Adaptive Designs in Phase II and Phase III Clinical Trials

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

Partnerships in Clinical Trials,

Amsterdam, November 2007

Wall Street Journal, July 2006

“FDA Signals It’s Open to Drug Trials That Shift Midcourse”

Ways adaptive designs may allow trials to be adjusted based on early results:

- Route a larger share of patients to the treatment that seems to work best
- Drop treatments that don’t seem to be effective
- Add more of the type of patients . . . reacting best to a particular treatment
- Merge two different phases of drug development into one trial

With views from:

Bob O’Neill , FDA

Paul Gallo, Novartis

Tom Fleming, Univ. Washington

Michael Krams, Wyeth

Don Berry, M. D. Anderson Cancer Center

Bruce Turnbull, Cornell University

Statements by Scott Gottlieb, MD

Quotes from a July 2006 speech by Scott Gottlieb, the FDA's Deputy Commissioner for Medical and Scientific Affairs:

> " . . . the advantages of these [adaptive] approaches, rigorously designed, are becoming more evident, including among the ranks of our experts at FDA. It is essential that we at the FDA do all we can to facilitate their appropriate use in modern drug development."

> " . . . the FDA's drug center is working on a series of guidance documents — up to five in all — that will help articulate the pathway for developing adaptive approaches to clinical trials."

See also

`http://www.fda.gov/oc/speeches/2006/trialdesign0710.html`

1. Motivation for adaptation in clinical trials

Adaptation to external factors

A rival treatment is withdrawn: a ***smaller effect size*** is now of clinical interest.

An improved financial position means ***sponsors can invest more*** in this trial.

Adaptation to internal factors

In-study estimates of response variance imply ***a greater sample size*** is needed to achieve the intended power.

Low adverse event rates indicate it may suffice to demonstrate ***non-inferiority***, rather than superiority, of the new treatment.

Treatment benefits a ***sub-group***: can the target population be re-defined to focus on these patients?

Adaptation in clinical trials

In a trial with multiple treatments or dose levels:

Eliminate weaker treatments as the study progresses.

Using a dose-response model, optimize treatment allocation in order to learn most efficiently about the best choice of dose level.

In Phase IIb (dose finding) and Phase III (confirmatory) trials:

Proceed directly from Phase IIb to Phase III and test the selected dose level, eliminating “white space” between phases.

Combine Phase IIb and Phase III data in the final statistical analysis.

Optimize the allocation of resource between Phase IIb and Phase III.

2. Adaptivity and flexibility

Rigid adaptive designs

Adaptivity does not necessarily imply flexibility.

Rules for adaptation can be completely specified in the protocol and implemented as such, with the advantages:

- Potential modifications are approved up front and there is no need to file protocol amendments.
- Logistics for changing treatments, patient eligibility, accrual rates, etc., can be planned for in advance.
- Credibility of results is maintained, especially with the DSMB as a “firewall”.

Adaptivity and flexibility

Totally flexible designs

If a suitable design system is put in place, it may be possible to allow modifications which were not thought about at all at the outset.

L. Fisher (*Biometrics*, 1998) refers to the “self-designing” clinical trial.

- Advantages
  - Investigators can learn about treatments and respond appropriately.
  - Complete flexibility to react to un-anticipated occurrences.
- Disadvantages
  - Ad hoc changes based on unblinded data may jeopardise credibility.
  - Test statistics have an unfamiliar form.

Adaptivity and flexibility

Partially flexible designs

A compromise is to provide modest flexibility for re-design within a clearly defined framework.

Example: Bauer & Köhne (*Biometrics*, 1994) two-stage design.

- Design and length of Stage 1 are fixed in advance.
- The design of Stage 2 is permitted to depend flexibly on Stage 1 results.
- Final inference is based on P-values from the two stages, according to a pre-specified rule.

3. An adaptive method protecting type I error

Bauer & Köhne (*Biometrics*, 1994)

Initial design

Stipulate that Bauer & Köhne's combination test will be used.

Define the null hypothesis H_0.

Design Stage 1, fixing the sample size and test statistic for this stage.

Stage 1

Observe the P-value P_1 for testing H_0.

Design Stage 2 in the light of Stage 1 data.

Stage 2

Observe the P-value P_2 for testing H_0.

Bauer & Köhne's two-stage method

Overall test

Under H_0, both P_1 and P_2 have the usual null distribution, uniform on $(0,\, 1)$.

It follows that, under H_0,

$$-\ln(P_1\, P_2) \;\sim\; \frac{1}{2}\,\chi^2_4.$$

Hence, the P-values can be combined in an overall test, rejecting H_0 if

$$-\ln(P_1\, P_2) \;>\; \frac{1}{2}\,\chi^2_{4,\,1-\alpha}.$$

This χ^2 test was originally proposed for combining results of several studies by R. A. Fisher (1932) in *Statistical Methods for Research Workers*.

4. Sample size re-estimation

In a study with normal responses, sample size needed to provide power at a particular effect size depends on the response variance, σ^2.

Initial design

A Bauer & Köhne two-stage design is specified.

Sample size n_0 is determined to give power $1 - \beta$ at effect size $\theta = \delta$ if response variance is equal to a preliminary estimate σ_0^2.

Stage 1 is planned with a sample size of $n_1 = n_0/2$.

Stage 1

Yields estimates $\hat{\theta}_1$ and $\hat{\sigma}_1^2$.

The t-statistic t_1 for testing H_0: $\theta \leq 0$ vs $\theta > 0$ is converted to a P-value, P_1.

Sample size re-estimation

Stage 1 . . .

Now use the variance estimate $\hat{\sigma}_1^2$ to re-calculate sample size.

One may simply use this value in the original calculation in place of σ_0^2.

Or, also take account of the interim estimate of treatment effect, $\hat{\theta}_1$.

This defines an additional sample size of n_2 in Stage 2.

Stage 2

Calculate the t-statistic t_2 for testing H_0 based on Stage 2 data alone and convert to a P-value, P_2.

The overall test — which has type I error rate exactly α — rejects H_0 if

$$-\ln(P_1\, P_2) \;>\; \frac{1}{2}\,\chi^2_{4,\,1-\alpha}.$$

5. Sample size modification to increase power

Investigators may start out optimistically and design a trial with power to detect a large treatment effect.

What if interim data suggest a smaller effect size — still clinically important but difficult to demonstrate with the chosen sample size?

A Bauer & Köhne two-stage design allows sample size, and hence power, to be increased after the first stage of the trial.

- Advantage

  Adaptive methodology allows ***rescue*** of an under-powered study.

- Disadvantages

  It is more transparent to design for the power curve that is really desired.

  Then, a ***group sequential design*** can be used to provide early stopping, and save sample size, when the effect size is large.

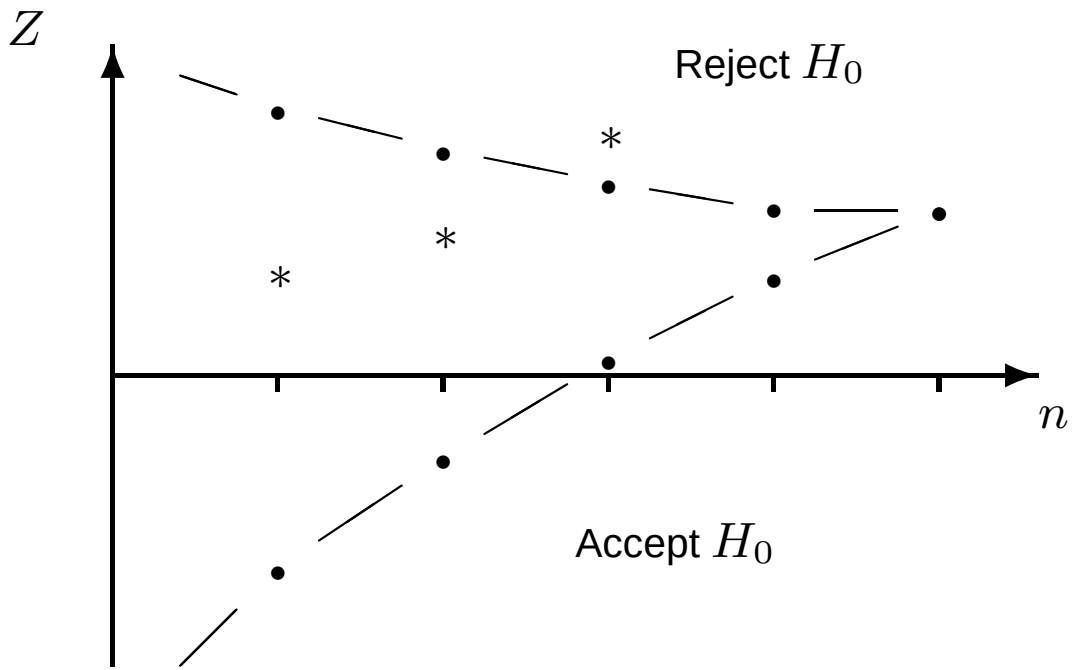
Group sequential tests

Accumulated data are examined at interim analyses and the trial is stopped:

If there is sufficient evidence to reject the null hypothesis,

If a positive outcome is no longer likely.

The procedure must protect the overall type I error rate and give the desired power.



6. Change of endpoint

Example: Trial of a new treatment for Altzheimer's disease.

Primary endpoint $=$ Decline in cognitive function 4 months from baseline.

The null hypothesis is

$H_0(4)$: New treatment does not reduce cognitive decline at 4 months.

Suppose the study is designed as a Bauer & Köhne two-stage design.

At Stage 1

We run this stage and calculate P_1 for testing $H_0(4)$.

Analysis of Stage 1 data suggests the treatment effect does better at 6 months.

Investigators decide to switch to a 6 month endpoint for Stage 2.

Change of endpoint: Example

At Stage 2

Primary endpoint $=$ Decline in cognitive function 6 months from baseline.

New null hypothesis is

$H_0(6)$: New treatment does not reduce cognitive decline at 6 months.

Run this stage and calculate P_2 for testing $H_0(6)$.

Overall analysis

Combine P_1 and P_2 in the overall test of the "null hypothesis".

Conclusion

With a significant P-value, we reject a *combination* of $H_0(4)$ and $H_0(6)$ —

conclude at least one of $H_0(4)$ and $H_0(6)$ is false, but we cannot say which!

Change of endpoint: Example

An alternative framework

Declare both $H_0(4)$ and $H_0(6)$ as hypotheses of interest at the outset.

Start with one, $H_0(4)$ say, as the primary endpoint.

Allow switching to $H_0(6)$ within an adaptive two-stage design and use a *multiple comparisons procedure* to account for considering two hypotheses.

Now, a positive result allows rejection of a specific H_0 (or possibly both).

The overall type I error probability of *any* false positive conclusion is protected.

7. Enrichment: Changing to a population subgroup

Suppose interim data show a strong treatment effect in a subgroup of patients only.

Can the study's goal be modified to test for an effect in that sub-population?

Account must be taken of the opportunity to consider several subgroups and choose the one for which results are most promising.

Multiple comparisons procedures can be combined with adaptive designs to allow such a change, but note:

> *All possible sub-populations must be specified at the outset.*
>
> *Because of the "adjustment" for testing multiple hypotheses, a small set of potential sub-populations is advisable.*

This approach may be of particular value where therapies are likely to be targetted to subgroups identified by genetic markers.

8. Switching between tests for superiority and non-inferiority

A trial may have two possible positive outcomes:

Showing the new treatment is *superior* to the current standard,

Showing the new treatment is *non-inferior* to the standard.

Investigators may start a trial intending to show superiority, then decide to adapt to a new goal of non-inferiority if results are not as good as expected.

Having two hypotheses is not a problem as the two tests are nested:

Superiority — effect size greater than zero,

Non-inferiority — effect size greater than $-\delta$.

Switching between superiority and non-inferiority

But, since the effect size used to power a superiority study is usually larger than the tolerance allowed for non-inferiority, quite different sample sizes are needed.

There have been proposals for adaptive trial designs that allow this change of objective from superiority to non-inferiority.

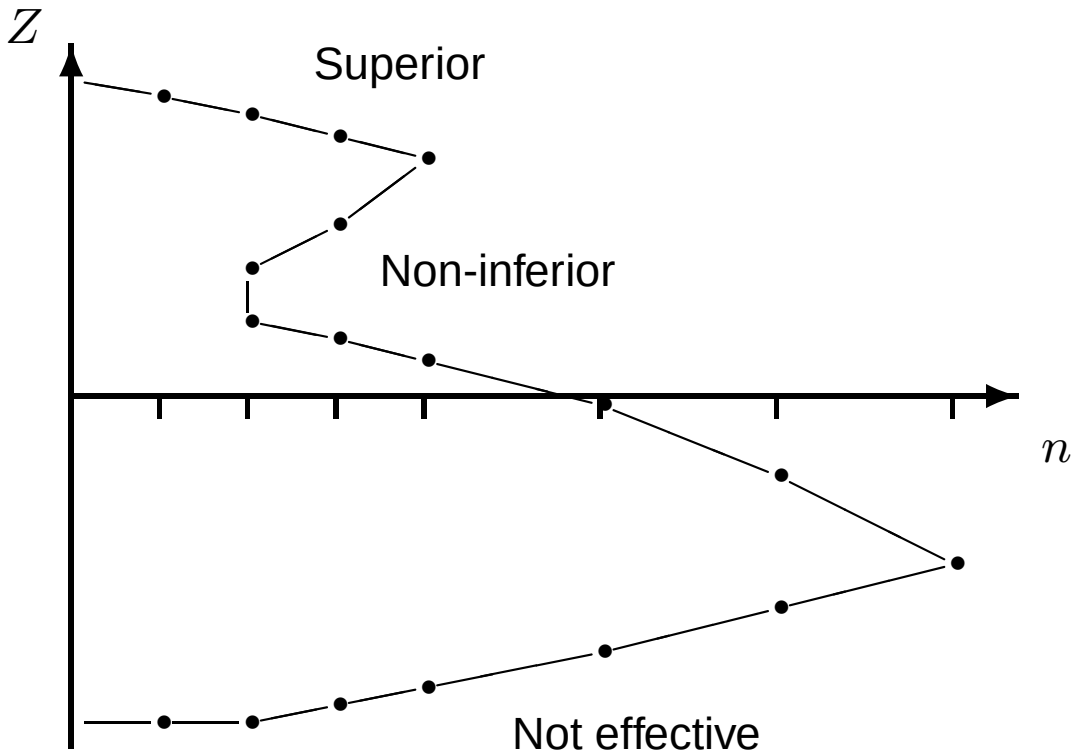
For example, one could consider such a change after the first stage of a Bauer & Köhne two-stage design.

Then, if interest shifts to establishing non-inferiority and a larger sample size is needed, the second stage sample size can be increased.

Switching between superiority and non-inferiority

A change of objective can be considered during a group sequential study.

In fact, a group sequential design can be pre-specified to accommodate these twin objectives:



9. Combined analysis of Phase IIb (dose finding) and Phase III (confirmatory) trials

There is currently a lively interest in combining phases of testing.

Elements of adaptive methods can play a role in the final analysis of “seamless” Phase II/III designs:

To account for the multiple hypotheses concerning different dose levels in Phase IIb,

To combine data from the two stages.

A combined treatment of these two phases of testing offers:

A speedier process as the “white space” between phases is removed,

The chance to use Phase IIb data on the selected dose in the final analysis.

10. Partnerships for adaptive designs

There has been plenty of development and discussion of adaptive methods in the Medical Statistics literature.

The novelty and complexity of these methods implies that ***partnerships with statisticians*** are key to their application in prospective clinical trials.

Regulators need to be reassured that novel methods will yield credible conclusions, meeting all the usual standards of scientific rigour.

Discussions in ***partnership with the regulators*** at an early stage of planning are essential in order to be sure that results and analyses will be acceptable when the trial is concluded.

References

Further detail on the topics of this talk is presented in the paper

"Adaptive seamless designs: Selection and prospective testing of hypotheses"

by C. Jennison and B.W. Turnbull,

to appear in the *Journal of Biopharmaceutical Statistics* (2007).

My web-page at

http://people.bath.ac.uk/mascj

has links to my other presentations and publications in this area and, hence, to many more statistical articles.