Group Sequential and Adaptive designs in East®

Cyrus Mehta
Cytel Corporation and Harvard TH Chan
April 25, 2016
Objectives of Training

This workshop will teach you how to best use East® to design, simulate, and monitor adequately and well controlled trials, while incorporating group sequential and adaptive elements into the design.

East® is the industry standard software package for the design of clinical trials. It directly addresses many of the requirements given in the FDA “Draft Guidance on Adaptive Design for Clinical Trials for Drugs and Biologics.”

- To provide the general background of adaptive designs in clinical trials
- To work through some practical examples using East
- To inform about regulatory positions regarding adaptive designs
Outline

1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
Warm-up Exercise; Recall Fixed Sample Size Session
Design a Phase 3 trial for a normal endpoint (difference of means for two independent populations). What is the total sample size, \( n \), required?

1. Given two-sided \( \alpha = 0.05 \),
2. power \( 1 - \beta = 0.85 \), so \( Z_{\alpha/2} + Z_\beta \approx 3 \),
3. treatment effect \( \delta = 0.3 \),
4. and nuisance parameter \( \sigma = 1 \),
5. compute the sample size \( n = 4 \left( \frac{Z_{\alpha/2} + Z_\beta}{\delta/\sigma} \right)^2 = ? \)
Sample Size Calculation 2

Confirm your answer in East:
Select Des1 in Output Preview, and click 'Output Summary' icon:

![Output Summary icon]

Select Des1 in Output Preview, and click 'Save in Workbook' icon:

![Save in workbook icon]
Select Des1 in the Library, and click 'Details' icon:

**Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means**

<table>
<thead>
<tr>
<th>Test Parameters</th>
<th>Design ID:</th>
<th>Des1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design Type:</td>
<td>Superiority</td>
<td></td>
</tr>
<tr>
<td>Number of Looks:</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Test Type:</td>
<td>2 Sided</td>
<td></td>
</tr>
<tr>
<td>Specified ( \alpha ):</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Power:</td>
<td>0.851</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Test Statistic:</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Method:</td>
<td>Difference of Means</td>
<td></td>
</tr>
<tr>
<td>( \delta = \mu_1 - \mu_c )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under H0:</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Under H1:</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation (( \sigma )):</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Allocation Ratio (( n_1/n_0 )):</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output:</th>
<th>Upper Critical Point</th>
<th>1.96</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Critical Point</td>
<td>-1.96</td>
</tr>
<tr>
<td></td>
<td>Sample Size (( n ))</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Sample Size Treatment (( n_t ))</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Sample Size Control (( n_c ))</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Max Information (( I_{max} ))</td>
<td>100</td>
</tr>
</tbody>
</table>
Introduction to Adaptive Designs
“...defined as a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study . . .”

(FDA, 2010)
Types of Designs

- **Traditional Design**: fix the sample size in advance, and perform one analysis after all subjects have been enrolled and evaluated.

- **Adaptive Design**: monitor the accruing data, and make interim decisions concerning the future course of the study (e.g., stop early, select dose, increase sample size).

- **Group Sequential Design**: a type of adaptive design in which the only interim decision is whether to stop early (for harm, efficacy, futility).
Background Readings

- Statistics for Biology and Health
  Michael A. Proschan
  K.K. Gordon Lan
  Janet Turk Wittes
  Statistical Monitoring of Clinical Trials
  A Unified Approach

- Group Sequential Methods with Applications to Clinical Trials
  Christopher Jennison
  Bruce W. Turnbull

- East
  Advanced Clinical Trial Design, Simulation and Monitoring System
Statistical Methods for Group Sequential Designs
CAPTURE (1997) was a parallel arm placebo-controlled trial, with binary primary endpoint

Use East to compute the sample size required to detect a reduction in event rates (difference of proportions) from 15% to 10% with 80% power using a two-sided level-0.05 test
### Single-Look Design

<table>
<thead>
<tr>
<th>Des1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mnemonics</td>
<td>PN-2S-DI</td>
</tr>
<tr>
<td><strong>Test Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Trial Type</td>
<td>Superiority</td>
</tr>
<tr>
<td>No. of Looks</td>
<td>1</td>
</tr>
<tr>
<td>Test Type</td>
<td>2-Sided</td>
</tr>
<tr>
<td>Variance</td>
<td>Unpooled Estimate</td>
</tr>
<tr>
<td>Specified $\alpha$</td>
<td>0.05</td>
</tr>
<tr>
<td>Specified Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Attained Power</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Model Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion under Control ($\pi_c$)</td>
<td>0.15</td>
</tr>
<tr>
<td>Proportion under Treatment ($\pi_t$)</td>
<td>0.1</td>
</tr>
<tr>
<td>Diff. in Prop. ($\pi_t - \pi_c$)</td>
<td>-0.05</td>
</tr>
<tr>
<td>Allocation Ratio ($nt/nc$)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1366</td>
</tr>
</tbody>
</table>
Ethical Concerns

Very large sample size commitment with no possibility of early termination for benefit, harm or futility

- Suppose Abciximab is actually beneficial: Do we really have to randomize 683 patients to placebo before we know for sure?
- What if Abciximab is no different than placebo or even harmful? Can we avoid randomizing 683 patients to a treatment that is no better or worse than placebo?

Interim monitoring can help in both these cases
Key concepts

1. Understand maximum information, and information fractions
2. Contrast Pocock and O’Brien-Fleming boundaries
3. Choose appropriate error spending functions
4. Compare designs (e.g., fixed vs group sequential) in East
5. Display boundaries on Z-scale and other scales
6. Calculate expected sample sizes for group sequential designs
1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
Let $\delta$ denote the true, unknown treatment effect. We always define $\delta$ to be a difference between the treatment and control groups. For instance:

- a difference of two means
- a difference of two binomial probabilities
- a log hazard ratio
- a log odds ratio
- any general coefficient in a regression model
Monitor the data $K$ times at calendar times $\tau_1, \tau_2, \ldots, \tau_K$

- For normal and binomial endpoints let

$$n_j = \text{sample size at calendar time } \tau_j$$

- For time-to-event endpoints let

$$d_j = \text{number of events at calendar time } \tau_j$$

- More generally, in terms of Fisher information let

$$l_j = \text{information at calendar time } \tau_j \approx \left[ se \left( \hat{\delta}_j \right) \right]^{-2}$$

where $\hat{\delta}_j$ is an efficient estimate of $\delta$ at calendar time $\tau_j$. 
The maximum information is the precision of \( \hat{\delta} \) needed to achieve the desired power.

- For normal and binomial endpoints let
  \[ n_{\text{max}} = \text{the maximum sample size required for the trial} \]

- For time-to-event endpoints let
  \[ d_{\text{max}} = \text{maximum number of events required for the trial} \]

- More generally, in terms of Fisher information, let
  \[ I_{\text{max}} \approx \left[ \text{se} \left( \hat{\delta}_{\text{max}} \right) \right]^{-2} = \text{maximum information to be collected} \]

In a group sequential design we will keep the trial open until either the maximum information is obtained or a stopping boundary is crossed.
Define the information fraction \( t_j \) at calendar time \( \tau_j \)

\[
t_j = \begin{cases} 
\frac{n_j}{n_{\text{max}}} & \text{for normal and binomial} \\
\frac{d_j}{d_{\text{max}}} & \text{for time-to-event} \\
\frac{l_j}{l_{\text{max}}} & \text{in general}
\end{cases}
\]

- If \( K \) is intended to be the last look, we will often denote \( l_{\text{max}} \) by \( l_K \), \( n_{\text{max}} \) by \( n_K \), and \( d_{\text{max}} \) by \( d_K \)
- We may regard the information fraction \( t \), \( 0 \leq t \leq 1 \), as the internal time axis of the clinical trial
1. **Introduction**

2. **Statistical Methods for Group Sequential Designs**
   - Distribution theory
   - **Efficacy stopping rules**
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. **Survival Designs in East**
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. **Workshops 1 - 2 (Group Sequential)**
Let \( \{c_1, c_2, \ldots, c_K\} \) be the corresponding two-sided stopping boundaries

Stop the trial and reject \( H_0 \) at the first \( t_j \) such that

\[
| Z(t_j) | \geq c_j
\]

We must select the \( c_j \)'s so as to preserve the type-1 error

\[
P_0\left\{ \bigcap_{j=1}^{K} | Z(t_j) | < c_j \right\} = 1 - \alpha
\]

Many \( c_j \)'s satisfy this condition.
Pocock (1977) was the first to propose a group sequential design: a constant boundary $c_j = C$ on the Wald scale.

- An equally spaced 5-look level-0.05 two-sided Pocock test utilizes constant boundaries $c_j = 2.413$ for $j = 1, 2, \ldots, 5$

The next proposal came from O’Brien and Fleming (1979) who introduced boundaries of the form $c_j = C/\sqrt{t_j}$.

- An equally spaced 5-look level-0.05 two-sided O’Brien-Fleming design utilizes the boundaries $c_j = 2.04/\sqrt{t_j}$ for $j = 1, 2, \ldots, 5$
Wang and Tsiatis (1987) proposed a family of boundary shapes:

\[ c_j = C t_j^{\Delta - 1/2} \]

where \( \Delta \) is known as a shape parameter.

Which values of \( \Delta \) yield the Pocock (1977) and O’Brien-Fleming (1979) boundaries, respectively?
Exercise: Pocock vs O’Brien-Fleming

- In East, plot the Pocock (1977) and O’Brien-Fleming (1979) boundaries on the Wald scale.

What key differences do you notice between these boundaries? Are they advantages or disadvantages?
Wang-Tsiatis boundaries depend on pre-specified values of the information fraction $t_1, t_2, \ldots, t_K$

In practice, it might be necessary to alter the spacing or number of looks after the trial has started, which requires changing remaining boundary values.

Lan and DeMets (1983) proposed the error spending function approach, where $\alpha$ is treated as a budgeted quantity to be spent.
The $\alpha$-Spending Function Approach

- Specify a monotone increasing function of $t$ for $t \in [0, 1]$, with $\alpha(0) = 0$ and $\alpha(1) = \alpha$
- Solve recursively for $c_1, c_2, \ldots, c_K$

\[ P_0\{| Z(t_1) | \geq c_1 \} = \alpha(t_1) \]

and for $j = 2, \ldots, K$

\[ \alpha(t_{j-1}) + P_0\{| Z(t_1) | < c_1, \ldots, | Z(t_{j-1}) | < c_{j-1}, | Z(t_j) | \geq c_j \} = \alpha(t_j) \]
Exercise: Cumulative $\alpha$ spent and Type I error

Consider a linear spending function ($\alpha = 0.05$) with 4 equally-spaced looks:

- What are the values of the information fractions: $t_1$, $t_2$, $t_3$, $t_4$?
- What are the values of cumulative alpha spent: $\alpha(t_1)$, $\alpha(t_2)$, $\alpha(t_3)$, $\alpha(t_4)$?
- What is the probability of committing a Type I error by Look 2?
- What is the probability of committing a Type I error at Look 3?
Lan and DeMets proposed two spending functions

- The $LD(OF)$ spending function

$$\alpha(t) = \begin{cases} 
4 - 4\Phi(z_{\alpha/4})/\sqrt{t} & \text{for two-sided tests} \\
2 - 2\Phi(z_{\alpha/2})/\sqrt{t} & \text{for one-sided tests}
\end{cases}$$

yields boundaries that approximate the O’Brien-Fleming boundaries

- The $LD(PK)$ spending function

$$\alpha(t) = \alpha \log\{1 + (e - 1)t\}$$

yields boundaries that approximate the Pocock boundaries
Which of these spending functions corresponds to LD(OF) and LD(PK)?
Gamma Family

- Hwang, Shih, & DeCani (1990) proposed

\[ \alpha(t) = \alpha \frac{(1 - e^{-\gamma t})}{(1 - e^{-\gamma})}, \text{ where } \gamma \neq 0 \]

- \( \gamma = -4 \) or \(-5\) (similar to O’Brien-Fleming); \( \gamma = 1 \) (similar to Pocock)
Introduction

2. **Statistical Methods for Group Sequential Designs**
   - Distribution theory
   - Efficacy stopping rules
   - **Power and sample size calculations**
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. **Survival Designs in East**
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. **Workshops 1 - 2 (Group Sequential)**
Three-Step Procedure for Sample Size / Events

1. Compute boundaries, \((c_1, c_2, \ldots, c_K)\), given \(\alpha\), spacing, number of looks and spending function.

2. Compute maximum information \(I_{\text{max}}\) needed to achieve the desired power, given boundaries, treatment effect \(\delta\) and desired power \(1 - \beta\).

3. Convert \(I_{\text{max}}\) into sample size \(n_{\text{max}}\) or events \(D_{\text{max}}\), for given endpoint, and nuisance parameters.
Maximum Information $I_{max}$ required to reject $H_1 : \delta = \delta_1$ with $1 - \beta$ power

- Jennison and Turnbull (1997) showed that $Z(t_j) \sim N(\eta \sqrt{t_j}, 1)$, where $\eta = \delta_1 \sqrt{I_{max}}$, and, for any $t_{j1} < t_{j2}$, $\text{cov}\{Z(t_{j1}), Z(t_{j2})\} = \sqrt{\frac{t_{j1}}{t_{j2}}}$
- Find the value of the drift parameter $\eta$ that satisfies the equation

$$P_{\eta}\left\{\bigcup_{j=1}^{K} |Z(t_j)| \geq c_j \right\} = 1 - \beta$$

- Solve for

$$I_{max} = \left[ \frac{\eta}{\delta_1} \right]^2$$
The effect size $\delta$ is estimated by the difference of the two group sample means

$$I_{\text{max}}^{-1} \equiv \text{var}[\hat{\delta}(t_K)] = \text{var}[\bar{X}_t(t_K) - \bar{X}_c(t_K)]$$

$$= \frac{\sigma^2}{(n_{\text{max}}/2)} + \frac{\sigma^2}{(n_{\text{max}}/2)}$$

$$= \frac{4\sigma^2}{n_{\text{max}}}$$

where $\sigma$ is the common standard deviation and assuming balanced allocation.

Therefore, the relationship between $n_{\text{max}}$ and $I_{\text{max}}$ is

$$n_{\text{max}} = 4\sigma^2 I_{\text{max}}$$

To compute $n_{\text{max}}$, we need $\sigma$, a “nuisance parameter”
The effect size \( \delta \) is estimated by the difference of the two group binomial response rates

\[
I^{-1}_{max} \equiv \text{var}[\hat{\delta}(t_K)] = \text{var}[\hat{\pi}_t(t_K) - \hat{\pi}_c(t_K)] = \frac{\pi_t(1-\pi_t)}{(n_{max}/2)} + \frac{\pi_c(1-\pi_c)}{(n_{max}/2)}
\]

where we have assumed balanced allocation

- Solve for \( n_{max} \) under \( H_1 : \pi_t - \pi_c = \delta_1 \) to obtain

\[
n_{max} = 2[\pi_c(1-\pi_c) + (\pi_c + \delta_1)(1-\pi_c - \delta_1)]I_{max}
\]

- To compute \( n_{max} \), we need \( \pi_c \), a “nuisance parameter”
For time-to-event endpoints, we usually assume the **proportional hazards** alternative. In this case, the effect size $\delta$ is the logarithm of the hazard ratio and the variance of $\hat{\delta}_j$ at any interim look $j$ is proportional to the number of events $D_j$.

For balanced randomization, it can be shown that, approximately

$$I_{\text{max}}^{-1} \equiv \text{var}[\hat{\delta}(t_K)] \approx [D_{\text{max}}/4]^{-1}$$

Thus

$$D_{\text{max}} = 4I_{\text{max}}$$
Topics

1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
CAPTURE (1997) was a parallel arm placebo-controlled trial, with binary primary endpoint

Use East to compute the sample size required to detect a reduction in event rates (difference of proportions) from 15% to 10% with 80% power using a two-sided level-0.05 test
## Single-Look Design

<table>
<thead>
<tr>
<th>Des1</th>
<th>Mnemonics</th>
<th>PN-2S-DI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial Type</td>
<td>Superiority</td>
<td></td>
</tr>
<tr>
<td>No. of Looks</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Test Type</td>
<td>2-Sided</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>Unpooled Estimate</td>
<td></td>
</tr>
<tr>
<td>Specified $\alpha$</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Specified Power</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Attained Power</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td><strong>Model Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion under Control ($\pi_c$)</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Proportion under Treatment ($\pi_t$)</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Diff. in Prop. ($\pi_t - \pi_c$)</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Allocation Ratio ($nt/nc$)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td>Maximum</td>
<td>1366</td>
</tr>
</tbody>
</table>
Ethical Concerns

Very large sample size commitment with no possibility of early termination for benefit, harm or futility

- Suppose Abciximab is actually beneficial: Do we really have to randomize 683 patients to placebo before we know for sure?
- What if Abciximab is no different than placebo or even harmful? Can we avoid randomizing 683 patients to a treatment that is no better or worse than placebo?

Interim monitoring can help in both these cases
The investigators planned to take up to two interim looks using the \( LD(OF) \) spending function. The three looks will be equally spaced.

Use East to compute the maximum sample size for this group sequential design.
## Summary of Three-Look Design

<table>
<thead>
<tr>
<th>Des8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mnemonics</strong></td>
<td>PN-2S-DI</td>
</tr>
<tr>
<td><strong>Test Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Trial Type</td>
<td>Superiority</td>
</tr>
<tr>
<td>No. of Looks</td>
<td>3</td>
</tr>
<tr>
<td>Test Type</td>
<td>2-Sided</td>
</tr>
<tr>
<td>Variance</td>
<td>Unpooled Estimate</td>
</tr>
<tr>
<td>Specified ( \alpha )</td>
<td>0.05</td>
</tr>
<tr>
<td>Specified Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Attained Power</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Model Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion under Control (( \pi_c ))</td>
<td>0.15</td>
</tr>
<tr>
<td>Proportion under Treatment (( \pi_t ))</td>
<td>0.1</td>
</tr>
<tr>
<td>Diff. in Prop. (( \pi_t - \pi_c ))</td>
<td>-0.05</td>
</tr>
<tr>
<td>Allocation Ratio (( nt/nc ))</td>
<td>1</td>
</tr>
<tr>
<td><strong>Boundary Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Efficacy Boundary</td>
<td>LD (OF)</td>
</tr>
<tr>
<td>Spacing of Looks</td>
<td>Equal</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1384</td>
</tr>
<tr>
<td>Expected Under H0</td>
<td>1378.32</td>
</tr>
<tr>
<td>Expected Under H1</td>
<td>1182.678</td>
</tr>
<tr>
<td><strong>Completers</strong></td>
<td></td>
</tr>
<tr>
<td>Expected Under H0</td>
<td>1378.32</td>
</tr>
<tr>
<td>Expected Under H1</td>
<td>1182.678</td>
</tr>
</tbody>
</table>
The sum of incremental boundary crossing probabilities under \([H_0 : \delta = 0]\) is 0.05, the type-1 error.

The sum of incremental boundary crossing probabilities under \([H_1 : \delta = -0.05]\) is 0.8, the power.
Exercise: Expected Sample Size

1. Let \(n_1\), \(n_2\), and \(n_3\) be the sample sizes at Looks 1, 2, and 3, respectively.

2. Let \(p_1\), \(p_2\), and \(p_3\) be the incremental boundary crossing probabilities under \(H_1\) at Looks 1, 2, and 3, respectively.

3. Write down an expression to calculate the expected sample size, from the variables: \(n_1\), \(n_2\), \(n_3\), \(p_1\), \(p_2\), \(p_3\). Confirm by calculation.
1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
An efficacy boundary offers possibility of early stopping and savings in sample size if $H_1$ is true, but what if $H_0$ is true?

Need to add a futility stopping rule.

Many options for futility boundaries in East (for one-sided test type):
(Pampallona, Tsiatis, and Kim, 2001)

- Just as we use an $\alpha$-spending function to generate efficacy boundaries that preserve type-1 error, we can also use a $\beta$-spending function to generate futility boundaries that control type-2 error.
- The probability of crossing the efficacy boundary under $H_0$ is $\alpha$.
- The probability of crossing the futility boundary under $H_1$ is $\beta$.
- We force the two boundaries to meet at the last look to ensure that either the null or the alternative hypothesis is rejected.
For the CAPTURE trial, we continue to use the $LD(OF)$ spending function for $\alpha$, but spend $\beta$ using the $\gamma(-2)$ spending function.
Both Efficacy and Futility Boundaries

- The two boundaries meet at $l_3 = u_3 = -1.993$, exactly as in the efficacy boundary only design.
- Thus the futility boundary can be safely ignored (Non-Binding) if crossed at an early look without inflating the study’s type-1 error.
Savings under both H0 and H1

- This design protects the sample size both if $\delta = 0$ and if $\delta = -0.05$

<table>
<thead>
<tr>
<th>Memonics</th>
<th>PN-25-D1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Trial Type</td>
<td>Superiority</td>
</tr>
<tr>
<td>No. of Looks</td>
<td>3</td>
</tr>
<tr>
<td>Test Type</td>
<td>1-Sided</td>
</tr>
<tr>
<td>Variance</td>
<td>Unpooled Estimate</td>
</tr>
<tr>
<td>Specified $\alpha$</td>
<td>0.025</td>
</tr>
<tr>
<td>Attained $\alpha$</td>
<td>0.023</td>
</tr>
<tr>
<td>Specified Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Attained Power</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Model Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion under Control ($\pi_c$)</td>
<td>0.15</td>
</tr>
<tr>
<td>Proportion under Treatment ($\pi_t$)</td>
<td>0.1</td>
</tr>
<tr>
<td>Diff. in Prop. ($\pi_t - \pi_c$)</td>
<td>-0.05</td>
</tr>
<tr>
<td>Allocation Ratio ($nt/nc$)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Boundary Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Efficacy Boundary</td>
<td>LD (OF)</td>
</tr>
<tr>
<td>Futility Boundary</td>
<td>Gm (~2) (NB)</td>
</tr>
<tr>
<td>Spacing of Looks</td>
<td>Equal</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1455</td>
</tr>
<tr>
<td>Expected Under H0</td>
<td>848</td>
</tr>
<tr>
<td>Expected Under H1</td>
<td>1174.624</td>
</tr>
<tr>
<td><strong>Completers</strong></td>
<td></td>
</tr>
<tr>
<td>Expected Under H0</td>
<td>848</td>
</tr>
<tr>
<td>Expected Under H1</td>
<td>1174.624</td>
</tr>
</tbody>
</table>

- Why is the attained $\alpha$ lower than the specified $\alpha$?
1. **Sensitivity analysis** for changes in assumptions (e.g., treatment effect \( \delta \) or nuisance parameters)

2. **Breakdown of asymptotics** when samples are small

3. **Operating characteristics** in complex survival designs, adaptive sample size re-estimation, or dose selection trials, are possible only via simulation

4. **Visual communication** of properties and variability of study design
Demo: Simulate the CAPTURE Trial

<table>
<thead>
<tr>
<th>Look #</th>
<th>Look Position</th>
<th>H0 -</th>
<th>H0 +</th>
<th>H1 -</th>
<th>H1 +</th>
<th>Latest Simul. Test Stat</th>
<th>Average Info.</th>
<th>Average Sample Size</th>
<th># Rejecting H0</th>
<th># Rejecting H1</th>
<th>Total Simul. Count</th>
<th>Total Simul. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>485</td>
<td>-2.71</td>
<td>0.216</td>
<td>-2.899</td>
<td>1119</td>
<td>485</td>
<td>34</td>
<td>57</td>
<td>91</td>
<td>4.55</td>
<td>2703...</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>870</td>
<td>-2.511</td>
<td>-0.923</td>
<td>-2.035</td>
<td>2235.33</td>
<td>970</td>
<td>840</td>
<td>128</td>
<td>977</td>
<td>48.85</td>
<td>1173.943</td>
<td>79</td>
</tr>
<tr>
<td>3</td>
<td>1455</td>
<td>-1.993</td>
<td>-1.993</td>
<td>-3.91</td>
<td>3354...</td>
<td>1455</td>
<td>697</td>
<td>235</td>
<td>932</td>
<td>46.6</td>
<td>1580</td>
<td>21</td>
</tr>
</tbody>
</table>

Total 2703... 1173.943 1580 420 2000 100

Stopping Boundaries

Rej. H0 and Rej. H1

Number of Trials Completed = 2000
Simulation Seed = 100
Elapsed Time = 00:00:12
Interim Monitoring of CAPTURE

The Capture trial was designed for three equally spaced looks with \( n_{\text{max}} = 1383 \) and stopping boundaries derived from the LD(OF) spending function. But the trial was actually monitored with unequal spacing as shown below:

<table>
<thead>
<tr>
<th>Look</th>
<th>( N_j )</th>
<th>( \frac{N_j}{1383} )</th>
<th>Resp. Plcbo</th>
<th>Resp. Abcix</th>
<th>Wald ( Z_j )</th>
<th>Old ( c_j )</th>
<th>New ( c_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>350</td>
<td>0.253</td>
<td>30/175</td>
<td>14/175</td>
<td>-2.6046</td>
<td>-3.71</td>
<td>-4.3048</td>
</tr>
<tr>
<td>2</td>
<td>700</td>
<td>0.506</td>
<td>55/353</td>
<td>37/347</td>
<td>-1.9332</td>
<td>-2.5114</td>
<td>-2.9429</td>
</tr>
<tr>
<td>3</td>
<td>1050</td>
<td>0.759</td>
<td>84/532</td>
<td>55/518</td>
<td>-2.485</td>
<td>-1.993</td>
<td>-2.3426</td>
</tr>
</tbody>
</table>

Look 3 was unplanned and occurred before the planned end of the trial.
## Demo: Monitor the CAPTURE Trial

![CAPTURE Trial Monitor](image)

### Table: Capture 3 Look-Fut Interim Monitoring

<table>
<thead>
<tr>
<th>Look</th>
<th>Information Fraction</th>
<th>Cumulative Sample Size</th>
<th>Test Statistic</th>
<th>δ</th>
<th>Standard Error</th>
<th>Efficacy</th>
<th>97.5% RCI for δ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repeated p-value</td>
</tr>
<tr>
<td>1</td>
<td>0.2529</td>
<td>350</td>
<td>-2.6046</td>
<td>-0.0914</td>
<td>0.0351</td>
<td>-4.3061</td>
<td>0.0597</td>
</tr>
<tr>
<td>2</td>
<td>0.5058</td>
<td>700</td>
<td>-1.9332</td>
<td>-0.0492</td>
<td>0.0254</td>
<td>-2.9439</td>
<td>0.0257</td>
</tr>
<tr>
<td>3</td>
<td>0.7587</td>
<td>1050</td>
<td>-2.485</td>
<td>-0.0517</td>
<td>0.0208</td>
<td>-2.3434</td>
<td>-0.0029</td>
</tr>
</tbody>
</table>
Designing and Simulating for Survival Endpoints
Survival Studies

- For studies with survival or time-to-event endpoints, the asymptotic distribution theory and the derivation of stopping boundaries remains the same.
- There are however some special considerations:
  - The information is directly proportional to the number of events.
  - Thus the number of events, not the number of patients, determines the power of the study.
  - If study duration is not fixed, there is a trade-off between sample size and study duration.

\[ S_a \quad (S_a + S_f) \]
If we recruit more patients to the study, we obtain the required number of events sooner and the total study duration is reduced.
1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) examined the question of whether treatment with 20 mg of rosuvastatin daily, as compared with placebo, would reduce the rate of first major cardiovascular events (Ridker et al., 2008)
JUPITER was a randomized, double-blind, placebo-controlled, multicenter trial conducted between 2003 and 2008 by AstraZeneca at 1315 sites in 26 countries.

Composite primary endpoint: occurrence of a first major cardiovascular event, defined as nonfatal myocardial infarction, nonfatal stroke, hospitalization for unstable angina, an arterial revascularization procedure, or confirmed death from cardiovascular causes.
Example: The JUPITER study (cont.)

- Designed for statistical power of 90% to detect a 25% reduction in the rate of the primary end point, with a two-sided significance level of 0.05

- We are interested in a 25% reduction in the rate of the primary endpoint, i.e. a hazard ratio $\lambda_t/\lambda_c = 0.75$; the effect size is thus $\delta = -\ln(\lambda_t/\lambda_c) = 0.2877$

- Baseline hazard rate in placebo arm of 0.0077

- Study to complete in 7.5 years, with 4 years accrual and 3.5 years of follow-up

- Two interim analyses are planned with $LD(OF)$ spending function defined boundaries at 37.5% and 75% of the information

- How do we design such a trial?
1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
To convert information into number of events, we use Schoenfeld’s approximation (Biometrika, 1981)

\[ D_{\text{max}} = \frac{l_{\text{max}}}{p(1 - p)} \]

where \( D_{\text{max}} \) is the maximum number of events required, \( p \) is the proportion of these events that occur on the treatment arm, and \((1 - p)\) is the proportion of these events that occur on the control arm.

If the null hypothesis holds, we can set \( p = r \), where \( r \) is the fraction of patients randomized to the treatment arm. For balanced randomization where \( p = 1/2 \) we have

\[ D_{\text{max}} = 4l_{\text{max}} \]
Equating Required and Expected Number of Events to Estimate Sample Size and Study Duration

- For a $K$ look design, if $D_{\text{max}}$ events are desired, then the study must remain open until a boundary is crossed or calendar time $\tau_K$ where $\tau_K$ satisfies the relationship

$$D_{\delta_1}(\tau_K) = D_{\text{max}}$$

where $D_{\delta_1}(\tau_K)$ is the expected number of events at time $\tau_K$

- We can derive an expression for $D_{\delta_1}(\tau_K)$ from exponential assumptions and thereby estimate sample size and study duration
We can use the Logrank Test Given Accrual Duration and Study Duration design in East to obtain a 3-look GSD for the JUPITER trial.

- East tells us that the required number of events is \( D_{\text{max}} = 517 \) as previously determined, and that we will require 14,229 subjects accrued over 4 years.
Early Stopping and Expected Study Duration

\[
E(\text{Study Duration} \mid H_1) = 0.0703 \times 4.042 \\
+ 0.6186 \times 6.109 \\
+ 0.2114 \times 7.5 \\
+ (1 - 0.0703 - 0.6185 - 0.2114) \times 7.5 \\
= 6.397
\]
The Events vs. Time Chart is useful in planning for interim analyses & DMC meetings; eg., after the second look, 389 events should occur roughly 6.1 years into the study assuming the treatment effect is $\delta_1$. 
1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
Dropouts: 5% per year in both the treatment and placebo arms

Like events: Dropouts assumed exponential, with corresponding hazard rates

Like Cumulative % survival: Cumulative % dropout calculated for patient time (from study entry), not calendar time (from study start)
JUPITER Design Accruals

- Non constant accrual (piece-wise linear): 15% by end of year 1; 35% by end of year 2; 65% by end of year 3

![Accrual Info Image]

<table>
<thead>
<tr>
<th>Period #</th>
<th>By Time</th>
<th>Cum. % Accrued</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>15.000</td>
</tr>
<tr>
<td>2</td>
<td>2.000</td>
<td>35.000</td>
</tr>
<tr>
<td>3</td>
<td>3.000</td>
<td>65.000</td>
</tr>
</tbody>
</table>
Note that since we have not changed the effect size $\delta_1$ that we are powering the trial for, $D_{\text{max}} = 517$ has not changed. However, the dropouts and slow starting accrual means we now need **17,344 patients** if we want to finish the study within 7.5 years.
Topics

1. Introduction

2. Statistical Methods for Group Sequential Designs
   - Distribution theory
   - Efficacy stopping rules
   - Power and sample size calculations
   - Example: The CAPTURE Trial
   - Futility stopping rules

3. Survival Designs in East
   - Example: The JUPITER Study
   - Calculating the required number of events
   - Handling drop-outs, variable accrual and non-constant hazards
   - Non-proportional hazards

4. Workshops 1 - 2 (Group Sequential)
Non-Proportional Hazards

- All survival designs in East assume the proportional hazards assumption holds.
- Non-proportional hazards may arise in clinical trials for many reasons (e.g., delayed treatment effects).
- In East, simulated data may be generated according to non-proportional hazards.
- Thus, it is possible to evaluate the impact of non-proportional hazards on the power of a study.
**Demo: Simulate the JUPITER Trial**

![Simulation Interface]

<table>
<thead>
<tr>
<th>Look</th>
<th>Look Position</th>
<th>H0+</th>
<th>H0-</th>
<th>H1+</th>
<th>H1-</th>
<th>Latest Simul...</th>
<th>Average Events</th>
<th># Rejecting... Up(H0+)</th>
<th># Rejecting... Low(H0-)</th>
<th># Unable... Reject...</th>
<th>Total Simul... Count</th>
<th>Total Simul... %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>194</td>
<td>3.477</td>
<td>-3.477</td>
<td>-1.201</td>
<td>194</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>41</td>
<td>6.833</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>388</td>
<td>2.342</td>
<td>-2.342</td>
<td>-1.745</td>
<td>388</td>
<td>0</td>
<td>381</td>
<td>0</td>
<td>381</td>
<td>63.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>517</td>
<td>2.012</td>
<td>-2.012</td>
<td>-2.24</td>
<td>517</td>
<td>0</td>
<td>110</td>
<td>0</td>
<td>68</td>
<td>29.667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total**

<table>
<thead>
<tr>
<th></th>
<th>413.013</th>
<th>532</th>
<th>68</th>
<th>600</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>0</td>
<td>88.667</td>
<td>11.333</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Stopping Boundaries**

![Stopping Boundary Graph]

**Rej. H0 and Unable to Rej. H0**

- Number of Trials Completed = 600
- Simulation Seed = 100
- Elapsed Time = 00:01:01

![Rejection and Unable to Rejection Graph]
Astrazeneca Stops Cholesterol Drug Trial Because Of Promising Results

01 Apr 2008  Click to Print

Pharmaceutical giant AstraZeneca announced yesterday, Monday 31st March, it was stopping the JUPITER clinical trial on its cholesterol-busting drug Crestor (rosuvastatin calcium) because early findings showed that the drug reduced deaths and risk of heart problems in patients compared to placebo.

A press statement on the company’s website said that the trial's Steering Committee and also the Independent Data Monitoring Board recommended that JUPITER be stopped early because there was "unequivocal evidence of a reduction in cardiovascular morbidity and mortality amongst patients" who took Crestor compared to patients who took placebo.
At the first interim analysis in September 2007, the efficacy boundary was crossed. However, the DMC voted to continue the trial for an additional 6 months. Thus, the next interim analysis in March 2008 was not originally planned.

Suppose that in March 2008, 142 events had been observed on the Rosuvastatin arm against 251 events on the placebo arm.

The estimated hazard ratio was 0.56 so that $\hat{\delta}_1 = \ln(0.56)$. The standard error can be estimated as $\text{se}(\hat{\delta}) = I_1^{-1/2} = \sqrt{4/D_1}$ where $D_1 = 142 + 251 = 393$.

The JUPITER study was terminated early for overwhelming efficacy in the primary endpoint: HR 0.56 (95% CI 0.46-0.69) $P < 0.00001$ (Ridker, 2008)
Demo: Monitor the JUPITER Trial

Test Statistic Calculator

Editing Look #1

- Set Current Look as Last
- Cumulative Events: 393
- Input for Survival end point:
  - Estimate of $\delta$: 0.58
  - $\delta = \ln(\lambda_t / \lambda_c)$
  - Standard Error of Estimate of $\delta$: 0.101

Output:
- Test Statistic: 5.747

Recalc  OK  Cancel
Demo: Monitor the JUPITER Trial

Cyrus Mehta

Harvard THC

April 25, 2016
East provides great flexibility, but in practice, most group sequential designs have a few common features:

1. **Conservative α-spending**: Discourages termination during early adjustments (patient treatment, trial management). Also, wait until enough safety data have been collected.

2. **Non-binding futility**: DMCs may wish to continue a trial even after the futility boundary has been crossed (e.g., to follow secondary endpoint).
3. **Delay time of first interim**: The first interim analysis is usually performed after enough data to provide stable estimates.

4. **Anticipate and avoid over-runs**: The final interim analysis should be performed while it is still possible to reduce the sample size.

5. **Fewer interim analyses**: Usually between 2 and 3. (Large trials with mortality endpoints might have as many as 6)
Questions?
Design of a Phase 3 Clinical Trial in Hypertension
You work for Lifebeat Inc., a multinational pharmaceutical company which has in-licensed a product called ANTIHYPE®.

Your team has been tasked with developing a phase 3 clinical study design for this product in the essential hypertension indication to serve in a New Drug Application submission for the US Food and Drug Administration regulatory agency.

Your team is told that the phase 3 trial must be completed within two years.
Fixed Sample Design

- Primary endpoint: Change from baseline to Week 6 (1.5 months) in the 24-hour mean systolic blood pressure by ambulatory blood pressure monitor
- Based on preliminary phase 2 study results and the commercial team's feedback, you should power your trial for a treatment effect of $\delta = -1.5$ mm Hg when compared to placebo in a 1 : 1 allocation
- A common standard deviation of $\sigma = 6$ mm Hg can be assumed
- Enrollment rate can be assumed to be 40 patients per month (Click “Include Options”, then “Accrual/Dropout Info”)

**Question 1**: Creating a fixed-sample design (Design 1) with 90% power and two-sided $\alpha = 0.05$ type-1 error. What is the sample size in each group? (Note that the sample size provided by East is the total sample size)

$$N_{Treatment} : _______ \quad N_{Placebo} : _______$$

**Question 2**: How long will the study last?
Three-Look Group Sequential Design

- Convert Design 1 to a 3-look group sequential study design with the default $LD(OF)$ spending function (Design 2)

- **Question 3:** What is the **maximum total** sample size of this 3-look group sequential design? (And what is the penalty in maximum sample size for Design 2, compared to Design 1)

  $N_{\text{max}}$:

- Examine the boundaries chart, the spending function chart, and the power chart by clicking on their respective icons.

- **Question 4:** From the spending function chart, what amount of alpha would have been spent by information fraction $= 0.5$?

  Alpha spent:
Observe that the boundaries chart can be displayed on various scales. The delta scale is particularly useful for communicating with clinicians (NB: contrast with design delta).

**Question 5:** What is final critical value on the Z-scale and delta-scale for the 3-look design (cf. for the 1-look design)?

**critical values : ____**
Further Exploration of Design Properties

- **Question 6**: What is the expected savings in sample size under $H_1$ for Design 2 compared to Design 1?
  
  Savings:

- Do you know how to calculate the expected sample size by hand?

- **Question 7**: If the treatment effect is indeed $\delta_1 = -1.5$, with what sample size are you most likely to stop the trial and reject $H_0$? With what probability does this occur?

  Sample Size: _____  probability: _____
Suppose you want to take only 2 looks, not 3. And suppose you want to space them unequally, after 75% and 100% of the information has accrued (Design 3).

**Question 8:** What is the maximum sample size of this 2-look group sequential design (Design 3)?

*Sample Size: _____*
Adding in a Futility Boundary

- Edit the original 3-look equally-spaced design (Design 2). Change to a one-sided with $\alpha = 0.025$. Add a non-binding futility boundary, with an aggressive $\beta$-spending function: $\gamma(-2)$. This will be Design 4

- **Question 9:** What is the maximum total sample size of this design?

  $N_{\text{max}} : \underline{______}$

- **Question 10:** At the first look, what is the futility boundary on the CP delta1 scale?

  CP boundary : $\underline{______}$
Simulation

- Let’s simulate Design 4 under different conditions.

- **Question 11**: Verify the operating characteristics of this design by simulating the study under $H_0 : \delta = 0$ and under $H_1 : \delta = -1.5$. Use the same fixed seed for each simulation.

  \[ \hat{\alpha} : \quad 1 - \hat{\beta} : \quad \]

- **Question 12**: Investigate each of the following scenarios (1) $\delta = -0.5$; (2) $\delta = -1.8$; (3) $\delta = -1.5; \sigma = 9.5$. Provide below the estimated power, the probability of stopping early for efficacy, and the probability of stopping for futility for each of these settings. You may need to complete some calculations by hand.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Power</th>
<th>Pr(early efficacy)</th>
<th>Pr(early futility)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $\delta = -0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) $\delta = -1.8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) $\sigma = 9.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Design of a Phase 3 Survival Study in Small Cell Lung Cancer
You now work for Cancergene Inc., a multinational pharmaceutical company which has developed an in-house product called ONCOBLAST®.

Your team has been tasked with developing a phase 3 clinical study design for this product in the small cell lung cancer indication to serve in a Market Authorisation Application submission for the European Medicines Agency regulatory agency.

Your final safety database must include at least 500 patients.

Your team is told that the phase 3 trial must be completed within five years.
Fixed Sample Design

- Primary endpoint: overall survival (OS)
- Based on early phase results, the medical literature, and the commercial team’s feedback, you should power your trial for a 30% improvement in median survival time for ONCOBLAST when administered with Standard of Care (SOC) relative to the SOC alone.
- Median survival time on the SOC is assumed to be 11 months.
- The study must enroll within 3 years and be completed within 5 years.
- Note that you are using Logrank Given Accrual Duration and Study Duration.

**Question 1**: Create a fixed-sample design (Design 1) with 90% power and one-sided $\alpha = 0.025$ type-1 error. How many events do you need to observe?

$$D : \text{______}$$

**Question 2**: How many subjects do you need to enroll in the 3 year period to complete the study within 5 years?
Three-Look Group Sequential Design

Convert this design to a 3-look group sequential study design. Choose the Lan-Demets (OF) boundary family with equal spacing of the looks (Design 2)

Question 3: What is the maximum number of events needed to power this 3-look group sequential design?

\[ D_{\text{max}} : \_\_\_\_\_ \]

Question 4: How many patients will you need to enroll to achieve study completion within 5 years?

\[ N_{\text{max}} : \_\_\_\_\_ \]
Additional Assumptions

- Add in a Lan-Demets(OF) non-binding futility boundary (Design 3)
- Plan for 3% dropout per year in each arm [NB: Use “Dropout Rates’’]
- Also, accrual is slower than expected: 20% by end year 1; 50% by end year 2; and the rest by end year 3
- Question 5: With these added assumptions, what is the maximum number of events needed to power this trial?

\[ D_{\text{max}} : \text{_____} \]

- Question 6: How many patients will you need to enroll to achieve study completion within 5 years?

\[ N_{\text{max}} : \text{_____} \]
Question 7: In Design 3, select Charts:Power vs. Sample size. Enter 500 in the Sample Size field. What power is achieved at 500 patients, the minimum needed for your safety database?

Power: ______

Question 8: In the table from Design Details, identify the timing from study start of your two interim analyses if the treatment effect is indeed $\delta_1$. Alternatively, find these times from the events/accruals vs. time Chart.

Look #1: ___  Look #2: ___
Survival simulations

- In the ‘Simulation Control Info’ tab, set a fixed random number seed.
- **Question 9**: Simulate Design 3 under the following three hypotheses: (1) \(HR = 0.9\); (2) \(HR = 0.5\), and (3) \(HR = 1\). What is the power and expected sample size under each scenario?

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Power</th>
<th>E(Sample Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) (HR = 0.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) (HR = 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) (HR = 1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Non-Proportional Hazards

- Now suppose the baseline hazard rate in the control arm remains unchanged, but that $HR = 1.0$ in the first 6 months, yet $HR = 0.72$ thereafter.

- **Question 10:** Simulate this study under the Design 3. What is the simulated power of the trial?

  \[
  \text{Power} : \
  \]

- **Question 11:** How often do we stop early for futility (1st or 2nd look)?

  \[\Pr(\text{futility}) : \]

Cyrus Mehta
Harvard THC
April 25, 2016