Adaptive Clinical Trials: What do they Offer?

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

ACRP Webinar

16 September 2009

Objectives for the training

Upon completion of this webinar, attendees should be able to:

State the key features of adaptive clinical trials that differentiate the adaptive design from the traditional clinical trial design,

Understand why adaptive clinical trials offer greater efficiency for your clinical trial budget,

Identify three applications of adaptive and group sequential designs ,

Gain awareness of the advantages of adaptive design,

Gain awareness of when it is most advantageous to employ an adaptive design and when the traditional group sequential design would be preferable,

Know some key links to further your knowledge and keep up-to-date on this topic.

The role of adaptive methods

Ordinarily, in a clinical trial one specifies at the outset:

Patient population,

Treatment,

Primary endpoint,

Hypothesis to be tested,

Power at a specific effect size.

Adaptive designs allow these elements to be reviewed during the trial.

Because . . . there may be limited information to guide these choices initially, but more knowledge will accrue as the study progresses.

Critical appraisal of adaptive designs

It is important to assess the benefits a new approach will be able to deliver.

How can the benefits be assessed?

Are these benefits real?

What "standard" methods should be used for comparison?

We shall consider adaptive methods for three applications:

1. Sample size modification in response to estimates of a nuisance parameter such as response variance,

2. Sample size when there is uncertainty about the likely treatment effect,

3. Switching to a patient sub-population.

1. Sample size modification for a nuisance parameter

(a) Internal pilots in studies with a single analysis

The sample size needed to satisfy a power requirement often depends on an unknown nuisance parameter.

Examples include

Normal response: Unknown variance, σ^2.

Binary response: Since variance depends on p, the sample size needed to detect a difference $p_1 - p_2 = \delta$ depends on $(p_1 + p_2)/2$.

Survival data: Information is governed by the number of observed deaths, which depends on the overall failure rate and degree of censoring.

"Over-interpretation of results from a small pilot study, positive or negative, may undermine support for the major investigation" (W. G. Cochran).

Internal pilots: Wittes & Brittain

Wittes & Brittain (*Statistics in Medicine*, 1990) suggest an "internal" pilot study.

Let ϕ denote a nuisance parameter, e.g., the response variance.

Suppose the sample size required under a particular value of ϕ is given by the function $n(\phi)$.

From a pre-study estimate, $\hat{\phi}_0$, calculate an initial sample size, $n(\hat{\phi}_0)$.

At an interim stage, find a new estimate $\hat{\phi}_1$ from the data observed so far and aim for the new target of $n(\hat{\phi}_1)$.

Variations on this scheme are possible, e.g., one might only allow an *increase* over the original target sample size.

Internal pilots: Properties

Wittes and Brittain's method has a complicated effect on the final test statistic. Variance estimates tend to be biased downwards but the type I error rate is only slightly perturbed.

Binary responses

Two-treatment comparison, H_0: $p_A = p_B$, $\alpha = 0.05$.

Internal pilots are used to achieve power at $p_B = p_A + \Delta$ for fixed Δ, or at $p_B = p_A/\rho$ for fixed ρ.

Pilot sample size per treatment	*Type I error rate*
10	0.057 – 0.059
30	0.049 – 0.057
50	0.049 – 0.053

Results from Jennison & Turnbull (2000) Ch. 14.

Internal pilots: Properties

Normal data, estimating σ^2

Two-treatment comparison, H_0: $\mu_A = \mu_B$, $\alpha = 0.05$.

Internal pilots are used to achieve power at $\mu_B - \mu_A = \pm\delta$ for fixed δ.

Degrees of freedom for s_1^2	*Type I error rate*
8	0.052 – 0.065
18	0.050 – 0.057
38	0.052 – 0.053
78	0.051

Results from Jennison & Turnbull (2000) Ch. 14.

Blinding: Finding s_1^2 may break the blinding and reveal the estimated effect, $\hat{\theta}$.
Instead, one can estimate σ^2 from pooled data without treatment labels.

Sample size modification for a nuisance parameter

(b) Two stage designs with combination tests

Combination tests (Bauer & Köhne, *Biometrics*, 1994).

Initial design

Define the null hypothesis H_0 (with a one-sided alternative).

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

Stage 1

Observe the P-value, P_1.

Under H_0, $P_1 \sim U(0, 1)$.

Design Stage 2 in the light of Stage 1 data.

Stage 2

Observe the P-value, P_2.

Under H_0, $P_2 \sim U(0, 1)$ and P_2 is independent of P_1.

Combination tests: The inverse χ^2 test

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

If $P \sim U(0,\, 1)$, then

$$-\ln(P) \;\sim\; \text{Exp}\,(1) = \frac{1}{2}\,\chi^2_2.$$

Thus, under H_0,

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

and a test combining the two P-values rejects 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) *Statistical Methods for Research Workers*.

Combination tests: The inverse normal test

Stipulate the ***inverse normal test*** with weights w_1 and w_2, where $w_1^2 + w_2^2 = 1$.

Stage 1

Compute $Z_1 = \Phi^{-1}(P_1)$.

Under H_0, $Z_1 \sim N(0, 1)$.

Design Stage 2 in the light of Stage 1 data.

Stage 2

Compute $Z_2 = \Phi^{-1}(P_2)$.

Under H_0, $Z_2 \sim N(0, 1)$ and Z_2 is independent of Z_1.

Overall test

Under H_0, $Z = w_1 Z_1 + w_2 Z_2 \sim N(0, 1)$.

Reject H_0 if $Z > \Phi^{-1}(1 - \alpha)$.

Sample size re-estimation using a combination test

In a two-treatment comparison with normal response, power $1-\beta$ at effect size $\theta = \delta$ requires sample size per treatment of

$$n \;=\; (z_\alpha + z_\beta)^2\, 2\,\sigma^2/\delta^2, \qquad (1)$$

where z_p denotes $\Phi^{-1}(1-p)$.

Initial design

A Bauer & Köhne two-stage design is specified using, say, the inverse χ^2 test.

Sample size n_0 is determined using a preliminary estimate σ_0^2 in (1).

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: $\theta \le 0$ vs $\theta > 0$ is converted to a P-value, P_1.

Sample size re-estimation using a combination test

Stage 1 . . .

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

One may simply substitute this value in (1).

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 rejects H_0 if

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

This test has type I error rate exactly equal to α.

Sample size modification for a nuisance parameter

(c) Sample size modification within group sequential tests

Group sequential tests

Reference: Jennison & Turnbull (2000)

Suppose we wish to test H_0: $\theta \leq 0$ against $\theta > 0$ with type I error probability α and power $1 - \beta$ at $\theta = \delta$.

In a Group Sequential Test, a decision is taken after each group of observations to:

Stop, reject H_0: $\theta \leq 0$ in favour of $\theta > 0$,

Continue to observe the next group of subjects, or

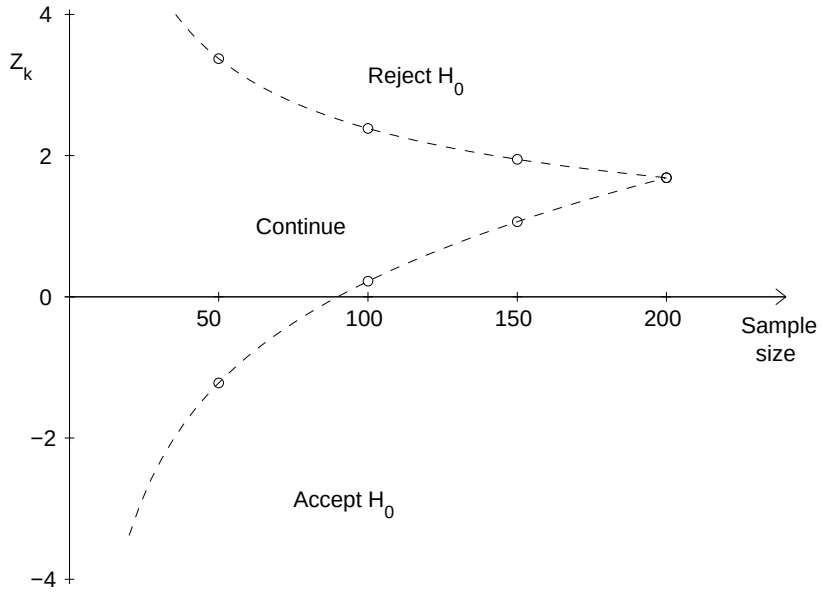
Stop, accept H_0.

Typically, group sizes are pre-specified.

"Error spending" designs are able to deal with unpredictable group sizes.

Group sequential tests

A group sequential test, specified in the trial protocol, can be described by a stopping boundary for the standardised statistic Z_k at each analysis k.



Early termination may be for a positive result (success) or a negative outcome (stopping for futility).

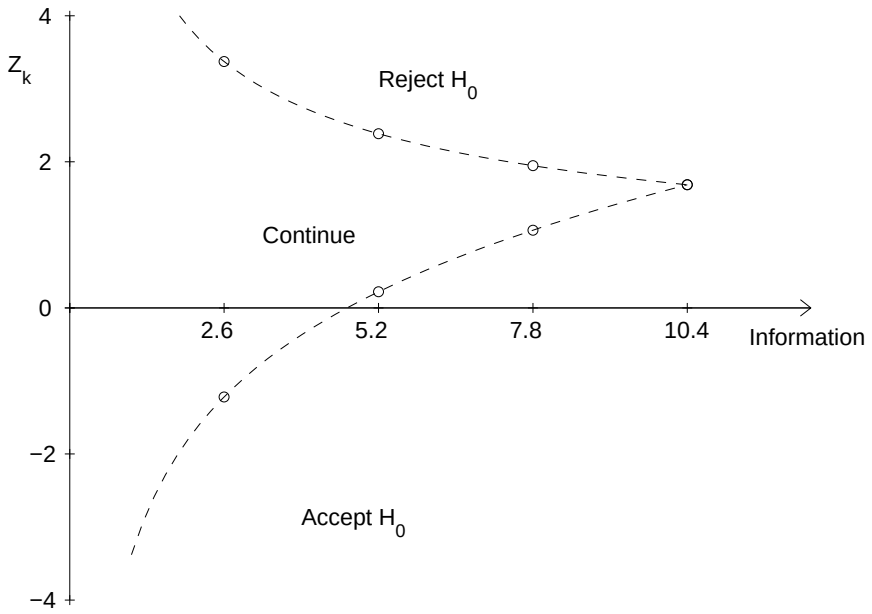
An efficient design ***reduces average sample size*** or time to a conclusion while protecting the type I error rate and maintaining desired power.

Error spending group sequential tests

Error spending designs offer a flexible way to deal with unpredictable group sizes.

Type I and type II error probabilities are “spent” as a function of the observed information at each analysis, as this increases towards a target $\mathcal{I}_{max}$.

It is natural to plot stopping boundaries against observed information.



Sample size modification for a nuisance parameter

Information monitoring

Mehta & Tsiatis (*Drug Inf. J.*, 2001) describe the "information monitoring" approach.

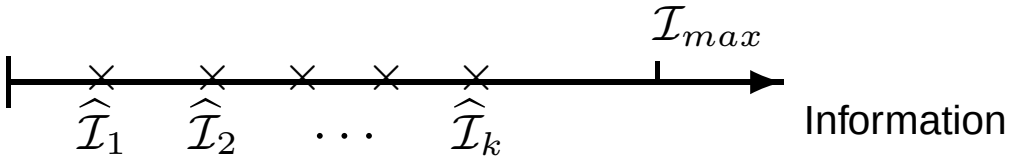
With n_A and n_B observations on treatments A and B at analysis k, estimate σ^2 by

$$s_k^2 = \frac{\sum(X_{Ai} - \bar{X}_A^{(k)})^2 + \sum(X_{Bi} - \bar{X}_B^{(k)})^2}{n_{Ak} + n_{Bk} - 2}$$

and estimate observed information by

$$\widehat{\mathcal{I}}_k = \frac{1}{\text{Var}(\hat{\theta})} = \left\{ \frac{s_k^2}{n_A} + \frac{s_k^2}{n_B} \right\}^{-1}.$$

Now "plug in" the estimated information sequence to create an error spending test with cumulative error at analysis k determined by $\widehat{\mathcal{I}}_k$.



Information monitoring

Updating the sample size

In a K–group design, at each analysis k, we can re-calculate the target for the final sample sizes n_{AK} and n_{BK} by solving the equation

$$\left\{ \frac{s_k^2}{n_{AK}} + \frac{s_k^2}{n_{BK}} \right\}^{-1} = \mathcal{I}_{\max}$$

and choose the next group size to work towards this target.

Approximations in this method can lead to inflation of type I error rates. In 3 and 5 group tests with $\alpha = 0.05$, we have found actual type I error rates:

Target total sample size	*Type I error rate*
50	0.054 – 0.063
100	0.052 – 0.056

More precise methods can attain error rates more closely (e.g., Denne & Jennison, *Biometrika*, 2000).

Sample size modification for a nuisance parameter

The method of Lehmacher and Wassmer (Biometrics, 1999)

This is a K–group version of Bauer & Köhne's inverse normal combination test.

Let Z_k be the Z-statistic based on data in group k alone.

Define pre-assigned weights $w_1, \ldots, w_K$ proportional to the square roots of the planned group sizes or information increments.

For $k = 1, \ldots, K$, create cumulative Z-statistics

$$Z_{(k)} \;=\; (w_1 Z_1 + \ldots + w_k Z_k) \,/\, (w_1^2 + \ldots + w_k^2)^{1/2}.$$

Future group sizes can be modified to achieve a target "information level" and the desired power at alternative $\theta = \delta$.

As long as each $Z_k \sim N(0,1)$ under H_0, the sequence $Z_{(1)}$, $Z_{(2)}$, $\ldots$ has a set distribution and can be used with a pre-specified group sequential boundary.

The target type I error probability will then be attained exactly.

2. Sample size modification to increase power

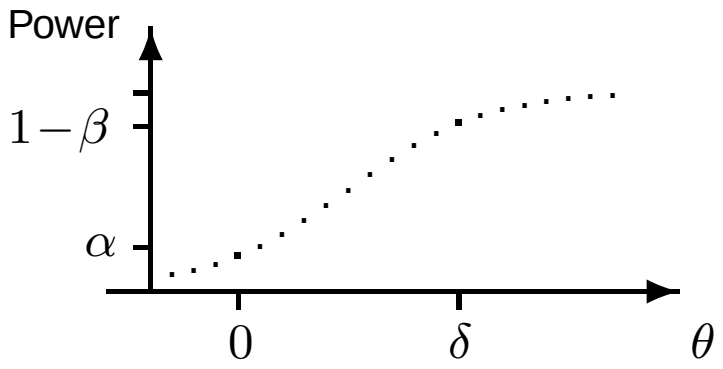
Type I error and power

Suppose θ represents the effect of a new treatment vs control.

A study is to test H_0: $\theta \leq 0$ against $\theta > 0$ with

one-sided type I error probability $\alpha = 0.025$, say.

Sample size is to be chosen to give a specific power curve.



(Assume now that there is no unknown "nuisance parameter" in this relationship, such as a normal variance or overall failure rate for survival data.)

Sample size modification to increase power

Investigators may start out optimistically and design a trial with power to detect a large treatment effect. Interim data may then suggest a smaller effect size — still clinically important but difficult to demonstrate with the chosen sample size.

- An adaptive design can allow sample size to be increased during the trial, ***rescuing*** an under-powered study.
- Some would advocate this ***wait and see*** approach as a way to "let the data say" what power and sample size should be chosen.
- Or, a ***group sequential design*** can achieve a desired power curve and save sample size through early stopping when the effect size is large.

Questions:

How should one set power and sample size?

Is there a down-side to the "wait and see" approach?

How are the adaptive and group sequential approaches related?

Sample size modification to increase power

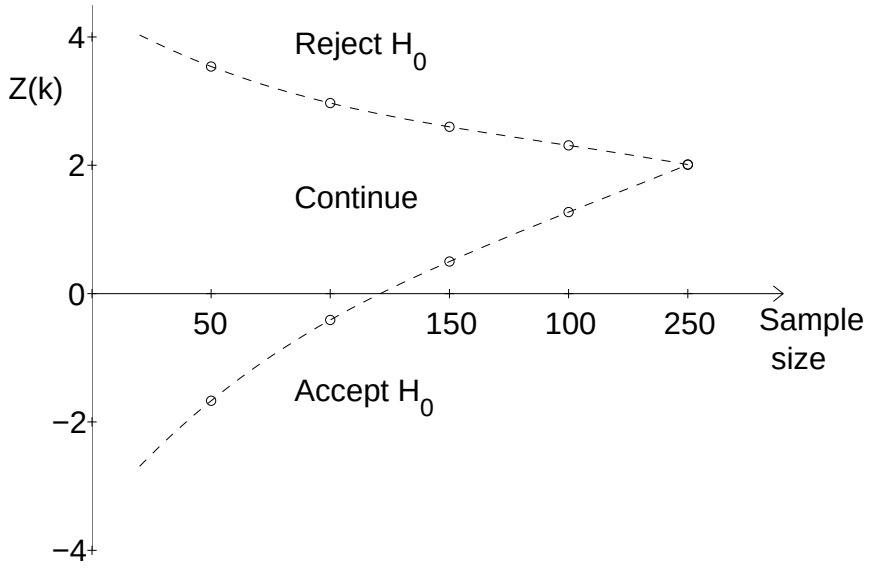
Example *(Jennison & Turnbull, Biometrika, 2006, Ex. 2)*

We start with a group sequential design with 5 analyses,

testing H_0: $\theta \leq 0$ against $\theta > 0$ with

one-sided type I error probability $\alpha = 0.025$ and

Initial design: power $1 - \beta = 0.9$ at $\theta = \delta$.



Sample size modification to increase power

Suppose, at analysis 2, a low interim estimate $\hat{\theta}_2$ prompts investigators to consider the trial's power at effect sizes below δ, where power 0.9 was originally set:

Lower effect sizes start to appear plausible,

Conditional power under these effect sizes, using the current design, is low.

Cui, Hung and Wang (*Biometrics*, 1999) cite instances of studies reporting to the FDA where such problems arose.

Special methods are needed in order to protect the type I error rate while making data-dependent modifications to sample size.

Cui, Hung and Wang developed a method which allows remaining group sizes to be increased in a group sequential design.

A variety of other methods for sample size modification is now available.

Sample size modification to increase power

Applying the method of Cui, Hung and Wang (*Biometrics*, 1999)

Following a decision at analysis 2 to increase sample size:

Sample sizes for groups 3 to 5 are multiplied by a factor γ.

Sample sums from these groups are down-weighted by $\gamma^{-1/2}$: this preserves the variance of this term but the mean is multiplied by $\gamma^{1/2}$.

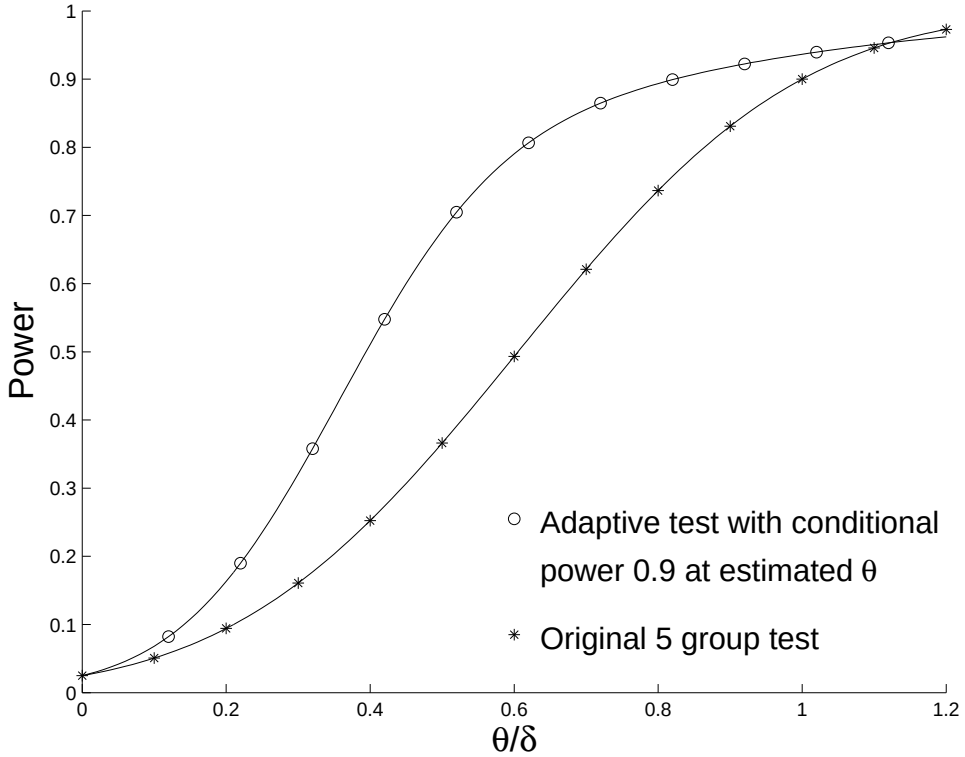
Using the new weighted sample sum in place of the original sample sum maintains the type I error rate and increases power.

In our example:

We choose the factor γ to give conditional power 0.9 **if $\boldsymbol{\theta}$ is equal to $\hat{\boldsymbol{\theta}}_{\mathbf{2}}$,** with the constraint $\gamma \leq 6$ so sample size can be at most 4 times the original maximum.

Sample size modification to increase power

Simulations show that re-design has raised the power curve at all effect sizes.



Overall power at $\theta = \delta/2$ has increased from 0.37 to 0.68.

Sample size modification to increase power

Reasons for re-design arose purely from observing $\hat{\theta}_2$. A group sequential design responds to such interim estimates — in the decision to stop the trial or to continue.

Investigators could have considered at the design stage how they would respond to low interim estimates of effect size.

If they had thought this through and chosen the above adaptive procedure, they could also have examined its overall power curve.

Assuming this power curve were acceptable, how else might it have been achieved?

An alternative group sequential design

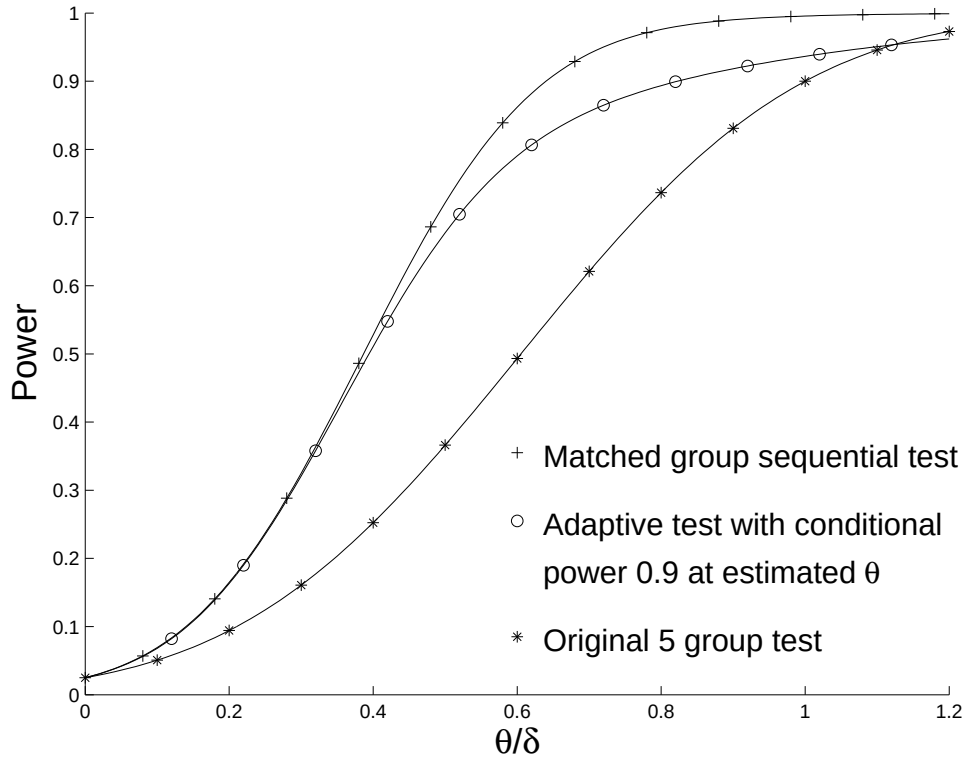
Five-group designs matching key features of the adaptive test can be found.

To be comparable, power curve should be as high as that of the adaptive design.

Can expected sample size be lower too?

Sample size modification to increase power

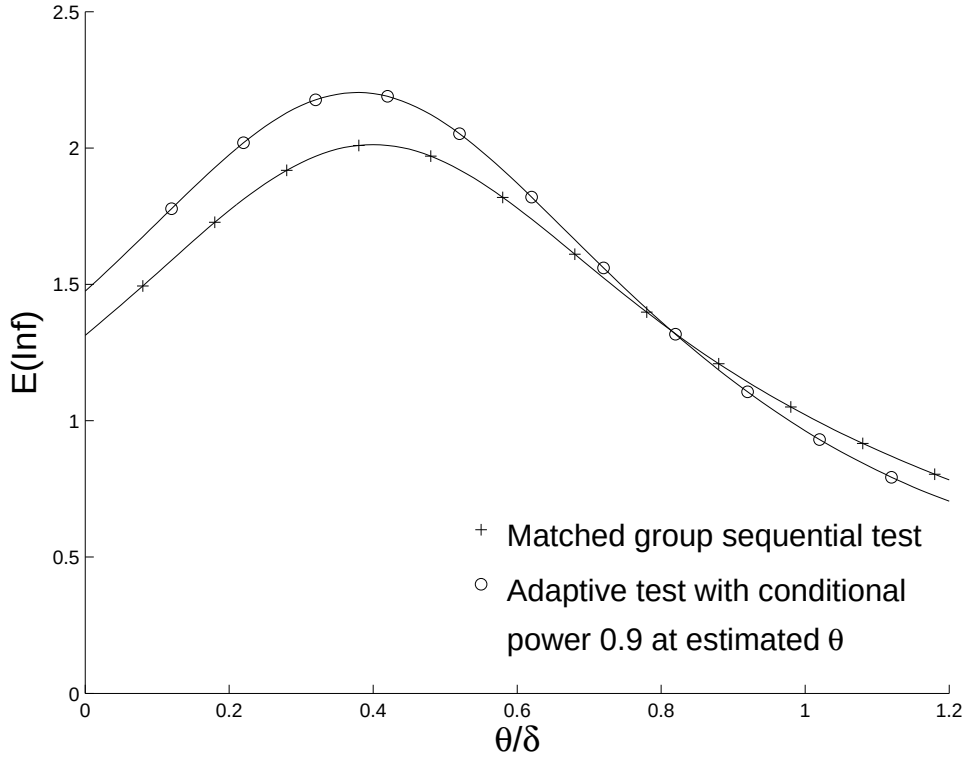
Power of our “matched” group sequential design is as high as that of the adaptive design at all effect sizes — and substantially higher at the largest θ values.



Sample size modification to increase power

The group sequential design has significantly lower expected information than the adaptive design over a range of effect sizes.

The group sequential design has slightly higher expected information for $\theta > 0.8\,\delta$, but this is where its power advantage is greatest.

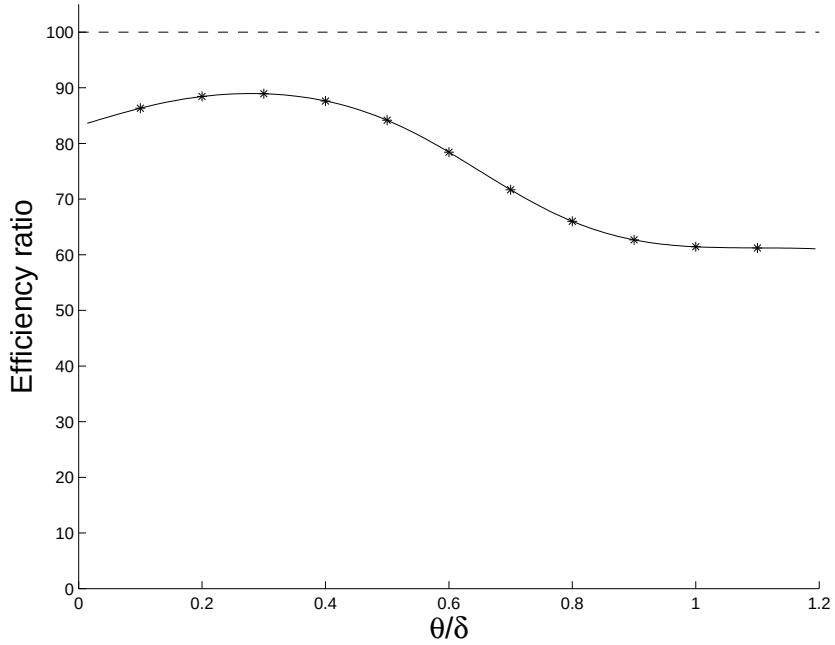


Sample size modification to increase power

Jennison & Turnbull (*Biometrika*, 2006) define an “Efficiency Ratio” to compare expected sample size, adjusting for differences in attained power.

By this measure, the adaptive design is up to 39% less efficient than the non-adaptive, group sequential alternative.

Efficiency ratio of adaptive design vs group sequential test



Sample size modification to increase power

We have found similar inefficiency relative to group sequential tests in a wide variety of proposed adaptive designs.

In general, adaptive designs have the advantage of extra freedom to choose group sizes in a response-dependent manner.

Jennison & Turnbull (*Biometrika*, 2006) show this adaptation can lead to gains in efficiency over non-adaptive group sequential tests — but the gains are very slight.

Sample size rules based on conditional power are far from optimal, hence the poor properties of adaptive designs using such rules.

Conclusion: Specify power properly at the outset, then group sequential designs offer a simple and efficient option.

3. Switching to a patient sub-population

A trial protocol defines a population of subjects who may benefit from the treatment.

Suppose it is believed the treatment could be particularly effective in a certain sub-population defined by a physiological or genetic biomarker.

Enrichment: Restricting recruitment to a sub-population

At an interim analysis, the options are:

Continue as originally planned, or

Restrict the remainder of the study to the defined sub-population.

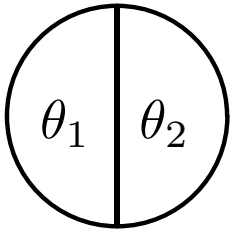
This choice will affect the licence a positive outcome can support.

The possibility of testing more than one null hypothesis means a multiple testing procedure must be used.

Enrichment: Example

Sub-population 1

(proportion λ_1)



Rest of the population

(proportion λ_2)

Overall treatment effect is $\theta = \lambda_1\theta_1 + \lambda_2\theta_2$.

We may wish to test:

The null hypothesis for the full population, H_0: $\theta \leq 0$ vs $\theta > 0$,

The null hypothesis for sub-population 1, H_1: $\theta_1 \leq 0$ vs $\theta_1 > 0$.

Multiple testing procedures

Suppose k null hypotheses, H_i: $\theta_i \leq 0$ for $i = 1, \ldots, k$, are to be considered.

A procedure's ***family-wise error rate*** under a set of values $(\theta_1, \ldots, \theta_k)$ is

$$Pr\{\text{Reject } H_i \text{ for some } i \text{ with } \theta_i \leq 0\} \;=\; Pr\{\text{Reject any true } H_i\}.$$

The family-wise error rate is controlled strongly at level α if this error rate is at most α for all possible combinations of θ_i values. Then

$$Pr\{\text{Reject any true } H_i\} \leq \alpha \quad \text{for all } (\theta_1, \ldots, \theta_k).$$

With such strong control, the probability of choosing to focus on the parameter θ_{i*} and then falsely claiming significance for null hypothesis H_{i*} is at most α.

Closed testing procedures (Marcus et al, *Biometrika*, 1976)

For each subset I of $\{1, \ldots, k\}$, we define the intersection hypothesis

$$H_I = \cap_{i \in I} H_i.$$

We construct a level α test of each intersection hypothesis H_I: this test rejects H_I with probability at most α whenever all hypotheses specified in H_I are true.

Closed testing procedure

The simple hypothesis H_j: $\theta_j \leq 0$ is rejected if, and only if, H_I is rejected for every set I containing index j.

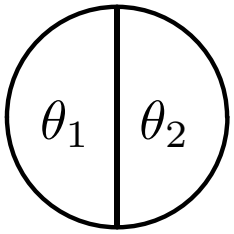
Proof of strong control of family-wise error rate

For a family-wise error to be committed, we must reject $H_{\tilde{I}}$ where $\tilde{I}$ is the set of indices of all true hypotheses H_i.

Since $H_{\tilde{I}}$ is true, $Pr\{\text{Reject } H_{\tilde{I}}\} = \alpha$ and, thus, the probability of a family-wise error is no greater than α.

Enrichment: Example

Sub-population 1 (proportion λ_1)



Rest of the population (proportion λ_2)

First, consider a design testing for a whole population effect, $\theta = \lambda_1\theta_1 + \lambda_2\theta_2$.

The design has two analyses and one-sided type I error probability 0.025.

Sample size is set to achieve power 0.9 at $\theta = 20$.

Data in each stage are summarised by a Z-value:

	Stage 1	*Stage 2*	*Overall*
H_0	$Z_{1,0}$	$Z_{2,0}$	$Z_0 = \frac{1}{\sqrt{2}}Z_{1,0} + \frac{1}{\sqrt{2}}Z_{2,0}$

Enrichment: Example

Two stage design, testing for a whole population effect, θ.

	Stage 1	*Stage 2*	*Overall*
H_0	$Z_{1,0}$	$Z_{2,0}$	$Z_0 = \frac{1}{\sqrt{2}} Z_{1,0} + \frac{1}{\sqrt{2}} Z_{2,0}$

Decision rules:

If $Z_{1,0} < 0$ — Stop at Stage 1, Accept H_0

If $Z_{1,0} \geq 0$ — Continue to Stage 2, then

- If $Z_0 < 1.95$ — Accept H_0
- If $Z_0 \geq 1.95$ — Reject H_0

Enrichment: Example

Assume the sub-population comprises half the total population, so $\lambda_1 = \lambda_2 = 0.5$.

Properties of design for the whole population effect, θ:

θ_1	θ_2	θ	*Power for* H_0: $\theta \leq 0$
20	20	20	0.90
10	10	10	0.37
20	0	10	0.37

Is it feasible to identify at Stage 1 that θ is low but θ_1 may be higher, so one might switch resources to test a sub-population?

Enrichment: A closed testing procedure

We wish to be able to consider two null hypotheses:

		On rejection, conclude:
H_0:	$\theta \leq 0$	Treatment is effective in the whole population
H_1:	$\theta_1 \leq 0$	Treatment is effective in sub-population 1

To apply a *closed testing procedure*, we also need a test of the intersection hypothesis:

$$H_{01}: \quad \theta \leq 0 \text{ and } \theta_1 \leq 0.$$

Note, since $\theta = 0.5\,\theta_1 + 0.5\,\theta_2$, either of H_0 and H_1 may be true on its own.

Enrichment: An adaptive design

At Stage 1, if $\hat{\theta} < 0$, stop to accept H_0: $\theta \leq 0$.

If $\hat{\theta} > 0$ and the trial continues:

If $\hat{\theta}_2 < 0$ and $\hat{\theta}_1 > \hat{\theta}_2 + 8$	Restrict to sub-population 1 and test H_1 only, needing to reject H_1 and H_{01}.
Else,	Continue with full population and test H_0, needing to reject H_0 and H_{01}.

The same *total* sample size for Stage 2 is retained in both cases, increasing the numbers for the sub-population when enrichment occurs.

Enrichment: An adaptive design

Each null hypothesis, H_i say, is tested in a 2-stage group sequential test.

With Z-statistics Z_1 and Z_2 from Stages 1 and 2, H_i is rejected if

$$Z_1 \geq 0 \quad \text{and} \quad \tfrac{1}{\sqrt{2}} Z_1 + \tfrac{1}{\sqrt{2}} Z_2 \geq 1.95.$$

When continuing with the full population, we use Z-statistics:

	Stage 1	*Stage 2*
H_0	$Z_{1,0}$	$Z_{2,0}$
H_{01}	$Z_{1,0}$	$Z_{2,0}$

where $Z_{i,0}$ is based on $\hat{\theta}$ from responses in Stage i.

So, there is no change from the original test of H_0.

Enrichment: An adaptive design

With Z-statistics Z_1 and Z_2 from Stages 1 and 2, H_i is rejected if

$$Z_1 \geq 0 \quad \text{and} \quad \tfrac{1}{\sqrt{2}} Z_1 + \tfrac{1}{\sqrt{2}} Z_2 \geq 1.95.$$

When switching to sub-population 1, we use:

	Stage 1	*Stage 2*
H_1	$Z_{1,1}$	$Z_{2,1}$
H_{01}	$Z_{1,0}$	$Z_{2,1}$

where $Z_{i,j}$ is based on $\hat{\theta}_j$ from responses in Stage i.

The need to reject the intersection hypothesis H_{01} adds an extra requirement to the simple test of H_1.

Simulation results: Power of non-adaptive and adaptive designs

				Non-adaptive	***Adaptive***		
	θ_1	θ_2	θ	*Full pop^n*	*Sub-pop 1 only*	*Full pop^n*	*Total*
1.	30	0	15	**0.68**	0.43	0.42	**0.85**
2.	20	0	10	**0.37**	0.24	0.26	**0.51**
3.	20	20	20	**0.90**	0.03	0.87	**0.90**
4.	20	10	15	**0.68**	0.11	0.60	**0.71**

Cases 1 & 2: Testing focuses (correctly) on H_1, but it is still possible to find an effect (wrongly) for the full population. Overall power is increased.

Case 3: Restricting to the sub-population reduces power for finding an effect in the full population.

Case 4: Adaptation improves overall power a little.

Increasing power for finding a sub-population effect

Greater power for the sub-population can be achieved by using $Z_{1,1}$ rather than $Z_{1,0}$ as the Stage 1 statistic in the test of H_{01}.

This gives the following results:

				Non-adaptive	***Adaptive***		
	θ_1	θ_2	θ	*Full pop^n*	*Sub-pop 1 only*	*Full pop^n*	*Total*
1.	30	0	15	**0.68**	0.47	0.40	**0.87**
2.	20	0	10	**0.37**	0.35	0.23	**0.58**
3.	20	20	20	**0.90**	0.04	0.74	**0.78**
4.	20	10	15	**0.68**	0.16	0.51	**0.56**

Benefits in Case 2 are balanced by loss of overall power in Cases 3 and 4.

Increasing power for finding a sub-population effect

As a compromise between the two previous methods, a combination* of $Z_{1,0}$ and $Z_{1,1}$ may be used as the Stage 1 statistic for the test of H_{01}.

This leads to the following results:

				Non-adaptive	***Adaptive***		
	θ_1	θ_2	θ	*Full pop^n*	*Sub-pop 1 only*	*Full pop^n*	*Total*
1.	30	0	15	**0.68**	0.47	0.41	**0.88**
2.	20	0	10	**0.37**	0.33	0.25	**0.58**
3.	20	20	20	**0.90**	0.04	0.83	**0.87**
4.	20	10	15	**0.68**	0.15	0.57	**0.72**

*Specifically, $(Z_{1,0} + Z_{1,1})/\sqrt{(2+\sqrt{2})}$, which is $N(0,1)$ under H_{01}.

Enrichment: Example

The rules for sticking or switching to a sub-population can be adjusted, but we cannot eliminate the probability of making an error in these decisions.

This is to be expected since the standard error of interim estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ is 12.3 — much higher than the differences between θ_1 and θ_2 that interest us.

Conclusion:

Restricting attention to a sub-population can be effective in improving power.

However, higher overall sample size is needed for accurate sub-population inference . . .

Increasing power for finding a sub-population effect

To match the non-adaptive test in cases 2 and 3, and obtain the benefits of adaptation elsewhere, increase the overall sample size by 30%.

With a combination* of $Z_{1,0}$ and $Z_{1,1}$ as the Stage 1 statistic for testing H_{01}, we obtain the following results:

				Non-adaptive	***Adaptive, 1.3 x sample size***		
	θ_1	θ_2	θ	*Full pop^n*	*Sub-pop 1 only*	*Full pop^n*	*Total*
1.	30	0	15	**0.68**	0.49	0.45	**0.94**
2.	20	0	10	**0.37**	0.38	0.30	**0.69**
3.	20	20	20	**0.90**	0.03	0.92	**0.94**
4.	20	10	15	**0.68**	0.15	0.68	**0.82**

*Using $(Z_{1,0} + Z_{1,1})/\sqrt{(2+\sqrt{2})}$.

Conclusions

1. *Adaptive Methods* provide a useful route to modifying sample size as a ***nuisance parameter*** is estimated.

2. However, they are an inferior option to *Group Sequential Tests* if one wishes to respond to estimates of the ***primary endpoint***.

3. Adaptive methods lead to moderate efficiency gains when ***restricting to a sub-population***. But, remember that interim estimates will have high variance.

We recommend adaptation as part of a pre-planned and *pre-tested* trial design — "flexible adaptation" brings risks as well as opportunities.