Estimation in Flexible Adaptive Designs

Werner Brannath
Section of Medical Statistics
Core Unit for Medical Statistics and Informatics
Medical University of Vienna

BBS and EFSPI

Scientific Seminar on
Adaptive Designs in Drug Development

Basel, June 2007

Contents

Review of Flexible Two Stage Tests

Flexible designs

Flexible designs allow for mid-trial design modifications based on all internal and external information gathered at interim analyses without compromising the type I error rate.
For a control of the type I error rate, the design modifications need **not** be specified in advance.

Examples for mid-trial design modifications:

Adaptation of the sample sizes, dropping (or adding) study doses, adapting the number of interim analyses, decision boundaries, test statistics, the endpoints, the study goal (non-inferiority and superiority), the multiple testing strategy, ...

Pre-specified Adaptivity versus Flexibility

Pre-specified adaptivity =

adapting design parameters according to a pre-specified adaptation rule

Aims: Increasing efficiency by optimizing specific cost functions.

Examples: Group sequential trials, play-the-winner allocation rules, . . .

Flexibility ("unscheduled" adaptivity) =

adapting design parameters without a (complete) specification of the adaptation rule

Flexibility

Aims of flexibility:

- Dealing with the unexpected.
- Dealing with the expected unpredictability of clinical trials.
- Improving the “quality” of the decision process as a whole in an environment where the parameter assumptions and also the *weighting of gains and costs are unclear* a priori and can change in the course of the trial.

Flexible Two Stage Tests

Step-wise procedure

Consist of two sequential stages:
Stage 1 (e.g. Phase II part) and *Stage 2* (e.g. Phase III part)

Stage 1 and Stage 2 data are from two independent cohorts.

Adaptivity

The design of Stage 2 (sample sizes, statistical test, . . .) can be chosen based on the data of Stage 1 as well as any other internal or external information.

Flexibility

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.

Flexible two stage combination tests

Notation: p and q the p-values from stage 1 and 2 for

$$H_0 : \theta \leq 0 \quad \text{versus} \quad H_1 : \theta > 0$$

p and q are independent under H_0.

Two stage combination test: Prefix a monotone combination function $C(p, q)$ and rejection bounds c and α_1.

$$\text{We reject } H_0 \text{ if} \quad \begin{array}{lrl} \text{either} & p \leq \alpha_1 & \text{(stage 1)} \\ \text{or} & C(p, q) \leq c & \text{(stage 2)} \end{array}$$

Level condition: We must prefix α_1, $C(p, q)$ and c such that

$$P_0(\{p \leq \alpha_1\} \cup \{C(p, q) \leq c\}) = \alpha$$

Examples for combination functions

- Fisher's product test:

$$C(p,q) = p \cdot q$$

- Inverse normal method:

$$C(p,q) = w_1\,\Phi^{-1}(p) + w_2\,\Phi^{-1}(q), \quad w_1^2 + w_2^2 = 1$$

Gives a two stage GSD with information times $t_1 \leq t_2$ if $w_1/w_2 = \sqrt{t_1/(t_2 - t_1)}$ and no adaptations are done.

Estimation

"Corresponding methods to estimate the size of the treatment effect and to provide confidence intervals with pre-specified coverage probability are additional requirements."

Reflection paper on flexible designs (Draft), EMEA 2006

Bias of conventional estimates

- Unblinded sample size adaptations may lead to (mean) biased estimates and invalid confidence intervals.
- Sample size adaptation rule unkown
  $\rightarrow$ bias and coverage probabilities **unknown.**

Confidence intervals for flexible designs

- Duality between confidence sets and significance tests:

  confidence set $=$ set of values where significance test accepts

- Flexible confidence interval:

  Use flexible tests at level α for all parameter values

- Flexible confidence intervals have coverage probability $\geq 1 - \alpha$ independently of the adaptation rule.

- Overall p-values can be constructed in a similar way.

Confidence intervals for flexible designs accounting for stopping rules

Two possible approaches:

- *Repeated confidence interval approach:*
  is very simple; flexibility also with regard to stopping rule; inevitable price is strict conservatism.

- *Exact confidence interval via stage wise ordering:*
  is more complicated; no flexibility with regard to stopping rule; exact coverage probability.

  Exact confidence interval at level 0.5 gives median unbiased point estimate which lies in the interior of the exact 95%-confidence interval.

Confidence intervals for flexible designs accounting for stopping rules

Two possible approaches:

- *Repeated confidence interval approach:*
  is very simple; flexibility also with regard to stopping rule; inevitable price is strict conservatism.

- *Exact confidence interval via stage wise ordering:*
  is more complicated; no flexibility with regard to stopping rule; exact coverage probability.

  Exact confidence interval at level 0.5 gives median unbiased point estimate which lies in the interior of the exact 95%-confidence interval.

Chapter 2: Repeated Confidence Intervals

(Lehmacher & Wassmer, 1999; Brannath et al. 2002, Lawrence & Hung, 2003; Proschan et al., 2003)

Repeated confidence intervals

Duality between hypothesis tests and confidence sets:

p_Δ stage 1 and q_Δ stage 2 p-values for $H_{0,\Delta} : \theta \leq \Delta$.

p_Δ and q_Δ independent under $H_{0,\Delta}$ and increasing in Δ.

Two stage combination test for $H_{0,\Delta}$: Use for all Δ the same combination test.

We reject $H_{0,\Delta}$ if either $p_\Delta \leq \alpha_1$ (stage 1)

or $C(p_\Delta, q_\Delta) \leq c$ (stage 2)

Remark: The rule "$p_\Delta \leq \alpha_1$" should *not* be understood as a *stopping rule*, but as *rejection rule* which we apply at stage 1.

Lower repeated confidence bounds

Stage 1: Solve the equation $p_\Delta = \alpha_1 \quad \rightarrow \quad \delta_1$ such that

$$p_\Delta \leq \alpha_1 \quad \Longleftrightarrow \quad \Delta \leq \delta_1$$

$\rightarrow \quad (\delta_1, \infty)$ one-sided confidence interval at first stage.

Stage 2: Solve $C(p_\Delta, q_\Delta) = c \quad \rightarrow \quad \delta_2$ such that

$$C(p_\Delta, q_\Delta) \leq c \quad \Longleftrightarrow \quad \Delta \leq \delta_2$$

$\rightarrow \quad (\delta_2, \infty)$ one-sided confidence interval at second stage.

Example I

Primary efficacy end point: Infarct size measured by the cumulative release of α-HDBH within 72 hours after administration of the drug (area under the curve, AUC).

θ the mean α-HDBH AUC difference between control c and treatment t, $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$

Inverse normal combination test:

$$C(p, q) = \sqrt{0.5} \cdot \Phi^{-1}(p) + \sqrt{0.5} \cdot \Phi^{-1}(q)$$

O'Brien and Flemming at one sided level $\alpha = 0.025$

$\rightarrow \alpha_1 = 0.0026$, $c = 0.024$.

Example I (cont.)

Stage 1: sample sizes: $n_{1c} = 88$, $n_{1t} = 91$,
standard deviation: $\hat{\sigma}_{1c} = 26.0$, $\hat{\sigma}_{1t} = 22.5$
treatment difference: $\hat{\theta}_1 = 4.0$, $\sigma_{\hat{\theta}_1} = 3.64$

p_Δ according to t-test for $H_0 : \theta = \Delta$.

Solving $p_\Delta = 0.0026 \quad \longrightarrow \quad$ classical CI at level 0.0026

$$\delta_1 = \hat{\theta}_1 - t_{\nu,0.9974} \cdot \sigma_{\hat{\theta}_1} = -6.3$$

First stage confidence interval is $(-6.3, \infty)$

Example I (cont.)

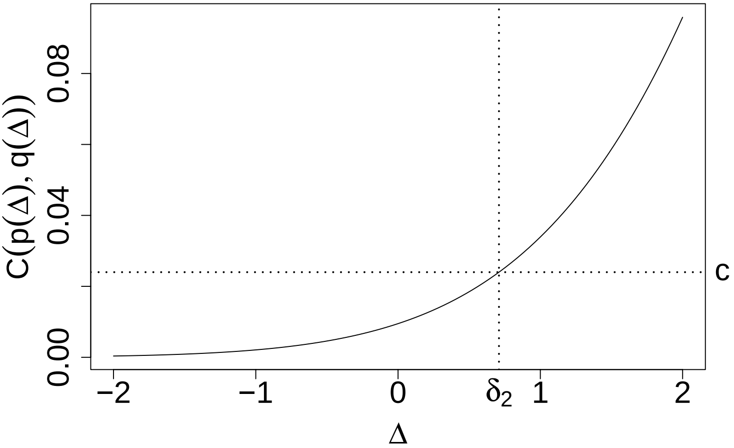
Stage 2: sample sizes $n_{2c} = 322$, $n_{2t} = 321$,
standard deviations: $\hat{\sigma}_{2c} = 26.1$, $\hat{\sigma}_{2t} = 28.5$,
treatment difference: $\hat{\theta}_2 = 4.8$, $\sigma_{\hat{\theta}_2} = 2.16$

q_Δ according to t-test for $H_0 : \theta = \Delta$ from second stage data.

Solving $C(p_\Delta, q_\Delta) = 0.024$ numerically $\longrightarrow$ $\delta_2 = 0.71$

Second stage confidence interval is $(0.71, \infty)$

Example I (cont.): Determination of δ_2



Properties of repeated confidence bounds

- One need not pre-specify the adaptation and stopping rule to keep the nominal coverage probability.
- Price for the flexibility with regard to stopping rule is strict conservatism: we must control the level for the worst case rule, also when actually not following this rule.
- H_0 is rejected with the combination test iff $\delta_L > 0$.
- The first stage bound δ_1 is the classical confidence bound at level α_1.

Normal approximations and inverse normal method (Lehmacher and Wassmer, 1999)

If the stage wise estimates $\hat{\theta}_i$ ($i = 1, 2$) for the treatment effect are asymptotically independent and normal with mean treatment effect Δ and variance $\sigma^2_{\hat{\theta}_i} = I_1^{-1}$, then

$$p(\Delta) = 1 - \Phi(\sqrt{I_1} \cdot (\hat{\theta}_1 - \Delta)) \quad \text{and} \quad q(\Delta) = 1 - \Phi(\sqrt{I_2} \cdot (\hat{\theta}_2 - \Delta))$$

are approximate independent p-values for $H_{0,\Delta} : \theta \leq \Delta$.

Inverse normal combination function

$$\delta_2 = \widehat{\theta}_w - \frac{\Phi^{-1}(1-c)}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_1}}, \quad \widehat{\theta}_w = \frac{w_1 \cdot \sqrt{I_1} \cdot \hat{\theta}_1 + w_2 \cdot \sqrt{I_2} \cdot \hat{\theta}_2}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_2}}$$

Example I with normal approximation

Stage 1: $\hat{\theta}_1 = 4.0, \quad I_1 = 0.076$

$$\delta_1 = 4.0 - \Phi^{-1}(0.9974) \cdot \sqrt{I_1} = -6.2 \quad \text{(before } -6.3\text{)}$$

Stage 2: $\hat{\theta}_2 = 4.8, \quad I_2 = 0.215, \quad w_1 = w_2 = \sqrt{0.5}$

$$\widehat{\theta}_w = \frac{\sqrt{I_1} \cdot \hat{\theta}_1 + \sqrt{I_2} \cdot \hat{\theta}_2}{\sqrt{I_1} + \sqrt{I_2}} = 4.5$$

$$\delta_1 = 4.5 - \frac{\Phi^{-1}(1 - 0.024)}{(\sqrt{I_1} + \sqrt{I_2})\sqrt{0.5}} = 0.70 \quad \text{(before } 0.71\text{)}$$

Extensions

- Repeated confidence intervals can be extended to multistage flexible designs, and can be computed even after adapting the number of interim looks

  (Lehmacher and Wassmer, 1999; Brannath et al., 2002)

- One can incorporate a futility boundary into the dual combination tests. However, one must carefully account for the futility bound in the determination of δ_2:

  One must accept all Δ for which stage 1 p-value p_Δ falls into stage 1 acceptance region even if the second stage data suggest rejection of $H_{0,\Delta}$.

Extensions

- We could use different α_1 and c for different Δ, however, one must be careful in our choice in α_1 and c in order to get nested dual rejection regions. (BRANNATH ET AL. 2003)
- Exact confidence intervals and median unbiased point estimates are available via the stage wise ordering. (BRANNATH ET AL. 2002)
- Confidence intervals and point estimates for adaptive GSD's following the principle of Müller and Schäfer have been derived only recently. (METHA, BRANNATH, POSCH AND BAUER 2006; SUBMITTED)

Two-sided tests and two-sided confidence intervals at level 2α

- One should not perform combination tests with two-sided p-values for $H_{0,\Delta} : \theta = \Delta$:

  Interpretation problem if the first and the second stage estimates point in conflictive directions.
- *Solution:* Intersection of two one-sided combination tests and corresponding repeated confidence intervals (one lower and one upper) each at level α.

Two-sided confidence intervals at level 2α

- At the first stage we get the classical two-sided interval at level $2\alpha_1$.
- With the normal approximation and normal inverse method we get at the second stage the interval

$$\left(\widehat{\theta}_w - \frac{\Phi^{-1}(1-c)}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_2}}, \widehat{\theta}_w + \frac{\Phi^{-1}(1-c)}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_2}}\right)$$

with

$$\widehat{\theta}_w = \frac{w_1 \cdot \sqrt{I_1} \cdot \hat{\theta}_1 + w_2 \cdot \sqrt{I_2} \cdot \hat{\theta}_2}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_2}}$$

Confidence intervals for the conditional error function approach

Conditional error function approach: Prefix a decreasing conditional error function $A(x)$ and first stage rejection level α_1.

$$\text{Reject } H_0 \text{ if } \quad p \le \alpha_1 \text{ (stage 1)} \quad \text{or} \quad q \le A(p) \text{ (stage 2)}.$$

Equivalent combination test (POSCH & BAUER 1999, WASSMER 1999):

$$e.g. \qquad \alpha_1, \quad C(p,q) = q - A(p), \quad \text{and} \quad c = 0$$

$\rightarrow$ One can use the same estimation methods as for combination tests

Chapter 3: Point Estimation

Maximum likelihood estimate (MLE)

Assuming normal data and balanced treatment groups the MLE can be written as

$$\hat{\theta}_{mle} = \frac{I_1}{I_1 + I_2} \cdot \hat{\theta}_1 + \frac{I_2}{I_1 + I_2} \cdot \hat{\theta}_2$$

(for small effect sizes approximatively also in other cases)

Mean Bias: $E_\Delta(\hat{\theta}_{mle} - \Delta) = Cov_\Delta(\frac{I_1}{I_1+I_2}, \hat{\theta}_1)$ (Liu et al. 2002)

One can show that always: $|E_\Delta(\hat{\theta}_{mle} - \Delta)| \leq 0.4 \cdot \sigma/\sqrt{n_1}$

Variance also depends on (unknown) adaptation/selection rule

Maximum likelihood estimate (MLE)

Mean bias of MLE for typical examples (qualitatively):

- *Stopping with early rejection:* the larger the effect size the smaller the sample size $\rightarrow$ positive mean bias.
- *Stopping for futility:* the smaller the effect size the smaller the sample size $\rightarrow$ negative mean bias.
- *Conditional or predictive power control:* the smaller the effect size the larger the sample size $\rightarrow$ positive mean bias.
- *Selecting promising treatments:* the larger the effect size the larger the sample size $\rightarrow$ negative mean bias.

Weighted maximum likelihood estimate

Center of a two sided repeated confidence interval:

$$\widehat{\theta}_w = \frac{w_1 \cdot \sqrt{I_1} \cdot \hat{\theta}_1 + w_2 \cdot \sqrt{I_2} \cdot \hat{\theta}_2}{w_1 \cdot \sqrt{I_1} + w_2 \cdot \sqrt{I_2}}$$

where $w_1, w_2 \geq 0$, $w_1^2 + w_2^2 = 1$ are the pre-specified weights.

Properties:

- If recruitment is stopped at stage 1 then $\widehat{\theta}_w = \hat{\theta}_1$.
- If recruitment is never stopped at the interim analysis, then $\hat{\theta}_w$ is *median unbiased*, i.e., $\hat{\theta}_w$ has median Δ.
- Median of $\hat{\theta}_w$ close to Δ also if recruitment can be stopped at interim analysis.

Cases for which the estimates are similar

The two estimates are equal or differ only slightly if

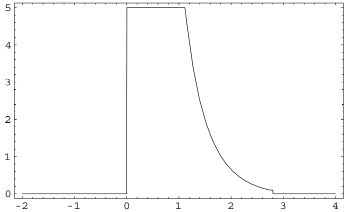
- recruitment is stopped at the interim analysis;
- recruitment is **not** stopped at the interim analysis, and
    - the first and second stage estimates are similar, $\hat{\theta}_1 \approx \hat{\theta}_2$;

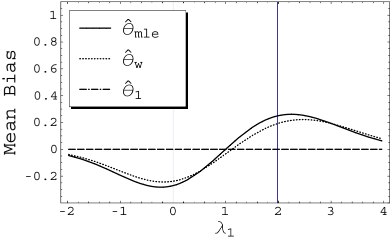
  or
    - the sample sizes are (almost) as pre-planned: $\sqrt{I_1/I_2} \approx w_1/w_2 = \sqrt{t_1/(t_2 - t_1)}$

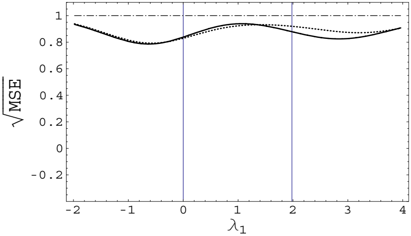
Numerical example

80% - Predictive power rule,
truncated $0.1 \cdot I_1 \leq I_2 \leq 5 \cdot I_1$.

$\widehat{\theta}_w$ with $w_1^2 = 0.5$
$\hat{\theta}_1$ first stage mean diffenrence
horizontal axis: $\lambda_1 = \sqrt{I_1} \cdot \theta$







Remarks on Seamless Phase II/III designs

- Start with a number of treatments (doses) and *select treatments* and reassess sample sizes at an adaptive interim analysis.
- Selection rule is typically **not** fully pre-specified.
- Flexible closed tests provide strong FWER control.

- Repeated confidence interval approach provides univariate confidence intervals (no multiplicity adjustment).
- Simultaneous confidence intervals which are consistent with the (multiple) test result may **not** be available.
- Mean or median unbiased estimates are currently **not** available.
- Selection bias can be an additional issue (ongoing research and discussion).

Summary

- Univariate confidence intervals are, in general, available for flexible adaptive designs.
- Using the normal approximation of stage wise estimates and the inverse normal combination function, we get explicit (and intuitive) formula for the confidence bounds.
- Maximum likelihood estimate is biased, however, seems to perform well in terms of the mean square error.
- The *weighted maximum likelihood estimate* is, in general, less biased and median unbiased in the case of an administrative interim look.
- With a stopping rule a median unbiased estimate can be obtained via the stage wise ordering.

Selected literature

- Brannath, König, Bauer (2006). Estimation in flexible two stage designs, *Statistics in Medicine* **25**: 3366–3381.
- Posch, Koenig, Branson, Brannath, Dunger-Baldauf, Bauer (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection, *Statistics in Medicine* **24**: 3697–3714.
- Brannath, Maurer, Posch and Bauer (2003). Sequential tests for non-inferiority and superiority. *Biometrics* **59**:106–114.
- Brannath, König and Bauer (2003). Improved repeated confidence bounds in trials with a maximal goal. *Biometrical Journal* **45**:311–324.
- Proschan, Liu, Hunsberger (2003). Practical midcourse sample size modification in clinical trials, *Controlled clinical trials* **24**:4–15.
- Lawrence and Hung (2003). Estimation and confidence intervals after adjusting the maximum information, *Biometrical Journal* **45**:143–152.
- Liu, Proschan and Pledger (2002). A unified theory of two-stage adaptive designs, JASA **97**:1034–1041.
- Brannath, Posch and Bauer (2002). Recursive combination tests. *JASA* **97**:236–244.
- Lehmacher and Wassmer (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* **55**:1286–1290.