ADAPTIVITY IN GROUP SEQUENTIAL DESIGNS

Professor Christopher Jennison, Dept of Mathematical Sciences, University of Bath, UK

Professor Bruce Turnbull, Cornell University, Ithaca, New York

Portland, Oregon, May 2005

26th Annual Meeting Society for Clinical Trials
Outline of presentation

1. Group sequential designs for clinical trials
   - Adapting to observed data

2. Error-spending tests
   - Adapting to unpredictable information
   - Adapting to nuisance parameters

3. Most efficient group sequential tests
   - Adapting optimally to observed data

4. Flexibility for unplanned design changes
   - Adapting super-optimally to observed data

5. An example of inefficiency in an adaptive design

6. Conclusions

- Adapting to new objectives
- Adapting to new objectives
- Adapting to new objectives
- Adapting to new objectives
A one-sided testing problem

1. Group sequential monitoring of clinical trials
One-sided group sequential tests

A typical group sequential testing boundary has the form:

<table>
<thead>
<tr>
<th>Reject</th>
<th>Accept</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H )</td>
<td>( H )</td>
</tr>
</tbody>
</table>

Adapting to data — stopping when a decision is possible.

Sample size can be around 50 to 70% of the fixed sample size.
Errorspends tests

Canonical joint distribution of parameter estimates

\[ \text{Let } \hat{\theta} \text{ be the estimate of } \theta \text{ based on data at analysis } \mathcal{I}. \]

\[ \text{For } \mathcal{I} = 1 \text{ and } \ldots \text{ the estimate of } \theta \text{ based on data at analysis } \mathcal{I} \text{ is} \]

\[ \mathcal{I} \text{ at analysis } \mathcal{I} \text{ is } \mathcal{I} \]

\[ \text{The information for } \theta \text{ at analysis } \mathcal{I} \text{ is } \mathcal{I} \]

\[ \text{In very many situations, } \mathcal{I} \text{ are approximately multivariate normal, } \mathcal{I} \]

\[ \text{Let } \mathcal{I} \text{ be the estimate of } \theta \text{ based on data at analysis } \mathcal{I}. \]

\[ \text{Canonical joint distribution of parameter estimates} \]

\[ \text{2. Error spending tests} \]
Spending type I and type II error probabilities.

To extend to one-sided tests, define two functions, (T)f and (T)g, for the cumulative type I error probability as a function of the observed information:

The current boundary point is set so that at analysis #t, the cumulative type I error probability is (T)f(t).

Thus, it may not be possible to predict the actual sequence of information levels, e.g., T1, T2, ..., in advance. Therefore, e.g., for survival data, the overall failure rate depends on the number of subjects and other factors.

To extend to two-sided tests with the Lan & DeMets (Biometrika, 1983) presented two-sided tests, define two functions, (T)f and (T)g, for spending type I and type II error probabilities as a function of the observed information.

Thus, it may not be possible to predict the actual sequence of information.
Adapting to unpredictable information

One-sided error spending tests

\[ \phi(\mathcal{I}/\mathcal{I}) \propto (\mathcal{I})^\beta \quad \text{and} \quad (\mathcal{I})^f \]

\[ (\cdot \mathcal{I})^\beta = \{ \cdot \text{accept by analysis} \}^{0=\theta, \mathcal{I}} \]

\[ (\cdot \mathcal{I})^f = \{ \cdot \text{reject by analysis} \}^{0=\theta, \mathcal{I}} \]

At analysis, set boundary values

Power family of error spending tests:

\[ \lambda \mathcal{I}, \lambda \mathcal{J}, \lambda \mathcal{K}, \lambda \mathcal{L}, \lambda \mathcal{M}, \lambda \mathcal{N}, \lambda \mathcal{O}, \lambda \mathcal{P}, \lambda \mathcal{Q}, \lambda \mathcal{R} \]

Adapting to unpredictable information
Maximum Information Designs

Design

Assume, say, \( \mathcal{I} \) equally spaced Information Levels.

\[ \mathcal{I} = \mathcal{I}_{max} \]

Find \( \mathcal{I} = \mathcal{I}_{max} \) such that boundaries meet up on reaching \( \mathcal{I}_{max} \).

Implementation

Use the error-spending construction with observed \( \mathcal{I} \)'s. Continue up to \( \mathcal{I}_{max} \) and make the boundaries converge, protecting type I error.

N.B. Changes affecting \( \mathcal{I} \) should not be influenced by \( \hat{\theta} \).

If necessary, extend patient accrual to reach \( \mathcal{I}_{max} \).

Information

\( \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5, \mathcal{I}_6 \)

\( x_{max} \mathcal{I} \)
Adapting to nuisance parameters

Error-spending designs handle these issues automatically.

\[ \frac{1}{\{ \text{Number of failures by analysis} \}} \approx \frac{\theta}{\mathcal{I}} \]

Information depends on the number of observed failures,

(2) Survival data, log-rank statistics

\[ \left( \frac{\theta}{\mathcal{I}}, \frac{\lambda}{\mathcal{I}} \right) \mathcal{N} \sim \chi^2 \left( \frac{\lambda}{\mathcal{I}} \right) \mathcal{N} \sim \chi^2 \]

(1) Normal responses with unknown variance

\[ \lambda \eta - X \eta = \theta \quad \text{and} \quad \left( \frac{\theta}{\mathcal{I}}, \frac{\lambda}{\mathcal{I}} \right) \mathcal{N} \sim \chi^2 \left( \frac{\lambda}{\mathcal{I}} \right) \mathcal{N} \sim \chi^2 \]

The target \( \mathcal{I} \) is fixed but the sample size needed to achieve this can...
3. Optimal group sequential tests

There is plenty of choice in defining a boundary to solve a particular testing problem. Thus, one can seek a boundary with an optimality property. Formulate the testing problem:

\[ \theta \]

Fix the design which minimises average sample size (information), if desired.

\[ \text{Fix number of analyses, } N, \]

\[ \text{Fix maximum sample size (information), if desired.} \]

This optimisation can be carried out by solving a related Bayes decision problem using backwards induction (dynamic programming).
Example of properties of optimal tests

Adapting optimally to observed data

Up to a point, \( H \) as \( (I)\bar{H} \)

but with diminishing returns, \( H \) as \( (I)\bar{H} \)

Note:

\[
\begin{array}{cccccc}
54.3 & 56.0 & 56.3 & 55.2 & 54.3 & 56.3 \\
69.1 & 62.1 & 59.0 & 55.2 & 56.3 & 55.2 \\
72.2 & 62.2 & 62.5 & 59.0 & 55.2 & 56.3 \\
72.2 & 72.8 & 73.2 & 75.3 & 72.2 & 72.8 \\
72.2 & 74.5 & 74.9 & 80.9 & 72.2 & 74.5 \\
72.2 & 72.8 & 73.2 & 75.3 & 80.9 & 74.9 \\
72.2 & 74.5 & 74.9 & 80.9 & 72.2 & 74.5 \\
72.2 & 72.8 & 73.2 & 75.3 & 80.9 & 74.9 \\
72.2 & 74.5 & 74.9 & 80.9 & 72.2 & 74.5 \\
72.2 & 72.8 & 73.2 & 75.3 & 80.9 & 74.9 \\
\end{array}
\]

Minimum values of \( T \):

5

\( \bar{H} \) over \( H \)

Minimum values of \( T \), as a percentage of \( H \):

\[
\frac{2}{\{ (I)\bar{H} + (I)^0\bar{H} \}}
\]

Equal group sizes, diminishing returns

One-sided tests, as \( \alpha = 0.05 \), \( \gamma = 0.1 \)
Squeezing a little extra efficiency

Schmitz (1993) proposed group sequential tests in which group sizes are chosen adaptively. We describe these on the score statistic scale:

\[(\mathcal{I} - \mathcal{Z}) (\mathcal{I} - \mathcal{Z}) \theta)^T \mathcal{N} \sim \mathcal{I}_S - \mathcal{Z}_S\]

then choose as a function of \(\mathcal{Z} + \mathcal{I}\) observe \(\mathcal{Z}\) where

\[(\mathcal{I} \mathcal{I} \theta)^T \mathcal{I}_S \sim \mathcal{I}_S\]

Initially, fix \(\mathcal{I}\), observe \(\mathcal{Z}\), etc, etc.

Specify sampling rule and stopping rule to achieve desired overall type I error rate and power.

Squeezing a little extra efficiency
Examples of "Schmitz" designs

Aim for low values of $q = \theta$ at $\theta = 0.0$, with type I error rate $\alpha = 0.025$.

To test $H_0: \theta = 0$ versus $H_1: \theta > 0$. Constraints:

Maximum sample information $\times 2 = I.2 \times$ fixed sample information.

Maximum number of analyses $Y = Y$.

$\theta p(\theta)f(I) \theta E \int$

where $f$ is the density of a $N(\theta, \theta)$ distribution.

Again, find optimal designs by solving related Bayes decision problems.
Efficiency of "Schmitz" designs

Varying group sizes adaptively makes for a complex procedure and the efficiency gains are slight.

<table>
<thead>
<tr>
<th>groupsizes</th>
<th>(Schmitz) design</th>
<th>K</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>equal group</td>
<td>non-adaptive, (Schmitz) optimised</td>
<td>10</td>
<td>Optimal</td>
</tr>
<tr>
<td>57.5</td>
<td>57.2</td>
<td>55.9</td>
<td>Optimal</td>
</tr>
<tr>
<td>59.8</td>
<td>59.4</td>
<td>58.0</td>
<td>Optimal</td>
</tr>
<tr>
<td>62.7</td>
<td>62.4</td>
<td>61.2</td>
<td>Optimal</td>
</tr>
<tr>
<td>66.1</td>
<td>66.6</td>
<td>64.8</td>
<td>Optimal</td>
</tr>
<tr>
<td>74.8</td>
<td>73.2</td>
<td>72.5</td>
<td>Optimal</td>
</tr>
</tbody>
</table>

Optimal average as a percentage of the fixed sample information.

Efficiency of "Schmitz" designs
Recent advances in flexible/adaptive methods

Mid-study re-design to increase power

Recently, advances in flexible/adaptive methods have been made. These methods allow for changes in the design of a study during its course, which can increase the power of the study. This is particularly useful in studies where reasons may arise to change the power. During the course of a study, re-designs may be necessary to increase power.

Denne (2001) and Muller & Schaefer (2001) show this is possible, as long as the re-design preserves the conditional type I error probability.

The methods of Bauer & Kohne (1994), Fisher (1998), Cui, Hung & Wang (1999) are described differently, but they also possess this property.

Suppose you design a study with power 0.9 at \( \theta \). If a competing treatment is withdrawn, you may wish to increase sample size to attain power 0.9 at \( \theta \). If a competing treatment is withdrawn, you may wish to increase sample size to attain power 0.9 at \( \theta \).

4. Recent advances in flexible/adaptive methods
Re-design in response to an interim estimate, \( \hat{\theta} \)

Motivation may be:

- to rescue an under-powered study,
- a "wait and see" approach to choosing a study's power requirement,
- trying to be efficient.

Schemes for modifying sample size have been proposed, often fixing conditional power under \( \hat{\theta} \).

But, group sequential tests already base the decision for early stopping on \( \hat{\theta} \) — and optimal GSTs do this optimally!
An adaptive design starts out as a fixed-sample test with
\[ q_0 = \frac{f u}{100}. \]
Suppose this requires a sample size \( q_0 = 200. \)
First, consider a fixed sample study attaining power 0.9 at \( \theta = 2.0. \)

Investigators are optimistic the effect size, \( \theta \), will be as high as \( \theta = 2.0. \)
But, effect sizes as low as \( \theta = 1.5 \) are clinically relevant and worth
detecting.

A test is to give type I error probability \( \alpha = 0.025. \)

Scenario (of the type described by Cui, Hung & Wang, 1999)

Example of inefficiency in an adaptive design
At an interim stage, after 50 observations, the estimated effect size is \( \dot{\theta} \).

If \( \dot{\theta} = \theta \), stop the trial for futility, accepting the conditional type I error rate given \( \theta \).

Otherwise, re-design the remainder of the trial, preserving the conditional power 0.9. Choose the remaining sample size to give conditional power 0.9. Truncate this additional sample size to the interval (50, 500) — no decrease in sample size is allowed and the total sample size is at most 550.

\[ H_0: \theta = 0, \quad \forall \alpha > 0.2 \Rightarrow \dot{\theta} \Rightarrow \theta \]

At an interim stage, after 50 observations, the estimated effect size is \( \dot{\theta} \).
If continuing past the first stage, total sample size ranges from 100 to 550. Achieving power 0.85 at $\theta = \frac{0.8}{\theta}$, i.e., $0.8 = 1.5 \Rightarrow \theta = 0.52$. The adaptive test improves on the power of the fixed sample test, achieving power 0.85 at $\theta = \frac{0.8}{\theta}$, i.e., $0.8 = 1.5 \Rightarrow \theta = 0.52$. If continuing past the first stage, total sample size ranges from 100 to 550.
A conventional group sequential test

compares to 550. Power and ASN. It also has a much lower maximum sample size—225

test dominates the Cui et al. adaptive design with respect to both

\[ \frac{q}{l} = \theta \]

After 225 gives a test with power 0.9 at \( j \), taking the first analysis after 68 observations and the second analysis

Type I error rate is \( \alpha = 0.025 \).

We have compared a power family, error spending test with

\[ = I \]

analyses, designed to attain power 0.9 when \( \theta = I \).

Similar overall power can be attained by a non-adaptive GST with \( K = 2 \).
Many other proposals for adaptive designs show similar inefficiencies.

much smaller maximum sample size.

lower average sample size function.

higher power.

The conventional GST has:

Cui et al. adaptive test vs non-adaptive GST
6. Conclusions

Group Sequential Tests — and they can be substantially inferior. But, these adaptive designs will not improve on the efficiency of "standard"

- an on-going approach to study design.
- re-sizing to rescue an under-powered study,
- re-design in response to external developments.

In addition, recent adaptive methods allow

- observed data, i.e., efficient stopping rules.
- nuisance parameters,
- unpredictable information levels.

Error Spending Tests using Information Monitoring can adapt to
References


(Jefferson, C. and Turnbull, B. W. (2005); Adaptable and non-adaptable group sequential tests. Submitted for publication.

Jefferson, C. and Turnbull, B. W. (2005). Efficient group sequential designs when there are several effect sizes under consideration.


References


Jennison, C. and Turnbull, B. W. (2005). Efficient group sequential designs when there are several effect sizes under consideration.


