Testing and estimation in adaptive group sequential designs with treatment selection

Martin Posch

Section of Medical Statistics
Medical University of Vienna, Austria

2008 Adaptive Design KOL Lecture Series
PhRMA Adaptive Design and Adaptive Dose-Ranging Studies Working Groups
January 11, 2008
Adaptive Phase II/III Trials

- Phase II trial is **internal part of a combined trial**.
- The design of the phase III part depends on data of the phase II part.
- Efficacy is demonstrated with data from **phase II + III part**.
Motivation

Exploratory and Confirmatory Phase in a Single Trial

Adaptive Phase II/Phase III Designs

Integrating both phases in a single trial

- saving of sample size
  Patients from the exploratory phase are included also in the confirmatory analysis.

- saving of time
  Preparation time for a second trial is spared.

- adds complexity
Regulatory Requirements for the Analysis of Adaptive Trials

“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.”

EMEA REFLECTION PAPER ON METHODOLOGICAL ISSUES IN CONFIRMATORY CLINICAL TRIALS PLANNED WITH AN ADAPTIVE DESIGN (2007).
Motivation

Types of Adaptive Designs

**Designs with completely predefined adaptation rule**

- The type I error rate is computed by simulation or numerical integration. Therefore, the adaptation rules
  - may depend only on the primary outcome variable
  - and must be strictly adhered to.
- Examples:
  - Designs with sample size reassessment rules
  - Bayesian adaptive designs

**Fully Flexible Designs Based on Combination Tests**

For the control of the Type I error rate

- the adaptation rule needs not to be preplanned.
- adaptations may depend on all interim data including secondary and safety endpoints.
However, in confirmatory trials:

“In all instances the interim analysis and the type of the anticipated design modification (change of sample size, discontinuation of treatment arms, etc.) would need to be described and justified in the study protocol.”

EMEA REFLECTION PAPER (2007)
Flexible Designs Based on Combination Tests
The Combination Test

First Stage

\[ p \]

\[ 0 \to \alpha_1 \]

Reject \( H_0 \)

\[ \alpha_1 \to \alpha_0 \to 1 \]

Accept \( H_0 \)

Adaptation

Second Stage

\[ C(p,q) \]

\[ 0 \to c \]

Reject \( H_0 \)

\[ c \to 1 \]

Accept \( H_0 \)

**References**

Design Adaptations with Combination Tests

- The combination function and stopping boundaries have to be laid down a priori.
- Examples of combination functions:
  - Fisher: \( C(p, q) = pq \)
  - Inverse Normal
    \[ C(p, q) = -w_1 \Phi^{-1}(1 - p) - w_2 \Phi^{-1}(1 - q) \]

Type I error control

The type I error is controlled if, under \( H_0 \), the p-values \( p \) and \( q \) are independent and uniformly distributed on [0,1].

This holds if, e.g., new patients are recruited for the second stage and conservative stage wise tests for \( H_0 \) are used.
Multiple Testing

If several treatments are compared in an adaptive trial the familywise error rate has to be controlled.

Adaptive multiple testing procedures

- Combine adaptive tests with the closed testing procedure
- Control the familywise Type I error rate in the strong sense

The Closure Principle

- Consider the null hypotheses $H_1, H_2$
- Define local level $\alpha$ tests for

$$H_1, \quad H_2, \quad H_{12} = H_1 \cap H_2.$$

The test at multiple level $\alpha$:
Reject $H_j, j = 1, 2$ if

$$H_{12} \text{ and } H_j \text{ are rejected at local level } \alpha.$$

This procedure controls the familywise error rate at level $\alpha$ in the strong sense.  

MARCUS ET AL., 1976
Adaptive Closed Tests

The elementary and intersection hypotheses are tested with combination tests.

Bauер & Kieser ’99, Hommel ’01, Posch et al. ’05

Example: Adaptive test of $H_1, H_2$ at multiple level $\alpha$ (I)

- Assume $\alpha_1 = 0, \alpha_0 = 1$ (no early stopping).
- $p_i, q_i \ldots$ (unadjusted) first and second stage p-values for $H_i$, $i = 1, 2$. 
Adaptive test of $H_1, H_2$ at multiple level $\alpha$ (II)

Test of the intersection hypothesis

- First stage p-value for $H_{12}$ (e.g., Bonferroni test)
  \[ p_{12} = \min\{2 \min(p_1, p_2), 1\} \]

- Second stage p-value for $H_{12}$
  - If both hypotheses are selected:
    \[ q_{12} = \min\{2 \min(q_1, q_2), 1\} \]
  - If only $H_i$ is selected: $q_{12} = q_i$

- Final Analysis: Reject $H_{12}$ if $C(p_{12}, q_{12}) \leq c$.

Adaptive Closed Test

Reject $H_i$ if $H_1 \cap H_2$ is rejected and $C(p_i, q_i) \leq c$. 
Comments

Flexibility

- The selection of hypotheses for the second stage may depend on the interim data in any way.
- The selection rule needs not to be specified in detail.
- The familywise error rate is controlled.

Extensions

- Instead of Bonferroni adjusted p-values, adjusted p-values of, e.g., Šidak, Dunnett, trend, likelihood ratio, or Hochberg tests can be used.
- Inclusion of new hypotheses
- Other adaptations (besides dropping of hypotheses)
Example

- Two experimental treatments, one control
- Parallel groups, treatment control comparisons
- Normally distributed endpoints with means $\mu_0$ (control) and $\mu_1$, $\mu_2$ (experimental treatments)
- The variance $\sigma^2$ is assumed to be known.
- We test the hypotheses

\[
H_{0,1} : \mu_1 - \mu_0 = 0 \quad H_{1,1} : \mu_1 - \mu_0 > 0
\]

\[
H_{0,2} : \mu_2 - \mu_0 = 0 \quad H_{1,2} : \mu_2 - \mu_0 > 0
\]
Allocation of Sample Size

Total Sample $n_T$
Allocation of Sample Size

Phase II Part

- Treatment 1
- Treatment 2
- Control

Total Sample $n_T$

$r \cdot n_T$
Allocation of Sample Size

Phase II Part

- Treatment 1
- Treatment 2
- Control

Phase III Part

- Selected Treatment
- Control

\[ r \cdot n_T \]
A Numerical Example

- Total sample size $n_T$:  
  *We fix the total sample size corresponding to that of a confirmative two armed trial with 90% power and effect size $\Delta$.*

- Sample size of phase II part: $r n_T$

- Sample size of phase III study: $(1 - r) n_T$

- Select treatment with higher interim estimate for the confirmatory part.

- One sided level: 0.025.

- Hypothesis testing with the adaptive combination test and the Šidak test for intersection Hypotheses.
Combination function: “inverse normal”

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

where $\Phi^{-1}$ denotes the quantiles of the standard normal distribution, and $w$ gives the weight of the exploratory part.

- $w = 0$
  Classical Phase II / Phase III design.
- $0 < w < 1$
  Adaptive Design integrating both phases.
- $w = 1$
  The pilot phase is already confirmatory.
Power for the Selected Treatment

\[ \mu_1 - \mu_0 = 0.0\Delta, \quad \mu_2 - \mu_0 = \Delta \]
**Power for the Selected Treatment**

\[ \mu_1 - \mu_0 = 0.2\Delta, \quad \mu_2 - \mu_0 = \Delta \]

- **r**: Proportion of Total Sample Size in Pilot Study
- **w**: Weight of Pilot Study

\[ \mu_1 - \mu_0 = 0.2\Delta, \quad \mu_2 - \mu_0 = \Delta \]

- Martin Posch (Medical University of Vienna)
Power for the Selected Treatment

$$\mu_1 - \mu_0 = 0.4\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Power for the Selected Treatment

\[ \mu_1 - \mu_0 = 0.6\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Power for the Selected Treatment

\[ \mu_1 - \mu_0 = 0.8\Delta, \quad \mu_2 - \mu_0 = \Delta \]

- \( r \): Proportion of Total Sample Size in Pilot Study
- \( w \): Weight of Pilot Study

\[ \mu_1 - \mu_0 = 0.8\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Power for the Selected Treatment

\[ \mu_1 - \mu_0 = 1.0\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Point Estimation
Univariate Estimates of all Treatments

Estimate of the Selected Treatment
Univariate Estimates

- Consider for each treatment $i$ the overall mean, $\bar{X}_i$
- The overall sample size for each treatment group $i$ is random.
- The bias and mean squared error (MSE) depend on the treatment selection rule and on $\mu_1 - \mu_2$.
- For the rule that selects the treatment with the highest interim effect estimate, the bias is given by

$$E(\bar{X}_1 - \mu_1) = -\frac{\sigma}{2\sqrt{\pi n_1}} (1 - t) \exp^{-n_1(\mu_1 - \mu_2)^2/(4 \sigma^2)},$$

where $t = n_1/(n_1 + n_2)$. 
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.0 \Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$\mu_1 - \mu_0 = 0.1\Delta$,  $\mu_2 - \mu_0 = \Delta$

Bias (% of $\Delta$) vs. $r$: Proportion of Total Sample Size in Pilot Study
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.2\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$\mu_1 - \mu_0 = 0.3\Delta$, \hspace{1em} $\mu_2 - \mu_0 = \Delta$

Bias (% of $\Delta$)

$r$: Proportion of Total Sample Size in Pilot Study

Bias (% of $\Delta$)

$r$: Proportion of Total Sample Size in Pilot Study
Univariate Bias of the Effect Estimate for $\mu_1$

$\mu_1 - \mu_0 = 0.4\Delta$, $\mu_2 - \mu_0 = \Delta$
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.5\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.6\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$\mu_1 - \mu_0 = 0.7 \Delta$, $\mu_2 - \mu_0 = \Delta$
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.8\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 0.9\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Univariate Bias of the Effect Estimate for $\mu_1$

$$\mu_1 - \mu_0 = 1.0\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Estimate of the Selected Treatment

- Typically, at the end of the trial one is only interested in the selected treatment.
- Let $S \in \{1, 2\}$ denote the index of the selected treatment. Define the selection bias:
  \[ E[(\bar{X}_S - \mu_S)] \]
  and the selection mean squared error by
  \[ E[(\bar{X}_S - \mu_S)^2] \]
Selection Bias and MSE

Consider the rule that selects the treatment with the highest interim effect estimate in the interim analysis.

- The selection bias is given by

\[ E(\bar{X}_S - \mu_S) = \frac{\sigma t}{\sqrt{\pi n_1}} \exp\left[-n_1(\mu_1 - \mu_2)^2/(4\sigma^2)\right]. \]

It has opposite sign than the univariate bias!

- The selection mean squared error is

\[ MSE_{sel} = \frac{\sigma^2}{n_1 + n_2}. \]

- One can show (POSCH ET AL. 2005) that for all selection rules that depend on the interim mean difference $\bar{X}_1 - \bar{X}_2$ the selection mean squared error is $\sigma^2/(n_1 + n_2)$.
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.0 \Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.1 \Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.2\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.3\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.4\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.5\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.6\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

$$\mu_1 - \mu_0 = 0.7\Delta, \quad \mu_2 - \mu_0 = \Delta$$
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.8\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 0.9\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias of the selected treatment’s estimate

\[ \mu_1 - \mu_0 = 1.0\Delta, \quad \mu_2 - \mu_0 = \Delta \]
Bias: More than two treatments

Example: All $\mu_i$ equal. Select the treatment with the highest interim estimate, $n_1 = n_2$. Bias in units of $\sigma/\sqrt{n_1}$.
MSE: More than two treatments

Example: All $\mu_i$ equal. Select the treatment with the highest interim estimate, $n_1 = n_2$. $\sqrt{\text{MSE}}$ in units of $\sigma/\sqrt{n_1}$.

![Graph showing the relationship between the number of treatments and the square root of MSE]

- selection MSE
- univariate MSE

Number of Treatments

$\sqrt{\text{MSE}}$
Summary & Comments

- $\bar{X}_i$: negative bias
- $\bar{X}_S$: positive bias

For fixed selection rule “select the treatment with highest interim effect”:

- Unbiased estimate for $\mu_S$ (Cohen and Sackrowitz, 1989). Also unbiased conditional on the treatment selected. MSE larger than for $\bar{X}_S$.
Simultaneous Confidence Intervals
Simultaneous Confidence Intervals

- Assume an adaptive test of the one sided hypotheses
  \[ H_i : \mu_i - \mu_0 \leq 0 \quad \text{against} \quad H'_i : \mu_i - \mu_0 > 0, \quad i = 1, 2. \]
  has been performed.
- For all parameter vectors \( \theta = (\theta_1, \theta_2) \) define
  \[ H_i(\theta_i) : \mu_i - \mu_0 \leq \theta_i \quad \text{against} \quad H'_i(\theta_i) : \mu_i - \mu_0 > \theta_i, \quad i = 1, 2, \]
  and let \( H_{12}(\theta) \) denote the corresponding intersection hypothesis.
- For each \( \theta \) denote the first stage p-value for \( H_{12}(\theta) \) by \( p_{12}(\theta) \).
A Confidence Region

Assume, that treatment 1 is selected in the interim analysis.

- Let $q_1(\theta_1)$ denote the second stage p-value for $H_1(\theta_1)$ which is also a p-value for $H_{12}(\theta)$.
- A 100%$(1 - \alpha)$ confidence region for $\mu_i - \mu_0$, $i = 1, 2$ is given by all vectors $\theta$ such that

$$C(p_{12}(\theta), q_1(\theta_1)) \geq c$$

- In general, this is not a cross product of confidence intervals!
- We need to embed the confidence region in a rectangle.
Adjusted p-values

- Define for each hypothesis stage wise adjusted p-values

\[ p_{i}^{\text{adj}}(\theta_i) = \sup_{\xi \in \mathbb{R}^2, \xi_i \leq \theta_i} p_{12}(\xi). \]

- e.g., for the Šidak test

\[ p_{i}^{\text{adj}}(\theta_i) = 1 - [1 - p_i(\theta_i)]^2. \]

- If, e.g., only Treatment 1 is continued

\[ q_{1}^{\text{adj}}(\theta_i) = q_1(\theta_1), \quad \text{and} \quad q_{2}^{\text{adj}}(\theta_2) = 1. \]
Simultaneous adaptive Confidence Bounds

The one-sided confidence intervals for $\theta_i$, $i = 1, 2$ are given by

$$I_i = \{\theta_i \mid C(p_i^{adj}(\theta_i), q_i^{adj}(\theta_i)) \geq c\}.$$
If early rejection boundaries are specified, non-trivial confidence bounds for dropped treatments can be computed.

- The confidence intervals are conservative.
- The confidence bound is consistent with the test decision if the treatment with the highest interim effect is selected. Otherwise, the confidence bound may be inconsistent with the test decision. This is not specific to flexible designs but also arises in stepwise multiple testing problems in fixed sample designs.
- The 50% confidence bounds are conservative point estimates (with non-positive median bias).
Summary

- Compared to separate Phase II and Phase III trials the adaptive design maybe
  - more powerful
  - more robust with respect to the planning assumptions
- The multiple type I error rate is controlled
- The mean bias of estimates is lower than in non-adaptive multiarmed trials.
- Conservative simultaneous confidence bounds are available
P. Bauer and M. Kieser.
Combining different phases in the development of medical treatments within a single trial.
*Statistics in Medicine, 18:1833–1848, 1999.*

F. Bretz, H. Schmidli, F. König, A. Racine, and W. Maurer.
Confirmatory seamless phase II/III clinical trials with hypothesis selection at interim: General concepts.

G. Hommel.
Adaptive modifications of hypotheses after an interim analysis.

M. Kieser, P. Bauer, and W. Lehmacher.
Inference on multiple endpoints in clinical trials with adaptive interim analyses.
F. Koenig, W. Brannath, F. Bretz, and M. Posch.
Adaptive Dunnett tests for treatment selection.
*Statistics in Medicine, 2007.*
In press.

M. Posch, F. König, M. Branson, W. Brannath, C. Dunger-Baldauf, and P. Bauer.
Adaptive treatment selection.
*Statistics in Medicine, 24:3697–3714, 2005.*

H. Schmidli, F. Bretz, A. Racine, and W. Maurer.
Confirmatory seamless phase II/III clinical trials with hypothesis selection at interim: Applications and practical considerations.

L. Shen.
An improved method of evaluating drug effect in a multiple dose clinical trial.
*Statistics in Medicine, 20:1913–1929, 2001.*

N. Stallard and S. Todd.
Point estimates and confidence regions for sequential trials involving selection.