Exploring changes in treatment effects across design stages in adaptive trials

Tim Friede
Warwick Medical School
The University of Warwick, UK
t.friede@warwick.ac.uk

Joint work with Robin Henderson (University of Newcastle, UK)

Heterogeneity in Treatment Effect Estimates

- When treatment effects differ across design stages . . .
  - results might be difficult to interpret
  - did information 'leak out' at interim???

- **minimum requirement** (CHMP guideline, Section 4.2.1)

  " [. . .] the same careful investigation of heterogeneity and justification to combine the results of different stages as is usually required for the combination of individual trials in a meta-analysis."

Investigation of Heterogeneity in Meta-Analyses

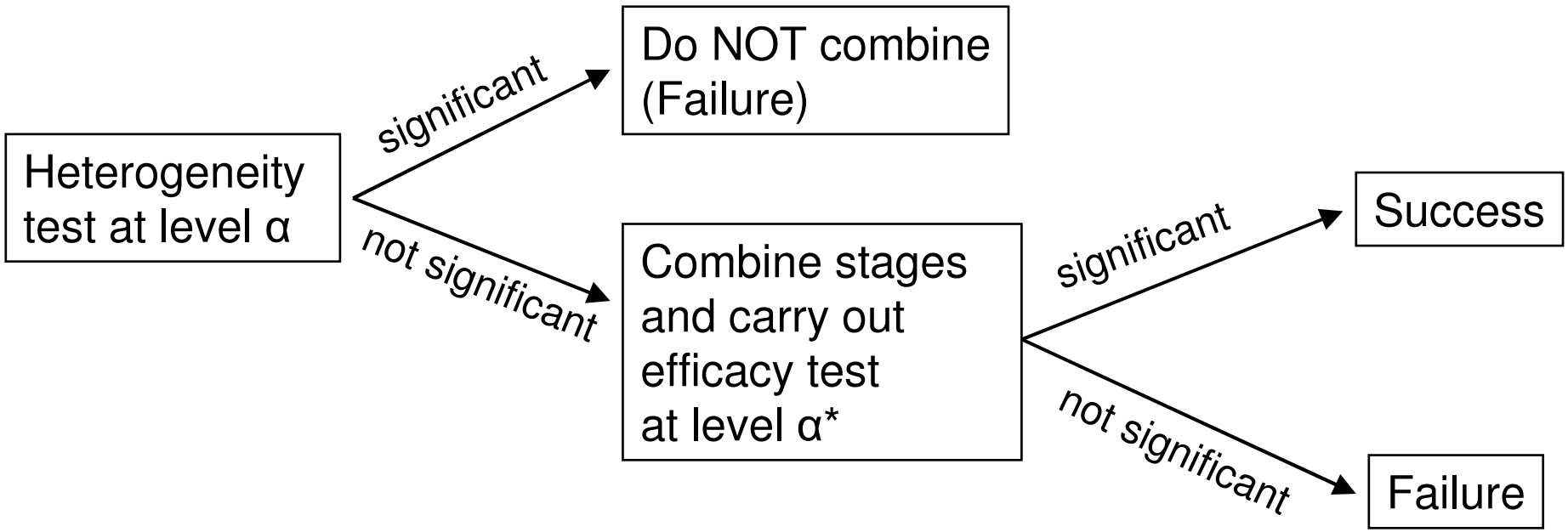
Basic procedure

- **formal hypothesis test**: do the treatment effects differ across stages?

- **if significant**, studies are not combined in meta-analysis

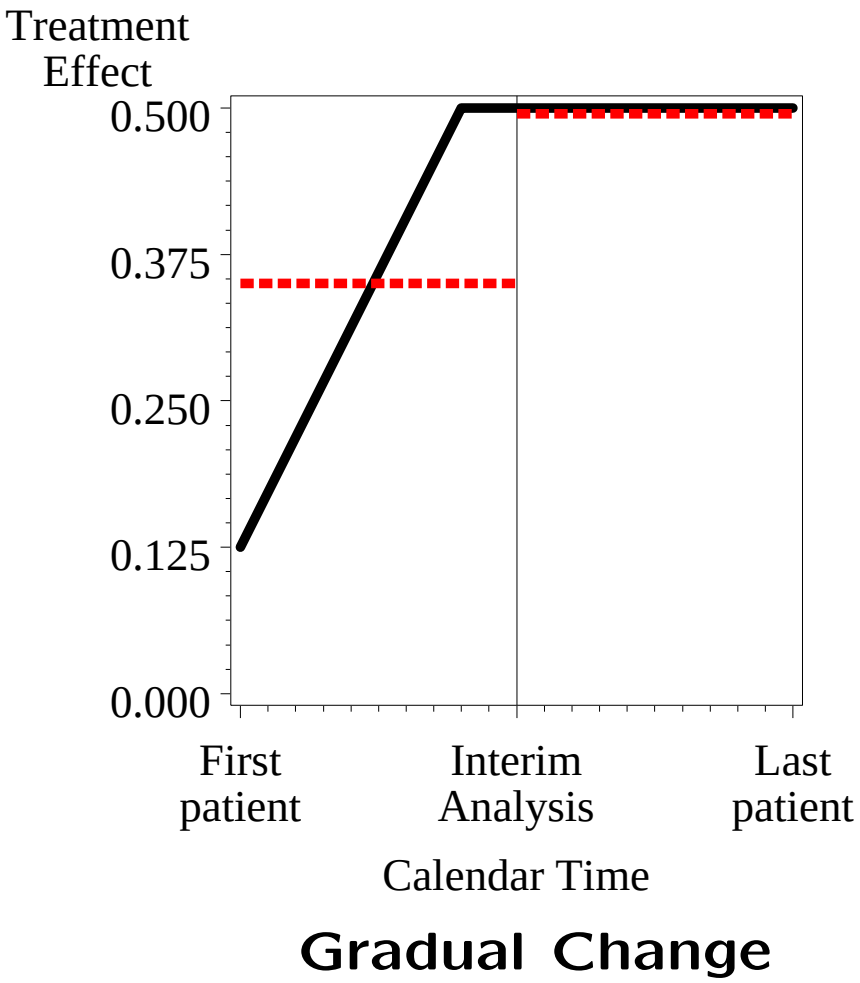
- **significance levels** $\alpha = 0.10$ or 0.15 common since power of heterogeneity test generally low

Applying the MA Procedure to Adaptive Trials

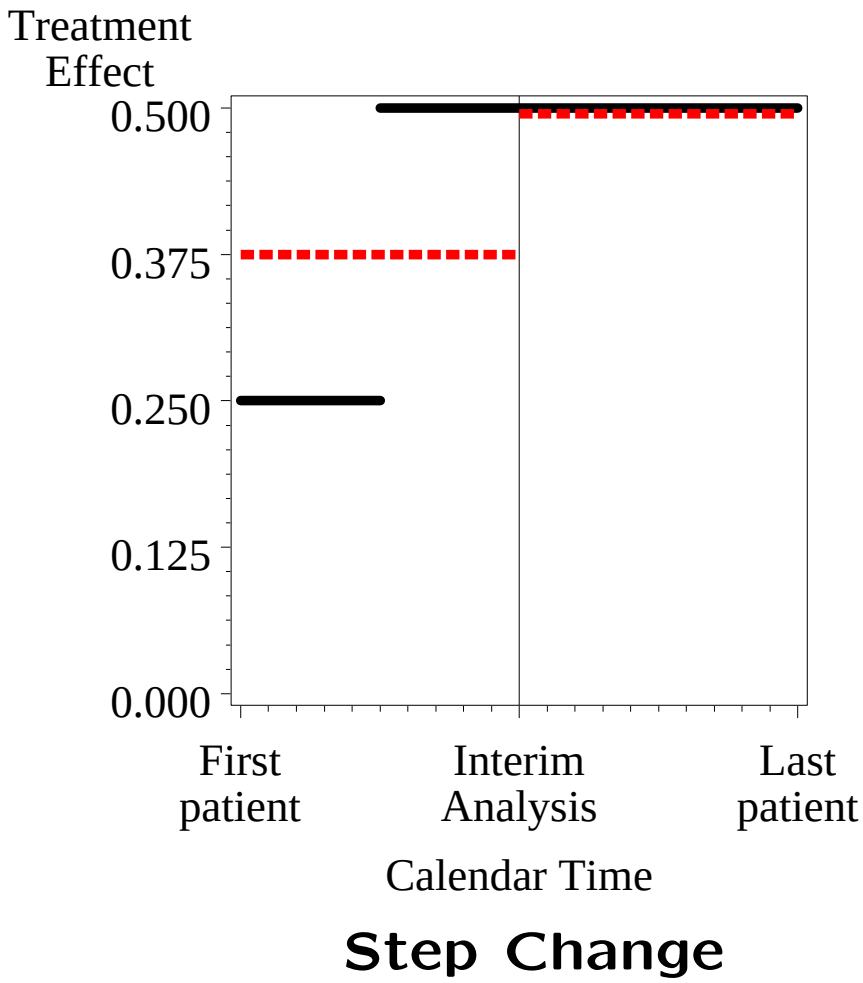


What are the consequences for adaptive trials?

Heterogeneity Test Confounded by Calendar Time



Gradual Change



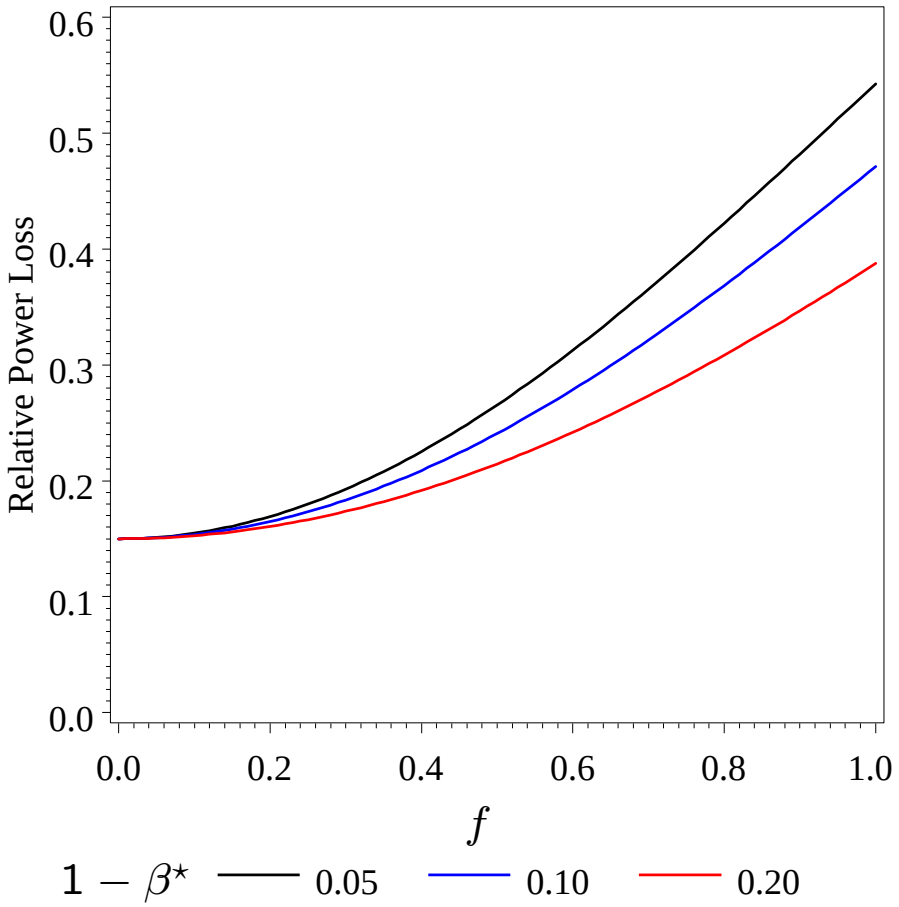
Step Change

An Investigation into Heterogeneity Testing in Adaptive Trials

- **situation considered**
  - two-stage trials with equally sized first and second stage
  - equally sized treatment arms
  - continuous (normal) outcomes
  - significance levels: heterogeneity $\alpha = 0.15$, efficacy $\alpha^\star = 0.025$

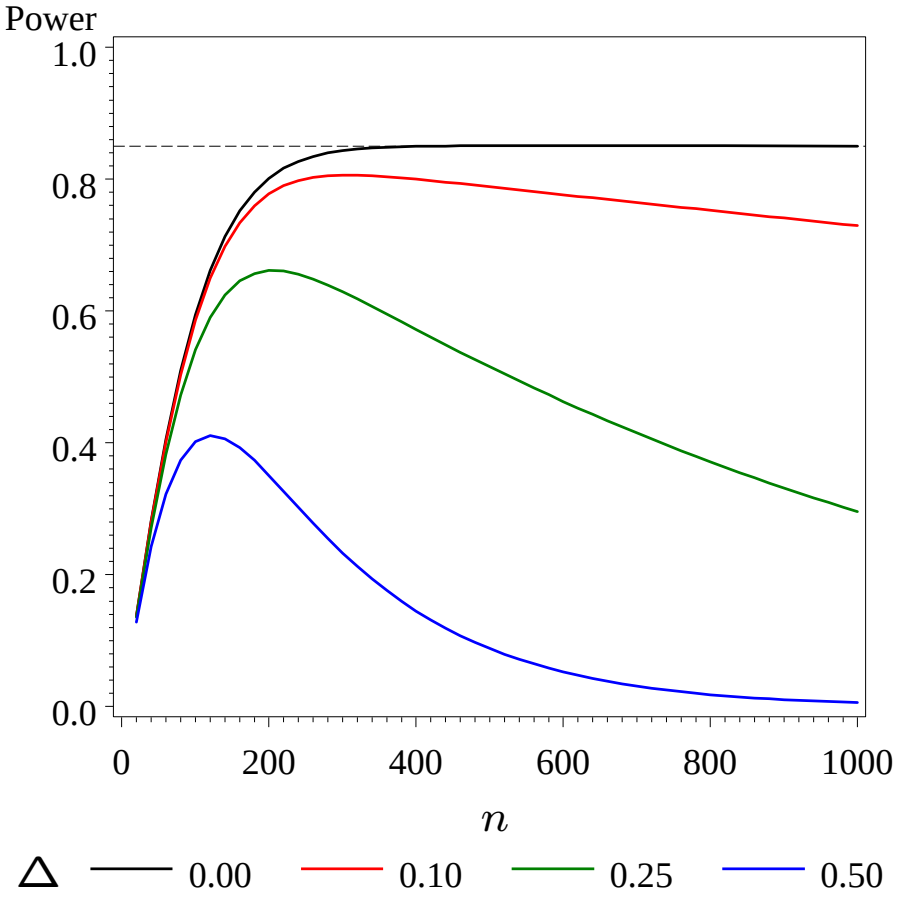
- **'successful study'**: non-significant heterogeneity test + significant efficacy test (probability of success is called **power** here)

Relative Loss in Power due to Heterogeneity Test



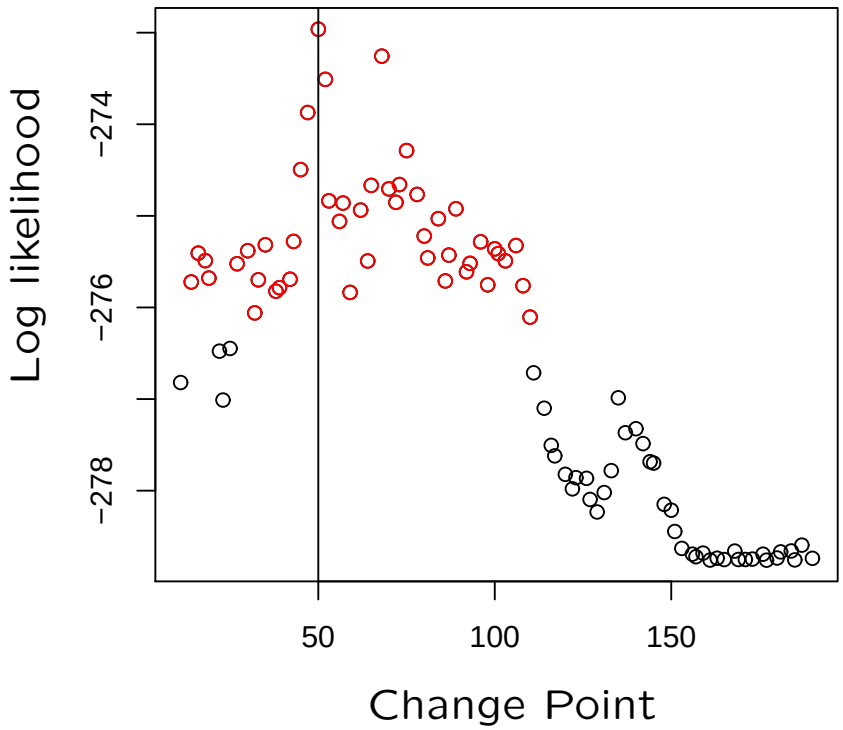
- loss in success probability (power) due to heterogeneity test
- relative power loss = power of heterogeneity test
- change in effect as fraction f of average effect
- effect change could be due to calendar time effects unrelated to interim analysis (e.g. learning effects)
- power $1 - \beta^\star$ of efficacy test

Can the power loss be compensated by larger samples?



- power of procedure with heterogeneity test
- total sample size n
- average trt effect $\theta = 0.5$
- change in treatment effect Δ
- effect change could be due to calendar time effects unrelated to interim analysis (e.g. learning effects)

Motivating the Use of Change Point Methods



CP method suggests change before IA.

- **simulated trial**
  - 100 patients per stage
  - step change after 50 patients with effect changing from 0.25 to 0.75
- **heterogeneity test**: $p = 0.01$
- **change point methods**
  - search for maximum test statistics
  - adjust critical value
  - calendar time confounding in studies with historic controls (Heuer & Abel 1998)

Alternative Testing Procedure

- initial heterogeneity test at level α_1: if significant, then . . .

- Considering only data of first stage: search for a change point and test whether it is significant at level $\alpha_{2.1}$.

  – if not, then conclude "change due to IA"

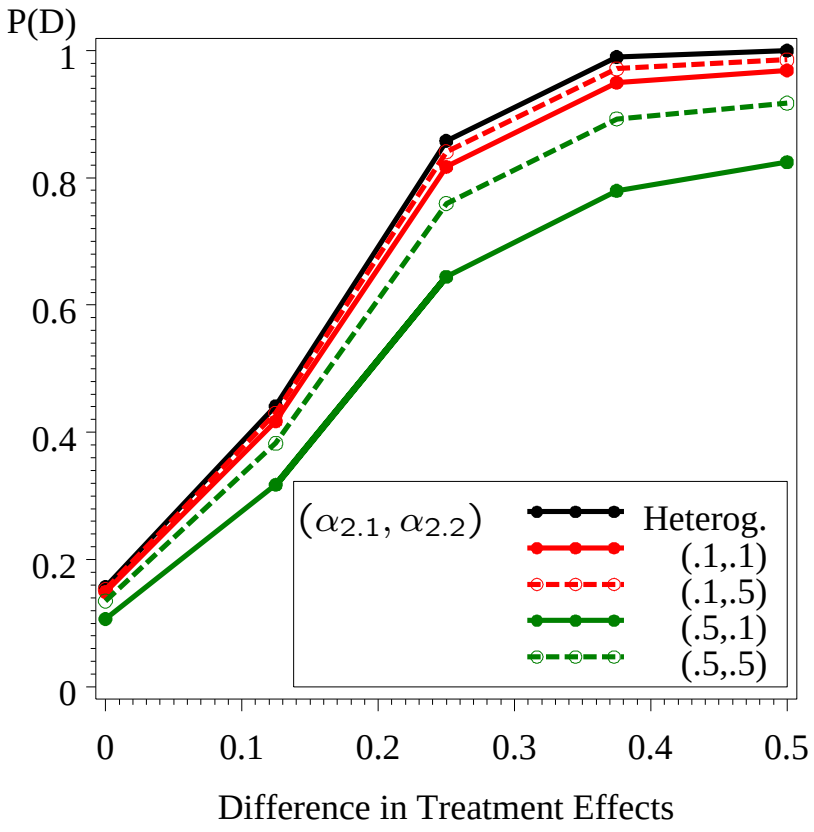
  – if yes, then . . .

- Carry out a test comparing treatment effects in the first stage after the change point and the second stage at level $\alpha_{2.2}$.

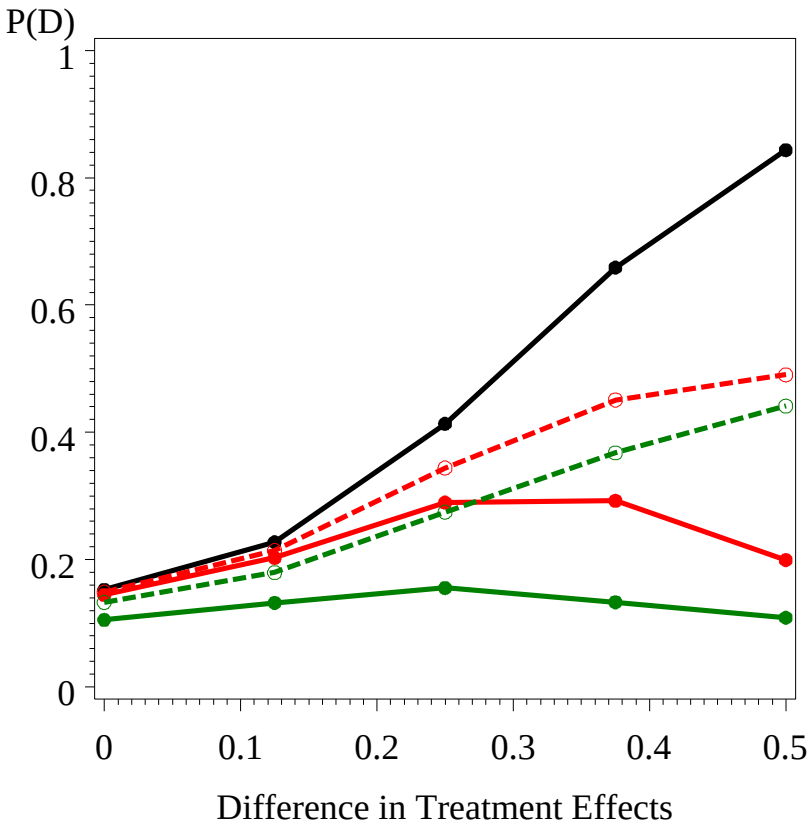
  – if (not) significant, then conclude "change (not) due to IA"

Simulated Probability of "Change due to IA" Conclusion

CP at IA



CP before IA



Investigation of Calendar Time Effects

- Altman & Royston (1988) suggest use of **CUSUM plots**
  - popular tool in quality control (Grigg et al 2003)
- **patient number as predictor** in linear model (Senn 2000)
- **critical issue in adaptive randomisation**
  - see e.g. Coad (1994), Hu & Rosenberger (2000)

What can be learned from meta-analysis?

- **investigating heterogeneity**: stage vs. patient level covariates
  - small number of stages $\Rightarrow$ investigation of stage-level covariates difficult
  - patient level data available in adaptive trials (unlike in publication based meta-analysis)
  - individual patient data: interactions of prognostic factors with treatment effect (subgroup analyses)
- **importance of treatment effect scale**: multiplicative vs. additive model (see Sutton et al 2000 Sec. 3.5.1, or Hand 1994 Ex.6)

Conclusions and Discussion

- **heterogeneity test approach**
  - leads to great loss and power that cannot be compensated for by larger sample sizes
  - calendar time effects unrelated to IA make matters worse
- **alternative approaches** allowing for calendar time effects need more attention
- **design**: careful consideration and discussion in planning phase

References

- Altman DG, Royston P (1988) The hidden effect of time. Statistics in Medicine 7: 629–637.

- Committee for Medicinal Products for Human Use (2007) Reflection paper on methodological issues in confirmatory clinical trials with flexible design and analysis plan. London, 18 October 2007, Doc. Ref. CHMP/EWP/2459/02.

- Coad DS (1994) Sequential estimation for two-stage and three-stage clinical trials. Journal of Statistical Planning and Inference 42: 343–351.

- Grigg OA et al (2003) Use of risk-adjusted CUSUM and RSPRTcharts for monitoring in medical contexts. Statistical Methods in Medical Research 12: 147–170.

- Hand DJ (1994) Deconstructing statistical questions. RSS Series A 157: 317–356.

- Heuer C, Abel U (1998) The analysis of intervention effects using observational data bases. In: Abel U, Koch A (eds.). Nonrandomized comparative clinical studies. Duesseldorf: Symposium Publishing; pp 101-107.

- Hu F, Rosenberger WF (2000) Analysis of time trends in adaptive designs with application to a neurophysiology experiment. Statistics in Medicine 19: 2067–2075.

- Senn S (2000) Consensus and controversy in pharmaceutical statistics. The Statistician 49: 135–176.

- Sutton AJ et al (2000) Methods for meta-analysis in medical research. Wiley.