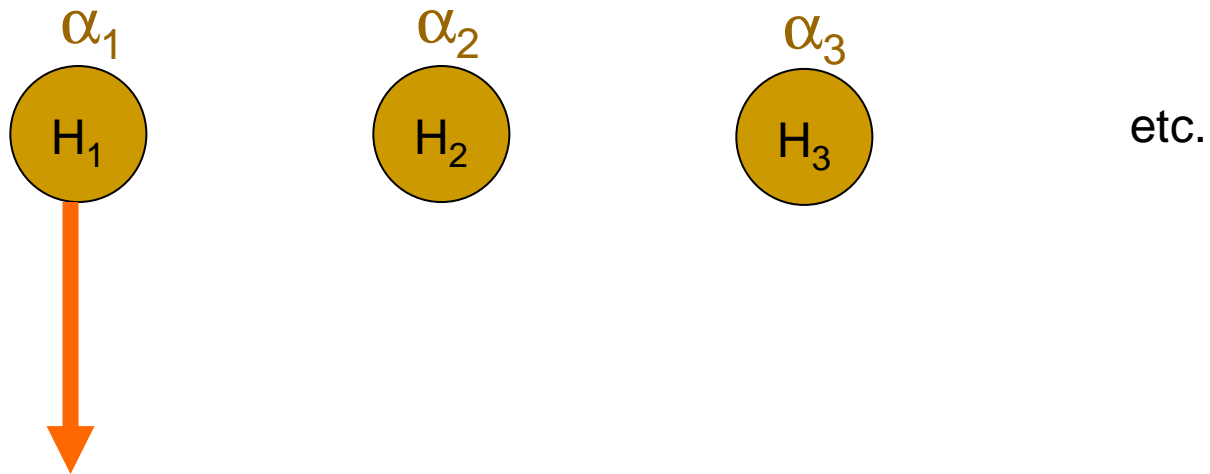
$$\alpha = \sum_{i=1}^n \alpha_i$$

---

control of the multiple type I error

primary and secondary objectives

fall back (start)



---

control of the multiple type I error

primary and secondary objectives

fall back (cont.)

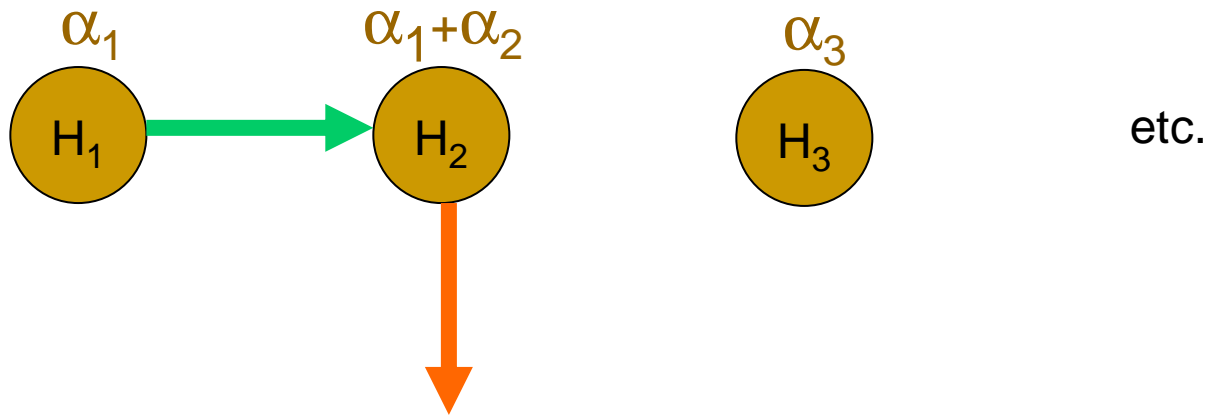


---

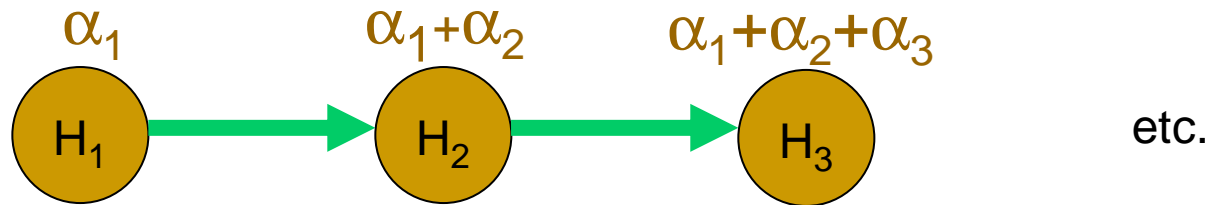
control of the multiple type I error

primary and secondary objectives

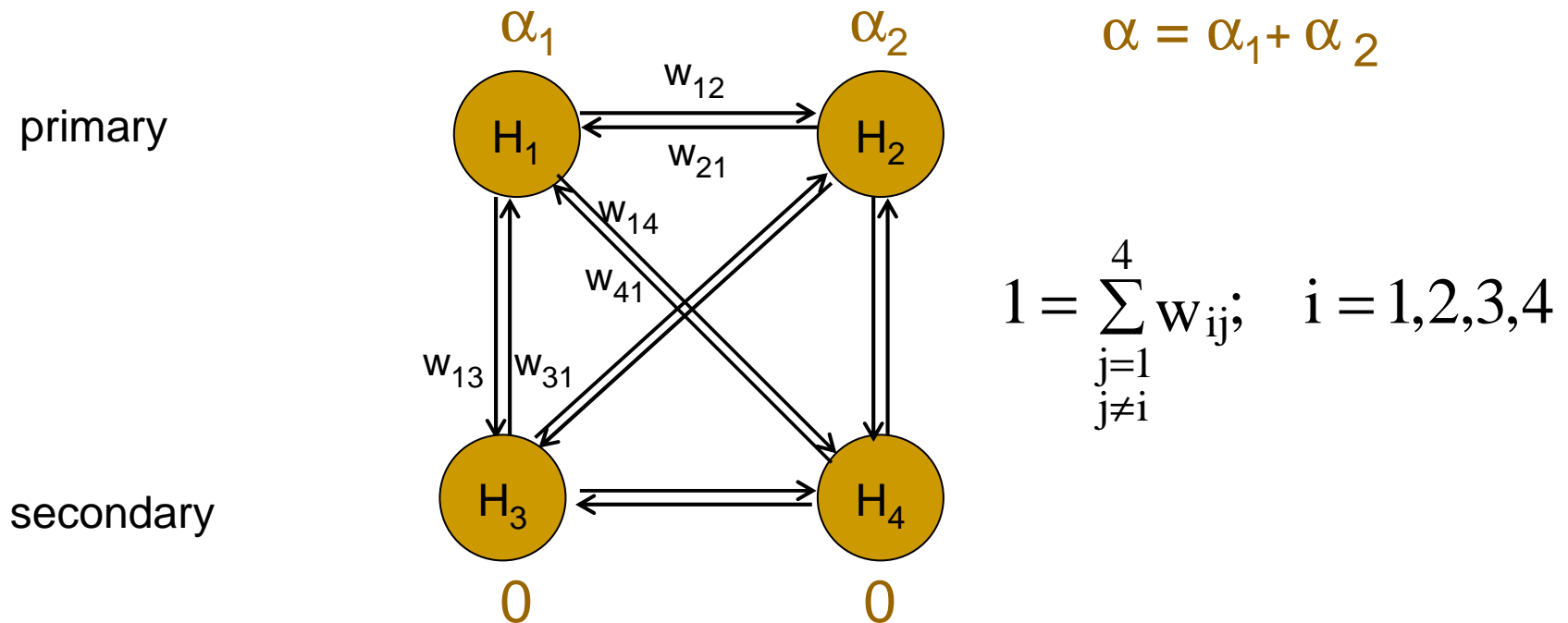
fall back (cont.)



control of the multiple type I error  
primary and secondary objectives  
fall back (cont.)



control of the multiple type I error  
 primary and secondary objectives  
 parallel gatekeeping (an example)



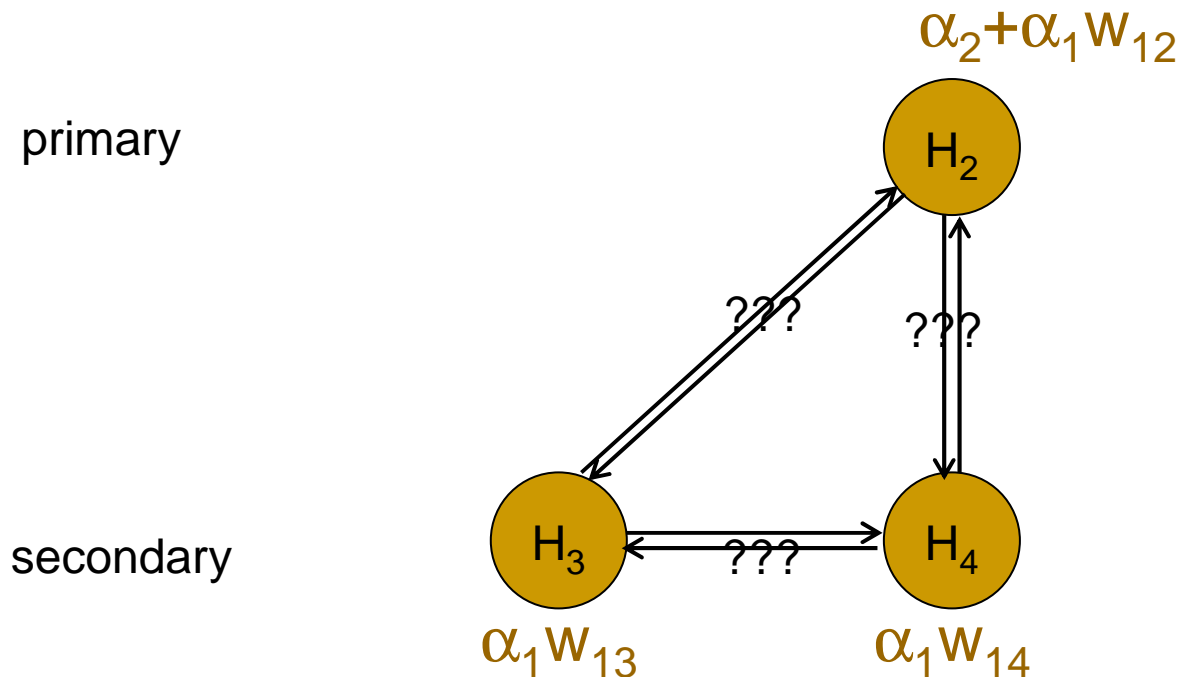
Be sure that no secondary hypothesis can be rejected unless at least one primary hypothesis is rejected

O'Neill RT: Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. CCT 18, 1997, 550-556

control of the multiple type I error

primary and secondary objectives

parallel gatekeeping – after rejection of  $H_1$



weights “???” have to satisfy the coherence principle (Hommel, Bretz, Maurer, 2007).

## References

- Maurer W, Hothorn L, Lehmacher W,. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. In Vollmar (Ed.) *Biometrie in der chemisch pharmazeutischen Industrie*, 3-18, 1995
- Bauer P, Röhmel J, Maurer W, Hothorn L. Testing strategies in multi-dose experiments including active control. *Statistics in Medicine* 17, 1998, 2133-2146
- Westfall P, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *J Stat Planning Inference* 99, 2001, 25-40
- Wiens B. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharaceutical Statistics* 2, 2003, 211-215
- Dmitrienko A, Offen WW, Westfall PH. gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Stat in Med* 22, 2003, 2387-2400
- Brian L. Wiens and Alexei Dmitrienko. The Fallback Procedure for Evaluating a Single Family of Hypotheses. *J Biopharm Stat* 15, 2005, 929 – 942
- Alex Dmitrienko, Ajit C. Tamhane, Xin Wang, Xun Chen. Correction: Stepwise Gatekeeping Procedures in Clinical Trial Applications. *Biom Journal* 48, 2006, 1041
- Alex Dmitrienko, Walt Offen, Ouhong Wang, Dan Xiao. Gatekeeping procedures in dose-response clinical trials based on the Dunnett test. *Pharm Stat* 5, 2006, 19-28

---

## more references

- Alex Dmitrienko, Brian L. Wiens, Ajit C. Tamhane, Xin Wang. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Stat in Med* 26, 2007, 2465-2478
- Olivier Guilbaud. Bonferroni Parallel Gatekeeping - Transparent Generalizations, Adjusted P -values, and Short Direct Proofs. *Biometrical Journal* 49, 2007, 917-927
- Alex Dmitrienko, Brian L. Wiens, Ajit C. Tamhane, Xin Wang. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Stat in Med* 26, 2007, 2465-2478
- Gerhard Hommel, Frank Bretz, Willi Maurer. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Stat in Med* 26, 2007, 4063-4073
- Alex Dmitrienko, Ajit C. Tamhane. Gatekeeping procedures with clinical trial applications. *Pharm Stat* 6, 2007, 171-180
- Alex Dmitrienko, Ajit C. Tamhane, Lingyun Liu, Brian L. Wiens. A note on tree gatekeeping procedures in clinical trials. *Stat in Med* 27, 2008, 3446-3451
- Alex Dmitrienko, Ajit C. Tamhane, Brian L. Wiens. General Multistage Gatekeeping Procedures. *Biom J* 50, 2008, 667-677
- Jianjun(David) Li, Devan V. Mehrotra. Gatekeeping Testing via Adaptive Alpha Allocation. *Biom J* 50, 2008, 704-715
- A Dunnett-Bonferroni based parallel gatekeeping procedure for dose-response clinical trials with multiple endpoints. Xu, Nuamah, Liu, Lim. *Pharm Stat* early view 2009

## Graphical solutions to parallel gatekeeping procedures

---

### Confidence sets

- Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat in Med* 28, 2009, 586-604
  - C.-F. Burman, C. Sonesson, O. Guilbaud. A recycling framework for the construction of Bonferroni-based multiple tests *Stat in Med* 28, 2009, 739-76
- 
- Olivier Guilbaud. Alternative Confidence Regions for Bonferroni-Based Closed-Testing Procedures that are not Alpha-Exhaustive *Biom J* 2009 Early view

---

## non-inferiority

### two arm versus three arm trial

- The three arm trial (Experimental E, Reference R, and Placebo P) is not controversial, because the concurrent placebo control
  - Pigeot I, Schäfer J, Röhmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Stat in Med* 22, 2003, 883-900
  - Koch A, Röhmel J. Hypothesis testing in the "gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. *J Biopharm Statistics* 14, 2004, 315-325
  - Röhmel J, Pigeot I. A comparison of multiple testing procedures for the gold standard non-inferiority trial. *J Biopharm Statistics* 2009 accepted

---

## non-inferiority

### two arm versus three arm trial

- The two arm trial (Experimental E, Reference R) is controversial, because no concurrent placebo control
  - references are many (also from regulators in Europe, USA and Japan)
  - two EMEA guideline (switching, choice of the non-inferiority margin)
  - awaiting draft FDA Guidance to Industry on this topic.

---

## non-inferiority multiple hypotheses

- Multiple testing with multiple primary endpoints has been the focus of ongoing research.
- Remember: what is a primary variable ?
  - Any variable that can express substantial benefit for the patients.
  - Therefore it can also express substantial harm in the opposite direction
- Since ICH E9 the majority of statistical tests has been one-sided
  - Harm (or suspicion of harm) in a primary variable may go undetected. It is just non-significant.
- A satisfactory situation for multiple primary variables is
  - some are clinically and statistically significant, and the remaining are at least not worrying.

## references

- Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics* 40, 549–567
- Tang, D. I., Gnecco, C., and Geller, N. (1989). An approximate likelihood ratio test for the normal mean vector with nonnegative components with application to clinical trials. *Biometrika* 76, 577–583.
- Bloch, D. A., Lai, T. L., and Tubert-Bitter, P. (2001). One-sided tests in clinical trials with multiple endpoints. *Biometrics* 57, 1039–1047.
- Tamhane, A. C. and Logan, B. R. (2002). Accurate critical constants for the one-sided approximate likelihood ratio test for a normal mean vector when the covariance matrix is estimated. *Biometrics* 58, 650–656.
- Perlman, M. D. and Wu, L. (2004). A note on one-sided tests with multiple endpoints. *Biometrics* 60, 276–280
- Tamhane, A. C. and Logan, B. R. (2004). A superiority-equivalence approach to one-sided tests on multiple endpoints in clinical trials. *Biometrika* 91, 715–727.
- Röhmel J, Gerlinger C, Benda N, Läuter J. On Testing Simultaneously Non-inferiority in Two Multiple Primary Endpoints and Superiority in at Least One of Them. *Biom J* 48, 2006,
- Bloch, D. A., Lai, T. L., Su, Z., and Tubert-Bitter, P. A combined superiority and non-inferiority approach to multiple endpoints in clinical trials. *Statistics in Medicine* 2006 DOI: 10.1002/sim.2611
- Logan BR, Tamhane AC. Superiority Inferences on Individual Endpoints Following Noninferiority Testing in Clinical Trials. *Biom J* 50, 2008, 693-703

---

## My conclusions

- Regulatory medical statistics has played an important role, especially since ICH E9 guideline was issued.
- Guidelines have motivated ongoing research to bring existing methods closer to maturity.
- New theory and methods help refining our statistical tool box and increase our understanding
- Science should always be the basis of regulatory requirements
- The more we know, the more we want.