Promises and Limitations of Flexible Designs for Clinical Research

Martin Posch    Werner Brannath    Peter Bauer

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

’05 FDA/Industry Workshop
Motivation

Flexible Versus Classical Frequentist Trials

Classical frequentist trials
lack the flexibility to react to information emerging from inside or outside the trial.

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.
Flexible Trials and Drug Development

Flexible Designs allow to...

- integrate Phase II and Phase III trials into a single trial
- formally integrate the data of exploratory and confirmatory phases
- speed up the drug development process
- react flexibly to unexpected events
Scenarios for Design Adaptations in Clinical Trials

Example

Consider the comparison of several treatments with a control with \( n \) observations per group. Assume that after \( m < n \) observations an adaptive interim analysis is performed.
## Design Adaptations I

<table>
<thead>
<tr>
<th>Interim Result</th>
<th>Adaptation strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large observed effect</td>
<td>Decreasing sample size</td>
</tr>
<tr>
<td></td>
<td>Adding interim analysis</td>
</tr>
<tr>
<td>Small observed effect</td>
<td>Adding higher doses</td>
</tr>
<tr>
<td></td>
<td>Increasing sample size</td>
</tr>
<tr>
<td></td>
<td>Stopping for futility</td>
</tr>
<tr>
<td></td>
<td>Adding interim analysis</td>
</tr>
<tr>
<td>Effect in only some endpoints</td>
<td>Changing the primary endpoints</td>
</tr>
</tbody>
</table>
Design Adaptations II

<table>
<thead>
<tr>
<th>Interim Result</th>
<th>Adaptation strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety Problems</td>
<td>Dropping high doses</td>
</tr>
<tr>
<td></td>
<td>Adding lower doses</td>
</tr>
<tr>
<td></td>
<td>Increasing sample size</td>
</tr>
<tr>
<td>No Adverse Events</td>
<td>Adding higher doses</td>
</tr>
<tr>
<td>Effect in a subgroup</td>
<td>Adapting the population</td>
</tr>
<tr>
<td>Important covariable not accounted for</td>
<td>Adapting the test statistics</td>
</tr>
</tbody>
</table>
Flexible Designs Based on Combination Tests
The Combination Test

First Stage

\[ p \]

\[ 0 \rightarrow \alpha_1 \rightarrow \text{Reject } H_0 \]

\[ \alpha_0 \rightarrow \text{Accept } H_0 \]

Adaptation

Second Stage

\[ C(p,q) \]

\[ 0 \rightarrow c \rightarrow \text{Reject } H_0 \]

\[ 1 \rightarrow \text{Accept } H_0 \]
Design Adaptations with Combination Tests

- The combination function and stopping boundaries have to be laid down a priori.
- All adaptations are allowed as long as new patients are recruited for the second stage and conservative stage wise tests are used.
- Examples of combination functions:
  - Fisher: \( C(p, q) = pq \)
  - Inverse Normal
    \[
    C(p, q) = -w_1 \Phi^{-1}(1 - p) - w_2 \Phi^{-1}(1 - q)
    \]

Extensions of the Combination Tests

The Recursive Combination Test

- As second stage test another combination test can be applied
- whose second stage test maybe another combination test ...
- whose second stage test maybe another combination test ...
- ...

Therefore the number of interim analyses can be chosen adaptively.

Brannath, Posch and Bauer, 2002
Flexible Designs Based on the Conditional Error Rate
Design Adaptations with the Conditional Error Rate

Definition of the Conditional Error Rate

Definition (The Conditional Error Rate)

Consider a test $\varphi$ with $n$ observations to test a hypothesis $H_0$.

The \textit{conditional error rate} after $m$ observations is given by

$$P_{H_0}(\varphi \text{ rejects } H_0 \mid \text{ first } m \text{ observations}),$$
Design Adaptations with the Conditional Error Rate

Example

One sample z-test at level $\alpha = 0.025$ for the mean of 5 normally distributed observations with known variance $\sigma^2 = 1$ to test the hypotheses

$$H_0 : \mu = 0 \quad \text{against} \quad H_1 : \mu > 0$$

A sample path of the Conditional Error Rate

<table>
<thead>
<tr>
<th>$m$</th>
<th>Obs.</th>
<th>Conditional Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>1</td>
<td>0.9</td>
<td>0.041</td>
</tr>
<tr>
<td>2</td>
<td>-0.2</td>
<td>0.017</td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>0.132</td>
</tr>
<tr>
<td>4</td>
<td>1.3</td>
<td>0.389</td>
</tr>
<tr>
<td>5</td>
<td>1.1</td>
<td>1.000 (rejection)</td>
</tr>
</tbody>
</table>
Design Adaptations with the Conditional Error Rate

The Adaptation Principle

Instead of completing the trial, after $m$ observations one can choose any other test for $H_0$ (with new observations) at the level of the conditional error rate.

The resulting procedure controls the level $\alpha$.

Proschan and Hunsberger, 1995, Müller & Schäfer 2001, 2004

Relation to Combination Tests

Combination tests and test based on the conditional error rate are equivalent.

Posch & Bauer, 1999, Wassmer 1999
Example (cont’d): Increasing Sample Size

<table>
<thead>
<tr>
<th>$m$</th>
<th>Obs.</th>
<th>Conditional Error $A_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>0.025</td>
</tr>
<tr>
<td>1</td>
<td>0.9</td>
<td>0.041</td>
</tr>
<tr>
<td>2</td>
<td>-0.2</td>
<td>0.017</td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>0.132</td>
</tr>
</tbody>
</table>

- After 3 observations it is decided to increase the sample size and to collect another 10 observations.
- Let $p'$ denote the p-value calculated from a test with the 10 additional observations.
- Then $H_0$ is rejected if $p' < 0.132$
## Design Adaptations with the Conditional Error Rate

### Properties

- In principle, after every observation adaptations can be performed.
- The interim analysis need not be preplanned.
- The adaptation rules need not be preplanned.
- The Type I error is controlled.

### Back to Reality...

- Nobody in a clinical trial realistically would calculate the conditional error rate after every observation.
- However, the general principle shows that an interim analysis can be performed whenever it seems appropriate without any preplanning.
Multiple Testing
Multiple Testing

Multiple testing issues arise if:

- treatments
- endpoints
- sub-populations

are added or dropped in an adaptive interim analysis.

Adaptive multiple testing procedures:

- adaptive tests & closed testing procedure
- adaptive tests for all intersection hypotheses are performed
- controls the familywise Type I error rate in the strong sense

Estimation
Point Estimates and Confidence Intervals

**Point Estimates**

The maximum likelihood estimates are typically mean biased. The mean bias depends on:

- the alternative
- the stopping rule
- the adaptation rule

Therefore the mean bias is not known.
Maximum likelihood estimate

Mean bias of MLE for typical examples

- **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

*Brannath et al. 2005, Posch et al. 2005*
Maximum Likelihood Estimate

Upper bound for the mean bias (normal case)

The mean bias due to sample size reassessment and early stopping is bounded by

\[
\text{mean bias} \leq 0.4 \cdot \text{standard error of first stage estimate}
\]

\text{Liu et al. (2001), Brannath et al. (2005)}

The bound can be improved, if minimum or maximum sample sizes are defined.
Confidence Intervals

### Confidence Intervals for Adaptive Designs

- **Repeated Confidence Intervals**
  
  Lehmacher and Wassmer, 1999; Brannath et al. 2002, Lawrence & Hung, 2003; Proschan et al., 2003

- **Monotone Confidence Intervals**
  
  Brannath et al. 2002

### Derived Point Estimates

- **Weighted Maximum Likelihood Estimate**
  
  Lawrence & Hung, 2003, Proschan et al., 2003, Brannath et al., 2002

- **Median Unbiased Point Estimate**
  
  Brannath et al., 2002
Limitations of Adaptive Designs
**All limitations of group sequential trials apply**

- Control of information flow
- Overrunning
- Suitable for trials with immediately available endpoint.

**Specific limitations of flexible designs**

- Extensive adaptations may affect the persuasiveness of results
- Deviation from maximum likelihood based test statistics
- Heterogeneity across stages can make it difficult to interpret the outcome
- Adaptations may be inefficient  
  
  Tsiatis, 2003, Jennison & Turnbull, 2003
Conclusion

- Testing and estimation in flexible designs have been extensively investigated and there is a wide field of applications beyond sample size reassessment.
- To construct persuasive flexible trial designs, careful planning is required.
- It has to be decided on a case by case basis if the gained flexibility justifies the increased complexity of flexible trials.
Selected References

M. Posch, F. König, M. Branson, W. Brannath, C. Dunger-Baldauf, and P. Bauer.
Adaptive treatment selection.

W. Brannath, F. König, and P. Bauer.
Estimation in flexible two stage designs.

H. H. Müller and H. Schäfer.
A general statistical principle for changing a design any time during the course of a trial.

M. Posch, P. Bauer, and W. Brannath.
Issues in designing flexible trials.

W. Brannath, M. Posch, and P. Bauer.
Recursive combination tests.