Using Group Sequential and Adaptive Designs to Improve the Efficiency of a Development Programme

Christopher Jennison

Dept of Mathematical Sciences,

University of Bath, UK

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

Mini-symposium

Vlaardingen, 12 September 2014

Experiences of drug development in the pharmaceutical industry

The process of developing a new medical treatment is long and expensive.

- Large numbers of molecules are screened in pre-clinical testing.
- Initial evidence is required that a compound can be tolerated by patients and has the potential to treat the target disease.
- In subsequent stages of testing a suitable dose is determined and the final form of the treatment is tested for efficacy against a control treatment or placebo.

A short remaining patent lifetime when a new treatment is approved limits the financial return to the developers.

Use of innovative experimental designs throughout the development process can:

- Improve the discovery rate of successful new treatments;
- Find successful treatments earlier, leaving a longer patent lifetime.

Outline of talk

1. Adaptive designs for Phase III clinical trials

    Nature of mid trial modifications

    Sample size re-estimation when testing a single hypothesis

    Testing multiple hypotheses: treatment selection, enrichment, . . .

2. Group sequential monitoring

    Controlling error rates when there are multiple looks at the data

    Benefits of early stopping

    Practical requirements and limitations

3. Joint planning for several stages of a development programme

4. Conclusions

1. Adaptive designs for Phase III clinical trials

Phase III trials are conducted at the end of the drug development process, or the development of a new medical treatment. Then,

The treatment has been refined and tested in earlier development and in Phase I and II trials,

A substantial body of work supports the investigators' belief that the new treatment is effective and safe.

The aim of the Phase III trial is to compare the new treatment with the current standard treatment or a placebo, when given to the target patient population.

The need for a clear, unambiguous comparison leads to the desire for a simple Phase III clinical trial.

All aspects of the Phase III trial design are pre-defined and written into the protocol and statistical analysis plan.

Traditional Phase III clinical trials (pre 2000 approx.)

A trial protocol specifies:

- The experimental treatment and the control or placebo treatment,
- The patient population (eligibility criteria, etc.),
- Sample size for the trial,
- Statistical analysis plan.

Interim analyses may be conducted to:

- Monitor safety,
- Stop early for futility if the new treatment is not effective,
- Stop early if there is overwhelming evidence of efficacy.

Many of the decisions taken in creating such a design would benefit from further knowledge of the treatment, the patients, or patient responses.

Early examples of adaptive methods

There is a long history of “Adaptive” statistical methods.

Adaptive randomisation

In a trial comparing two treatments, adaptive randomisation can be used to increase the proportion of patients allocated to the better of two treatments.

However, once randomisation becomes unequal, ethical issues may arise as to whether it is permissible to randomise at all.

Adaptive randomisation highlights the role of “equipoise” in a randomised clinical trial and just what this term should mean.

Ethical and statistical concerns were clearly evident in two Harvard trials in the 1970s and 1980s which investigated ECMO treatment of critically ill, new-born babies (Ware, *Statistical Science*, 1989).

Early examples of adaptive methods

Sample size re-estimation

The sample size needed to achieve a specific power under a given treatment effect is proportional to the response variance — which is typically unknown when planning a trial.

Wittes & Brittain (*Statistics in Medicine*, 1990) suggested choosing an initial sample size based on a plausible response variance, then updating the sample size as better estimates of response variance are obtained.

The same approach can be used to handle an unknown baseline hazard rate for survival data.

Sample size re-estimation in the light of estimates of “nuisance parameters” is still one of the most commonly used adaptive methodologies.

Adaptive designs for Phase III clinical trials

We noted that many of the decisions taken in designing a clinical trial would benefit from further knowledge of the treatment, the patients, or patient responses.

What is the best dose for the new treatment?

What is the best method of delivery for the new treatment?

Does the treatment have greater benefit for a sub-population of patients?

For a normally distributed response, what is the variance?

Or, for time-to-event data, what is the baseline hazard rate?

How large a treatment effect is clinically significant?

How large a treatment effect is anticipated?

Such questions are addressed throughout the development of a new treatment.

Adaptive designs allow final changes to be made as new information is gathered during a Phase III trial.

Adaptive designs for Phase III clinical trials

In the early 2000s, industry and regulators were aware of falling success rates in late stage trials — a “statistical” solution would be very welcome indeed!

The idea of shifting from rigidly defined Phase III clinical trials to a flexible, adaptive approach was both attractive and challenging.

Sceptics asked:

- Can the results of an adaptive trial be statistically valid and credible?
- Will regulators accept adaptive designs?
- What features of a trial should be adapted?
- What are the benefits of adaptation?

Critical appraisal of the new methodologies and experience of their use has demonstrated their benefits – and the pitfalls that should be avoided.

The statistical building blocks of adaptive clinical trials

(i) A ***Combination test*** allows testing of a null hypothesis by combining data across stages of an adaptive trial (Bauer & Köhne, *Biometrics*, 1994).

Initial design

Define the null hypothesis, H_0: $\theta \leq 0$, and say a combination test will be used.

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

Stage 1

Observe P_1, the one-sided P-value for testing H_0 based on Stage 1 data.

Design Stage 2 in the light of Stage 1 data.

Stage 2

Observe P_2, the one-sided P-value for testing H_0 based on Stage 2 data.

Now combine P_1 and P_2 as if they were from independent sets of data.

Despite the data-dependent adaptation, the overall type I error rate is still protected at level α under H_0.

The statistical building blocks of adaptive clinical trials

(ii) ***Testing multiple hypotheses***

There may be reasons to change the null hypothesis or to select from a set of possible null hypotheses during a clinical trial:

Selecting one out of several versions of a treatment,

Restriction to a sub-group of the patient population,

Switching from a test of superiority to a test of non-inferiority.

Or, one may test a list of possible claims to be used in *labelling a new treatment*.

Selecting a hypothesis after observing data can inflate the false positive rate.

To avoid misleading results, tests which protect the "family-wise" type I error rate should be used when multiple hypotheses are (or may be) considered.

Then, the probability of choosing to test a particular null hypothesis and then falsely claiming significance is at most α.

Case study: A clinical trial with a survival endpoint and treatment selection

Consider a trial of cancer treatments comparing

Experimental Treatment 1: Intensive dosing

Experimental Treatment 2: Slower dosing

Control treatment

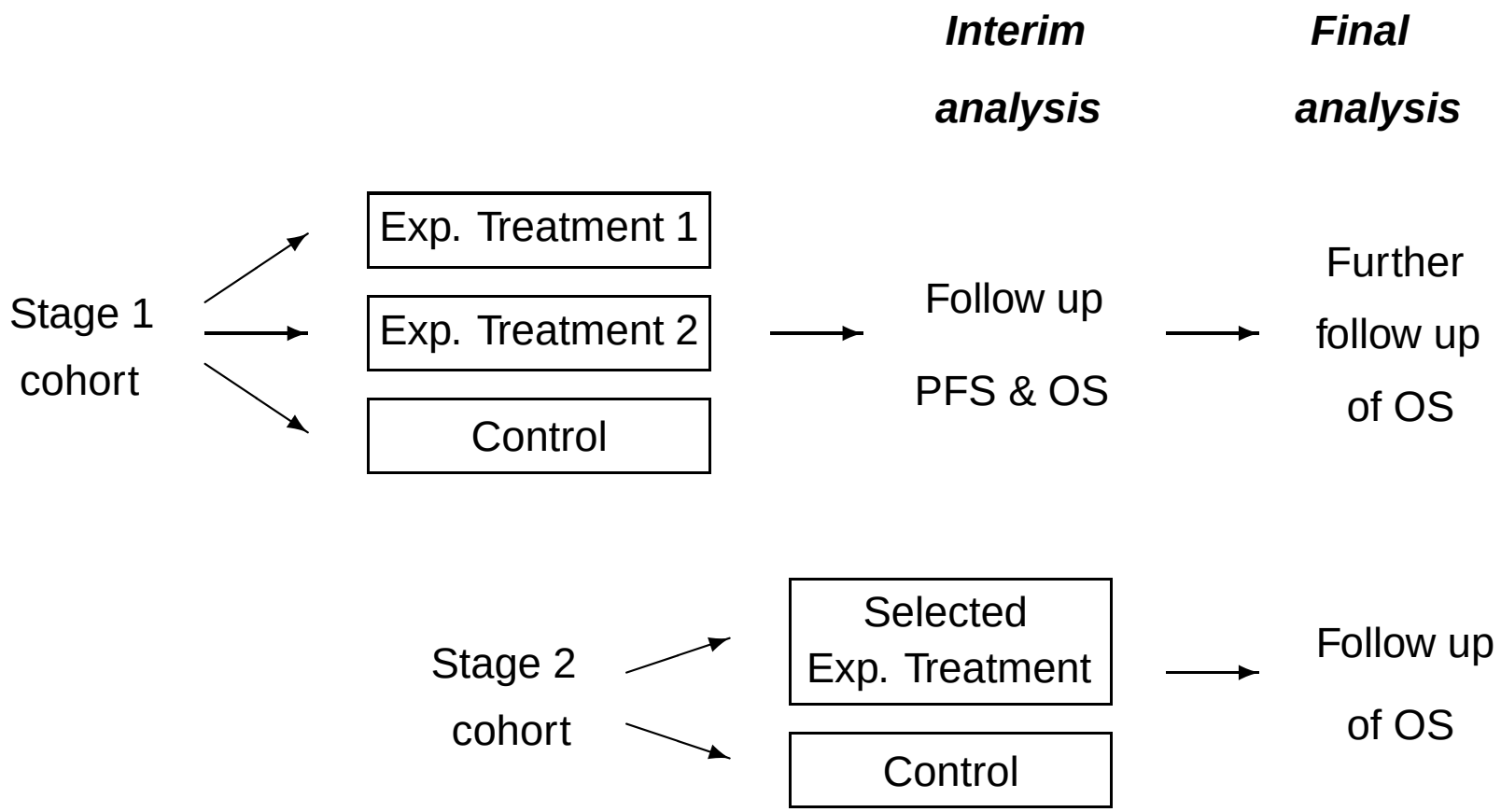
The primary endpoint is Overall Survival (OS).

Information on OS, Progression Free Survival (PFS) and safety will be used at an interim analysis to choose between the two experimental treatments.

Note that PFS is useful here as it is more rapidly observed.

After the interim analysis, patients will only be recruited to the selected treatment and the control.

Overall plan of the trial



At the final analysis, we test the null hypothesis that overall survival (OS) on the selected treatment is no better than OS on the control.

Multiple testing procedures and combination tests protect overall type I error.

Properties of the adaptive trial design

We have found this form of adaptive design to have good performance.

After studying the options for how to define the combination test and conduct multiple hypothesis tests and selecting the best of these, we found:

Compared to a non-adaptive design testing two treatments against a control, the adaptive design can have superior power to find the better treatment and conclude that this treatment is effective;

Performance depends critically on having a good volume of data when making the treatment selection at the interim analysis — this could include pharmaco-kinetic measurements and the limited OS information, as well as PFS data.

This illustrates the general principle that novel study designs should be carefully assessed before use, to optimise the design and to make sure the innovative features are worthwhile.

2. Monitoring clinical trials

A clinical trial is run to compare a new treatment with an existing treatment or placebo.

As the trial progresses, a Data and Safety Monitoring Board (DSMB) monitors patient recruitment, treatment administration, and the responses observed at interim points.

The DSMB can take actions in view of safety variables or secondary endpoints, for example, to drop a treatment arm with a high dose level if this appears unsafe.

Response on the primary endpoint may indicate early termination of the study is desirable, for either a positive or negative conclusion.

The need for special methods

Multiple looks at accumulating data can lead to over-interpretation of interim results.

Armitage, McPherson & Rowe (*J. Royal Statist. Soc, A*, 1969) report the overall false positive rate when applying repeated significance tests each with error rate 5% to accumulating data:

Number of tests	*Overall error rate*
1	5%
2	8%
3	11%
5	14%
10	19%

Clearly, a different approach is needed to avoid inflation of the type I error rate.

Formulating the problem

Let θ denote the "effect size", a measure of the improvement in the new treatment over the standard.

We shall test the null hypothesis H_0: $\theta \leq 0$ against the alternative $\theta > 0$.

Then, rejecting H_0 allows us to conclude the new treatment is better than the standard.

We allow type I error probability α for rejecting H_0 when it is actually true.

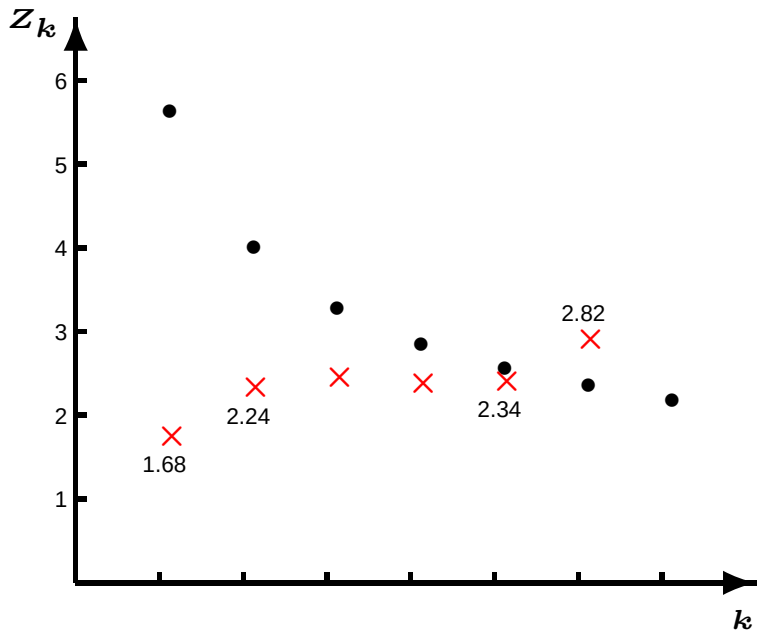
We specify power $1 - \beta$ for the probability of (correctly) rejecting H_0 when $\theta = \delta$. Here, δ is, typically, the minimal clinically significant treatment difference.

The trial design, including the method of analysis and stopping rule, must be set up to attain these error rates.

Sequential analysis of clinical trials

Group sequential methods, introduced in the late 1970s, allow early stopping for a positive or negative final decision — with proper protection of overall error rates.

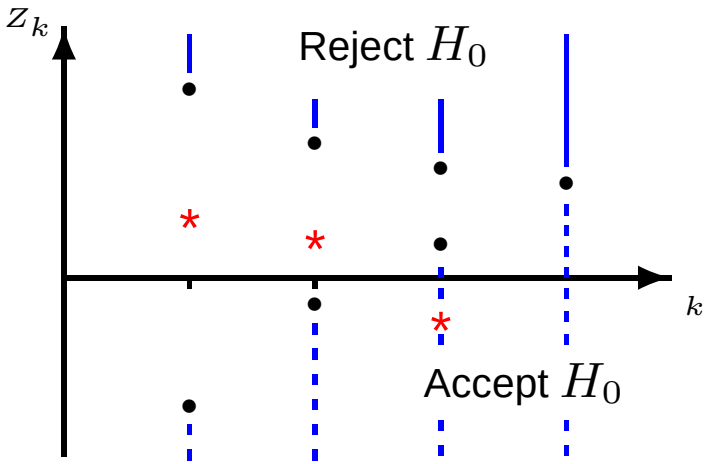
An early example, the the Beta-Blocker Heart Attack Trial, compared propanolol with placebo. (DeMets et al., *Controlled Clinical Trials*, 1984)



The trial stopped with a positive outcome after the 6th of 7 planned analyses.

Group sequential tests: Stopping for futility

Adding a lower boundary allows a trial to be stopped when there is little chance of a positive conclusion.



Rosner & Tsiatis (*Statistics in Medicine*, 1989) carried out retrospective analyses of 72 cancer studies of the U.S. Eastern Co-operative Oncology Group.

If group sequential stopping rules had been applied, early stopping (mostly for futility, i.e., to accept H_0) could have occurred in around 80% of cases.

Benefits of interim monitoring for early stopping

Consider experimental designs that provide a test of H_0: $\theta \leq 0$ against $\theta > 0$ with one-sided type I error probability 0.025 and power 0.9 when $\theta = \delta$

Suppose the response variance and the value of δ are such that a fixed sample size design needs 100 observations.

With just one interim analysis, the average sample size can be reduced to 75 – 80, depending on the true value of the treatment effect θ.

With five analyses, the average sample size can be reduced to 60 – 65, ***saving a third of the cost of the trial***.

The time until a positive result is declared is reduced accordingly.

Sequential designs need a larger maximum sample size, but 5% or 10% more than the fixed sample size is perfectly sufficient.

Sequential analysis of clinical trials

In their book *Group Sequential Methods with Applications to Clinical Trials*, Jennison & Turnbull (2000) give a unified treatment of group sequential methods, including:

General theory of group sequential analysis

Early stopping for futility or for a positive outcome

Families of designs that can be used for many types of response variable, including survival data and longitudinal data

Error spending designs that adapt to unpredictable group sizes (or, more generally, information levels)

Sample size re-estimation as nuisance parameters, such as the response variance, are estimated

Multiple endpoints or multiple treatments

Practical considerations

If a certain sample size is required to show that a new treatment is safe, one should not stop for a positive result before this sample size is reached — ***but allowing early stopping for futility is still advisable***.

Group sequential designs which monitor both safety and efficacy have been proposed (Jennison & Turnbull, *Biometrics*, 1993).

If the endpoint is observed some time after treatment, there will be "pipeline" subjects at each interim analysis who have been treated but not yet observed.

If recruitment proceeds rapidly, early stopping may not reduce sample size as the remaining subjects have already been recruited and treated (but not observed).

Hampson & Jennison (*J. Royal Statist. Soc, B*, 2013) introduced new types of group sequential design for a delayed response, and proposed ways to achieve efficiency gains closer to those seen for an immediate response.

Practical considerations

In creating a clinical trial design, one needs to balance costs and benefits.

If time until a positive outcome is of prime importance, there may be less concern about the cost of “pipeline” subjects for a delayed response.

The general theory of group sequential design applies to a wide variety of data types and experimental designs.

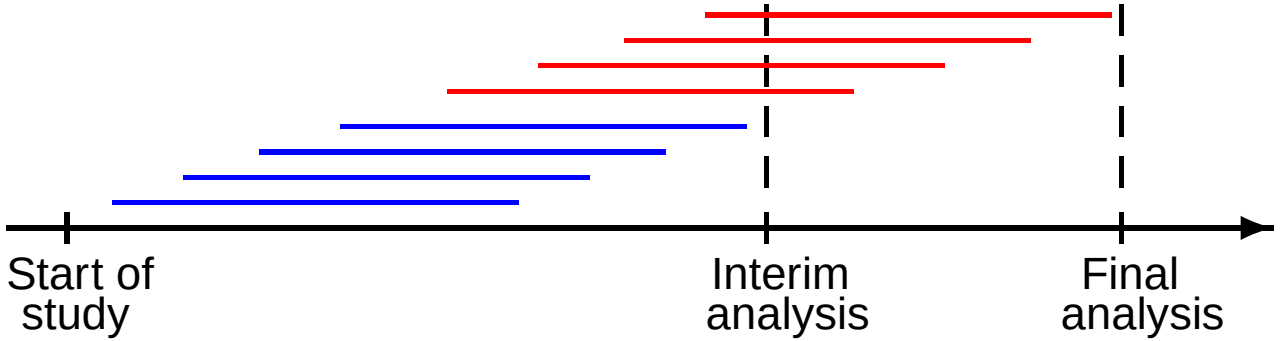
Longitudinal studies and crossover trials fall into this general framework, but there are fewer examples of the application of group sequential designs to draw on.

Practical issues of patient recruitment and ethical considerations have implications for how clinical trials can be conducted.

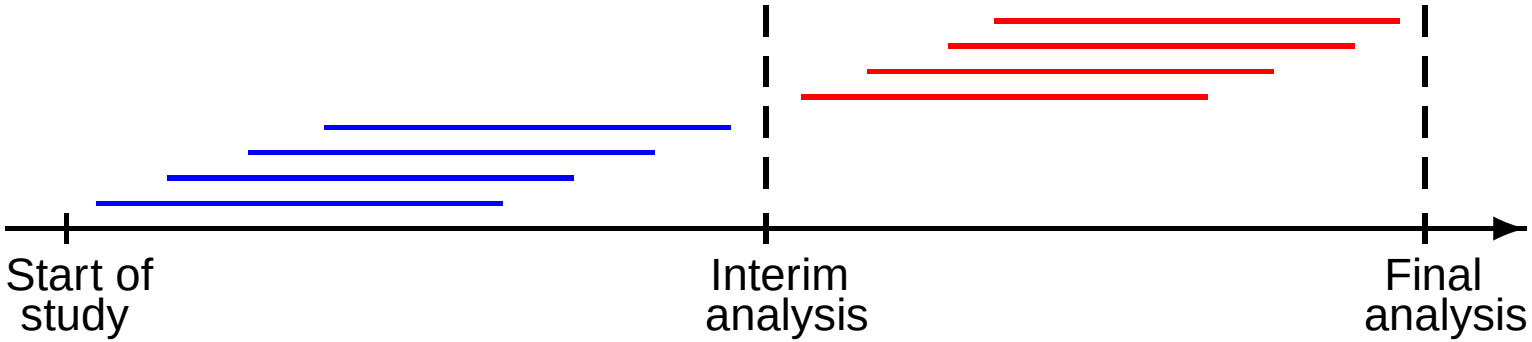
However, there may be greater freedom in the choice of trial design when testing other types of product.

Pausing recruitment: A block sequential design

In multi-centre clinical trials, a break in recruitment is problematic as centres may not return to the study when recruitment re-commences — but with continuous recruitment, there are pipeline subjects at the interim analysis.



If recruitment can be halted to observe responses, we avoid pipeline subjects: the trial will take a little longer, but have smaller average sample size.



3. Planning several stages of a development programme jointly

Trials in the different phases of drug development can be optimised individually.

It is also sensible — and perhaps much more important — to consider the design of the overall development process.

In drug development, the dose selected in Phase II determines the likelihood of success in Phase III.

The confidence in a selected dose affects the choice of Phase III sample size, and also whether more than one dose should be tested in Phase III.

Knowledge gained prior to Phase II should help in setting the degree of positive evidence required to proceed to Phase III.

Several research groups are tackling these over-arching design problems.

Problem formulation requires many assumptions. It may not be easy to give the probabilities or costs and benefits needed for this analysis — but companies do currently make decisions that appear to depend on these unknowns!

4. Conclusions

1. Experience over 50 or 60 years has led to reliable and well understood methodology for the successive stages of drug development.

2. Use of some types of experimental design can be constrained by logistical factors or the special issues in trials that treat patients with a serious illness.

3. At the turn of the millennium, adaptive methods caused great excitement and offered hope of a step change in the success rate of clinical trials.

4. We now have

   Clearer understanding of the benefits of adaptive designs,

   Practical experience of conducting adaptive trials.

5. Many new types of clinical trial design have potential for other applications.

   For some methods, further development may be needed; in other cases with fewer constraints, application may be simpler and benefits greater.