Group Sequential Tests for Delayed Responses

Lisa Hampson
Department of Mathematics and Statistics,
Lancaster University, UK

Chris Jennison
Department of Mathematical Sciences,
University of Bath, UK

Basler Biometric Section
Basel, July 2012

Outline

Outline

Outline

Outline

Outline

Superiority trials

We conduct a clinical trial comparing a new treatment versus control. As the trial progresses, we accumulate responses

- $X_{A,i} \sim N(\mu_A, \sigma^2)$, $i = 1, 2, \ldots$, on the new treatment
- $X_{B,i} \sim N(\mu_B, \sigma^2)$, $i = 1, 2, \ldots$, on the control treatment.

Superiority trials

We conduct a clinical trial comparing a new treatment versus control. As the trial progresses, we accumulate responses

- $X_{A,i} \sim N(\mu_A, \sigma^2)$, $i = 1, 2, \ldots$, on the new treatment
- $X_{B,i} \sim N(\mu_B, \sigma^2)$, $i = 1, 2, \ldots$, on the control treatment.

We assume that all responses are independent and σ^2 is known.

Define $\theta = \mu_A - \mu_B$ to be the “effect size” for the new treatment.

Superiority trials

We conduct a clinical trial comparing a new treatment versus control. As the trial progresses, we accumulate responses

- $X_{A,i} \sim N(\mu_A, \sigma^2)$, $i = 1, 2, \ldots$, on the new treatment
- $X_{B,i} \sim N(\mu_B, \sigma^2)$, $i = 1, 2, \ldots$, on the control treatment.

We assume that all responses are independent and σ^2 is known.

Define $\theta = \mu_A - \mu_B$ to be the "effect size" for the new treatment.

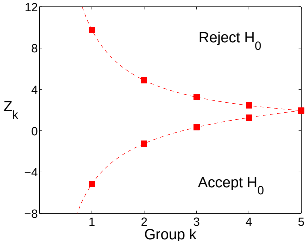
We wish to test

$$H_0 : \theta \leq 0 \quad \text{vs} \quad \theta > 0$$

with type I error rate α at $\theta = 0$ and power $1 - \beta$ at $\theta = \delta > 0$.

One-sided group sequential tests

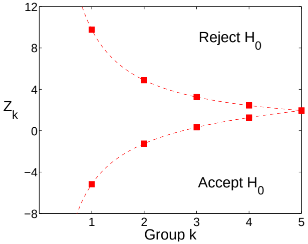
A one-sided group sequential test of $H_0 : \theta \leq 0$ against $\theta > 0$ is of the form



- Decision of whether to reject or accept H_0 is made on basis of responses observed at time of interim analysis.

One-sided group sequential tests

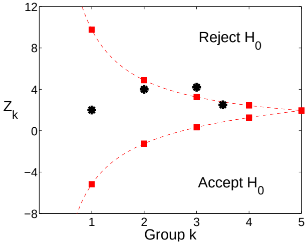
A one-sided group sequential test of $H_0 : \theta \leq 0$ against $\theta > 0$ is of the form



- Decision of whether to reject or accept H_0 is made on basis of responses observed at time of interim analysis.
- "Immediate responses" fit nicely into this framework since the flow of data stops immediately on termination of recruitment.

One-sided group sequential tests

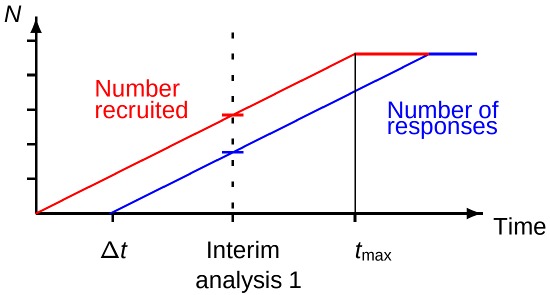
A one-sided group sequential test of $H_0 : \theta \leq 0$ against $\theta > 0$ is of the form



- Decision of whether to reject or accept H_0 is made on basis of responses observed at time of interim analysis.
- "Immediate responses" fit nicely into this framework since the flow of data stops immediately on termination of recruitment.

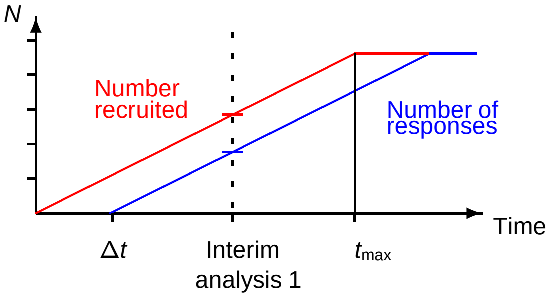
Incorporating delayed responses into GSTs

Consider a trial where response is observed time Δ_t after treatment.



Incorporating delayed responses into GSTs

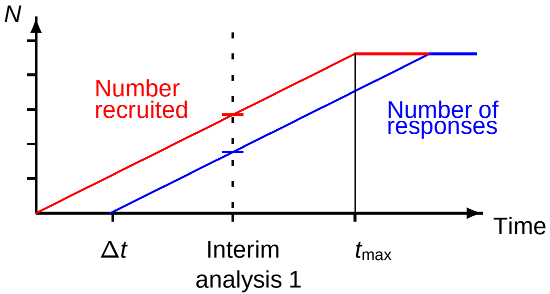
Consider a trial where response is observed time Δ_t after treatment.



We assume information is proportional to the observed number of responses.

Incorporating delayed responses into GSTs

Consider a trial where response is observed time Δ_t after treatment.

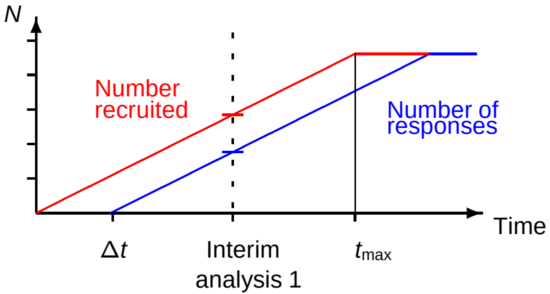


We assume information is proportional to the observed number of responses.

We will equally space interim analyses between times Δ_t and t_{max}.

Incorporating delayed responses into GSTs

Consider a trial where response is observed time Δ_t after treatment.



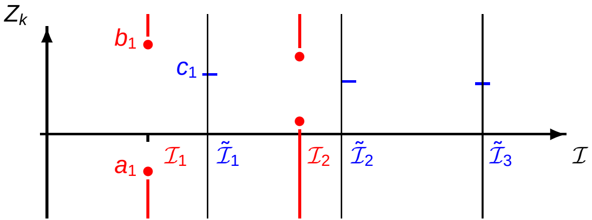
We assume information is proportional to the observed number of responses.

We will equally space interim analyses between times Δ_t and t_{max}.

T.W. Anderson (*JASA*, 1964) considers sequential tests for delayed responses. We follow this basic structure to construct GSTs.

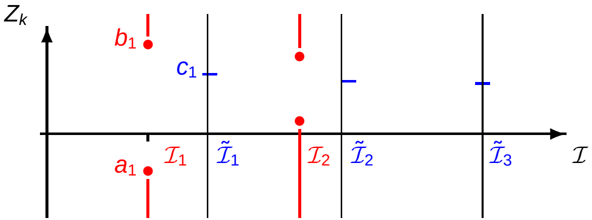
Boundaries for a Delayed Response GST

At interim analysis k, Z_k is associated with information level $\mathcal{I}_k = \text{Var}(\hat{\theta}_k)$.



Boundaries for a Delayed Response GST

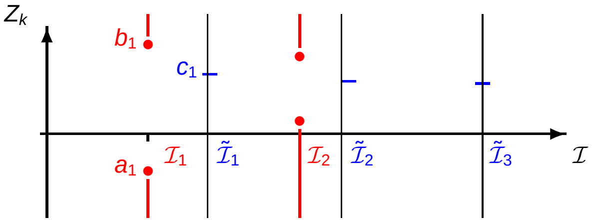
At interim analysis k, Z_k is associated with information level $\mathcal{I}_k = \text{Var}(\hat{\theta}_k)$.



If $Z_k > b_k$ or $Z_k < a_k$, cease enrollment of future patients and follow-up all recruited subjects.

Boundaries for a Delayed Response GST

At interim analysis k, Z_k is associated with information level $\mathcal{I}_k = \text{Var}(\hat{\theta}_k)$.



If $Z_k > b_k$ or $Z_k < a_k$, cease enrollment of future patients and follow-up all recruited subjects.

At the decision analysis, based on information $\tilde{\mathcal{I}}_k$, reject H_0 if $\tilde{Z}_k > c_k$.

Calculating properties of Delayed Response GSTs

Calculations of test properties (type I error rate, power, $\mathbb{E}_\theta(N)$) require the joint distributions of test statistic sequences:

- $\{Z_1, \ldots, Z_k, \tilde{Z}_k\}, \qquad$ for $k = 1, \ldots, K-1$,
- $\{Z_1, \ldots, Z_{K-1}, \tilde{Z}_K\}$.

Calculating properties of Delayed Response GSTs

Calculations of test properties (type I error rate, power, $\mathbb{E}_\theta(N)$) require the joint distributions of test statistic sequences:

- $\{Z_1, \dots, Z_k, \tilde{Z}_k\}$, for $k = 1, \dots, K-1$,
- $\{Z_1, \dots, Z_{K-1}, \tilde{Z}_K\}$.

Each sequence is based on accumulating datasets.

Given $\{\mathcal{I}_1, \dots, \mathcal{I}_k, \tilde{\mathcal{I}}_k\}$, the sequence $\{Z_1, \dots, Z_k, \tilde{Z}_k\}$ follows the canonical distribution for statistics generated by a GST for immediate responses (Jennison & Turnbull, *JASA*, 1997).

Calculating properties of Delayed Response GSTs

Calculations of test properties (type I error rate, power, $\mathbb{E}_\theta(N)$) require the joint distributions of test statistic sequences:

- $\{Z_1, \dots, Z_k, \tilde{Z}_k\}$, for $k = 1, \dots, K - 1$,
- $\{Z_1, \dots, Z_{K-1}, \tilde{Z}_K\}$.

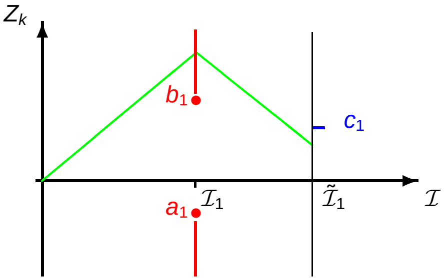
Each sequence is based on accumulating datasets.

Given $\{\mathcal{I}_1, \dots, \mathcal{I}_k, \tilde{\mathcal{I}}_k\}$, the sequence $\{Z_1, \dots, Z_k, \tilde{Z}_k\}$ follows the canonical distribution for statistics generated by a GST for immediate responses (Jennison & Turnbull, *JASA*, 1997).

Properties of Delayed Response GSTs can therefore be calculated using numerical routines devised for standard designs.

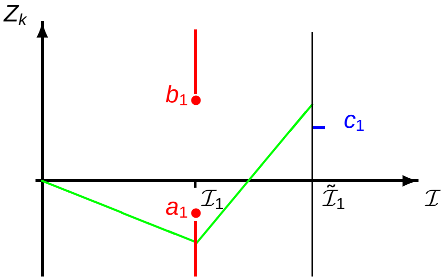
Reversals of anticipated final decisions

Stopping with $Z_k > b_k$ or $Z_k < a_k$ indicates our *likely* final decision but there there may be a reversal. We could observe



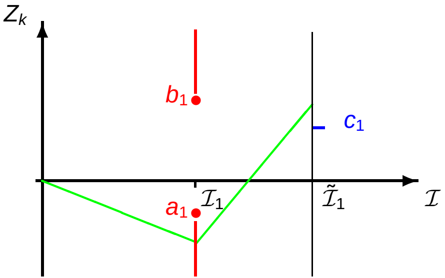
Reversals of anticipated final decisions

Stopping with $Z_k > b_k$ or $Z_k < a_k$ indicates our *likely* final decision but there there may be a reversal. We could observe



Reversals of anticipated final decisions

Stopping with $Z_k > b_k$ or $Z_k < a_k$ indicates our *likely* final decision but there there may be a reversal. We could observe



We optimise our designs to maximise the value of the additional pipeline responses for increasing the test's power.

Optimal Delayed Response GSTs

Let N represent the total number of subjects recruited.

Let r be the fraction of a test's maximum sample size in the pipeline at each interim analysis.

Optimal Delayed Response GSTs

Let N represent the total number of subjects recruited.

Let r be the fraction of a test's maximum sample size in the pipeline at each interim analysis.

Objective: For a given r, maximum sample size n_{max}, stages K and analysis schedule, we find the Delayed Response GST minimising

$$F = \int \mathbb{E}_{\theta}(N) \, f(\theta) \, d\theta$$

with type I error rate α at $\theta = 0$ and power $1 - \beta$ at $\theta = \delta$. Here $f(\theta)$ is the density of a $N(\delta/2, (\delta/2)^2)$ distribution.

Optimal Delayed Response GSTs

Let N represent the total number of subjects recruited.

Let r be the fraction of a test's maximum sample size in the pipeline at each interim analysis.

Objective: For a given r, maximum sample size n_{max}, stages K and analysis schedule, we find the Delayed Response GST minimising

$$F = \int \mathbb{E}_\theta(N) \, f(\theta) \, d\theta$$

with type I error rate α at $\theta = 0$ and power $1 - \beta$ at $\theta = \delta$. Here $f(\theta)$ is the density of a $N(\delta/2, (\delta/2)^2)$ distribution.

We create an unconstrained Bayes problem by adding a prior on θ and costs for sampling and for making incorrect decisions. We search for the combination of prior and costs which gives a solution with frequentist error rates α and β.

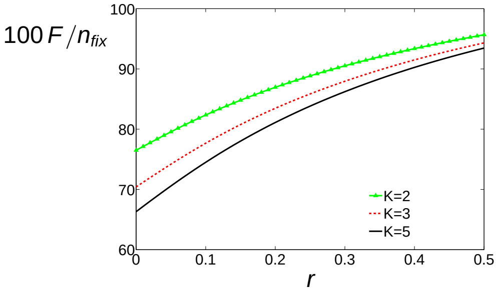
Efficiency loss when there is a delay in response

It is required to test $H_0 : \theta \leq 0$ against $\theta > 0$ with $\alpha = 0.025$ and $\beta = 0.1$. Suppose the fixed sample test requires n_{fix} subjects and set $n_{max} = 1.1\, n_{fix}$.

Efficiency loss when there is a delay in response

It is required to test $H_0 : \theta \leq 0$ against $\theta > 0$ with $\alpha = 0.025$ and $\beta = 0.1$. Suppose the fixed sample test requires n_{fix} subjects and set $n_{max} = 1.1\, n_{fix}$.

We plot the minima of F attained by optimal tests with $K = 2, 3$ and 5 stages.



When $r = 0.1$, almost 25% of the gains of group sequential testing are lost. When $r = 0.3$, this increases up to 60%.

Designing a Delayed Response GST

Example A: Cholesterol reduction after 4 weeks of treatment

Responses are assumed normally distributed with variance $\sigma^2 = 2$.

It is required to test $H_0 : \theta \leq 0$ against $\theta > 0$ with

- type I error rate $\alpha = 0.025$ at $\theta = 0$,
- power $1 - \beta = 0.9$ at $\theta = \delta = 1.0$.

Designing a Delayed Response GST

Example A: Cholesterol reduction after 4 weeks of treatment

Responses are assumed normally distributed with variance $\sigma^2 = 2$.

It is required to test $H_0 : \theta \leq 0$ against $\theta > 0$ with

- type I error rate $\alpha = 0.025$ at $\theta = 0$,
- power $1 - \beta = 0.9$ at $\theta = \delta = 1.0$.

The fixed sample test needs $n_{fix} = 86$ subjects divided between the two treatments.

Designing a Delayed Response GST

Example A: Cholesterol reduction after 4 weeks of treatment

Responses are assumed normally distributed with variance $\sigma^2 = 2$.

It is required to test $H_0 : \theta \leq 0$ against $\theta > 0$ with

- type I error rate $\alpha = 0.025$ at $\theta = 0$,
- power $1 - \beta = 0.9$ at $\theta = \delta = 1.0$.

The fixed sample test needs $n_{fix} = 86$ subjects divided between the two treatments.

We consider designs with a maximum sample size of 96, assuming a recruitment rate of 4 per week, giving $4 \times 4 = 16$ pipeline subjects at each interim analysis.

Designing a Delayed Response GST

Once the trial is underway, data start to accrue after 4 weeks. Recruitment will close after 24 weeks.

Interim analyses are planned after $n_1 = 28$ and $n_2 = 54$ observed responses.

Designing a Delayed Response GST

Once the trial is underway, data start to accrue after 4 weeks. Recruitment will close after 24 weeks.

Interim analyses are planned after $n_1 = 28$ and $n_2 = 54$ observed responses.

A decision analysis will be based on

- $\tilde{n}_1 = 44$ responses if recruitment stops at interim analysis 1
- $\tilde{n}_2 = 70$ responses if recruitment stops at interim analysis 2
- $\tilde{n}_3 = 96$ responses in the absence of early stopping.

Designing a Delayed Response GST

Once the trial is underway, data start to accrue after 4 weeks. Recruitment will close after 24 weeks.

Interim analyses are planned after $n_1 = 28$ and $n_2 = 54$ observed responses.

A decision analysis will be based on

- $\tilde{n}_1 = 44$ responses if recruitment stops at interim analysis 1
- $\tilde{n}_2 = 70$ responses if recruitment stops at interim analysis 2
- $\tilde{n}_3 = 96$ responses in the absence of early stopping.

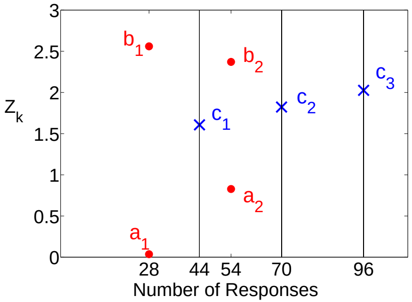
We derive a Delayed Response GST minimising

$$F = \int \mathbb{E}_{\theta}(N) f(\theta) \, d\theta,$$

where $f(\theta)$ is the density of a $N(0.5, 0.5^2)$ distribution.

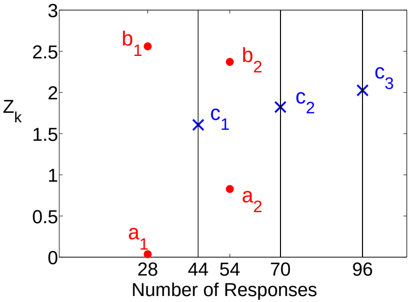
Designing a Delayed Response GST

Critical values for the optimised Delayed Response GST are shown below.



Designing a Delayed Response GST

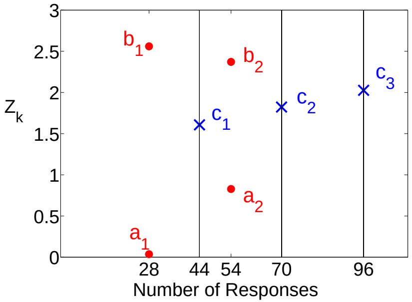
Critical values for the optimised Delayed Response GST are shown below.



Critical values c_1 and c_2 are well below b_1 and b_2, so the probability of a reversal is small.

Designing a Delayed Response GST

Critical values for the optimised Delayed Response GST are shown below.



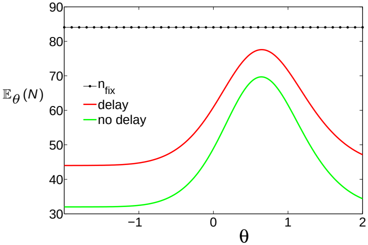
Critical values c_1 and c_2 are well below b_1 and b_2, so the probability of a reversal is small.

Both c_1 and c_2 are less than 1.96. If desired, these can be raised to 1.96 with little change to the design's power curve.

Designing a Delayed Response GST

The figure shows expected sample size curves for

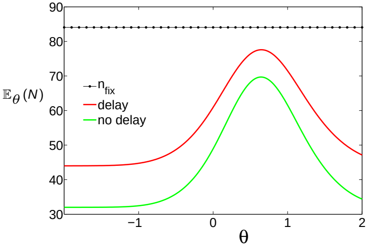
- the fixed sample test with $n_{fix} = 85$ patients,
- the Delayed Response GST minimising F,
- the GST for immediate responses with analyses after 32, 64 and 96 responses, also minimising F.



Designing a Delayed Response GST

The figure shows expected sample size curves for

- the fixed sample test with $n_{fix} = 85$ patients,
- the Delayed Response GST minimising F,
- the GST for immediate responses with analyses after 32, 64 and 96 responses, also minimising F.



The delay in response means savings in $\mathbb{E}_\theta(N)$ are smaller than they would be if response were immediate.

Making inferences on termination

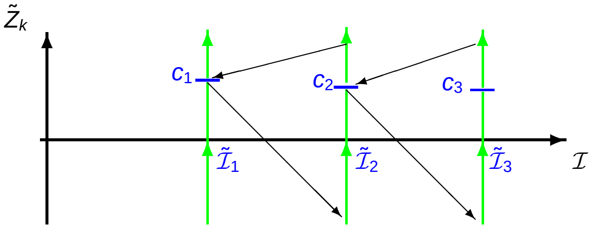
How can we calculate a p-value for $H_0 : \theta \leq 0$ and a CI for θ?

On termination of the test at stage T, $(\tilde{\mathcal{I}}_T, \tilde{Z}_T)$ is a sufficient statistic for θ. We base inferences on a "stage-wise" ordering of the test's sample space for this pair.

Making inferences on termination

How can we calculate a p-value for $H_0 : \theta \leq 0$ and a CI for θ?

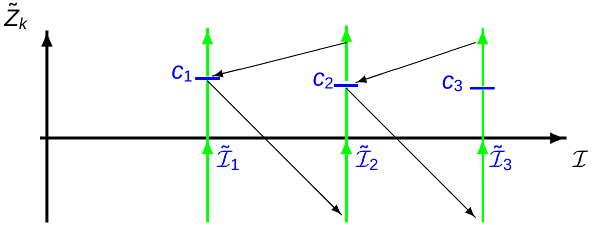
On termination of the test at stage T, $(\tilde{\mathcal{I}}_T, \tilde{Z}_T)$ is a sufficient statistic for θ. We base inferences on a “stage-wise” ordering of the test's sample space for this pair.



Making inferences on termination

How can we calculate a p-value for $H_0 : \theta \leq 0$ and a CI for θ?

On termination of the test at stage T, $(\tilde{\mathcal{I}}_T, \tilde{Z}_T)$ is a sufficient statistic for θ. We base inferences on a "stage-wise" ordering of the test's sample space for this pair.

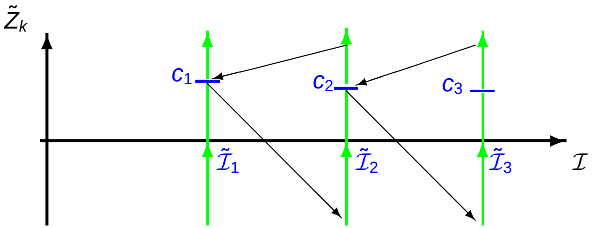


The sample space at $\tilde{\mathcal{I}}_T = \tilde{\mathcal{I}}_k$ is partitioned by c_k into "high" and "low" sets.

Making inferences on termination

How can we calculate a p-value for $H_0 : \theta \leq 0$ and a CI for θ?

On termination of the test at stage T, $(\tilde{\mathcal{I}}_T, \tilde{Z}_T)$ is a sufficient statistic for θ. We base inferences on a "stage-wise" ordering of the test's sample space for this pair.



The sample space at $\tilde{\mathcal{I}}_T = \tilde{\mathcal{I}}_k$ is partitioned by c_k into "high" and "low" sets.

This ordering ensures p-value calculations do not depend on future, possibly *unpredictable*, information levels.

Error spending Delayed Response GSTs

We design error spending Delayed Response GSTs which

- reach a target information level $\mathcal{I}_{max}$ in absence of early stopping,
- spend error probabilities as a function of $\mathcal{I}/\mathcal{I}_{max}$.

Error spending Delayed Response GSTs

We design error spending Delayed Response GSTs which

- reach a target information level $\mathcal{I}_{max}$ in absence of early stopping,
- spend error probabilities as a function of $\mathcal{I}/\mathcal{I}_{max}$.

Let π_k and γ_k be cumulative type I and II error rates to be spent by stage k.

Error spending Delayed Response GSTs

We design error spending Delayed Response GSTs which

- reach a target information level $\mathcal{I}_{max}$ in absence of early stopping,
- spend error probabilities as a function of $\mathcal{I}/\mathcal{I}_{max}$.

Let π_k and γ_k be cumulative type I and II error rates to be spent by stage k.

Choosing c_k to balance reversal probabilities under $\theta = 0$ implies we may choose (a_k, b_k) to satisfy

$$\mathbb{P}_{\theta=0}\{Z_1 \in \mathcal{C}_1, \ldots, Z_{k-1} \in \mathcal{C}_{k-1}, Z_k \geq b_k\} = \pi_k - \pi_{k-1}$$
$$\mathbb{P}_{\theta=\delta}\{Z_1 \in \mathcal{C}_1, \ldots, Z_{k-1} \in \mathcal{C}_{k-1}, Z_k \leq a_k\} = \gamma_k - \gamma_{k-1},$$

and control the type I error rate at level α, and the type II error rate at a level just below β.

Error spending Delayed Response GSTs

We design error spending Delayed Response GSTs which

- reach a target information level $\mathcal{I}_{max}$ in absence of early stopping,
- spend error probabilities as a function of $\mathcal{I}/\mathcal{I}_{max}$.

Let π_k and γ_k be cumulative type I and II error rates to be spent by stage k.

Choosing c_k to balance reversal probabilities under $\theta = 0$ implies we may choose (a_k, b_k) to satisfy

$$\mathbb{P}_{\theta=0}\{Z_1 \in \mathcal{C}_1, \ldots, Z_{k-1} \in \mathcal{C}_{k-1}, Z_k \geq b_k\} = \pi_k - \pi_{k-1}$$
$$\mathbb{P}_{\theta=\delta}\{Z_1 \in \mathcal{C}_1, \ldots, Z_{k-1} \in \mathcal{C}_{k-1}, Z_k \leq a_k\} = \gamma_k - \gamma_{k-1},$$

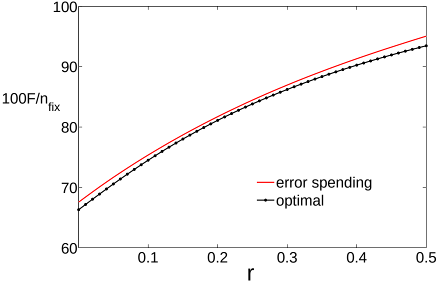
and control the type I error rate at level α, and the type II error rate at a level just below β.

Under this construction, the stage k stopping rule can be set without knowledge of $\tilde{\mathcal{I}}_k$.

Efficiency of error spending tests

In the figure below, error spending tests are designed using the ρ-family of error spending functions.

Values of F are attained by tests designed and conducted with $K = 5$, $n_{max} = 1.1\, n_{fix}$, $\alpha = 0.025$ and $\beta = 0.1$.



Error spending Delayed Response GSTs are flexible and closely match the optimal tests for savings in $\mathbb{E}_\theta(N)$.

Dealing with unexpected overrunning

Suppose a standard GST designed with $\mathcal{I}_k$ and boundaries (a_k, b_k) stops at analysis $k^\star < K$ with $Z_{k^\star} > b_{k^\star}$ or $Z_{k^\star} < a_{k^\star}$.

Dealing with unexpected overrunning

Suppose a standard GST designed with $\mathcal{I}_k$ and boundaries (a_k, b_k) stops at analysis $k^\star < K$ with $Z_{k^\star} > b_{k^\star}$ or $Z_{k^\star} < a_{k^\star}$.

Question: If additional data are observed, how can these be incorporated into the final analysis while preserving the type I error rate?

Dealing with unexpected overrunning

Suppose a standard GST designed with $\mathcal{I}_k$ and boundaries (a_k, b_k) stops at analysis $k^\star < K$ with $Z_{k^\star} > b_{k^\star}$ or $Z_{k^\star} < a_{k^\star}$.

Question: If additional data are observed, how can these be incorporated into the final analysis while preserving the type I error rate?

Solution: We partition the sample space at $\tilde{\mathcal{I}}_{k^\star}$ such that

- if $\tilde{Z}_{k^\star} \geq c_{k^\star}$, reject H_0,
- if $\tilde{Z}_{k^\star} \leq c_{k^\star}$, accept H_0.

Requiring $c_{k^\star}$ to balance the probabilities of reversing decisions under $\theta = 0$ at stage $k^\star$ preserves the test's overall type I error rate.

Dealing with unexpected overrunning

Suppose a standard GST designed with $\mathcal{I}_k$ and boundaries (a_k, b_k) stops at analysis $k^\star < K$ with $Z_{k^\star} > b_{k^\star}$ or $Z_{k^\star} < a_{k^\star}$.

Question: If additional data are observed, how can these be incorporated into the final analysis while preserving the type I error rate?

Solution: We partition the sample space at $\tilde{\mathcal{I}}_{k^\star}$ such that

- if $\tilde{Z}_{k^\star} \geq c_{k^\star}$, reject H_0,
- if $\tilde{Z}_{k^\star} \leq c_{k^\star}$, accept H_0.

Requiring $c_{k^\star}$ to balance the probabilities of reversing decisions under $\theta = 0$ at stage $k^\star$ preserves the test's overall type I error rate.

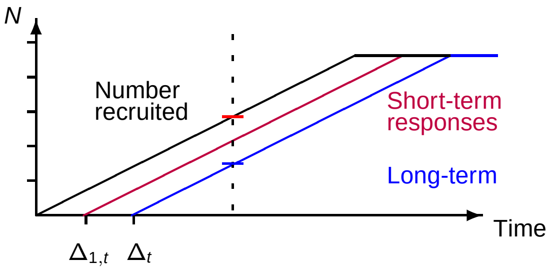
In addition, p-value calculations do not depend on $\tilde{\mathcal{I}}_1, \ldots, \tilde{\mathcal{I}}_{k^\star-1}$, nor on information levels beyond stage $k^\star$.

Using a short term endpoint to recover efficiency

Suppose a second endpoint, correlated with the primary response, is available soon after treatment.

Using a short term endpoint to recover efficiency

Suppose a second endpoint, correlated with the primary response, is available soon after treatment.

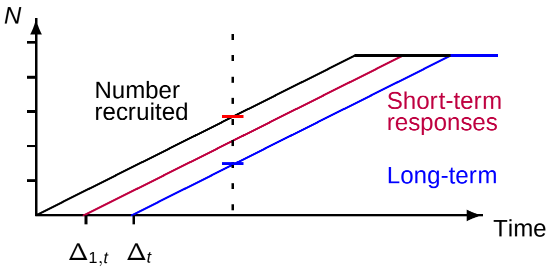


For each patient i on treatment $T = A$ or B, we measure

- a short-term response $Y_{T,i}$

Using a short term endpoint to recover efficiency

Suppose a second endpoint, correlated with the primary response, is available soon after treatment.

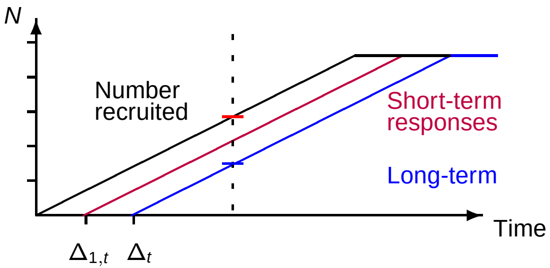


For each patient i on treatment $T = A$ or B, we measure

- a short-term response $Y_{T,i}$
- a long-term response $X_{T,i}$.

Using a short term endpoint to recover efficiency

Suppose a second endpoint, correlated with the primary response, is available soon after treatment.



For each patient i on treatment $T = A$ or B, we measure

- a short-term response $Y_{T,i}$
- a long-term response $X_{T,i}$.

Using a short term endpoint to recover efficiency

Suppose each pair $(Y_{T,i}, X_{T,i})$ has joint distribution

$$\begin{pmatrix} Y_{T,i} \\ X_{T,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{T,1} \\ \mu_{T,2} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

Using a short term endpoint to recover efficiency

Suppose each pair $(Y_{T,i}, X_{T,i})$ has joint distribution

$$\begin{pmatrix} Y_{T,i} \\ X_{T,i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{T,1} \\ \mu_{T,2} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

At interim analysis k, we estimate $\theta = \mu_{A,2} - \mu_{B,2}$ from all available data, using maximum likelihood estimation to fit the full model then extracting $\widehat{\theta}_k$ and $\mathcal{I}_k = \text{Var}(\widehat{\theta}_k)$.

Using a short term endpoint to recover efficiency

Suppose each pair $(Y_{T,i}, X_{T,i})$ has joint distribution

$$\begin{pmatrix} Y_{T,i} \\ X_{T,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{T,1} \\ \mu_{T,2} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

At interim analysis k, we estimate $\theta = \mu_{A,2} - \mu_{B,2}$ from all available data, using maximum likelihood estimation to fit the full model then extracting $\widehat{\theta}_k$ and $\mathcal{I}_k = \mathrm{Var}(\widehat{\theta}_k)$.

Given $\{\mathcal{I}_1, \ldots, \mathcal{I}_k, \tilde{\mathcal{I}}_k\}$, the sequence of estimates $\{\widehat{\theta}_k\}$ follows the canonical joint distribution for a group sequential trial.

Using a short term endpoint to recover efficiency

Suppose each pair $(Y_{T,i}, X_{T,i})$ has joint distribution

$$\begin{pmatrix} Y_{T,i} \\ X_{T,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{T,1} \\ \mu_{T,2} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

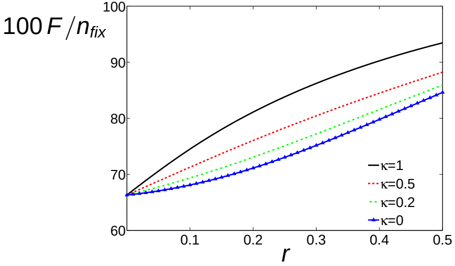
At interim analysis k, we estimate $\theta = \mu_{A,2} - \mu_{B,2}$ from all available data, using maximum likelihood estimation to fit the full model then extracting $\widehat{\theta}_k$ and $\mathcal{I}_k = \text{Var}(\widehat{\theta}_k)$.

Given $\{\mathcal{I}_1, \ldots, \mathcal{I}_k, \tilde{\mathcal{I}}_k\}$, the sequence of estimates $\{\widehat{\theta}_k\}$ follows the canonical joint distribution for a group sequential trial.

At decision analysis k when all subjects are fully observed, short-term responses don't contribute any additional information for θ.

Revisiting Example A

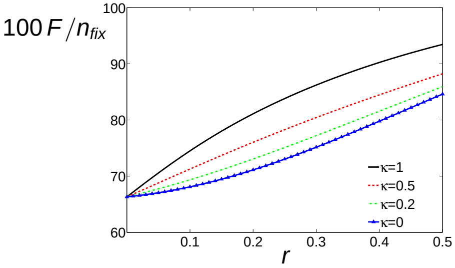
Example A: Incorporating a second, short-term endpoint



We assume $Y_{T,i}$ and $X_{T,i}$ have correlation 0.9.

Revisiting Example A

Example A: Incorporating a second, short-term endpoint

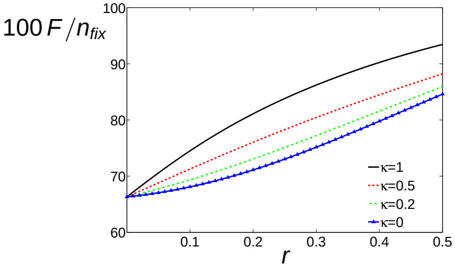


We assume $Y_{T,i}$ and $X_{T,i}$ have correlation 0.9.

The ratio of time to short-term and long-term endpoints is κ.

Revisiting Example A

Example A: Incorporating a second, short-term endpoint



We assume $Y_{T,i}$ and $X_{T,i}$ have correlation 0.9.

The ratio of time to short-term and long-term endpoints is κ.

The solid line for $\kappa = 1$ is the case of no short-term endpoint.

Conclusions

In this presentation, we have presented

- Delayed Response GSTs as a coherent approach to handling delayed data in a sequential setting.
- Versions of Delayed Response GSTs that can accommodate unpredictable group sizes and unexpected overrunning.
- P-values and confidence intervals on termination.

The impact on efficiency of a delay in response can be ameliorated by

- incorporating information on correlated short-term endpoints
- slowing recruitment rates
- ensuring rapid data cleaning before an analysis.