Exploring changes in treatment effects across design stages in adaptive trials

Tim Friede
Warwick Medical School
University of Warwick, UK

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

CHMP Guideline on Flexible Trials

CHMP (2006), Section 4.2.1

> " [. . .] the applicant must pre-plan methods to ensure that results from different stages of the trial can be justifiably combined. In this respect, studies with adaptive designs need at least 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 metaanalysis."

Heterogeneity across Design Stages

- **two-stage design considered**

- **model**: independent responses with $Y_i \sim \mathrm{N}(\mu_i, \sigma^2)$

$$\mu_i = \begin{cases} \mu_1 + \tau_1 \; T_i & \text{before IA} \\ \\ \mu_2 + \tau_2 \; T_i & \text{after IA} \end{cases}$$

  with treatment group indicator T_i

- **assumptions**

  – variance homogeneity across design stages

Testing for Heterogeneity across Design Stages

- **hypotheses**: $H_0 : \tau_1 = \tau_2$ vs. $H_1 : \tau_1 \neq \tau_2$

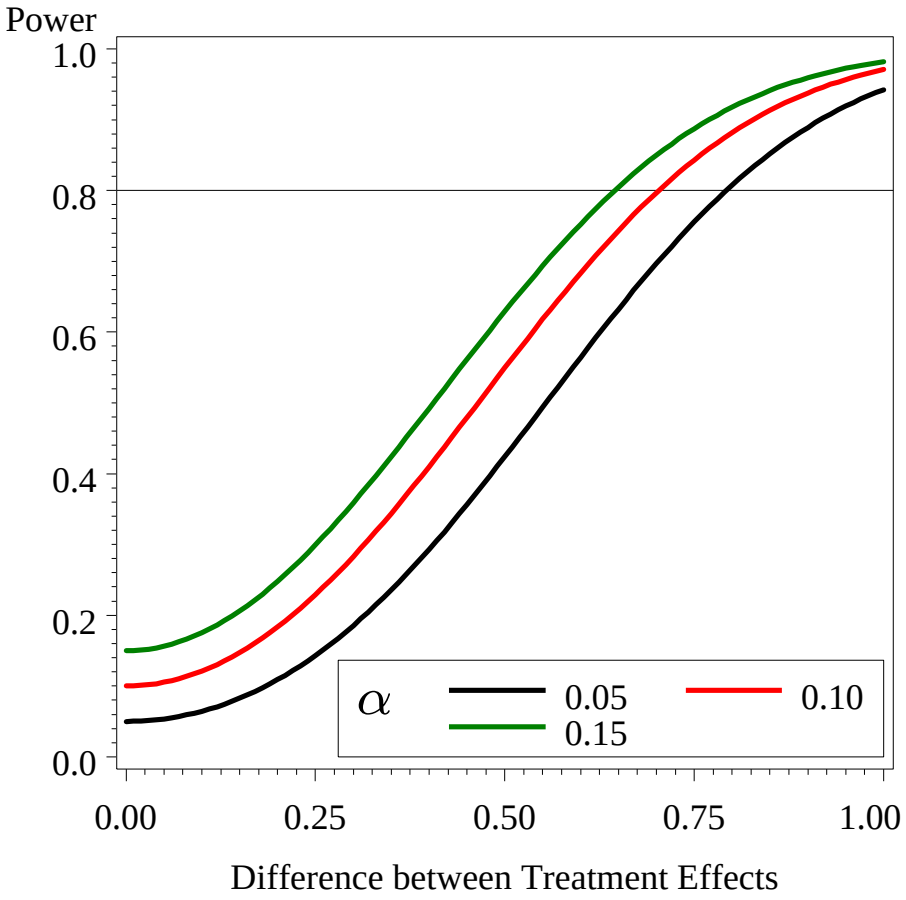
- **hypothesis test**: likelihood ratio test (LRT)

  - test statistic: $-2 \log \Lambda = \dfrac{(\hat{\tau}_1 - \hat{\tau}_2)^2}{4\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

  - if σ^2 unknown, estimated by MLE

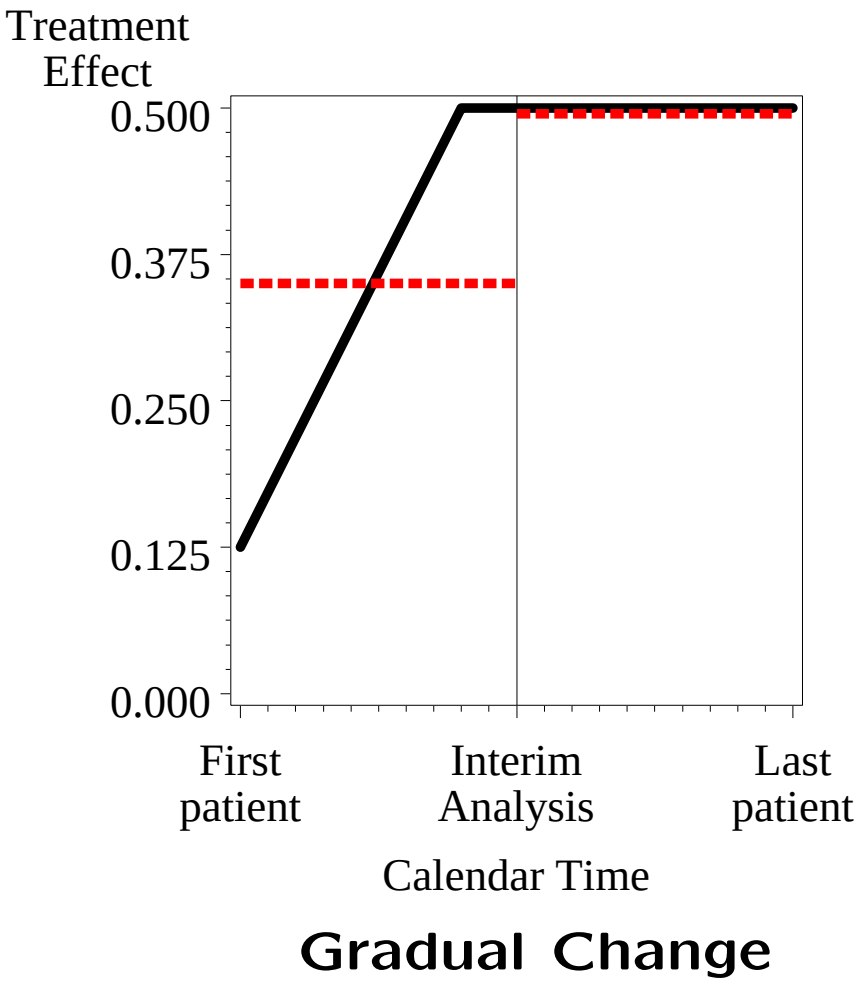
  - critical value from χ^2 distribution with 1 df (exact if σ^2 known)

Power of Heterogeneity Test

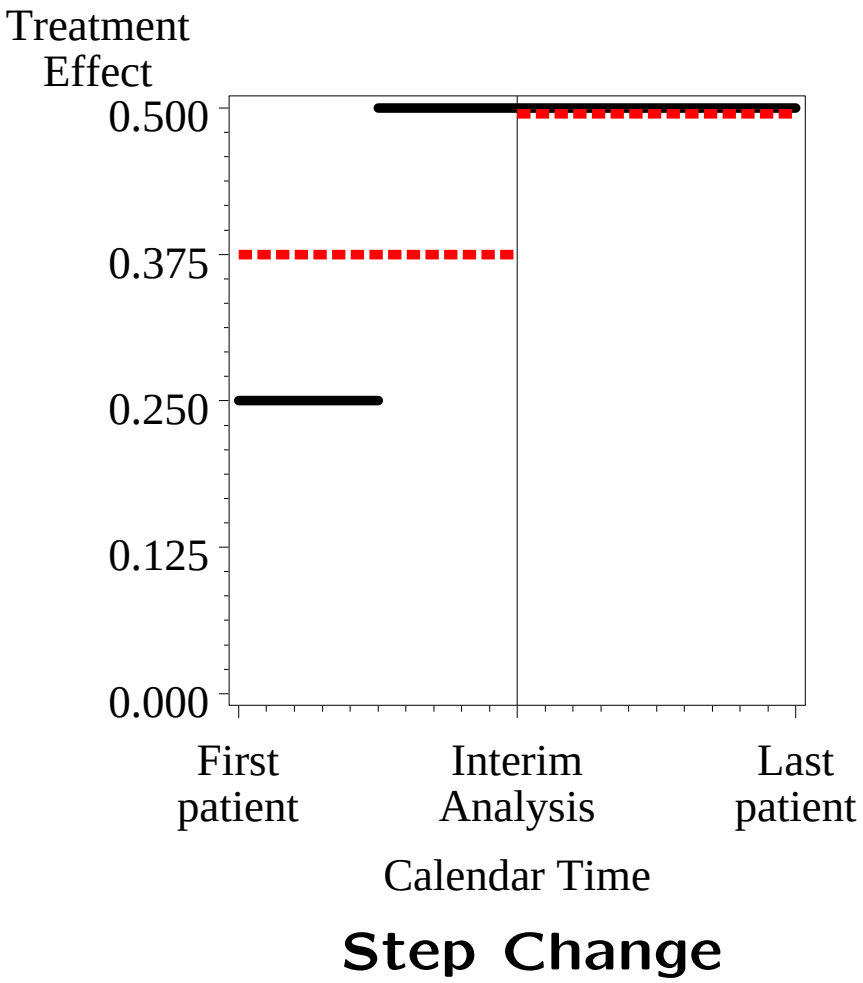


- **example**
  - 100 patients per stage
  - one-sided sign.level 0.025
  - std. treatment effect of 0.5 (0.4)
  - power of test for treatment effect 95% (80%)
- in MA **usually** $\alpha > .05$ **used**
  - e.g. Cochrane Handbook suggests $\alpha = 0.10$

Heterogeneity Test Confounded by Calendar Time

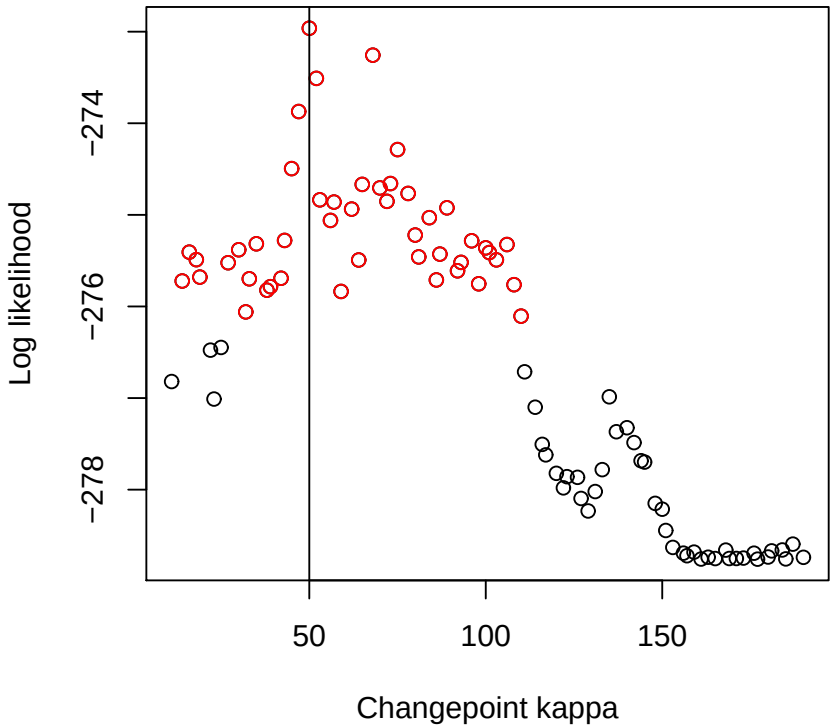


Gradual Change



Step Change

A Simulated Trial: Motivating the Use of CP Methods



Significant LRT in red.

- **considered situation**
  - 100 patients per stage
  - step change at $\kappa = 50$
  - $\tau_1^{(S)} = 0.25$, $\tau_2^{(S)} = 0.75$
- **heterogeneity test**: $p = 0.01$
- **changepoint methods**
  - search for maximum over LRT stats
  - correct critical value / p-value (e.g. Hansen 1997)
  - 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 changepoint 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 $\widehat{\kappa}$ and the second stage at level $\alpha_{2.2}$.

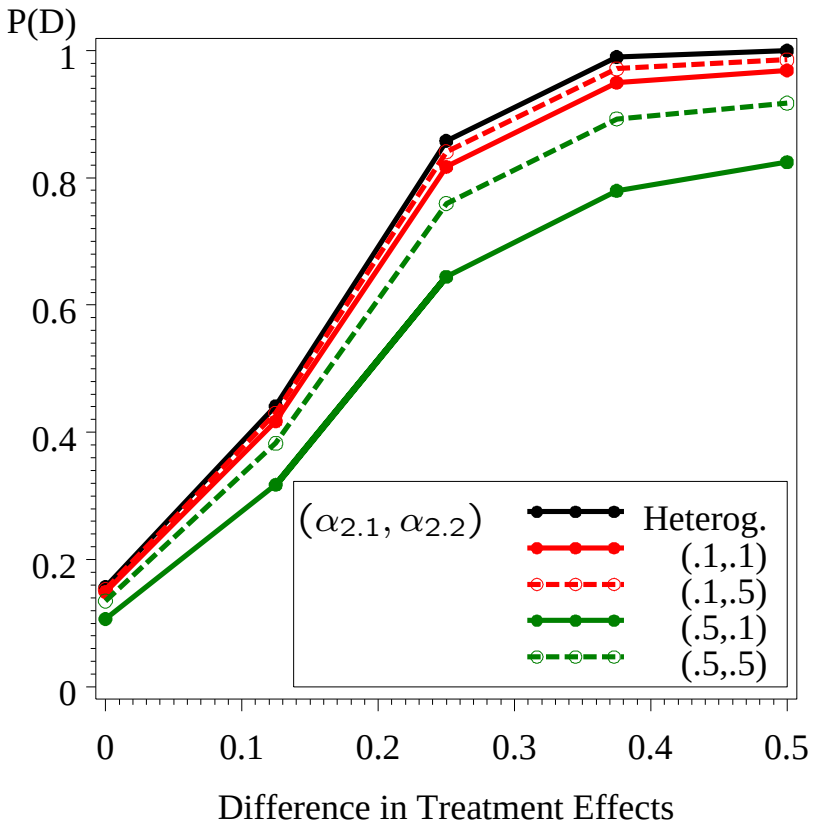
  – if (not) significant, then conclude "change (not) due to IA"

Simulation Study

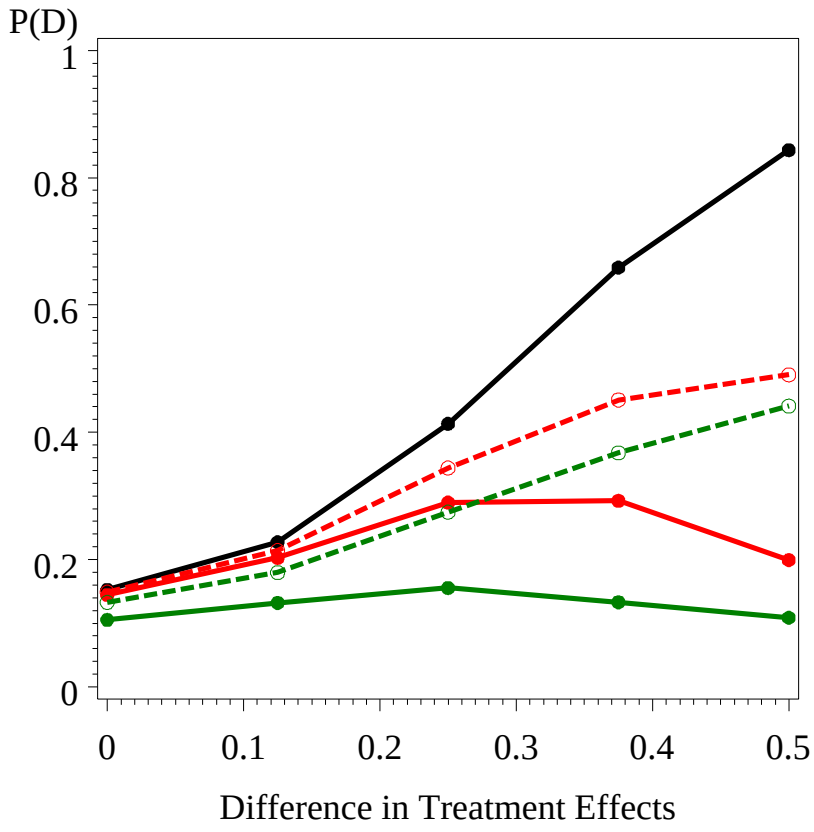
- **number of replications**: 10,000 trials per situation
- 200 patients in total with IA after 100 patients
- **step change** at κ of size $\Delta = \tau_2^{(S)} - \tau_1^{(S)}$
  - treatment effect $\tau = (\tau_1^{(S)} + \tau_2^{(S)})/2 = 0.5$
  - change in treatment effect $\Delta = 0, 0.125, \ldots, 0.5$
  - changepoint $\kappa = 50, 75, 100, 125$

Simulated Probability of "Change due to IA" Conclusion

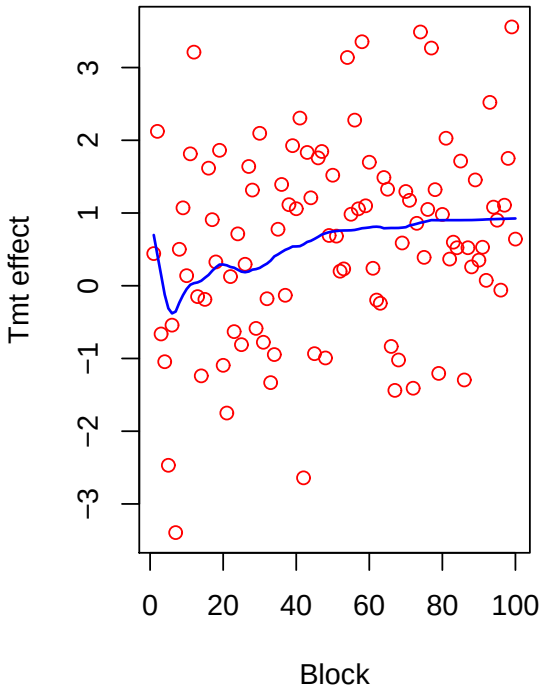
$\kappa = 100$



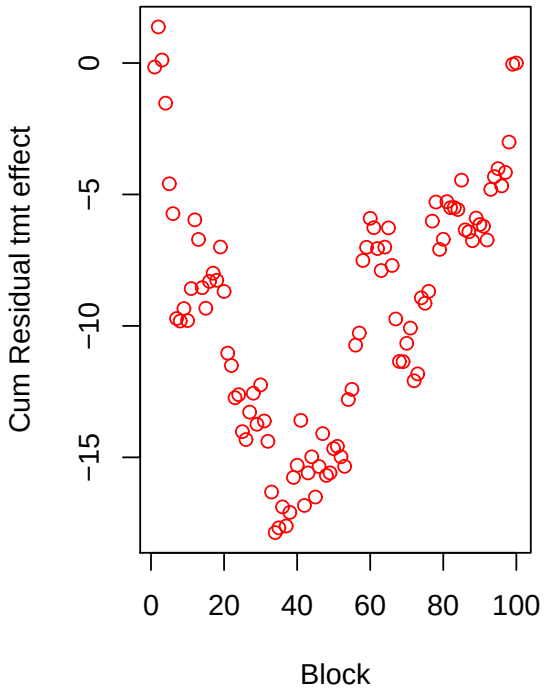
$\kappa = 50$



Estimating the treatment effect as a function of calendar time



Smoother



Cum. Residuals

- **simulated example**
  - as above
  - heterogeneity test at IA: $p = 0.01$
  - cp test: significant with $\widehat{\kappa} = 50$
- **smoother**: treatment effect estimated in blocks of 2, then smoothed
- **conclusion**: difficult to pick up trends in estimated treatment effect

Discussion

- **optimal choice** of $\alpha_{2.1}$ ($\alpha_{2.2}$ might be set to α_1)
- **gradual change** (rather than step change)
- designs with **more than 2 stages**
- adaptations create **dependence**: ignored here
- **two-sided heterogeneity tests** used, but problem is one-sided
  - problematic: effect of stage 2 larger than of stage 1

Conclusions

- **heterogeneity test approach** wasteful (therefore unethical)
  - even with no change over time: studies reduced in size / discarded with probability α
  - calendar time effects unrelated to IA lead to far higher probabilities
- **alternative approach** favourable, but doesn't solve the issues
- **estimation** seems difficult given limited data
- **design**: careful consideration and discussion in planning phase (like with cary-over effects in crossover trials)

References

- Committee for Medicinal Products for Human Use (2006) Reflection paper on methodological issues in confirmatory clinical trials with flexible design and analysis plan. London, 23 March 2006, Doc. Ref. CHMP/EWP/2459/02.

- Friede T, Henderson R (2003) Intervention effects in observational studies with an application in total hip replacements. Statistics in Medicine 22: 3725-3737.

- Friede T, Henderson R, Kao CF (2006) A note on testing for intervention effects on binary responses. Methods of Information in Medicine 45: 435-440.

- Hansen BE (1997) Approximate asymptotic p values for structural-change tests. Journal of Business & Economic Statistics 15: 60-67.

- 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.

- Cochrane Handbook for Systematic Reviews of Interventions 4.2.6. http://www.cochrane.org/resources/handbook/hbook.htm.

- Whitehead A (2002) Meta-analysis of controlled clinical trials. Wiley.

Backup Slides

Two-stage Adaptive Designs

- **design stages**
  - first (second) stage: patients recruited before (after) IA
- **interim analysis**
  - hypothesis test
  - adaptations: sample size adjustment, treatment selection, ...
- **final analysis**: combining information across design stages

Meta-Analysis: Testing for Heterogeneity across Studies (Whitehead, 2002)

- **fixed effect model** (r studies): $\widehat{\theta}_i \sim N(\theta, w_i^{-1})$, $i = 1, \ldots, r$

- **test statistic**: $Q = \sum_i w_i(\widehat{\theta}_i - \widehat{\theta})^2$ with $\widehat{\theta} = \dfrac{\sum_i w_i \widehat{\theta}_i}{\sum_i w_i}$

  for $r = 2$ studies: $Q = \dfrac{w_1 w_2}{w_1 + w_2}(\widehat{\theta}_1 - \widehat{\theta}_2)^2$

- **reference distribution**: if $E(\widehat{\theta}_i) = \theta$ for all i, then $Q \sim \chi^2_{r-1}$

- **note**: variance of $\widehat{\theta}_i$ estimated $\Rightarrow$ large-sample test

Heterogeneity across Design Stages

- **model**: independent responses with $Y_i \sim \mathsf{N}(\mu_i, \sigma^2)$

$$\mu_i = \begin{cases} \mu_1 + \tau_1 \; T_i & \text{before IA} \\ \\ \mu_2 + \tau_2 \; T_i & \text{after IA} \end{cases}$$

  with treatment group indicator T_i

- **assumptions**

  – variance homogeneity across design stages

Simulated Probability of "Change due to IA" Conclusion – Gradual Change up to κ ($\Delta = 0.5$) –

		P(D)	
α_{21}	α_{22}	$\kappa = 25$	$\kappa = 50$
Heterog.		0.224	0.425
0.1	0.1	0.157	0.217
0.1	0.5	0.195	0.312
0.5	0.1	0.087	0.101
0.5	0.5	0.167	0.258