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

1. Group sequential tests
Outline

1. Group sequential tests
2. Optimal designs
Outline

1. Group sequential tests
2. Optimal designs
3. Extensions
Outline

1. Group sequential tests
2. Optimal designs
3. Extensions
4. Recovering efficiency
Outline

1. Group sequential tests
2. Optimal designs
3. Extensions
4. Recovering efficiency
5. Summary
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 $\sigma^2$ is known.

Define $\theta = \mu_A - \mu_B$ to be the “effect size” for the new treatment.
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 \( \sigma^2 \) 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 \( \alpha \) at \( \theta = 0 \) and power \( 1 - \beta \) at \( \theta = \delta > 0 \).
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.
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.
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 $\Delta t$ after treatment.

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
Incorporating delayed responses into GSTs

Consider a trial where response is observed time $\Delta t$ 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 $\Delta t$ after treatment.

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

We will equally space interim analyses between times $\Delta t$ and $t_{max}$. 
Consider a trial where response is observed time $\Delta t$ after treatment.

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

We will equally space interim analyses between times $\Delta t$ and $t_{\text{max}}$.

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 $\mathcal{I}_k = \text{Var}(\hat{\theta}_k)$. 

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
At interim analysis $k$, $Z_k$ is associated with information level $I_k = \text{Var}(\hat{\theta}_k)$.

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 $\mathcal{I}_k$, reject $H_0$ if $\tilde{Z}_k > c_k$. 
Calculations of test properties (type I error rate, power, $\mathbb{E}_\theta(N)$) require the joint distributions of test statistic sequences:

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

Lisa Hampson and Chris Jennison

Group Sequential Tests for Delayed Responses
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:

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

Each sequence is based on accumulating datasets.

Given $\{I_1, \ldots, I_k, \tilde{I}_k\}$, the sequence $\{Z_1, \ldots, Z_k, \tilde{Z}_k\}$ follows the canonical distribution for statistics generated by a GST for immediate responses (Jennison & Turnbull, JASA, 1997).
Calculations of test properties (type I error rate, power, $E_\theta(N)$) require the joint distributions of test statistic sequences:

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

Each sequence is based on accumulating datasets.

Given $\{I_1, \ldots, I_k, \tilde{I}_k\}$, the sequence $\{Z_1, \ldots, Z_k, \tilde{Z}_k\}$ 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 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 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 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 $n_{\text{max}}$, 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 $\alpha$ 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 $n_{\text{max}}$, stages $K$ and analysis schedule, we find the Delayed Response GST minimising

$$F = \int E_\theta(N) f(\theta) \, d\theta$$

with type I error rate $\alpha$ 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 $\theta$ 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 $\alpha$ and $\beta$. 

Lisa Hampson and Chris Jennison

Group Sequential Tests for Delayed Responses
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 \( n_{\text{fix}} \) subjects and set \( n_{\text{max}} = 1.1 \, n_{\text{fix}} \).

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%. 

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
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 $\theta = \delta = 1.0$. 
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.
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.
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.

Lisa Hampson and Chris Jennison

Group Sequential Tests for Delayed Responses
Designing a Delayed Response GST

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

![Graph showing critical values](image)

Critical values $c_1$ and $c_2$ are well below $b_1$ and $b_2$, so the probability of a reversal is small.
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 $n_{\text{fix}} = 85$ 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 $\theta$?

On termination of the test at stage $T$, $(\tilde{I}_T, \tilde{Z}_T)$ is a sufficient statistic for $\theta$. 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 $\theta$?

On termination of the test at stage $T$, $(\tilde{I}_T, \tilde{Z}_T)$ is a sufficient statistic for $\theta$. 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 $\theta$?

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

The sample space at $\tilde{I}_T = \tilde{I}_k$ 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 $\theta$?

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

The sample space at $\tilde{I}_T = \tilde{I}_k$ 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.
We design error spending Delayed Response GSTs which

- reach a target information level $I_{\text{max}}$ in absence of early stopping,
- spend error probabilities as a function of $I/I_{\text{max}}$. 
We design error spending Delayed Response GSTs which
- reach a target information level $I_{\text{max}}$ in absence of early stopping,
- spend error probabilities as a function of $I/I_{\text{max}}$.

Let $\pi_k$ and $\gamma_k$ 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 $I_{\text{max}}$ in absence of early stopping,
- spend error probabilities as a function of $I/I_{\text{max}}$.

Let $\pi_k$ and $\gamma_k$ 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

$$
P_{\theta=0}\{Z_1 \in C_1, \ldots, Z_{k-1} \in C_{k-1}, Z_k \geq b_k\} = \pi_k - \pi_{k-1}
$$

$$
P_{\theta=\delta}\{Z_1 \in C_1, \ldots, Z_{k-1} \in C_{k-1}, Z_k \leq a_k\} = \gamma_k - \gamma_{k-1},
$$

and control the type I error rate at level $\alpha$, and the type II error rate at a level just below $\beta$. 

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
We design error spending Delayed Response GSTs which
- reach a target information level $I_{\text{max}}$ in absence of early stopping,
- spend error probabilities as a function of $I/I_{\text{max}}$.

Let $\pi_k$ and $\gamma_k$ 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

$$
\Pr_{\theta=0}\{Z_1 \in C_1, \ldots, Z_{k-1} \in C_{k-1}, Z_k \geq b_k\} = \pi_k - \pi_{k-1} \\
\Pr_{\theta=\delta}\{Z_1 \in C_1, \ldots, Z_{k-1} \in C_{k-1}, Z_k \leq a_k\} = \gamma_k - \gamma_{k-1},
$$

and control the type I error rate at level $\alpha$, and the type II error rate at a level just below $\beta$.

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

In the figure below, error spending tests are designed using the $\rho$-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 $\beta = 0.1$.

![Graph showing error spending and optimal tests]

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

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
Dealing with unexpected overrunning

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

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
Dealing with unexpected overrunning

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

**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 $I_k$ and boundaries $(a_k, b_k)$ stops at analysis $k^* < K$ with $Z_{k^*} > b_{k^*}$ or $Z_{k^*} < a_{k^*}$.

**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 $\tilde{I}_{k^*}$ such that

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

Requiring $c_{k^*}$ to balance the probabilities of reversing decisions under $\theta = 0$ at stage $k^*$ preserves the test’s overall type I error rate.
Dealing with unexpected overrunning

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

**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 $\tilde{I}_{k^*}$ such that

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

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

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

Lisa Hampson and Chris Jennison
Group Sequential Tests for Delayed Responses
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

\[
\begin{pmatrix}
Y_{T,i} \\
X_{T,i}
\end{pmatrix} \sim N
\begin{pmatrix}
\mu_{T,1} \\
\mu_{T,2}
\end{pmatrix},
\begin{pmatrix}
\sigma_1^2 & \tau \sigma_1 \sigma_2 \\
\tau \sigma_1 \sigma_2 & \sigma_2^2
\end{pmatrix}
\]

Lisa Hampson and Chris Jennison

Group Sequential Tests for Delayed Responses
Using a short term endpoint to recover efficiency

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

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

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 \(\hat{\theta}_k\) and \(\mathcal{I}_k = \text{Var}(\hat{\theta}_k)\).
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 $\hat{\theta}_k$ and $I_k = \text{Var}(\hat{\theta}_k)$.

Given $\{I_1, \ldots, I_k, \tilde{I}_k\}$, the sequence of estimates $\{\hat{\theta}_k\}$ 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 \(\hat{\theta}_k\) and \(I_k = \text{Var}(\hat{\theta}_k)\).

Given \(\{I_1, \ldots, I_k, \tilde{I}_k\}\), the sequence of estimates \(\{\hat{\theta}_k\}\) 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 \(\theta\).
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 $\kappa$. 
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 $\kappa$.
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.