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

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

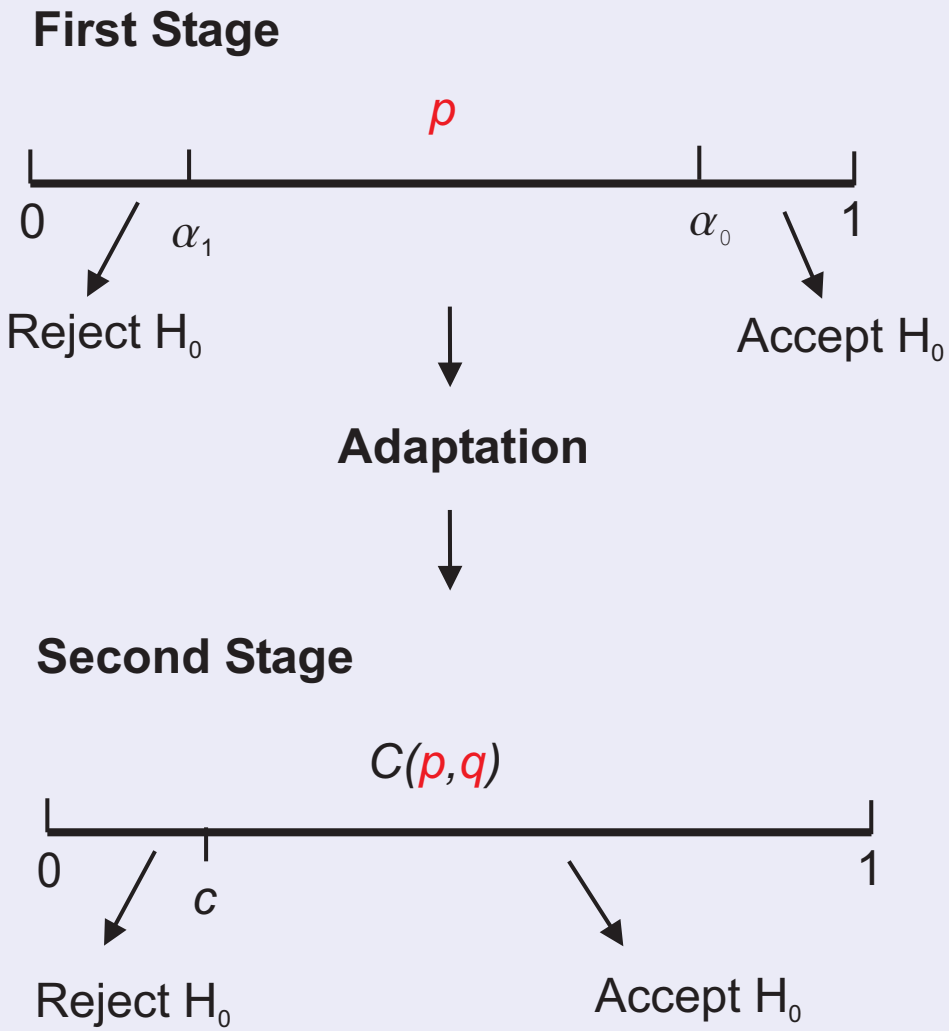
Interim Result		Adaptation strategy
Large observed effect	$\rightarrow$	Decreasing sample size
	$\rightarrow$	Adding interim analysis
Small observed effect	$\rightarrow$	Adding higher doses
	$\rightarrow$	Increasing sample size
	$\rightarrow$	Stopping for futility
	$\rightarrow$	Adding interim analysis
Effect in only some endpoints	$\rightarrow$	Changing the primary endpoints

Design Adaptations II

Interim Result		Adaptation strategy
Safety Problems	$\rightarrow$	Dropping high doses
	$\rightarrow$	Adding lower doses
	$\rightarrow$	Increasing sample size
No Adverse Events	$\rightarrow$	Adding higher doses
Effect in a subgroup	$\rightarrow$	Adapting the population
Important covariable not accounted for	$\rightarrow$	Adapting the test statistics

Flexible Designs Based on Combination Tests

The Combination Test



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: $C(p, q) = p\,q$
  - Inverse Normal
    $C(p, q) = -w_1\,\Phi^{-1}(1-p) - w_2\,\Phi^{-1}(1-q)$

BAUER 1989, BAUER & KÖHNE 1994, LEHMACHER & WASSMER 1999,
CUI ET AL. 1999

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 φ with n observations to test a hypothesis H_0.

The *conditional error rate* after m observations is given by

$$P_{H_0}\left(\varphi \text{ rejects } H_0 \mid \text{ first } m \text{ observations}\right),$$

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

m	Obs.	Conditional Error Rate
0		0.025
1	0.9	0.041
2	-0.2	0.017
3	2.1	0.132
4	1.3	0.389
5	1.1	1.000 (rejection)

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 α.

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

Design Adaptations with the Conditional Error Rate

Example (cont'd): Increasing Sample Size

m	Obs.	Conditional Error A_m
0		0.025
1	0.9	0.041
2	-0.2	0.017
3	2.1	0.132
4		
5		

- 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

BAUER & KIESER 1999, HOMMEL, 2001, HOMMEL & KROPF, POSCH ET AL. 2005

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.

Skip Examples

Maximum likelihood estimate

Mean bias of MLE for typical examples

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

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

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.

Statistics in Medicine (to appear), 2005.

W. Brannath, F. König, and P. Bauer.

Estimation in flexible two stage designs.

Statistics in Medicine (to appear), 2005.

H. H. Müller and H. Schäfer.

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

Statistics in Medicine, 23:2497–2508, 2004.

M. Posch, P. Bauer, and W. Brannath.

Issues in designing flexible trials.

Statistics in Medicine, 23:953–969, 2003.

W. Brannath, M. Posch, and P. Bauer.

Recursive combination tests.

Journal of the American Statistical Association, 97:236–244, 2002.