Improving Adaptive Designs

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

8th Annual Adaptive Designs in Clinical Trials

London, April 2016

Adaptive designs

Adaptive trial designs have been proposed for a number of important applications:

Sample size modification,

Treatment selection and testing (seamless Phase 2/3 trials),

Population selection and testing (enrichment designs).

There are usually options to choose from within such a design.

How should one make such choices and assess the end result?

Choosing an adaptive design

The first requirement for any Phase 3 trial is to protect the type I error rate.

This can be a complex problem when there are multiple null hypotheses under consideration — and multiple parameters, so the type I error rate must be controlled over a high-dimensional region.

One then wishes to be efficient, gaining high power with low sample size.

Question

How should one make decisions:

At interim analyses,

At the final analysis.

Outline of talk

1. Sample size modification

   A two-stage trial design with delayed response,

   Optimising a sample size modification rule.

2. Seamless Phase 2/3 designs

   Designs that protect family-wise error rate,

   Optimising decision rules and sample size allocation.

3. Enrichment designs

   Adaptive enrichment in response to interim data.

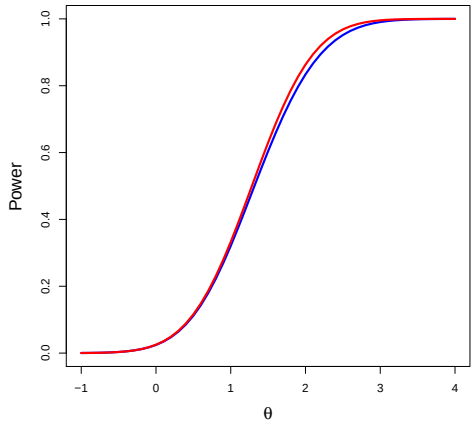
   Optimising the decision rule for when to enrich.

4. Conclusions

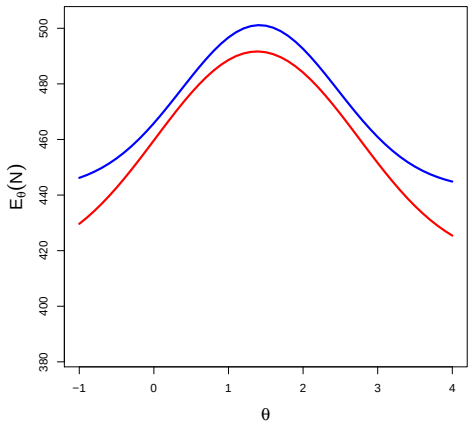
1. Deriving an efficient rule for sample size modification

All designs have overall power and $E_\theta(N)$ curves.

Power curve



$E_\theta(N)$ curves



Designs with similar power curves can be compared in terms of their average sample size functions, $E_\theta(N)$.

Even if they are uncertain about the likely treatment effect, investigators can usually specify values of θ under which early stopping is most desirable.

Thus, we shall define efficiency in terms of power and $E_\theta(N)$.

An example with a delayed response

We consider the design of a clinical trial that forms Example 1 of Mehta & Pocock (*Statistics in Medicine*, 2011)

> "Adaptive increase in sample size when interim results are promising: A practical guide with examples".

In this example, response is measured some time after treatment.

Thus, at an interim analysis, many patients have been treated but are yet to produce a response.

Delayed responses are common — and not easily dealt with by standard group sequential tests (but see Hampson & Jennison, *JRSS B*, 2013).

For an extended discussion of Mehta & Pocock's example, see Jennison & Turnbull (*Statistics in Medicine*, 2015).

Research Article

Statistics in Medicine

Received 10 November 2014, Accepted 2 June 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6575

Adaptive sample size modification in clinical trials: start small then ask for more?

Christopher Jennison[a*†] and Bruce W. Turnbull[b]

We consider sample size re-estimation in a clinical trial, in particular when there is a significant delay before the measurement of patient response. Mehta and Pocock have proposed methods in which sample size is increased when interim results fall in a 'promising zone' where it is deemed worthwhile to increase conditional power by adding more subjects. Our analysis reveals potential pitfalls in applying this approach. Mehta and Pocock use results of Chen, DeMets and Lan to identify when increasing sample size, but applying a conventional level α significance test at the end of the trial does not inflate the type I error rate: we have found the greatest gains in power per additional observation are liable to lie outside the region defined by this method. Mehta and Pocock increase sample size to achieve a particular conditional power, calculated under the current estimate of treatment effect: this leads to high increases in sample size for a small range of interim outcomes, whereas we have found it more efficient to make moderate increases in sample size over a wider range of cases. If the aforementioned pitfalls are avoided, we believe the broad framework proposed by Mehta and Pocock is valuable for clinical trial design. Working in this framework, we propose sample size rules that apply explicitly the principle of adding observations when they are most beneficial. The resulting trial designs are closely related to efficient group sequential tests for a delayed response proposed by Hampson and Jennison.

Keywords: group sequential test; sample size re-estimation; adaptive design; clinical trial; optimal design; promising zone

Example 1 of Mehta & Pocock (MP)

MP's Example 1 concerns a Phase 3 trial of a new treatment for schizophrenia, comparing the new drug to an active control.

The efficacy endpoint is improvement in the Negative Symptoms Assessment score from baseline to week 26.

Responses are

$$Y_{Bi} \sim N(\mu_B, \sigma^2), \;\; i = 1, 2, \ldots, \;\; \text{on the new treatment,}$$

$$Y_{Ai} \sim N(\mu_A, \sigma^2), \;\; i = 1, 2, \ldots, \;\; \text{on the control arm,}$$

where $\sigma^2 = 7.5^2$.

The treatment effect is

$$\theta \;=\; \mu_B - \mu_A.$$

and we estimate θ by

$$\widehat{\theta} \;=\; \widehat{\mu}_B - \widehat{\mu}_A \;=\; \overline{Y}_B - \overline{Y}_A.$$

Mehta & Pocock's Example 1

The initial plan is for $n_2 = 442$ patients, 221 on each treatment.

In testing H_0: $\theta \leq 0$ vs $\theta > 0$ at the final analysis, we reject H_0 if

$$Z_2 = \frac{\hat{\theta}(n_2)}{\sqrt{\{4\sigma^2/n_2\}}} > 1.96.$$

This design and analysis gives type I error rate 0.025 and power 0.8 at $\theta = 2$.

Higher power, e.g., power 0.8 at $\theta = 1.6$, would be desirable.

The sponsors will increase sample size if interim results are "promising".

An interim analysis is planned after observing $n_1 = 208$ responses.

At this time a further 208 subjects will have been admitted to the trial, but treated for less than 26 weeks.

Increasing the sample size

At the interim analysis with $n_1 = 208$ observed responses, the estimated treatment effect is

$$\widehat{\theta}_1(n_1) \;=\; \overline{Y}_B(n_1) - \overline{Y}_A(n_1)$$

and

$$Z_1 = \frac{\widehat{\theta}_1(n_1)}{\sqrt{\{4\sigma^2/n_1\}}}.$$

A further 208 subjects will have been treated for less than 26 weeks. So at least 416 responses will be observed in due course.

MP use the values of $\widehat{\theta}_1(n_1)$ and Z_1 in choosing a new total sample size — between the original 442 and a maximum of 884.

In deciding whether to increase the sample size, MP consider conditional power of the original test with $n_2 = 442$ observations under $\theta = \widehat{\theta}_1(n_1)$, given the observed value of Z_1.

Increasing the sample size

Definition

The conditional power $CP_\theta(z_1)$ is the probability the final test, with 442 observations, rejects H_0, given $Z_1 = z_1$ and effect size θ,

$$CP_\theta(z_1) \;=\; P_\theta\{Z_2 > 1.96 \,|\, Z_1 = z_1\}.$$

MP's adaptive design is based on conditional power under $\theta = \hat{\theta}_1$.

They divide the range of z_1 into three regions:

Favourable	$CP_{\hat{\theta}_1}(z_1) \geq 0.8$	*Continue to* $n_2 = 442$,
Promising	$0.365 \leq CP_{\hat{\theta}_1}(z_1) < 0.8$	*Increase* n_2,
Unfavourable	$CP_{\hat{\theta}_1}(z_1) < 0.365$	*Continue to* $n_2 = 442$.

Mehta & Pocock refer to this as the "Promising Zone" approach.

Protecting the type I error rate

When the final sample size is n_2^*, MP carry out a standard test, rejecting H_0 if

$$Z_2(n_2^*) \;=\; \frac{\widehat{\theta}(n_2^*)}{\sqrt{\{4\sigma^2/n_2^*\}}} \;>\; 1.96.$$

The "Promising Zone" and sample size rule are defined so that sample size changes can only decrease the type I error rate.

For proof that the type I error rate is protected, see

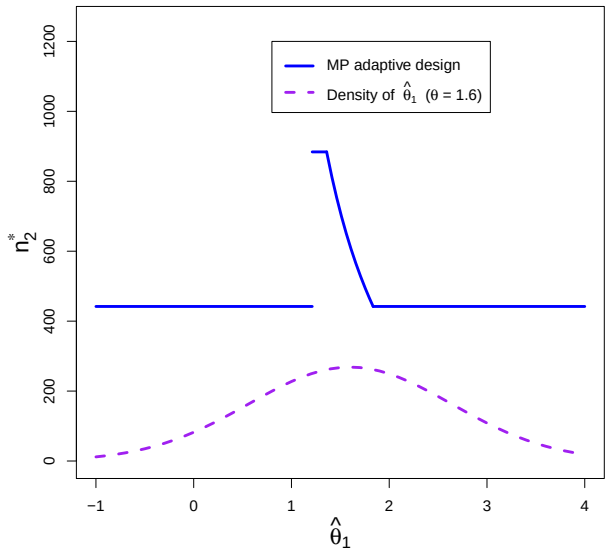
Chen, DeMets & Lan, *Statistics in Medicine* (2004),

Gao, Ware & Mehta, *J. Biopharmaceutical Statistics* (2008).

In general, changes to sample size may increase or decrease the type I error rate — use of the Chen, DeMets & Lan result restricts the values of $\widehat{\theta}_1(n_1)$ for which sample size can be increased.

The MP design

In their "promising zone", MP increase n_2 to achieve conditional power 0.8 under $\theta = \widehat{\theta}_1$, truncating this value to 884 if it is larger than that.

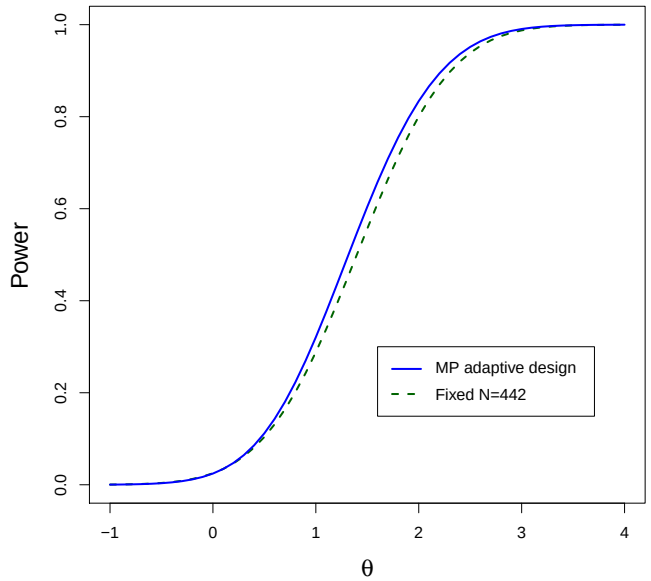


Comparison with the distribution of $\widehat{\theta}_1$ under $\theta = 1.6$ shows that increases in n_2 occur in a region of quite small probability.

(The distribution of $\widehat{\theta}_1$ under other values of θ is shifted but has the same variance.)

Properties of the MP design

The increase in n_2 in the "promising zone" has increased the power curve a little.

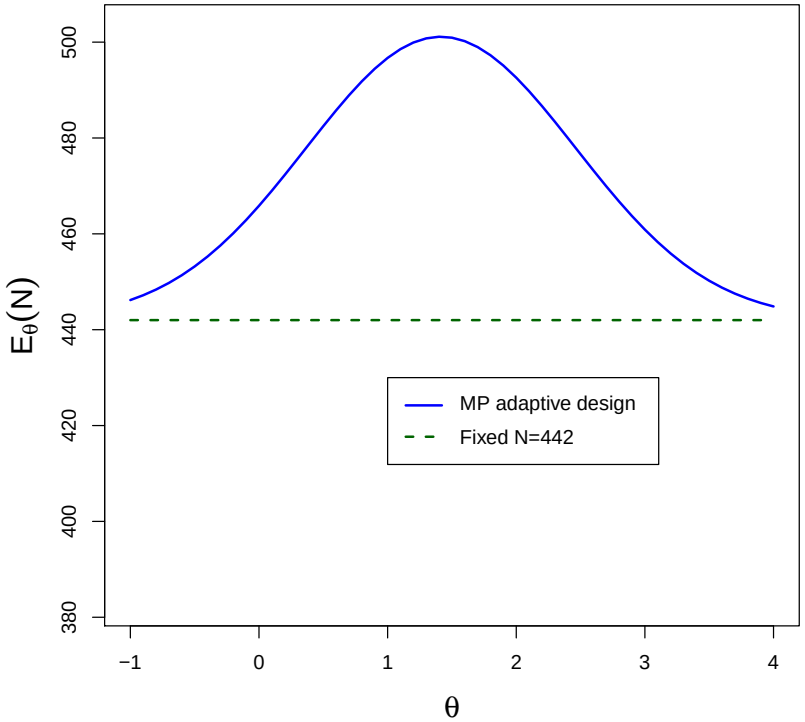


Given the limited range of values of $\widehat{\theta}_1$ for which n_2 is increased, only a small improvement in power can be expected.

Although it was stated that power 0.8 at $\theta = 1.6$ would be desirable, power at this effect size has only risen from 0.61 to 0.66.

Properties of the MP design

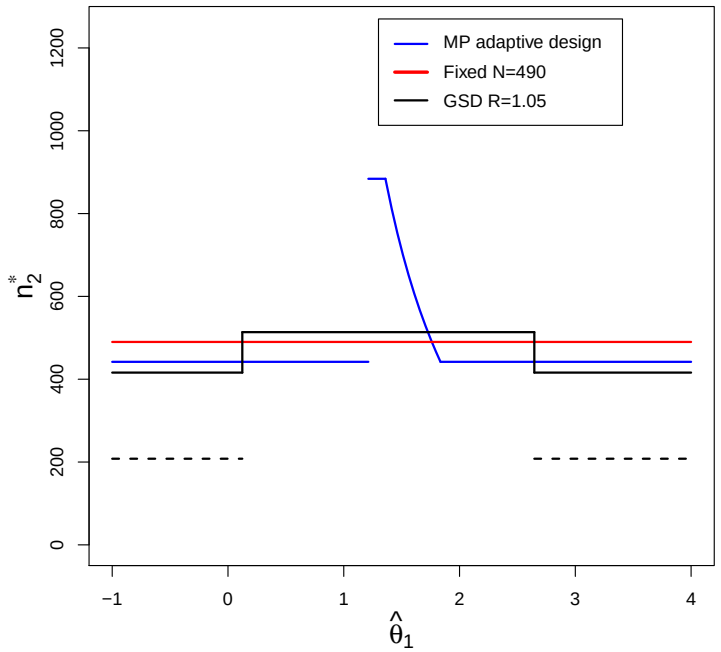
The cost of higher power is an increase in expected sample size, $E(N)$.



Aiming for higher conditional power under $\theta = \widehat{\theta}_1$ or raising the sample size beyond 884 gives small increases in power at the cost of large increases in $E(N)$.

Alternatives to the MP design

Here are sample size rules for two other trial designs that achieve the same power as MP's design.

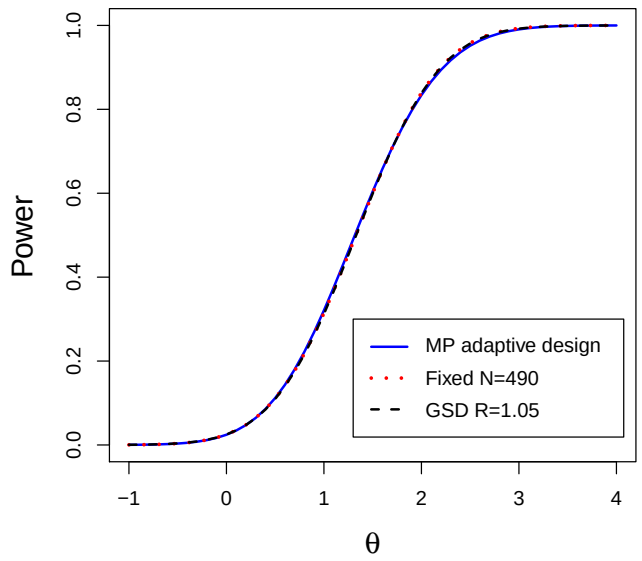


1. A fixed sample size design with 490 observations.

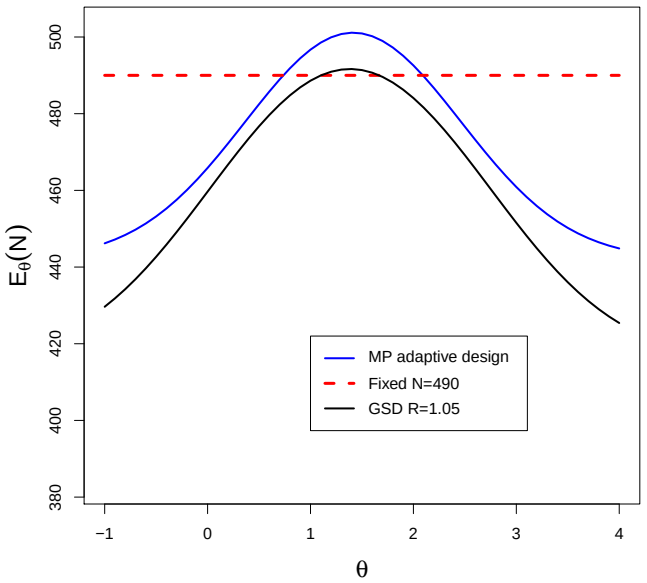
2. A group sequential test that stops after with a sample size of 416 or 514 (but only 208 responses are available when making the decision at the first analysis).

Comparison of designs

Power curve



$E_\theta(N)$ curves



All three designs have essentially the same power curve.

Clearly, it is possible to improve on the MP design's $E_\theta(N)$ curve.

How should we go about finding an efficient design for specific objectives?

Deriving an efficient sample size rule

We specify γ, a "rate of exchange" between sample size and power.

Focusing, for now, on properties under $\theta = \tilde{\theta} = 1.6$, we shall aim to maximise

$$P_{\theta=\tilde{\theta}}(\text{Reject } H_0) - \gamma E_{\tilde{\theta}}(N).$$

Consider the class of two-stage designs with:

- An interim analysis after $n_1 = 208$ responses are observed,
- Total sample size n_2^* chosen in the range $(416, 884)$, based on $\widehat{\theta}_1$,
- After stage 2, a normal combination test, rejecting H_0 if

$$\frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2 > 1.96,$$

where Z_1 and Z_2 are based on stage 1 and stage 2 data, respectively.

Deriving an efficient sample size rule

Suppose $Z_1 = z_1$ and the total sample size is set at n_2^*. Denote the conditional power of the combination test under $\theta = \tilde{\theta}$, given $Z_1 = z_1$ and this choice of n_2^*, by

$$CP_{\tilde{\theta}}(z_1, n_2^*) \;=\; P_{\tilde{\theta}}\{\frac{1}{\sqrt{2}}\, Z_1 + \frac{1}{\sqrt{2}}\, Z_2 \;>\; 1.96 \mid Z_1 = z_1,\, n_2^*\}.$$

We aim to find the sample size function $n_2^*(z_1)$ that maximises $P_{\theta=\tilde{\theta}}\,(\text{Reject } H_0) \;-\; \gamma E_{\tilde{\theta}}(N)$, which can be written as

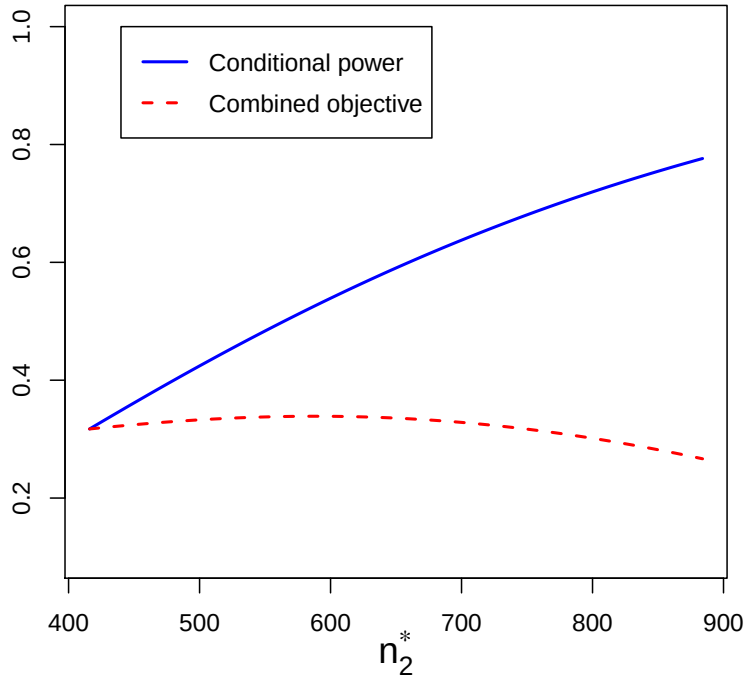
$$\int \{CP_{\tilde{\theta}}(z_1, n_2^*(z_1)) - \gamma n_2^*(z_1)\}\; f_{\tilde{\theta}}(z_1)\, dz_1,$$

where $f_{\tilde{\theta}}(z_1)$ denotes the density of Z_1 under $\theta = \tilde{\theta}$.

So, for each z_1, we need to choose $n_2^*(z_1)$ to maximise

$$CP_{\tilde{\theta}}(z_1, n_2^*(z_1)) - \gamma n_2^*(z_1).$$

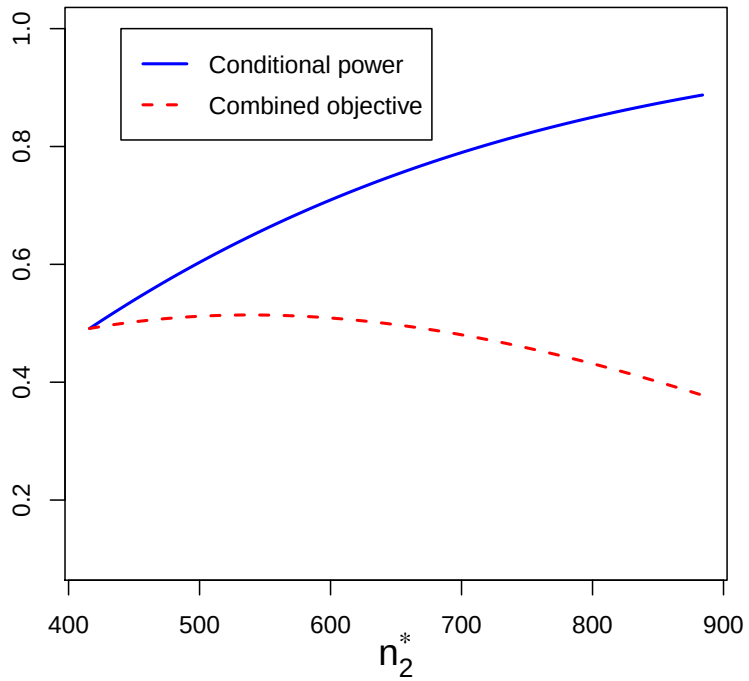
Plots for $\tilde{\theta} = 1.6$, $\gamma = 0.245/(4\,\sigma^2)$ and $\widehat{\theta}_1 = 0.75$



The function $CP_{\tilde{\theta}}(z_1, n_2^*) - \gamma(n_2^* - 442)$ attains its maximum at $n_2^* = 589$.

MP's design has $n_2^* = 442$ when $\widehat{\theta}_1 = 0.75$ — the Chen, DeMets & Lan construction does not allow n_2 to increase for this value of $\widehat{\theta}_1$, which lies below the "promising zone".

Plots for $\tilde{\theta} = 1.6$, $\gamma = 0.245/(4\,\sigma^2)$ and $\widehat{\theta}_1 = 1.25$

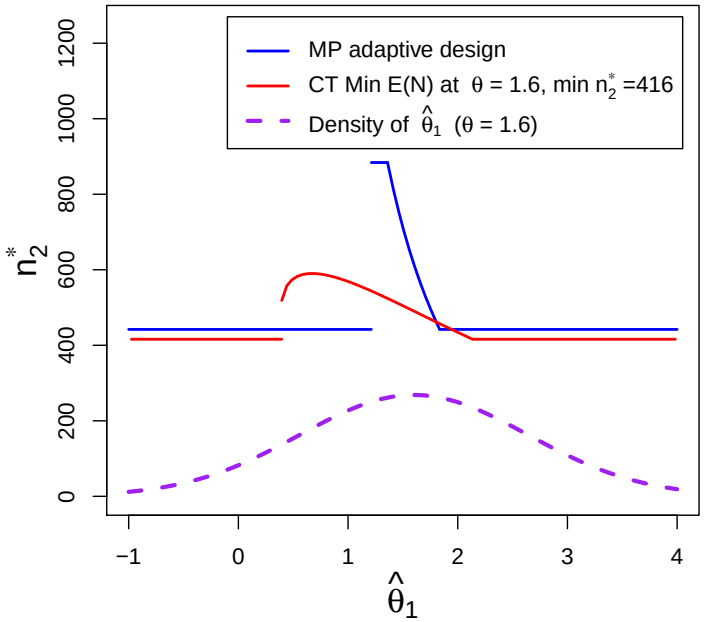


Now, the function $CP_{\tilde{\theta}}(z_1, n_2^*) - \gamma(n_2^* - 442)$ has its maximum at $n_2^* = 570$.

In this case, MP's design takes the maximum permitted value of $n_2^* = 884$.

Optimal sample size rule for combination test design

with n_2^* in $(416, 884)$, $\tilde{\theta} = 1.6$, $\gamma = 0.245/(4\,\sigma^2)$



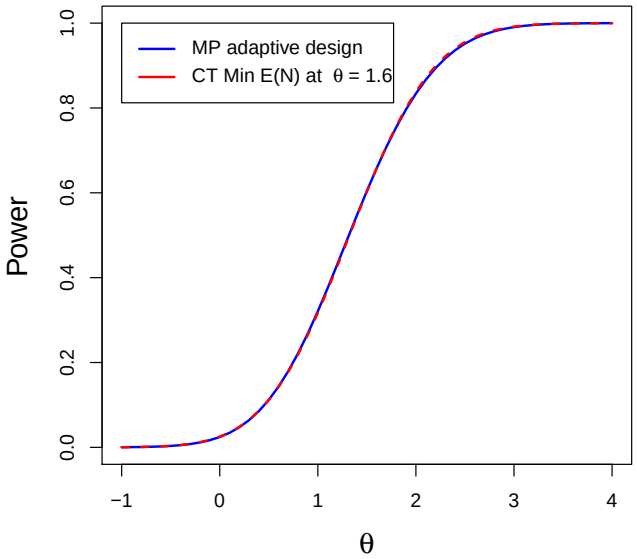
With $\gamma = 0.245/(4\,\sigma^2)$, overall power is 0.658 at $\theta = 1.6$, the same as for the MP design.

By construction, the procedure has minimum $E_{\theta=1.6}(N)$ among all normal combination test designs with $n_1 = 208$ and $n_2^* \geq 416$ that achieve the same power.

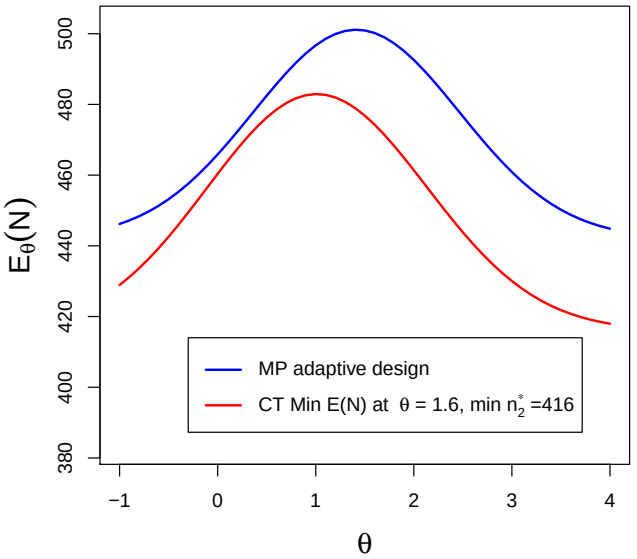
Optimal sample size rule for combination test design

with n_2^* in $(416, 884)$, $\tilde{\theta} = 1.6$, $\gamma = 0.245/(4\,\sigma^2)$

Power curve



$E_\theta(N)$ curves



Our optimised design has essentially the same power curve as the MP design and lower $E_\theta(N)$ at all θ values.

In fact, optimising over general designs test with $n_1 = 208$ and $n_2^* \geq 416$ leads to barely perceptible reductions in $E_{\theta=1.6}(N)$.

Other options

1. We could increase sample size further and achieve higher power, now we can do this efficiently.

2. We could optimise other criteria, replacing $E_{\tilde{\theta}}(N)$ in $P_{\theta=\tilde{\theta}}\,(\text{Reject } H_0) - \gamma E_{\tilde{\theta}}(N)$ by a weighted sum or integral,

$$\sum_i w_i \, E_{\theta_i}(N) \quad \text{or} \quad \int w(\theta) \, E_\theta(N) \, d\theta.$$

In the integral case, treating the power function in a similar way, we seek to maximise

$$\int w(\theta) \, P_\theta \, (\text{Reject } H_0) \, d\theta \; - \; \gamma \int w(\theta) \, E_\theta(N) \, d\theta.$$

If $w(\theta)$ is a prior distribution, representing investigators' beliefs about likely values of θ, we have a Bayes decision problem.

Other options

Suppose we wish to maximise

$$\int w(\theta)\, P_\theta\,(\text{Reject } H_0)\, d\theta \;-\; \gamma \int w(\theta)\, E_\theta(N)\, d\theta.$$

For each z_1, we must choose $n_2^*(z_1)$ to maximise

$$CP(z_1, n_2^*(z_1)) - \gamma n_2^*(z_1),$$

where $CP(z_1, n_2^*(z_1))$ is the conditional power *integrated over the posterior distribution of* θ given the stage 1 data summary z_1.

By working within the class of normal combination tests, we automatically protect the type I error rate at level α.

As before, we can choose γ so that a specific power condition is met — and after such calibration, we will return to almost exactly the same power curve as before.

Comments on this example

Controlling the frequentist type I error rate

Use of the combination test guarantees control of type I error for any sample size rule.

We can optimise within this framework.

Solving a Bayes decision problem

First, specify a "gain function" or "utility" to be maximised.

This may involve a prior distribution for unknown parameters.

Then, optimise within the class of permitted rules.

An outer layer

Consider varying the "framework" or constraints that were set in the initial formulation of the problem.

2. Optimising a Phase 2/3 seamless design

During Phase 2 and Phase 3 of the drug development process,

The final decision is made on the treatment specification, including the dose level,

The selected treatment is tested against control.

A seamless Phase 2/3 trial design combines these two phases:

In stage 1

Compare K "treatments" against control

Select the best treatment and, if it has performed sufficiently well, proceed to stage 2.

In stage 2

Compare the selected treatment against the control.

Optimising a Phase 2/3 seamless design

After both stages are completed, we test the null hypothesis that the selected treatment is no better than the control.

Since this treatment was selected based on data that will also be used in the final analysis, care must be taken to avoid inflating the overall type I error rate.

Design issues

We would like to optimise:

1. The way in which data on all treatments are combined in the final hypothesis test,
2. The way in which the total sample size is divided between the two stages.

Optimising a Phase 2/3 seamless design

Denote the K treatment effects vs control by $\theta_1, \ldots, \theta_K$.

Stage 1

Randomise m_1 subjects to each of the K treatments and the control and observe their responses.

Denote the estimated treatment effects by $\widehat{\theta}_{1,i},\ i = 1, \ldots, K$.

Treatment i^* with the highest $\widehat{\theta}_{1,i}$ is selected for stage 2.

Stage 2

Treatment i^* is compared against control, with m_2 observations on each. The estimated treatment effect is $\widehat{\theta}_{2,i^*}$.

Conclusion

A final decision is made, based on $\widehat{\theta}_{1,1}, \ldots, \widehat{\theta}_{1,K}$ and $\widehat{\theta}_{2,i^*}$.

Optimising a Phase 2/3 seamless design

There are K null hypotheses, H_i: $\theta_i \leq 0$, $i = 1, \ldots, K$.

If dose i^* is selected for Phase 3, we focus on testing H_{i^*}: $\theta_{i^*} \leq 0$.

Family-wise error

We want strong control of the **family-wise error** rate. Then, for all vectors $\theta = (\theta_1, \ldots, \theta_K)$,

$$Pr\{\text{Reject } \textit{any} \text{ true } H_i\} \leq \alpha.$$

Power

When some θ_i are greater than zero, we can define power as

$$Pr\{\text{Select treatment } j \text{ with maximum } \theta_i \text{ and reject } H_j\text{: } \theta_j \leq 0\}.$$

More generally, we can define a gain function or utility that is positive when H_{i^*} is rejected, whichever treatment is selected, but the gain increases with θ_{i^*}.

Optimising a Phase 2/3 seamless design

The family-wise error rate can be controlled by using a "closed testing procedure".

This requires level α tests of each null hypothesis H_i, and of all intersections of sets of these hypotheses.

Each of these tests can be constructed as a combination test across the two stages of the trial.

Then, general theory implies that the family-wise type 1 error rate is controlled at level α.

There are still choices to be made:

- How should we test the intersection hypotheses in stage 1?
- What type of combination test is best?

The best choice may depend on the K-dimensional parameter θ.

Research Article

Statistics in Medicine

Received 20 November 2013, Accepted 8 September 2014 Published online 15 October 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6316

Optimizing the data combination rule for seamless phase II/III clinical trials

Lisa V. Hampson[a*†] and Christopher Jennison[b]

We consider seamless phase II/III clinical trials that compare K treatments with a common control in phase II then test the most promising treatment against control in phase III. The final hypothesis test for the selected treatment can use data from both phases, subject to controlling the familywise type I error rate. We show that the choice of method for conducting the final hypothesis test has a substantial impact on the power to demonstrate that an effective treatment is superior to control. To understand these differences in power, we derive decision rules maximizing power for particular configurations of treatment effects. A rule with such an optimal frequentist property is found as the solution to a multivariate Bayes decision problem. The optimal rules that we derive depend on the assumed configuration of treatment means. However, we are able to identify two decision rules with robust efficiency: a rule using a weighted average of the phase II and phase III data on the selected treatment and control, and a closed testing procedure using an inverse normal combination rule and a Dunnett test for intersection hypotheses. For the first of these rules, we find the optimal division of a given total sample size between phases II and III. We also assess the value of using phase II data in the final analysis and find that for many plausible scenarios, between 50% and 70% of the phase II numbers on the selected treatment and control would need to be added to the phase III sample size in order to achieve the same increase in power. © 2014 The Authors. ***Statistics in Medicine*** **published by John Wiley & Sons Ltd.**

Keywords: Bayes decision problem; combination test; closed testing procedure; multiple hypothesis testing; seamless phase II/III trial; treatment selection

Optimising a Phase 2/3 seamless design

Hampson & Jennison (*Statistics in Medicine*, 2013) found optimal final decision rules that maximise power when $\theta = \delta\, v$, for various choices of vector v.

Interestingly, two procedures were close to 100% efficient across a wide range of scenarios.

1. In the framework we have described, use a Dunnett test for each intersection hypothesis in stage 1 and combine Z values across stages with a weighted normal combination test.

2. Use the procedure proposed by Thall, Simon and Ellenberg (*Biometrika*, 1988).

We were surprised that procedures with such robust efficiency exist.

However, this deals conveniently with the problem that the best choice of design may depend on a high-dimensional, unknown vector θ.

Optimising a Phase 2/3 seamless design

Hampson & Jennison also considered how best to divide a total sample size between stage 1 (m_1 observations on K treatments and control) and stage 2 (m_2 on selected treatment and control).

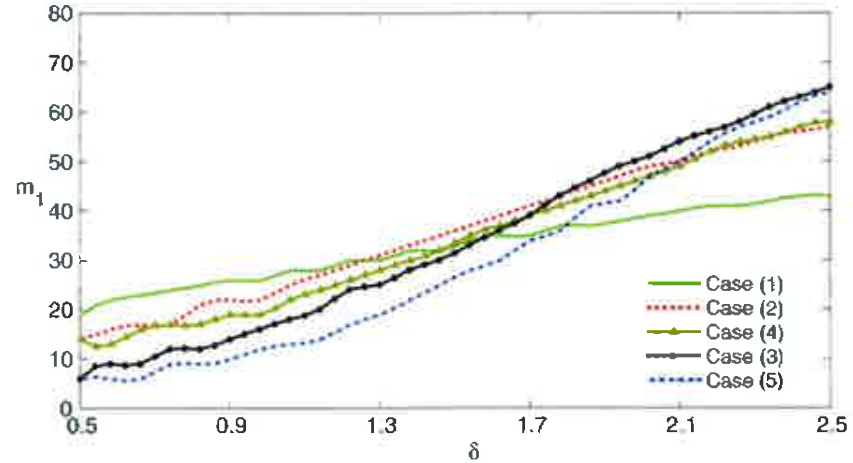
The choice that maximises power depends on the vector of treatment effects, θ, with the largest treatment effect playing a leading role.

If the highest treatment effect is large, one can afford a high m_1, increasing the probability of selecting this treatment.

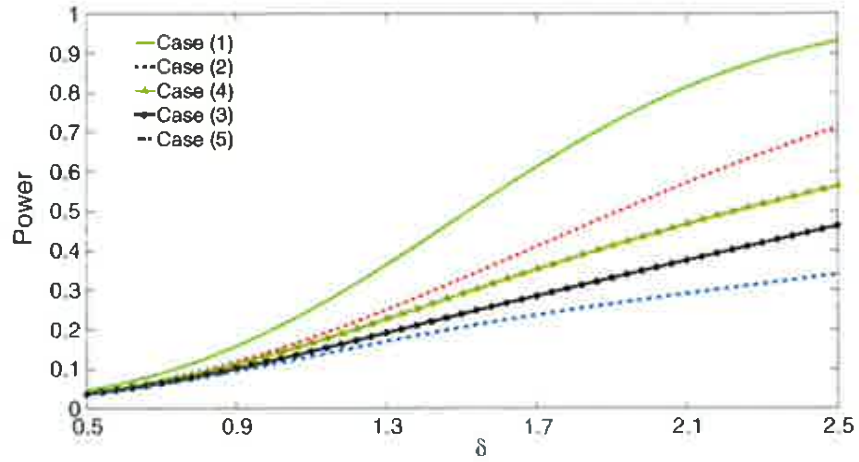
If the highest treatment effect is smaller, a high m_2 is needed to give power in stage 2 when the best treatment is selected.

Advice:

Express your expectations as a distribution for θ and choose a design with good average properties across this distribution.



(a)



(b)

Figure 4. (a) Stage 1 group sizes maximizing the power of the TSE procedure when the total sample size is fixed at 448 and $\boldsymbol{\theta}$ is a random permutation (1) of $(0, 0, 0, 0, 1)\,\delta$, (2) of $(0.5, 0.5, 0.5, 0.5, 1)\,\delta$, (3) of $(0.75, 0.75, 0.75, 0.75, 1)\,\delta$, (4) of $(0.3, 0.475, 0.65, 0.825, 1)\,\delta$ and (5) of $(0.75, 0.8125, 0.875, 0.9375, 1)\,\delta$. (b) Power achieved by the optimized TSE procedures. Decision rules are listed in order of decreasing power. Designs are specified with $K = 5$, $\ell = 0$, $\sigma = 5.0$ and $\alpha = 0.025$. Results are based on 1 million simulations for each scenario.

Benefits of Phase 2/3 seamless designs

Regulators require a seamless Phase 2/3 trial to be conducted as a single trial, with a firewall between the data monitoring committee and the investigators.

Efficiency gains from using "Phase 2" data in the final hypothesis test must balance extra planning and organisational requirements.

With m_1 observations on each treatment and control in stage 1 and m_2 on the selected treatment and control in stage 2, what are the benefits of using the stage 1 data in the final analysis?

Hampson & Jennison show that:

> If only stage 2 data are used in the final analysis, then in many plausible scenarios, m_2 needs to be increased by between $0.5m_1$ and $0.7m_1$, in order to achieve the same power as the seamless design.

Comments on this example

Controlling the frequentist type I error rate

Use of a closed testing procedure and combination tests guarantees control of type I error.

Optimising within this class of designs

We can, essentially, optimise the choice of closed testing procedure and combination test for all treatment effect vectors, θ.

However, the best choice of sample sizes in stage 1 and stage 2 does depend on the vector θ.

The Bayes solution is to specify a prior distribution for the unknown θ and optimise performance integrated over this distribution.

An outer layer

If the value of m_1 appears unacceptably small, consider a higher total sample size for the two stages.

3. Creating an efficient enrichment design: Switching to a sub-population in response to interim data

Consider a treatment developed to disrupt a disease's biological pathway. Patients with high levels of a biomarker for this pathway should gain particular benefit.

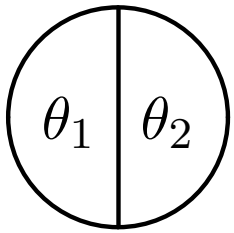
In a clinical trial with *enrichment* we

- Start by comparing the new treatment against control in the full population.
- Examine responses at an interim stage and decide whether to:
  - Stop for futility,
  - Continue recruiting from the full population,
  - Continue, but recruit only from the subgroup — and increase their numbers.

Results may support a licence for the full population or just the sub-population.

Creating an efficient enrichment design

Sub-population (proportion λ_1)



Rest of the population (proportion λ_2)

The treatment effect (difference in mean response between new treatment and control) is θ_1 in the sub-population and θ_2 in the complement of this sub-population.

The treatment effect over the full population is $\theta_3 = \lambda_1\theta_1 + \lambda_2\theta_2$.

We may wish to test either or both of:

- The null hypothesis for the full population, H_3: $\theta_3 \leq 0$ vs $\theta_3 > 0$,
- The null hypothesis for the sub-population, H_1: $\theta_1 \leq 0$ vs $\theta_1 > 0$.

Creating an efficient enrichment design

As in the adaptive seamless Phase 2/3 design, we want to control strongly the **family-wise error** rate.

Then, for all values of θ_1 and θ_3,

$$Pr\{\text{Reject } \textit{any} \text{ true } H_i\} \leq \alpha.$$

This can be achieved by a "closed testing procedure", involving level α tests of H_1, H_3 and the intersection hypothesis $H_1 \cap H_3$.

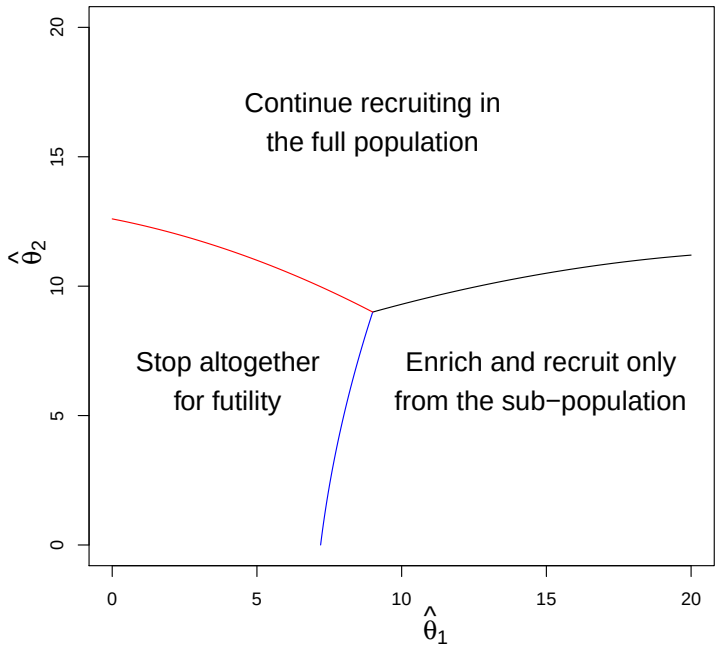
Each of these tests can be constructed as a combination test across the two stages of the trial.

Then, general theory implies that the family-wise type 1 error rate is controlled at level α.

This leaves freedom to define the rule for deciding whether or not to enrich at the interim analysis.

Creating an efficient enrichment design

At the interim analysis, there are three options.



Optimising this decision rule requires specification of:

- Benefits from rejecting H_1 or H_3 for parameter values θ_1 and θ_3,
- The costs saved when the trial stops for futility,
- A prior distribution for (θ_1, θ_3).

Creating an efficient enrichment design

PhD student Thomas Burnett, at Bath, has been working on the derivation of optimal adaptive designs for enrichment trials.

The computation can be demanding, but Thomas has developed code to find optimal rules.

The appropriate adaptive decision rule depends strongly on the prior for (θ_1, θ_3).

Once such a prior is specified, it is natural to compare simpler trial designs, that do not involve adaptation:

Recruit from the full population throughout the trial,

Recruit only from the sub-population throughout the trial.

For many examples of gain function and prior, the best adaptive design is not necessarily superior to both simple designs.

Comments on this example

Controlling the frequentist type I error rate

Use of a closed testing procedure and combination tests guarantees control of type I error.

Optimising within this class of designs

Given gain and cost functions, and a prior distribution for (θ_1, θ_3), one can compute Bayes-optimal adaptive enrichment designs.

An outer layer

Other design features that merit investigation include:

- Details of the closed testing procedure and combination tests.
- The timing of the interim analysis.
- Preferential sampling of one population when the proportions λ_1 and λ_2 are away from 0.5.

Overall conclusions

Controlling the frequentist type I error rate

We can apply closed testing procedures and combination tests to protect family-wise error in complex, high-dimensional settings.

We can then work on optimising other aspects of a given design.

Optimising within a class of designs

Before trying to optimise, we need to understand which properties of a design are important to the investigators.

Typically, this is done through the elicitation of their gain function, cost function, and prior distribution for unknown parameters.

Then, we can optimise by analysis, calculation or simulation.

An outer layer

Once we can optimise the central component of a design, we may re-visit higher level aspects and question initial assumptions.