Optimising Group Sequential Designs:

Special considerations for small populations

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

European Statistical Forum:

Statistical Methods for Rare Diseases

and Special Populations

November 2017

Motivation

Consider how you would formulate a hypothesis testing problem when designing a clinical trial.

Should the formulation be different when the trial concerns treatments for members of a small population?

We shall consider how to find efficient group sequential designs for a traditional testing formulation and optimality criterion.

We then propose a different formulation for small populations.

We obtain qualitatively designs in the two cases, raising questions about some fundamental issues.

Plan of talk

- Monitoring clinical trials
- Sequential distribution theory
- An optimal stopping problem
- Numerical evaluation of stopping boundaries
- Finding optimal group sequential designs
- Designs for rare diseases
- Conclusions

For further reading, see Jennison and Turnbull (*Kuwait J. of Science*, 2013)

CHRISTOPHER JENNISON[1] AND BRUCE W. TURNBULL[2]

[1]*Department of Mathematical Sciences, University of Bath, Bath, U.K.*

[2]*Department of Operations Research and Information Engineering, Cornell University, Ithaca, U.S.A*

ABSTRACT

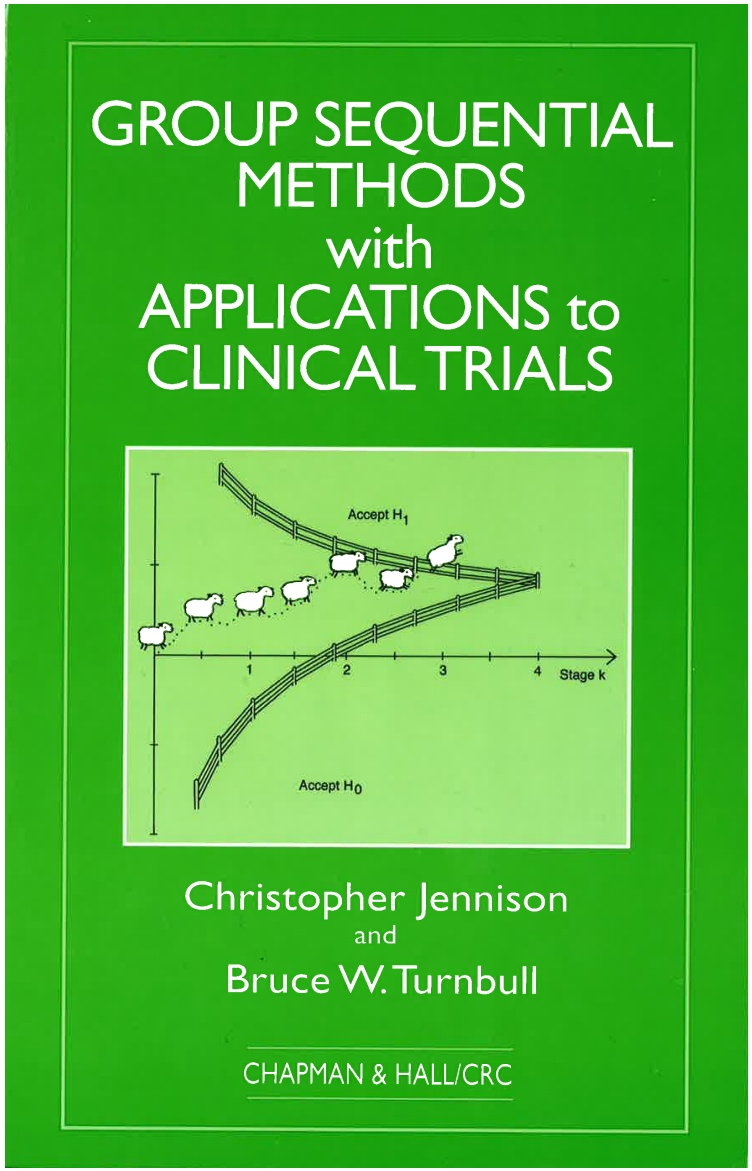
It is standard practice to monitor clinical trials with a view to stopping early if results are sufficiently compelling. We explain how the properties of stopping boundaries can be calculated numerically and how to optimise boundaries to minimise expected sample size while controlling type I and II error probabilities. Our optimisation method involves the use of dynamic programming to solve Bayes decision problems with no constraint on error rates. This conversion to an unconstrained problem is equivalent to using Lagrange multipliers. Applications of these methods in clinical trial design include the derivation of optimal adaptive designs in which future group sizes are allowed to depend on previously observed responses; designs which test both for superiority and non-inferiority; and group sequential tests which allow for a delay between treatment and response.

Keywords: Clinical trial; group sequential test; Bayes decision problem; dynamic programming; optimal stopping.

INTRODUCTION

It is natural to wish to examine data as they accumulate during the course of a long-term clinical trial. However, with frequent looks at the data, there is greater opportunity to make an erroneous decision. Armitage *et al.* (1969) report the overall type I error rate when applying repeated two-sided significance tests at $\alpha = 0.05$ to accumulating data and show this rises to 0.11 with 3 analyses and 0.14 with 5 analyses. Thus, special statistical methods are required to avoid

And for even more, see



1. 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 that 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 et al. (*JRSS, A*, 1969) report the overall type I error rate when applying repeated significance tests at level $\alpha = 0.05$ to accumulating data:

Number of tests	*Error rate*
1	0.05
2	0.08
3	0.11
5	0.14
10	0.19

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 $\theta > 0$.

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.

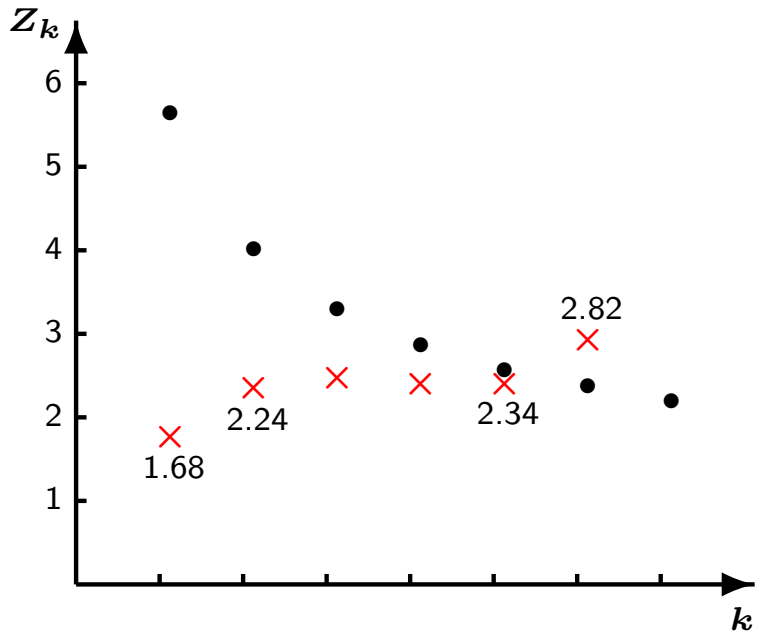
We specify power $1 - \beta$ for the probability of (correctly) rejecting H_0 when $\theta = \delta$. Here, δ is, typically, the minimal clinically significant treatment difference.

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

An early example: The BHAT trial

DeMets et al. (*Cont. Clin. Trials*, 1984) report on the Beta-Blocker Heart Attack Trial, that compared propanolol with placebo.

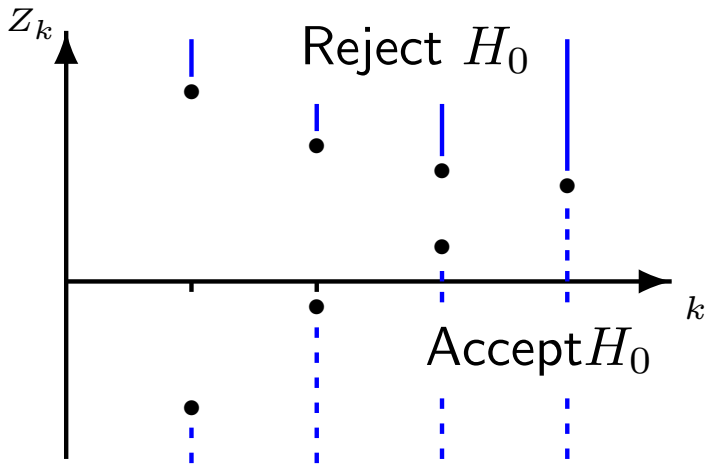
An "O'Brien and Fleming" stopping boundary was defined with overall type I error probability 0.025.



The trial stopped after the 6th of 7 planned analyses.

Group sequential tests: Stopping for futility

Adding a lower boundary allows stopping 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.

Had group sequential stopping rules been applied, early stopping (mostly to accept H_0) would have occurred in $\sim$80% of cases.

Requirements for clinical trial designs

We seek designs which:

Achieve specified type I error rate and power,

Stop early, on average, under key parameter values,

Can be applied to a variety of response types.

We shall present distribution theory which shows that a common set of methods can be applied to many data types.

To define efficient tests, we shall formulate and solve an optimal stopping problem.

2. Sequential distribution theory

Our interest is in the parameter for the treatment effect, θ.

Let $\widehat{\theta}_k$ denote the estimate of θ based on data at analysis k.

The information for θ at analysis k is

$$\mathcal{I}_k = \{\mathsf{Var}(\widehat{\theta}_k)\}^{-1}, \quad k = 1, \ldots, K.$$

Canonical joint distribution of $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$

In many situations, $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$ are approximately multivariate normal,

$$\widehat{\theta}_k \sim N(\theta, \{\mathcal{I}_k\}^{-1}), \quad k = 1, \ldots, K,$$

and

$$\mathsf{Cov}(\widehat{\theta}_{k_1}, \widehat{\theta}_{k_2}) = \mathsf{Var}(\widehat{\theta}_{k_2}) = \{\mathcal{I}_{k_2}\}^{-1} \quad \text{for } k_1 < k_2.$$

Sequential distribution theory

The preceding result about the joint distribution of $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$ can be demonstrated directly for:

θ *a single normal mean,*

$\theta = \mu_A - \mu_B$, *comparing two normal means.*

The results also apply when θ is a parameter in:

a general normal linear model,

a model fitted by maximum likelihood (large sample theory),

a Cox proportional hazards regression model for survival data.

Thus, theory supports general comparisons, including:

crossover trials, studies with longitudinal data,

analyses with covariate adjustment.

Canonical joint distribution of score statistics

The general theory implies that *score statistics*, $S_k = Z_k\sqrt{\mathcal{I}_k}$, are multivariate normal with

$$S_k \sim N(\theta\,\mathcal{I}_k, \mathcal{I}_k), \quad k = 1, \ldots, K.$$

The score statistics have the "independent increments" property

$$\text{Cov}(S_k - S_{k-1},\, S_{k'} - S_{k'-1}) = 0 \quad \text{for } k \neq k'.$$

It can be helpful to know that the score statistics behave as Brownian motion with drift θ observed at times $\mathcal{I}_1, \ldots, \mathcal{I}_K$.

References:

Jennison & Turnbull, *JASA*, 1997; Scharfstein et al, *JASA*, 1997.

3. An optimal stopping problem

Consider a trial designed to test H_0: $\theta \leq 0$ vs $\theta > 0$, with:

- Type I error rate α,
- Power $1 - \beta$ at $\theta = \delta$,
- Up to K analyses.

A fixed sample test needs information

$$\mathcal{I}_{fix} = \{\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)\}^2/\delta^2.$$

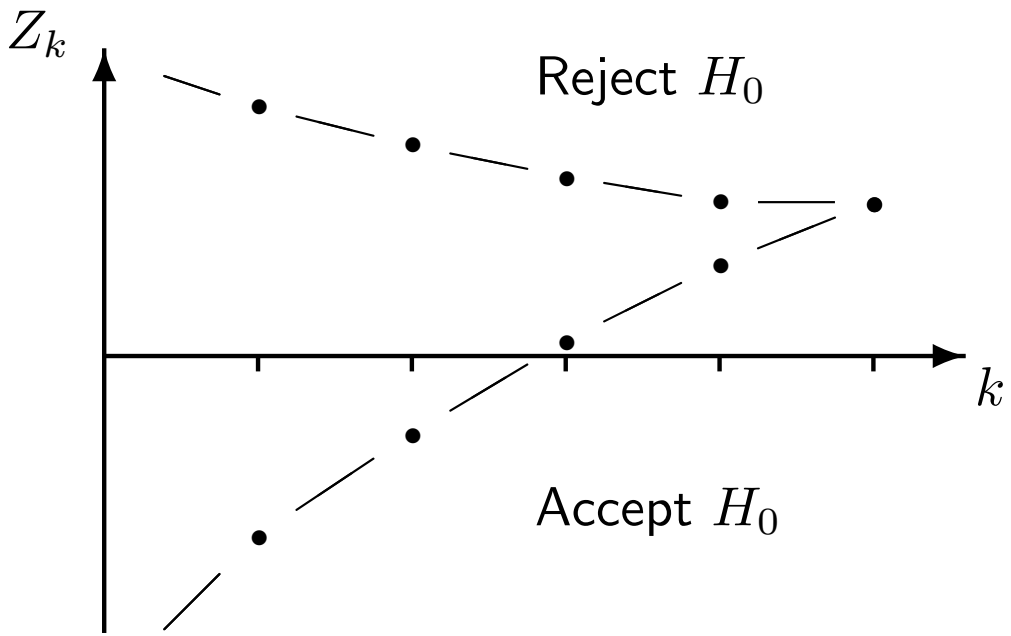
We set the maximum information to be

$$\mathcal{I}_{max} = R\mathcal{I}_{fix},$$

where $R > 1$, with equal increments between analyses.

Optimal group sequential tests

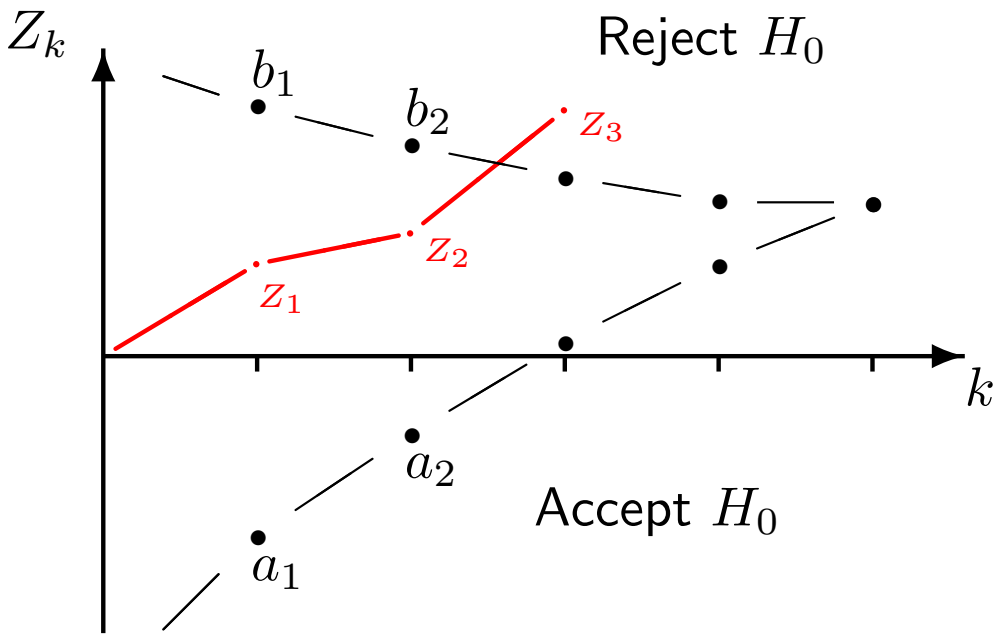
The error rates impose two constraints on the $2K - 1$ boundary points — leaving a high dimensional space of possible boundaries.



We shall look for a boundary with an optimality property, specifically, minimising

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2.$$

4. Computations for group sequential tests



We need to be able to calculate the probabilities of basic events such as

$$a_1 < Z_1 < b_1, \quad a_2 < Z_2 < b_2, \quad Z_3 > b_3.$$

Combining such probabilities gives key properties, such as $Pr_\theta\{\text{Reject } H_0\}$ and $E_\theta(\mathcal{I})$.

Numerical integration

We can write probabilities as nested integrals, e.g.,

$$Pr\{a_1 < Z_1 < b_1,\, a_2 < Z_2 < b_2,\, Z_3 > b_3\} =$$

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{b_3}^{\infty} f_1(z_1)\, f_2(z_2|z_1)\, f_3(z_3|z_2)\, dz_3\, dz_2\, dz_1.$$

Applying numerical integration, we replace each integral by a sum of the form

$$\int_a^b f(z)\, dz \;=\; \sum_{i=1}^{n} w(i)\, f(z(i)),$$

where $z(1), \ldots, z(n)$ is a grid of points from a to b.

Numerical integration

Thus, we have

$$Pr\{a_1 < Z_1 < b_1,\, a_2 < Z_2 < b_2,\, Z_3 > b_3\} \approx$$

$$\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$w_3(i_3)\; f_3(z_3(i_3)|z_2(i_2)).$$

Multiple integrations and summations will arise, e.g., for an outcome at analysis k,

$$\sum_{i_1=1}^{n_1} \ldots \sum_{i_k=1}^{n_k} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$\ldots\; w_k(i_k)\; f_k(z_k(i_k)|z_{k-1}(i_{k-1})).$$

Numerical integration

In the multiple summation

$$\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_k=1}^{n_k} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$\cdots\; w_k(i_k)\; f_k(z_k(i_k)|z_{k-1}(i_{k-1})),$$

the structure of the k nested summations is such that the computation required is of the order of $k-1$ double summations.

Using Simpson's rule with 100 to 200 grid points per integral can give accuracy to 5 or 6 decimal places.

For details of efficient sets of grid points, see Ch. 19 of *Group Sequential Methods with Applications to Clinical Trials* by Jennison and Turnbull (2000).

5. Finding optimal group sequential tests

Recall, we want a group sequential test of H_0: $\theta \leq 0$ vs $\theta > 0$ with

$$Pr_{\theta=0}\{\text{Reject } H_0\} = \alpha,$$

$$Pr_{\theta=\delta}\{\text{Accept } H_0\} = \beta,$$

Analyses at $\mathcal{I}_k = (k/K)\,\mathcal{I}_{max}, \quad k = 1, \ldots, K,$

Minimum possible value of $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2.$

We deal with constraints on error rates by introducing Lagrangian multipliers to create the *unconstrained problem* of minimising

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2 + \lambda_1 Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\}.$$

We shall find a pair of multipliers (λ_1, λ_2) such that the solution has type I and II error rates α and β, then this design will solve the *constrained problem* too.

Bayesian interpretation of the Lagrangian approach

If we put a prior on θ with $Pr\{\theta = 0\} = Pr\{\theta = \delta\} = 0.5$ and specify costs of

$$\begin{array}{ll} 1 & \text{per unit of information observed,} \\ 2\,\lambda_1 & \text{for rejecting } H_0 \text{ when } \theta = 0, \\ 2\,\lambda_2 & \text{for accepting } H_0 \text{ when } \theta = \delta, \end{array}$$

then, the total Bayes risk is

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2 + \lambda_1\, Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\},$$

just as in the Lagrangian problem.

An advantage of the Bayes interpretation is that it can give insight into solving the problem by techniques of "Dynamic Programming" or "Backwards Induction".

Solution by Dynamic Programming

Denote the posterior distribution of θ given $Z_k = z_k$ at analysis k by

$$p^{(k)}(\theta|z_k), \quad \theta = 0,\, \delta.$$

At the final analysis, K

There is no further sampling cost, so compare decisions

$$\text{Reject } H_0: \qquad E(\text{Cost}) = 2\,\lambda_1\, p^{(K)}(0|z_K),$$

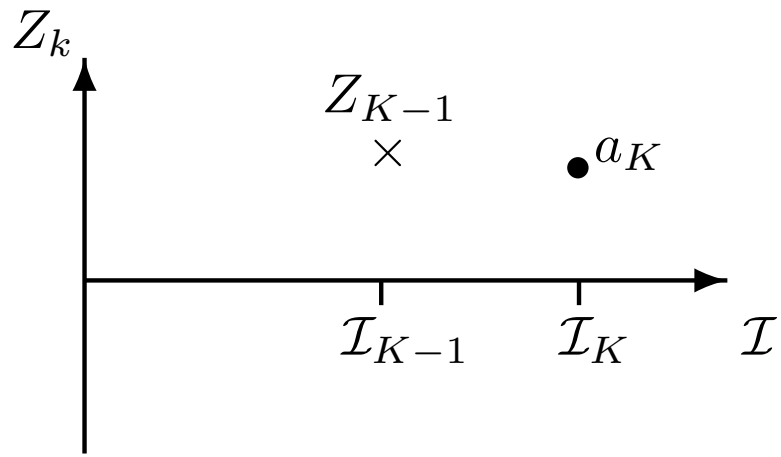
$$\text{Accept } H_0: \qquad E(\text{Cost}) = 2\,\lambda_2\, p^{(K)}(\delta|z_K).$$

The boundary point a_K is the value of z_K where these expected losses are equal.

The optimum decision rule is to reject H_0 for $Z_K > a_K$.

Dynamic Programming

At analysis $K-1$



If the trial stops at this analysis, there is no further cost of sampling and the expected additional cost is

$$\text{Reject } H_0: \quad 2\,\lambda_1\,p^{(K-1)}(0|z_{K-1}),$$

$$\text{Accept } H_0: \quad 2\,\lambda_2\,p^{(K-1)}(\delta|z_{K-1}).$$

At analysis $K-1$

If the trial continues to analysis K, the expected additional cost is

$$1 \times (\mathcal{I}_K - \mathcal{I}_{K-1})$$

$$+ 2\,\lambda_1\, p^{(K-1)}(0|z_{K-1})\, Pr_{\theta=0}\{Z_K > a_K | Z_{K-1} = z_{K-1}\}$$

$$+ 2\,\lambda_2\, p^{(K-1)}(\delta|z_{K-1})\, Pr_{\theta=\delta}\{Z_K < a_K | Z_{K-1} = z_{K-1}\}.$$

We can now define the optimal boundary points:

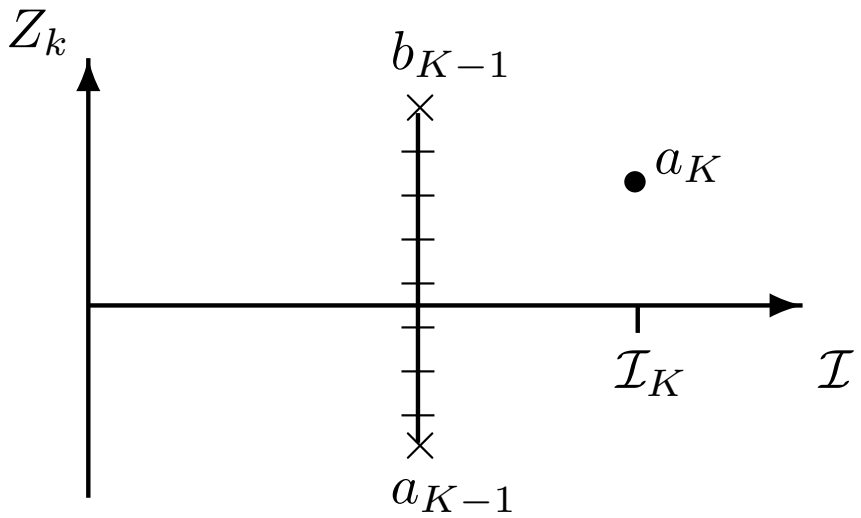
Set b_{K-1} to be the value of z_{K-1} where

$$E(\text{Cost of continuing}) \;=\; E(\text{Cost of stopping to reject } H_0).$$

Set a_{K-1} to be the value of z_{K-1} where

$$E(\text{Cost of continuing}) \;=\; E(\text{Cost of stopping to accept } H_0).$$

At analysis $K-1$



Before leaving analysis $K-1$, we set up a grid of points for use in numerical integration over the range a_{K-1} to b_{K-1}.

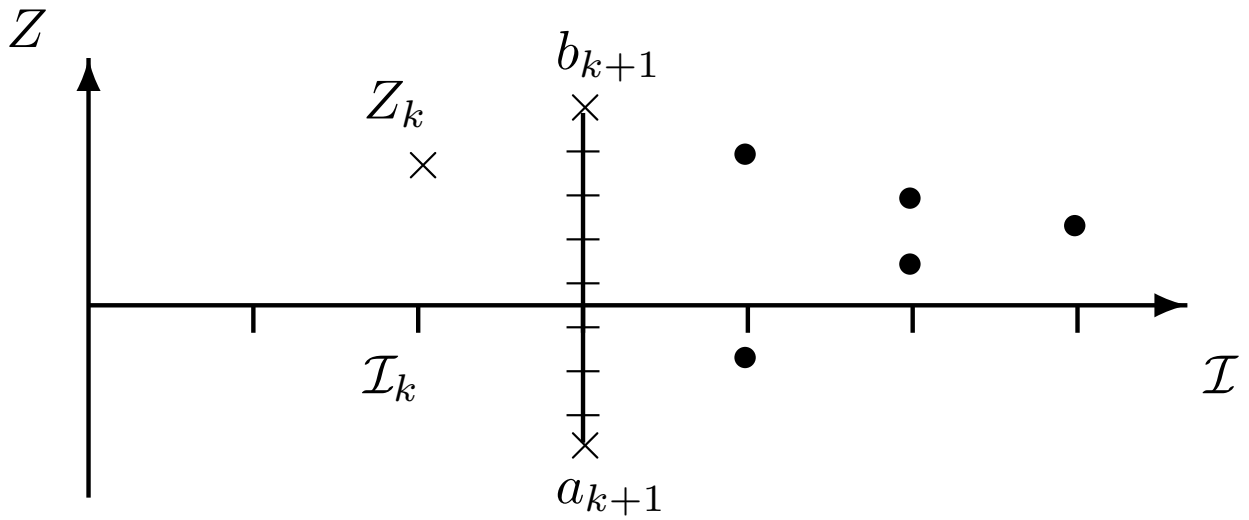
For each point, we sum over the posterior distribution of θ to calculate

$$\beta^{(K-1)}(z_{K-1}) = E(\text{Additional cost when continuing} \,|\, Z_{K-1} = z_{K-1}).$$

We are now ready to move back to analysis $K-2$.

Analyses 1 to $K-2$

We work back through analyses $k = K-2,\ K-3,\ \ldots,\ 1$.



At each analysis, we find the optimal stopping boundary using knowledge of the optimal stopping rule at future analyses.

Then, for a grid of values of z_k, compute

$$\beta^{(k)}(z_k) = E(\text{Additional cost when continuing} \,|\, Z_k = z_k)$$

to use in evaluating the option of continuing at analysis $k-1$.

Solving the original problem

Now, for any given (λ_1, λ_2) we can find the Bayes optimal design and compute its type I and II error rates.

We add another layer to the problem and search for a pair (λ_1, λ_2) for which type I and type II error rates of the optimal design equal α and β respectively.

The resulting design will be the optimal group sequential test, with the specified frequentist error rates, for our original problem.

Notes

1. The method of solving the overall, constrained problem provides an explicit demonstration that good frequentist procedures should be similar to Bayes procedures.

2. The prior and costs in the final Bayes problem may not reflect investigators' beliefs about the treatment effect, costs of type I and type II errors, or costs of treating patients in the trial.

Properties of optimal designs

Tests with $\alpha = 0.025$, $1 - \beta = 0.9$, K analyses, $\mathcal{I}_{max} = R\mathcal{I}_{fix}$, equal group sizes, minimising $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2$.

Minimum values of $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2$, as a percentage of $\mathcal{I}_{fix}$

	R					
K	1.01	1.05	1.1	1.2	1.3	*Minimum over R*
2	80.8	74.7	73.2	73.7	75.8	73.0 at R=1.13
5	72.2	65.2	62.2	59.8	59.0	58.8 at R=1.38
10	69.2	62.2	59.0	56.3	55.1	54.2 at R=1.6
20	67.8	60.6	57.5	54.6	53.3	51.7 at R=1.8

Observe: $E(\mathcal{I}) \searrow$ as $K \nearrow$ but with diminishing returns,

$E(\mathcal{I}) \searrow$ as $R \nearrow$ up to a point.

Generalisations

- Other optimality criteria such as a weighted sum

$$\sum_i w_i E_{\theta_i}(\mathcal{I})$$

or an integral

$$\int f(\theta)\, E_\theta(\mathcal{I})\, d\theta$$

- Optimising a set of fixed group sizes in a group sequential test
- Data dependent group sizes in a group sequential test
- Group sequential tests for a delayed response
- Testing for either superiority or non-inferiority

6. Design of trials for rare diseases

Joint work with Sebastian Jobjörnsson, Carl-Fredrik Burman and Stephen Senn

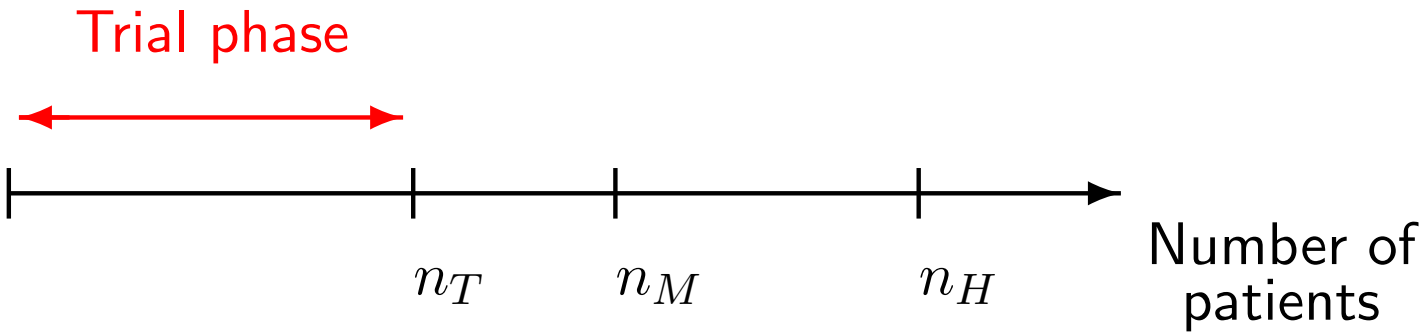
In the case of a rare disease, patient numbers may not be high enough to conduct a clinical trial with low type I error probability and high power.

Moreover, for a very rare disease, almost all patients with the disease may be in the trial.

Then, attention may focus on treating the patients in the trial as well as possible — rather than providing overwhelming evidence that one or other treatment is superior.

We can formulate a decision problem that has such a goal, and use dynamic programming to solve this problem.

Design of trials for rare diseases



Consider the treatment of the next n_H patients.

A group sequential clinical trial, with a maximum of n_M patients, will be conducted to compare treatments A and B.

On the conclusion of the trial, after observing responses for n_T patients, one treatment will be selected and administered to the remaining $n_H - n_T$ patients.

Here, H stands for "horizon". The horizon is intended to represent a future time point when a new treatment, which supersedes both treatments A and B, is likely to have been developed.

Design of trials for rare diseases: Problem formulation

As before, let θ denote the difference in $E(\text{Response})$ between treatments A and B , with $\theta > 0$ when B is superior to A.

We assume a prior for θ has been elicited.

We specify the objective of maximising the expected sum of the responses for all n_H patients.

Let N_A and N_B be random variables denoting the numbers of patients receiving treatments A and B, so $N_A + N_B = n_H$.

If the average of the expected responses on treatments A and B is μ, we wish to maximise

$$E\{N_A(\mu - \theta/2)\} + E\{N_B(\mu + \theta/2)\}$$

$$= \text{ constant } - \frac{1}{2} E(N_A\theta) + \frac{1}{2} E(N_B\theta),$$

where expectation is over the prior distribution of θ as well as the distributions of N_A and N_B.

Design of trials for rare diseases: Example

Suppose patient responses are independent, normally distributed with variance $\sigma^2 = 5$.

The prior distribution for θ is

$$\theta \sim N(0, 1).$$

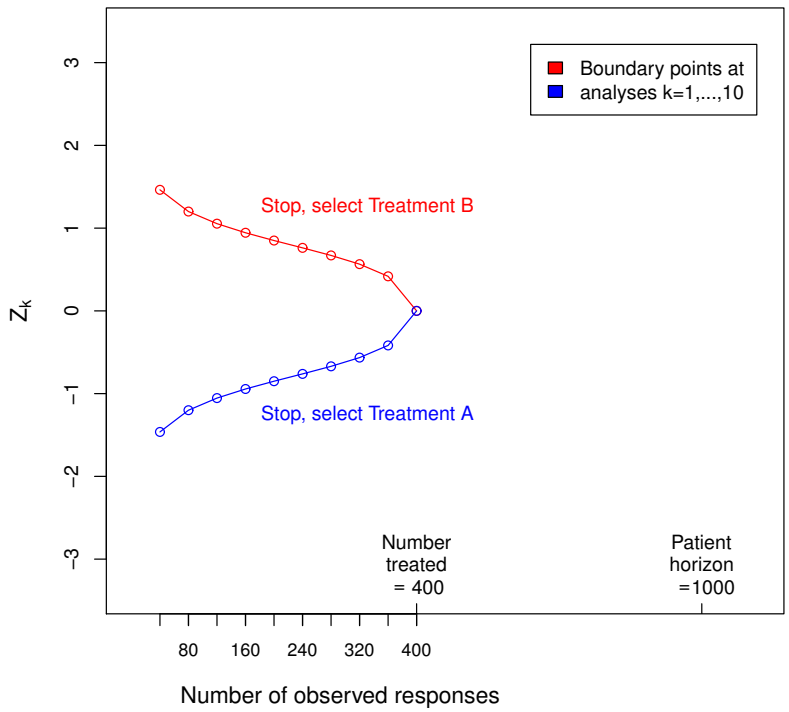
The total number of patients to be treated is $n_H = 1000$.

The trial stage has a maximum of $n_M = 400$.

The trial is conducted group sequentially with a maximum of 10 analyses.

We have applied dynamic programming to find the optimal stopping rule for the trial.

Design of trials for rare diseases: Optimal stopping rule



Note that the Z values at the boundary are smaller than one would expect for a typical test of H_0: $\theta = 0$.

Anscombe (*JASA*, 1963) proposed the "horizon problem".

Petkau (*IMS Lecture Notes, Vol 42*, 2003) presented group sequential horizon designs.

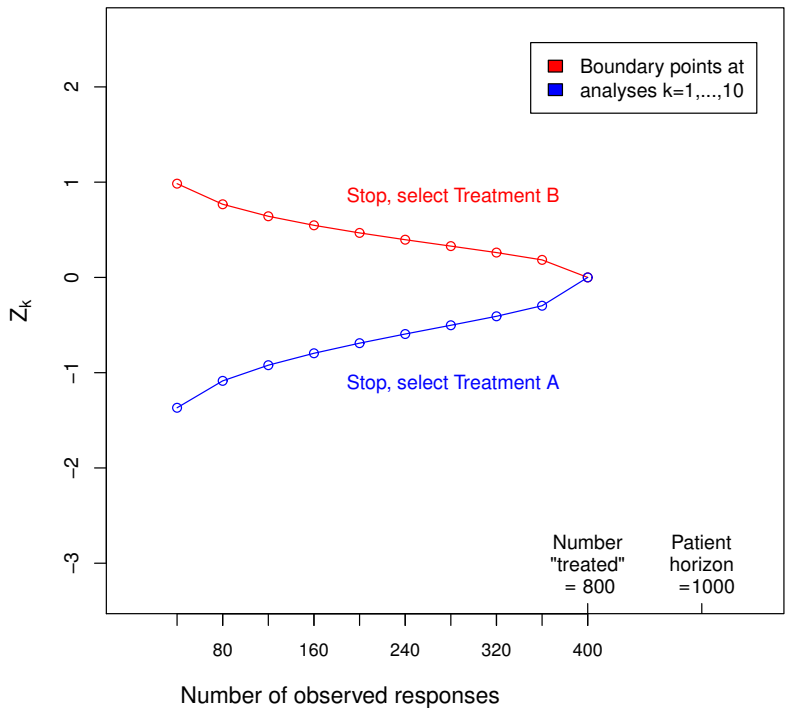
Clinical trials for rare diseases could be a realistic application area for horizon problem designs.

Design of trials for rare diseases: Optimal stopping rule

Now suppose that for every patient in the trial, there is another patient not in the trial, who receives treatment A.

The horizon of 1000 includes all patients, in or out of the trial.

Once the trial stage ends, **all** future patients receive the treatment chosen in the trial.



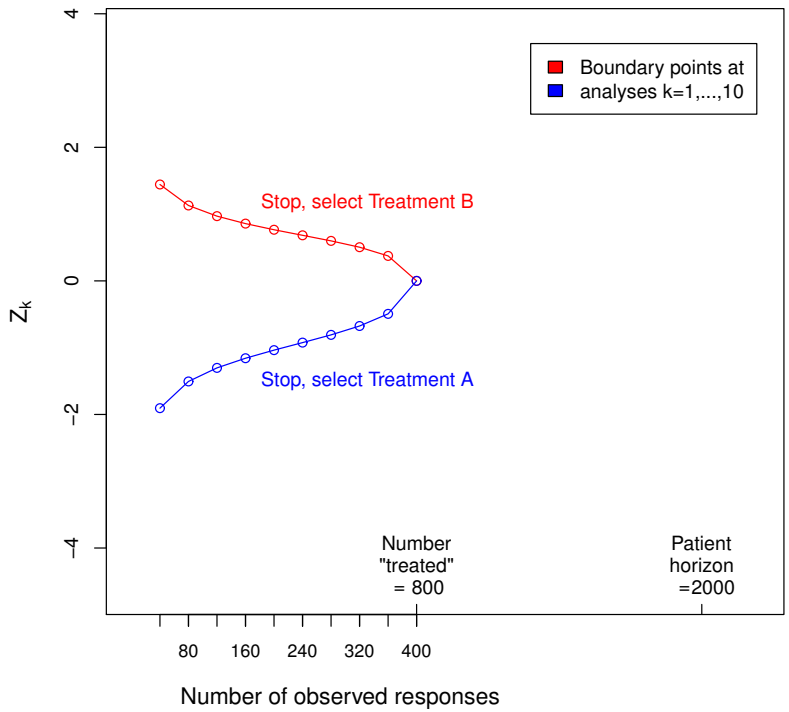
Boundaries for Z values are lower than in the previous case which had no patients outside the trial.

The boundaries are no longer symmetric about zero.

Design of trials for rare diseases: Optimal stopping rule

Again, we suppose that for every patient in the trial, there is another patient not in the trial, who receives treatment A.

But now we increase the horizon to 2000 patients.



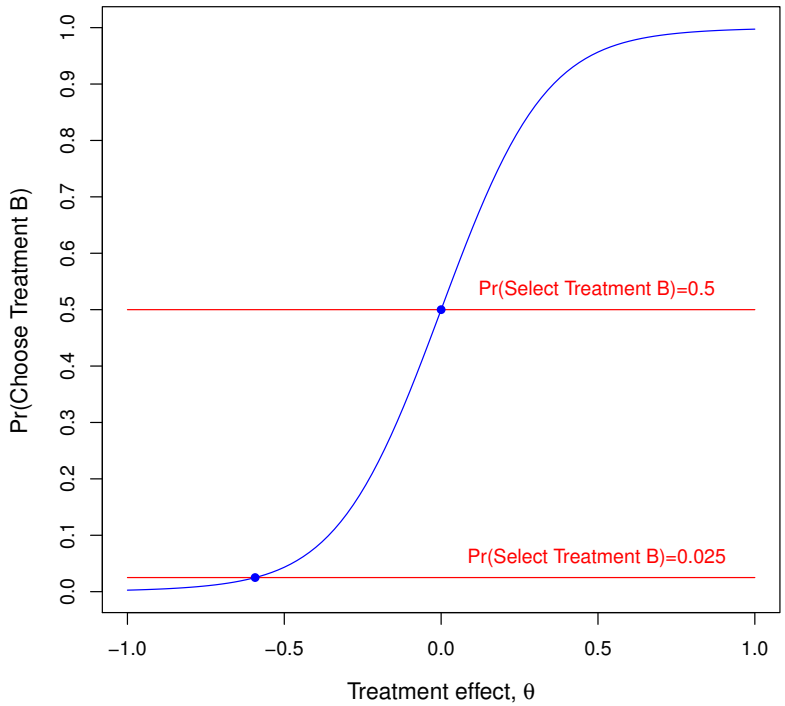
Boundaries for Z values are similar to those in the first example, which had no patients outside the trial and a horizon of 1000.

Design of trials for rare diseases: Optimal stopping rule

We return to the first example with no patients being treated outside of the trial and a horizon of $n_H = 1000$.

Rather than a "power curve", we plot the probability of choosing Treatment B as a function of θ.

When $\theta = 0$, there is no difference between the two treatments and Pr(Select Treatment B)=0.5.



When $\theta = -0.62$,

Pr(Choose Tr B)=0.025.

We might view this as a level $\alpha = 0.025$ test of non-inferiority of B vs A, with non-inferiority margin $= 0.62$.

7. Conclusions

- The monitoring of clinical trials poses a range of problems of statistical inference and optimal design, but general distribution theory provides a basis for generic methodology.
- Using Dynamic Programming to solve specially constructed Bayes decision problems provides a route to deriving optimal frequentist group sequential designs.
- Such optimal procedures serve as benchmarks for other methods which have additional useful features.
- The "horizon problem" formulation appears to be suited to trials for rare diseases.
- The resulting procedures are not easily interpreted as tests of H_0: $\theta = 0$, but they may be viewed as testing between the hypotheses $\theta = -\delta^*$ and $\theta = +\delta^*$ for certain δ^*.